ARIPUC No. 2/1967

ANNUAL REPORT
OF THE
INSTITUTE OF PHONETICS
UNIVERSITY OF COPENHAGEN

ARIPUC No.2/1967

ANNUAL REPORT OF THE INSTITUTE OF PHONETICS
OF THE UNIVERSITY OF COPENHAGEN

Copenhagen, 1968

CONTENTS.

PERSONNEL OF THE INSTITUTE OF PHONETICS

1967.

Permanent Staff:

Eli Fischer-Jørgensen (professor, director of the Institute)
Jørgen Rischel (amanuensis of general phonetics)
Hans Peter Jørgensen (amanuensis of general phonetics)
Oluf M. Thorsen (amanuensis of general phonetics)
Børge Frøkjær-Jensen (amanuensis of general phonetics)
Svend-Erik Lystlund (technician)
Inger Østergaard (secretary)

Part Time Teachers of General Phonetics:

Mogens Baumann Larsen (amanuensis of Danish phonetics)
Hans Basbøll (teaching assistant)
Karen Landschultz (teaching assistant)
Walter Wilfried Schuhmacher (teaching assistant)

Guest Research Workers:

Radhekant Dave (India)
Ved Kumari Ghai (India)
Hideo Mase (Japan)

INSTRUMENTAL EQUIPMENT OF THE LABORATORY.

The following is a list of the instruments that have been purchased or built since January 1st, 1967.

Instrumentation for Speech Analysis.

1 Kay-Electric Sona-Graph, type 6061 A.

2 Air-Pressure Manometers, Simonsen & Weel type HB 66.

1 Photo-Electric Glottograph (see this report, pp.5ff.).

Instrumentation for Speech Synthesis.

1 Voice-Source Generator (also see this report, p.34).

Filters.

1 Heterodyne Bandpass Filter (modification of filter mentioned in ARIPUC 1/1966, see this report, pp.20ff.).

Instrumentation for Visual Recording.

1 Oscilloscope, Tektronix type 564 storage.

1 Dual-Trace Amplifier, Tektronix type 3A1.

1 Time Base, Tektronix type 3B3.

Tape Recorders.

4 Tape Recorders, Tandberg type 92 SL.

Gramophones.

2 Gramophones, Delphon (mono, Ortofon pick-up).

General-Purpose Electronic Instrumentation.

1 DC Millivoltmeter, Danameter type 205.

1 DC Nanoammeter, Danameter type 206.

1 Component Bridge C/L/R, Wayne Kerr type B522.

2 Stabilized Rectifiers (5 - 30 V and -6 - 0 - +12 V)

LECTURES AND COURSES IN 1967.

1. Elementary phonetics courses.

One-semester courses (two hours a week) in elementary phonetics (intended for all students of foreign languages) were given by Hans Basbøll, Børge Frøkjær-Jensen, Hans Peter Jørgensen, Karen Landschultz, Mogens Baumann Larsen, Jørgen Rischel, W.W. Schuhmacher, and Oluf M. Thorsen. There were 4 parallel classes in the spring semester and 17 in the autumn semester (some 500 students in all).

2. Practical training in sound perception and transcription.

Courses for beginners as well as more advanced students (each course two hours a week) were given through 1967 by Eli Fischer-Jørgensen, Jørgen Rischel, and Oluf M. Thorsen (the courses form a cycle of three semesters).

3. Instrumental phonetics.

The courses form a cycle of three semesters with two hours a week.

Spring semester: Methods for the investigation of intensity and fundamental frequency.

Autumn semester: (a) General introduction and spectrography.
(b) Physiological methods.

The courses were given by Eli Fischer-Jørgensen and Børge Frøkjær-Jensen.

4. Phonemics.

The courses form a cycle of three semesters with two hours a week.

Spring semester: Tone and stress from a functional point of view.

Autumn semester: Elementary course in phonemic analysis, 1st semester (general principles).

The phonemics courses were given by Jørgen Rischel.

5. Seminars.

The following seminars were held in the spring semester:

a. Ole Kongsdal Jensen reported on his instrumental analysis of French nasal vowels (cp. ARIPUC 1/1966, pp.59ff.).

b. Dr Wayne Slawson (Yale) gave a seminar on "A comparison of vowel quality and musical timbre".

c. Dr odont. Eigild Møller (Dental College, Copenhagen) reported on electro-myography of the speech-organs.

The following were held in the autumn semester:

d. Karen Landschultz reported on her instrumental analysis of vowel length in French (cp. this report, pp.109ff.).

e. Eli Fischer-Jørgensen reported on her study of murmured vowels in Gujarati (cp. this report, pp.35ff.).

It may be added that a meeting was held in the Linguistic Circle of Copenhagen on 9 May, at which Ved Kumari Ghai gave a paper on "Tones in Dogri and Punjabi" (cp. this report, pp.133ff.), and Radhekant Dave and Eli Fischer-Jørgensen gave papers on "Murmured vowels in Gujarati" (cp. this report, pp.35ff. and 119ff.).

6. Staff-classes.

Jørgen Rischel gave a brief survey of methods in speech synthesis by means of electrical analogs.

7. Participation in congresses.

All members of the teaching staff took part in the International Congress of Phonetic Sciences in Prague September 1967. Jørgen Rischel gave a paper on "Acoustic features of syllabicity in Danish".

A PHOTO-ELECTRIC GLOTTOGRAPH

Børge Frøkjær-Jensen

In the past year a photo-electric glottograph was built in the Institute of Phonetics by the present writer.

Different types of glottographs.

The first glottograph was constructed by Dr Sonesson, Sweden (1), (2). This photo-electric equipment is destined for observations of the larynx vibrations, but it is not possible to use it in connected speech because of interference from the plastic rod which is inserted through the mouth and placed in the pharynx. One can mainly pronounce or sing long vowels like [ɛː] and [œː].

The Fabre glottograph (3), (4), and (5), which measures the electric impedance in the larynx by means of a pair of electrodes placed in contact with the skin on both sides of the larynx, is considered to be very good for recording glottis movements in connected speech. But it is doubtful whether it is exclusively glottis movements that are recorded. Contractions of the pharynx and raising and lowering of the larynx obviously influence the result (6).

In the last few years a good deal of work has been made in different laboratories to develop the glottograph. Some experiments have been done at the Haskins Laboratories. The light source was inserted through the mouth via a fibre optics bundle, and the photo transducer that picked up the light passing through the glottis was placed in contact with the skin just below the larynx (7). Under these conditions some difficulties arose when the subject made velopharyngeal closures and when the cartilages of the larynx were moving up and down.

A better solution is apparently to place the light source below the glottis (as Sonesson did) and to pick up the light variations by means of a small photo transducer inserted through the nose into the pharynx. This has been done by Malécot whose glottograph (called GLOTLUX) was primarily intended for investigations of the fortis/lenis problem (8).

The same principles have been employed by John Ohala (9). He has tried to fix the photo transducer in the pharynx by mounting it 7 cm from the lower end in a transparent plastic tube (4 mm in diameter). The lower end of this tube was fixed by swallowing it down in the oesophagus.

Slis and Damsté describe a glottograph (1o) with fibre optics inserted through the nose. They are able to move the flat end of the optics by means of a nylon thread which goes through the protecting tube and through a hole in the tube outside of the nose.

Sawashima, Hirose, and Fujimura have made some experiments with photography of the glottis by means of a flexible fibre optics cable inserted through the nasal passage, too (11). Camera and light source are looking at the glottis through the same fibre optics cable containing both light guide and image guide. The optical cable is pushed in position behind the epiglottis. - It is now possible to connect a fibre optics cable to a video camera for making tape recordings or moving pictures of the glottis (as far as we know this has not yet been done).

Our glottograph.

So far, our experiments with the glottograph are based on the most common principle, i.e. an external light source and a light sensitive cell pushed down through the nasal passage and placed behind the epiglottis in the pharynx.

The light source consists of a 5oo watts projector lamp with blower. The light is led to the skin through a 6o cm long acryl rod with a diameter of 4o mm. This acryl rod makes it more comfortable for the subject because of the greater distance between the subject's face and the hot light house of the projector. Only a few per cent of the light is lost in the transmission rod which is mounted in the lense holder instead of the front lenses (objectives) of the projector.

In order to avoid the 5o cps ripple of the light, which will interfere very strongly with the oscillations from the vocal cords, the current for the projector lamp is rectified and smoothed. After a normal bridge rectifier and an LC smoothing filter about 2oo volts are left for the lamp, i.e. the lamp consumes an effect of about 4oo watts.

Our mains supply of 22o volts AC is not very stable. When the mains supply varies, it causes variations in the light level. The supply varies over a range of up to 25 volts. We intend, therefore, to stabilize the mains supply with a 22o/22o volts stabilizer, and connect the bridge rectifier to that stabilized voltage. However, this had not been done when we made the illustrations for this paper, as we had not got the stabilizer yet.

In the first experiments, which are described in this article, the light sensitive transducer in the pharynx was a photo transistor (Texas Instruments type LS4oo) (9). This photo transistor is very small (diameter 2 mm) and has a shape which makes it very well suited for mounting in a polyethylene tube with an internal diameter of 2 mm and an external diameter of 3 mm. Thus mounted in the end of the small tube it goes quite easily down into the pharynx. If the subject has very sensitive mucous membranes and strong reflexes, it can be necessary to give a little amount of local anaesthesia, but this is not normally needed (it has not been necessary to do so for making any of the illustrations in this paper). The distance from the nostrils to the lense of the photo transistor is for a normal subject 18o cm tall between 13 and 16 cm (9). Therefore we have placed a little mark on the polyethylene tube at a distance of 13 cm from the photo transistor. The photo transistor must be fixed by means of glue in the tube in order to avoid that saliva percolates into the tube thereby causing partial short cut between the terminals.

The photo transistor is connected through a shielded cable to an amplifier, which is mounted near the subject's face on the acryl rod extending from the projector. In this way the distance between the light sensitive cell and the amplifier is the smallest possible, which reduces the possibility of picking up hum (and our local broacast station) which would disturb the glottogram curves.

The amplifier itself is based on an integrated circuit, "Fairchild" type /uA 7o2 C. This integrated circuit gives sufficient amplification for driving an oscilloscope or an ink writer (mingograph).

However, we have supplied the photo transistor amplifier with an output stage 2N 3o54 (RCA). In this way the amplifier delivers an output power sufficient for driving nearly all the different recorders on the market.

When we have got some experience through the practical work with the glottograph, we shall be able to determine whether or not some filtering of the signal is necessary. Possibly an HP-filter could stabilize the zero line when making analyses of the larynx oscillation.

Technical specifications of the glottograph.

If the photo transistor is placed in total darkness the input voltage of the amplifier is only 0.7 μV. If the photo transistor is placed in position over a closed glottis which is illuminated from below, the input voltage is 3.7 μV. A fully open glottis causes an input voltage of 6 mV, whereas the input level only varies up to 3 mV under normal conditions of phonation (subject BFJ).

The maximum output from the amplifier is 3 volts/500 mA which is obtained with an input voltage of 1o mV (input impedance 2000 ohms). During this measurement the output load resistance has been 6 ohms, which is the minimum permissible load resistance.

In the normal working range: 0 - 3 mV input voltage which corresponds to 0 - 1.6 volts/0-200 mA output, the amplification is linear within $\pm$ 2 per cent. The power gain is 81 dB (voltage gain 56 dB and current gain 1o6 dB).

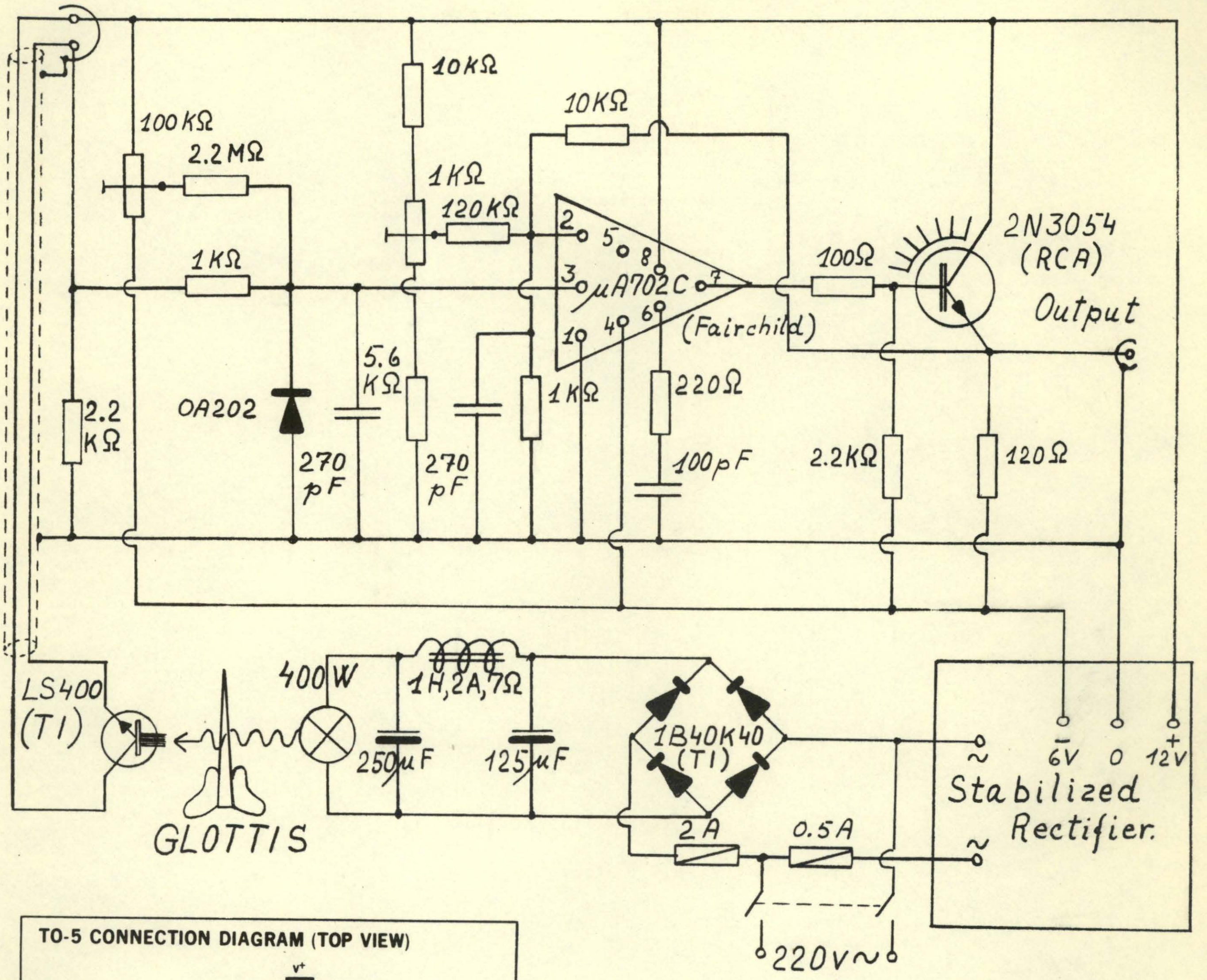
The frequency range has been tested by means of a Philips stroboscope type 91o3. The light pulses and the output pulses of the glottograph have been measured with a Tektronix scope type 5o2A. Calculations show that the photo transistor with amplifier has a rise time of 10 μs and a fall time better than 50 μs, i.e. the glottograph has a frequency range from DC up to at least 20,000 cps.

Picking up the light from the glottis:

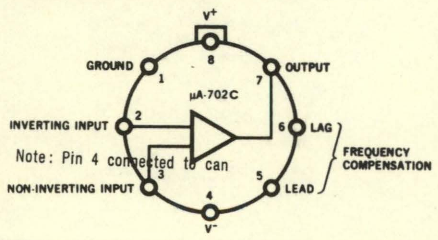
As yet, our experience with the glottograph is limited to a few subjects. Only some of the subjects (all phoneticians at the Institute) were able to fix the photo transducer in such a way that glottograms were produced on the oscilloscope screen. The subjects accomplished this by looking at the screen while moving the transducer up and down with open glottis until a maximum deflection was obtained on the screen.

It is, however, difficult to fix this position of the transducer for most subjects. The articulatory movements disturb it, especially movements of the soft palate and movements of the epiglottis. We have tried to fix the transducer by mounting it midway

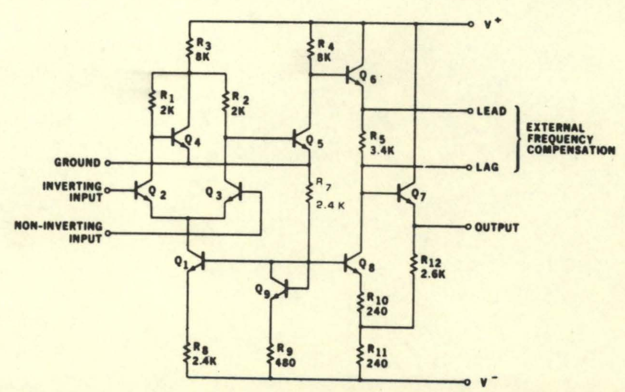
AMPLIFIER FOR GLOTTOGRAPH



TO-5 CONNECTION DIAGRAM (TOP VIEW)



SCHEMATIC DIAGRAM OF THE μA-702C



in the polyethylene tube, which then was swallowed down into the oesophagus (9). However, after trying myself for several days to produce glottograms in that way I made sure that this was not the best method. The photo transducer was placed too far back in the pharynx.

Now we want to test different types and shapes of photo transistors and photo diodes. The type LS400 which has hitherto been used has a light sensitive angle of only 15^{o}, which is probably too little when the transducer is moved by the articulations.

After making the illustrations for this paper I have succeeded in removing the convex lense which formed the front part of the photo transistor house. As a result of this it is now much easier to pick up the light, and the position of the photo transistor in the pharynx is not so critical now as before.

Different ways of fixing the transducer and shaping the polyethylene tube, e.g. by fixing a steel spring inside the tube and moving it into position by means of a nylon thread (lo), will be tried too.

Results:

The two main purposes we had in mind when building the glottograph were: (I) Investigations of changes in the voice spectrum at high voice effort, and changes in voice quality in different registers and at different pitch levels in singing. Some investigations of these problems have been started by the present writer (12). (II) Investigations of the position and movements of the vocal cords in Danish stop consonants as part of the comprehensive studies of stop consonants started some years ago by Eli Fischer-Jørgensen (13), (14).

The photos shown on the following pages present some preliminary results. The photos are all made from the storage screen of a Tektronix oscilloscope type 564. The same vertical sensivity has been used for all the photos, whereas different time settings have been used for the horizontal sweep time.

An opening movement of the glottis is recorded as a raising of the glottogram curve. Unfortunately, it is not possible to show a zero line indicating the curve position for the closed glottis. Even if the glottis is closed more or less light may penetrate through the vocal cords and the mucous membrane.

Comparison between low and high voice effort in chest register

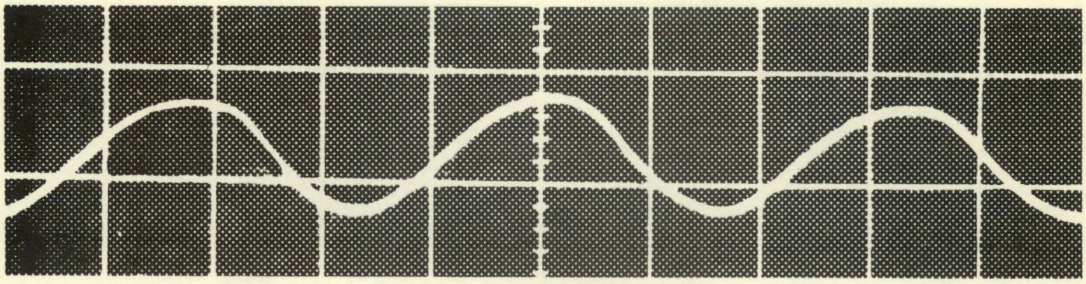


Fig. 1

Glottogram of low voice effort at F_o = 12o cps in chest register. (Vowel [i:] , subject JR).

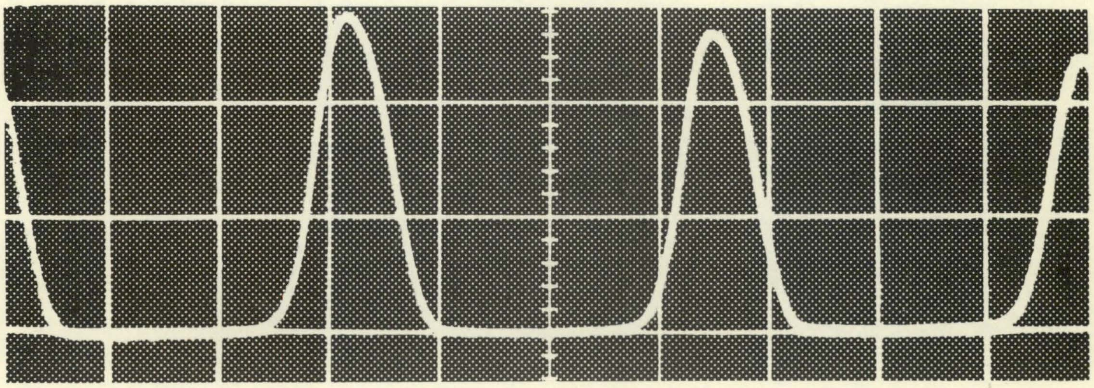


Fig. 2

Glottogram of high voice effort at F_o = 120 cps in chest register. Notice the long closure phase in the vowel spoken with high voice effort. (Vowel [i:] , subject JR).
For further explanations of the difference on the acoustic level between low voice effort (Fig. 1) and high voice effort (Fig. 2), see reference (12).

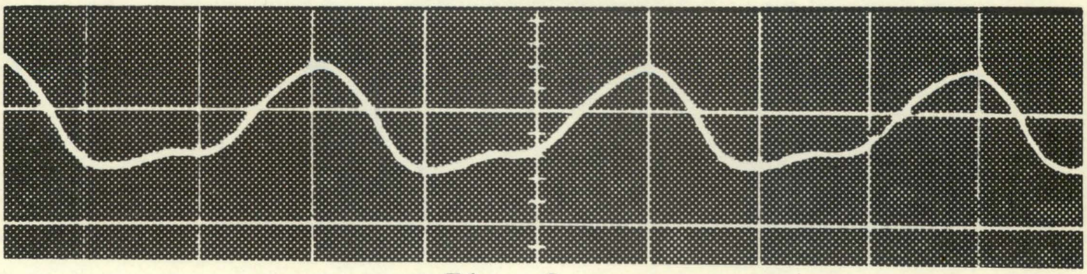


Fig. 3

Glottogram of high voice effort at F_o = 110 cps in chest register.
The subject has deliberately reinforced the 2. harmonic. This physiological waveform with the strong 2. harmonic may be explained by means of the upward travelling wave in the mucous membrane of the vocal cords (15). (Vowel [ɛ:] , subject JR).

Comparison between low and high voice effort in chest register and in head register

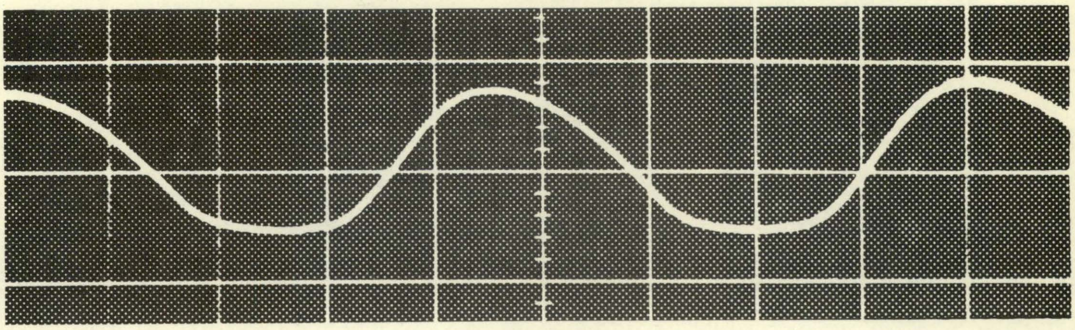


Fig. 4

Glottogram of low voice effort at F_o = 120 cps in chest register. (Vowel [ε:], subject BFJ). Compare Fig. 1, subject JR.

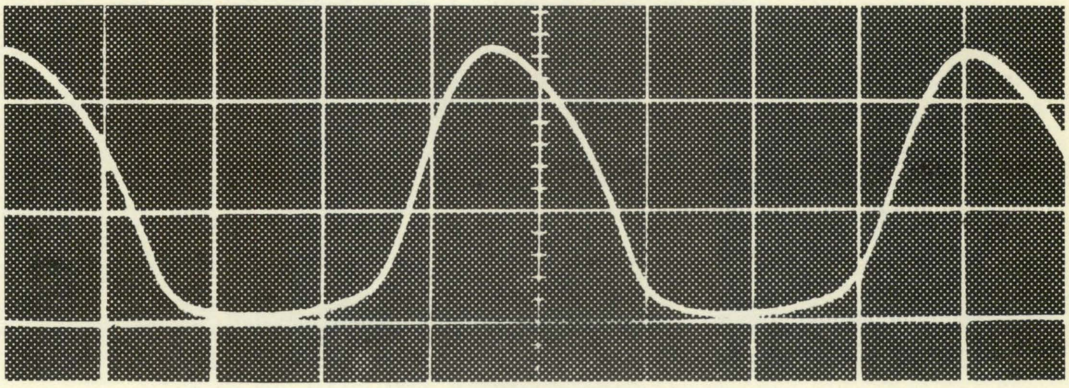


Fig. 5

Glottogram of high voice effort at F_o = 120 cps in chest register. (Vowel [ε:], subject BFJ). Compare Fig. 2, subject JR.

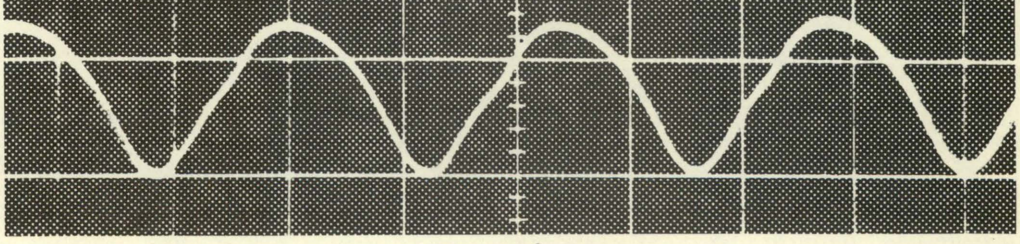


Fig. 6

Glottogram of low voice effort at F_o = 400 cps in head register. Notice that the closure phase, if it is at all present, is very short. (Consonant [m:], subject BFJ).

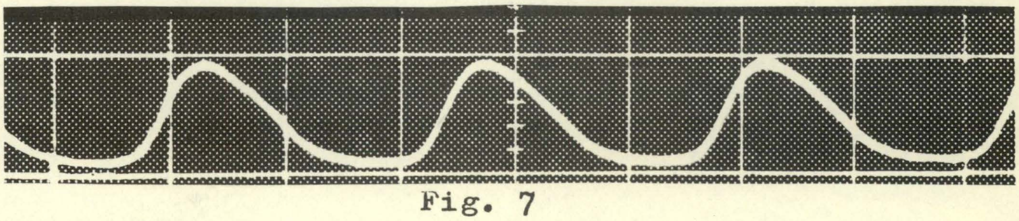


Fig. 7

Glottogram of high voice effort at F_o = 400 cps in head register. Notice that the closure phase is much longer than in Fig. 6, and that the slope is steeper in the opening than in the closing movement. (Consonant [m:], subject BFJ).

Comparison between head register and chest register at the same fundamental frequency F_o = 225 cps

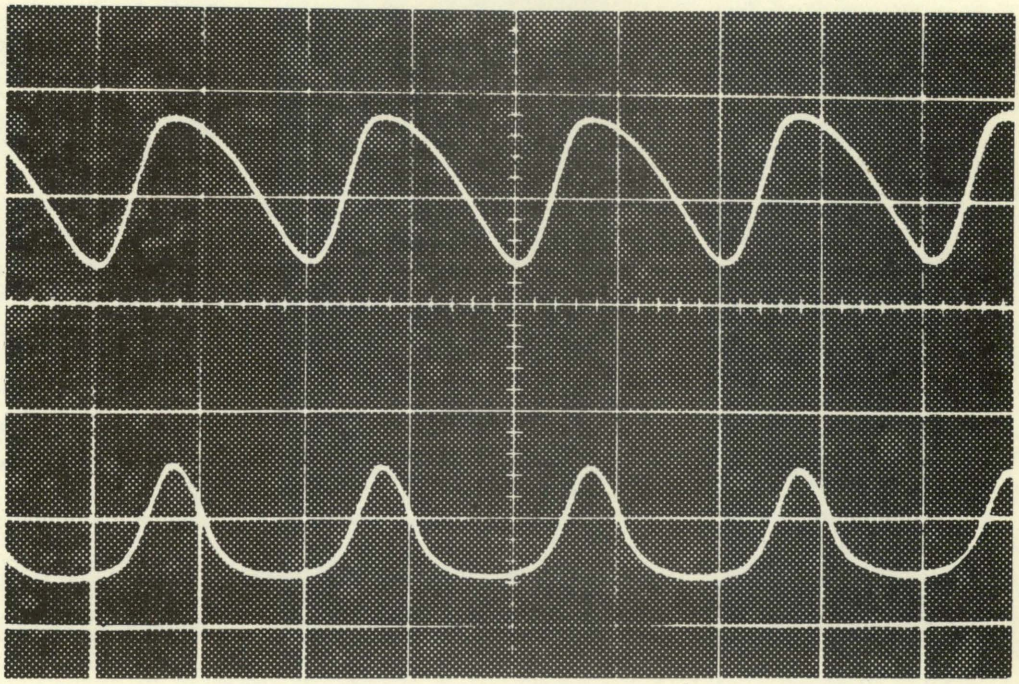


Fig. 8

Both glottograms illustrate medium voice effort. Notice the longer closing phase in the lower curve (chest register) compared to the upper curve (head register). This is a typical difference. (Vowel [i:], subject BFJ).

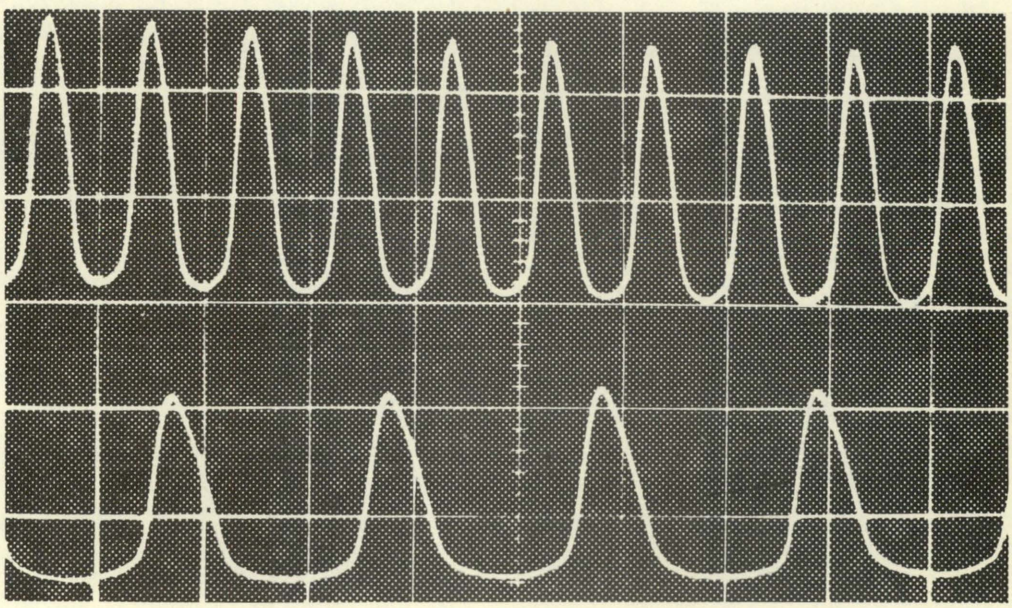


Fig. 9

Upper curve: Fundamental frequency 200 cps.
Lower curve: Fundamental frequency 100 cps.
High voice effort in <u>chest register</u>. (Vowel [ɛ:], song, subject BFJ).

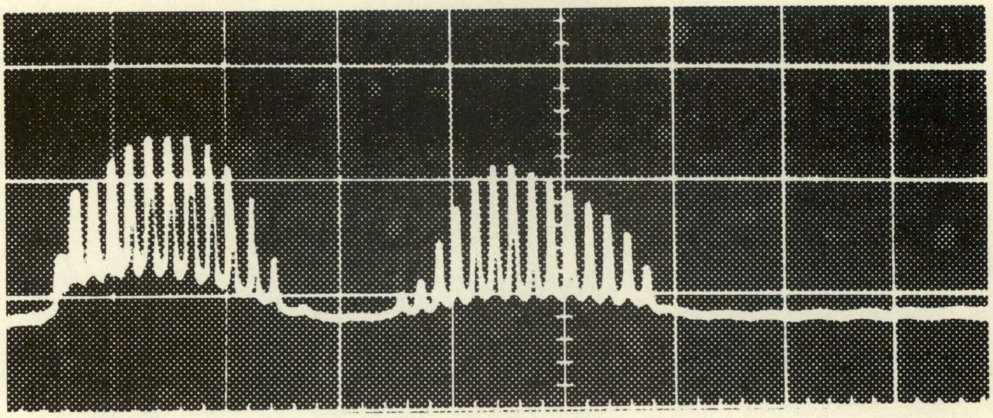


Fig. 10

Glottal attack in the word [ʔaʔa] 'a - a' (baby talk) (subject BFJ).

Compare this glottal attack with the examples of the Danish "stød" in Figs. 12 and 13.

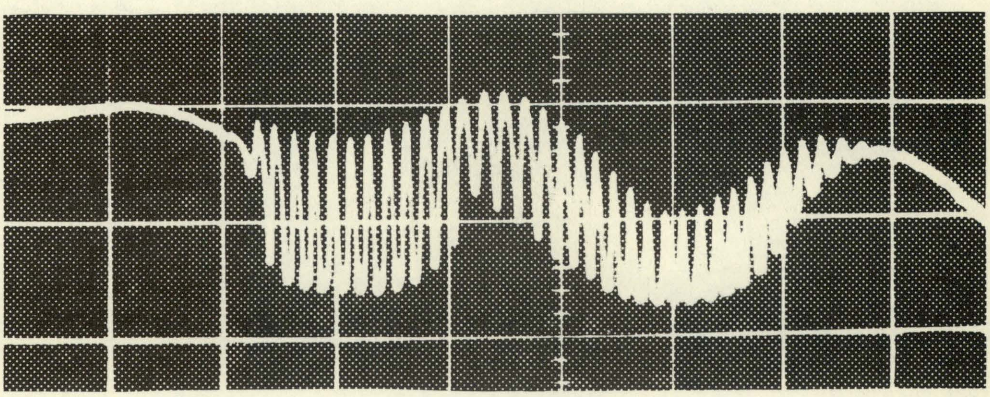


Fig. 11

Unvoiced and voiced [h] in the word [haɦah] , 'ha-ha' (expression of surprise) (subject BFJ).

Notice in these two glottograms (Figs. 10 and 11) the difference between a final vowel ending with a glottal closure and a final vowel ending with an aspiration.

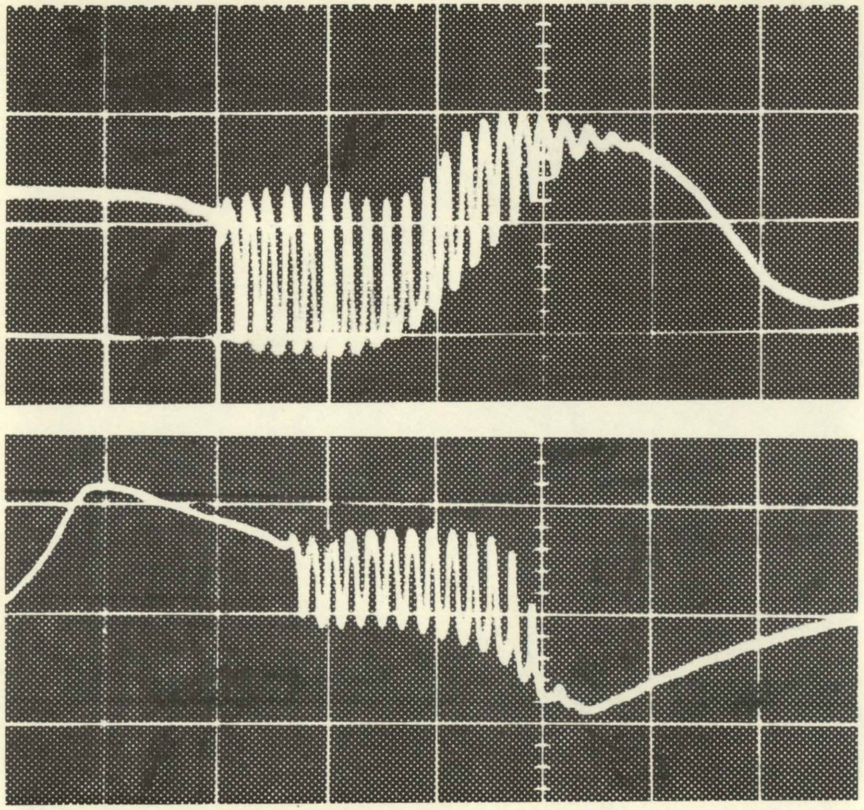


Fig. 12

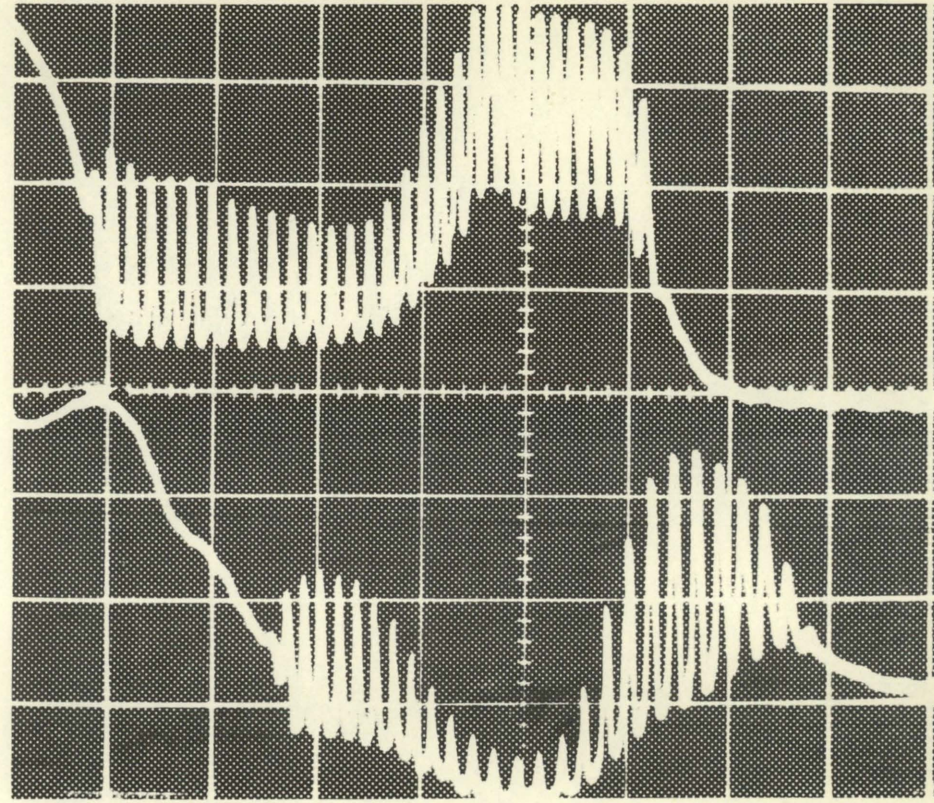


Fig. 13

Danish "Stød"

One-syllable minimal pair and two-syllable minimal pair illustrating the "stød" in Danish.

Upper curve: [hun]	hun	'she'	
Lower curve: [hunʔ]	hund	'dog'	
Upper curve: [hu:ən]	huen	'the cap'	
Lower curve: [huʔən]	huen	'the inclination'	

(The "stød" is not a glottal closure in standard Danish (16), but possibly in the Jutland dialect (17). The lack of a zero line in the glottograms makes it impossible to determine whether or not the glottis is closed in the "stød"-phase. Subject BFJ is talking standard Danish with a slight Jutland dialect substratum.)

The Danish p t k

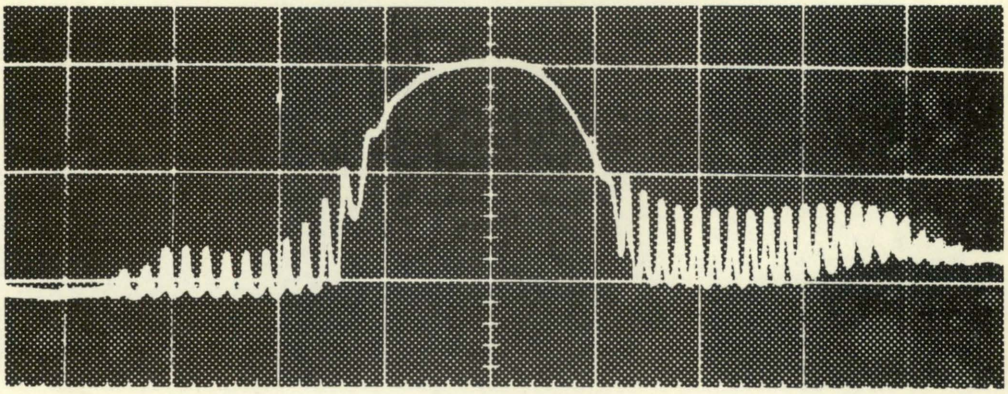


Fig. 14

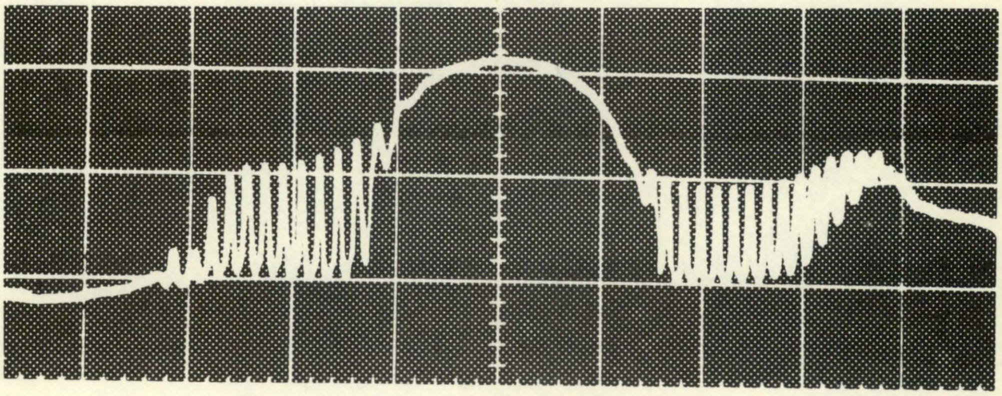


Fig. 15

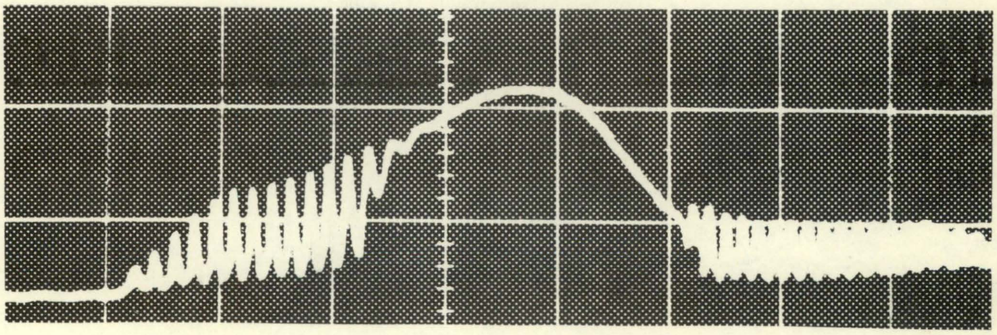


Fig. 16

Glottograms of the Danish tenues ptk in the words [ɛpɛ], [ɛtɛ], and [ɛkɛ] show in all cases a fully open glottis (as observed earlier in an endoscopic examination by Eli Fischer-Jørgensen (14)). The vocal cords do not vibrate during the closure. In our last report (3) we published some curves of the same stops made with the FABRE GLOTTOGRAPH. These curves, too, show an open glottis during the closure phase of the articulation of Danish ptk. - The main difference between the tenues ptk and the mediae bdg in Danish is one of aspiration. Tenues are aspirated and mediae are not. (Subject BFJ.)

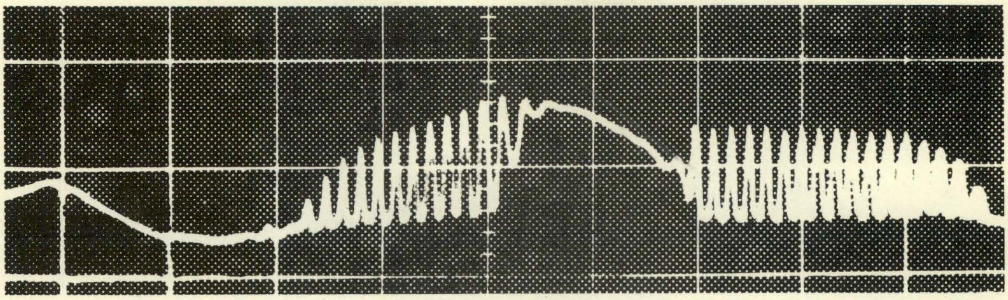


Fig. 17

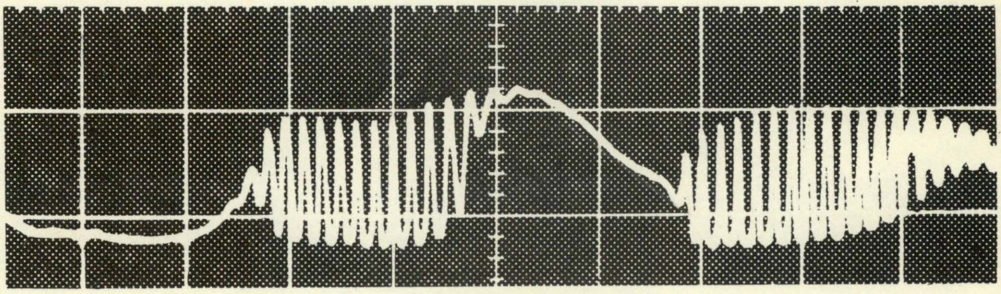


Fig. 18

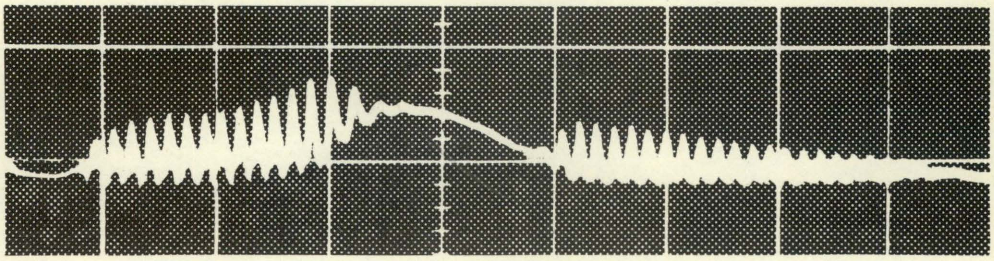


Fig. 19

Fig. 17 [ɛbɛ], Fig. 18 [ɛdɛ], and Fig. 19 [ɛgɛ] may be compared with Figs. 14, 15, and 16. (Subject BFJ.)

At first we observe that the glottis is on the whole more closed in the mediae than in the tenues, which confirms our visual observations in the endoscope. Secondly, the glottis is most open in the beginning of bdg when the air pressure is being built up in the pharyngeal and oral cavities. Then the glottis gradually closes in order to avoid aspiration after the explosion. As the glottis is nearly closed during the explosion the vocal cords will start to vibrate immediately after the explosion.

(This may be confirmed with more subjects by means of simultaneous recordings of glottogram, oscillogram, subglottal air pressure, and pharyngeal air pressure.)

Future tasks.

We hope that it will later be possible to make simultaneous recordings of the subglottal pressure, the pharyngeal pressure, the air flow, and the glottogram. Such combinations of physiological parameters would be extremely useful for investigations of the stop mechanism and of whisper. For research on changes in the voice spectrum at high voice effort and for research on register shift and changes in pitch levels in singing it is also necessary to know how the primary voice spectrum changes. As for the relations between the glottal area curve recorded with the glottograph and the primary voice source, see references (18), (19), and (2o).

Finally we may employ the waveform observed as voice source input for our synthesizer when imitating a special voice quality.

Acknowledgement:

Our physiological investigations in 1967 as well as the development of the glottograph have been carried out by means of grants from "Statens almindelige videnskabsfond" (The Danish State Research Foundation).

References:

(1) B. Sonesson, "On the Anatomy and Vibratory Pattern of the Human Vocal Folds. With Special Reference to a Photo-Electrical Method for Studying the Vibratory Movements". Acta Oto-Laryngologica Suppl. 156 (196o), Lund, Sweden.

(2) G. Fant and B. Sonesson, "Indirect Studies of Glottal Cycles by Synchronous Inverse Filtering and Photo-Electrical Glottography". QPSR 4 (1962), Speech Transmission Laboratory, Stockholm, Sweden.

(3) Ph. Fabre, "Glottographie respiratoire", Comptes rendus des séances de l'Académie des Sciences No. 6 (1961), pp. 1386-88; id., "La Glottographie en haute fréquence, particularités de l'appareillage", Comptes rendus de la Société de Biologie CLIII,7 (1959), pp. 1361-64.

(4) R. Husson, "The Fabre Glottograph", Physiologie de la phonation (1962).

(5) G. Fant, J. Ondráčková, J. Lindqvist, and B. Sonesson, "Electrical Glottography", QPSR 4 (1966), Speech Transmission Laboratory, Stockholm, Sweden.

(6) Eli Fischer-Jørgensen, Børge Frøkjær-Jensen, and Jørgen Rischel, "Preliminary Experiments with the Fabre Glottograph", ARIPUC 1 (1966), pp. 22-3o.

(7) A. S. Abramson, L. Lisker, F. S. Cooper, "Laryngeal Activity in Stop Consonants", Status Report on Speech Research, SR-4 (1965), Haskins Laboratories, New York.

(8) André Malécot and Kenneth Peebles, "An Optical Device for Recording Glottal Adduction-Abduction during Normal Speech", _Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung_, Bd. 18, No. 6 (1965), pp.545-55o.

(9) John Ohala, "A New Photo-Electric Glottograph", _Working Papers in Phonetics_ 4 (1966), pp. 44 ff., University of California, Los Angeles.

(1o) I. H. Slis and P. H. Damsté, "Transillumination of the Glottis during Voiced and Voiceless Consonants", _I.P.O. Annual Progress_ Report No. 2 (1967), Eindhoven, Holland.

(11) M. Sawashima, H. Hirose, and O. Fujimura, "Observations of the Larynx by a Fiberscope Inserted through the Nose", _Paper GG-5_, 74th Meeting of the Acoustical Society of America (1967).

(12) B. Frøkjær-Jensen, "Changes in Formant Frequencies and Formant Levels at High Voice Effort", _ARIPUC_ 1 (1966), pp. 47-54.

(13) Eli Fischer-Jørgensen, "Acoustic Analysis of Stop Consonants", _ARIPUC_ 1 (1966), pp. 31-34.

(14) Eli Fischer-Jørgensen, "Phonetic Analysis of Danish Stop Consonants", _Miscellanea Phonetica_ II (1954), pp. 42-59.

(15) Svend Smith, "Chest Register versus Head Register in the Membrane Cushion Model of the Vocal Cords", _Folia Phoniatrica_ 9, 32 (1957), pp. 32-36.

(16) Svend Smith, _Stødet i Dansk Rigssprog_ (1944), English Summary.

(17) K. Ringgaard, "The Pronunciation of a Glottal Stop", _Phonetica_ 8 (1962), pp. 2o3-2o8.

(18) R. L. Miller, "Nature of the Vocal Cord Wave", _The Journal of the Acoustic Society of America_, Vol. 31, No. 6 (1959), pp. 667-677.

(19) J. L. Flanagan, "The Source for Voiced Sounds", _Speech Analysis, Synthesis, and Perception_ (Kommunikation und Kybernetik in Einzeldarstellungen, Bd. 3 (1965), pp. 43-44.

(2o) J. L. Flanagan, "Some Properties of the Glottal Sound Source", _Journal of Speech and Hearing Research_, No. 1 (1958), pp. 99 - 116.

THE HETERODYNE FILTER OF THE INSTITUTE OF PHONETICS

Jørgen Rischel and Svend-Erik Lystlund

In the 1966 report of the Institute a brief mention was made of a heterodyne bandpass filter under work (1). The filter has now by and large reached the stage of completion aimed at in the said report, and it has been taken in use in our research (see this volume, pp. 35 ff.). On account of the rather special nature of the device it seems reasonable to give a more detailed description of it here.

A great variety of electronic wave filters are used in phonetic research, serving a number of different purposes: frequency selective analysis of the speech-wave (the sonagraph, etc.), pre-shaping of the signal applied to intensity meters or pitch extracting devices, synthesis of speech, amplitude distortion of natural speech (for perceptual tests), and so on. Obviously, the specifications to be met by a filter vary according to the purpose, and the concept of a "standard filter" has limited validity. For purposes involving rather unselective filtering (pre-emphasis) it is generally a relatively simple task to construct individual networks meeting the individual demands. However, when very sharp filtering is required as it may be the case in experiments with distorted speech or in frequency analysis of the speech-wave, construction of the devices becomes quite complicated, and it may be most expedient to have access to a versatile unit which can perform a sharp filtering (highpass or lowpass or bandpass) at any frequency desired. The heterodyne filter is intended as such a standard device for sharp filtering.

It is somewhat difficult to argue in a general way about the merits or drawbacks of a wave filter, since ideal filters do not exist. If a filter is entirely adequate from one point of view it is likely to be insatisfactory from another point of view (although the drawbacks may escape the attention of the user). In most cases one aspect of filter performance (e.g., the "sharpness" of cutoff) is so essential that the phonetician may choose to pay much less attention to other aspects (e.g. phase or amplitude distortion in the passband), but it goes without saying that the filter cannot be characterized in terms of one such property.

This is particularly true of a complicated system like the heterodyne filter.

In the following survey the first section explains the general principle of the heterodyne system, and the second section describes our actual setup. In the last two sections an attempt is made to characterize the performance of the filter from different aspects and to motivate various details of the design.

1. Filtering according to the heterodyne principle.

Fant has given a detailed description (2) of a filter employing the heterodyne principle. However, the strategy used in our setup differs somewhat from his, so it is necessary to go into some detail also in this report.

The motivation for choosing the heterodyne principle is that it makes it possible to combine a very sharp cutoff with a free choice of cutoff frequencies. The idea is that the filter is designed to work at one fixed frequency (which enormously facilitates the design of the filter itself): instead of moving the cutoff frequencies of the filter up and down one moves the signal up and down in frequency so that different parts of the spectrum are allowed to pass through the fixed filter. Bandpass filtering with mutually independent highpass and lowpass functions requires that the procedure of signal frequency shift be performed twice: first the signal is shifted to one frequency location and applied to a filter performing the highpass filtering, and afterwards it is shifted to another frequency location and applied to a filter performing the lowpass filtering. Finally the signal is shifted back to its original frequency location.

Each frequency shift is performed by heterodyning the audio signal with a sinewave from a variable oscillator, the cutoff frequencies of the filtering processes being controlled directly by the settings of the oscillators. The heterodyning of the speechwave is performed in each step by means of a double-balanced modulator. If the frequency limits of the audio signal are denoted by f_{min} (the lowest frequency of interest) and f_{max} (the highest frequency of interest), and the (variable) frequency of the oscillator is denoted by f_{osc}, the output of such a modulator is essentially composed of two frequency bands: the upper sideband reaching

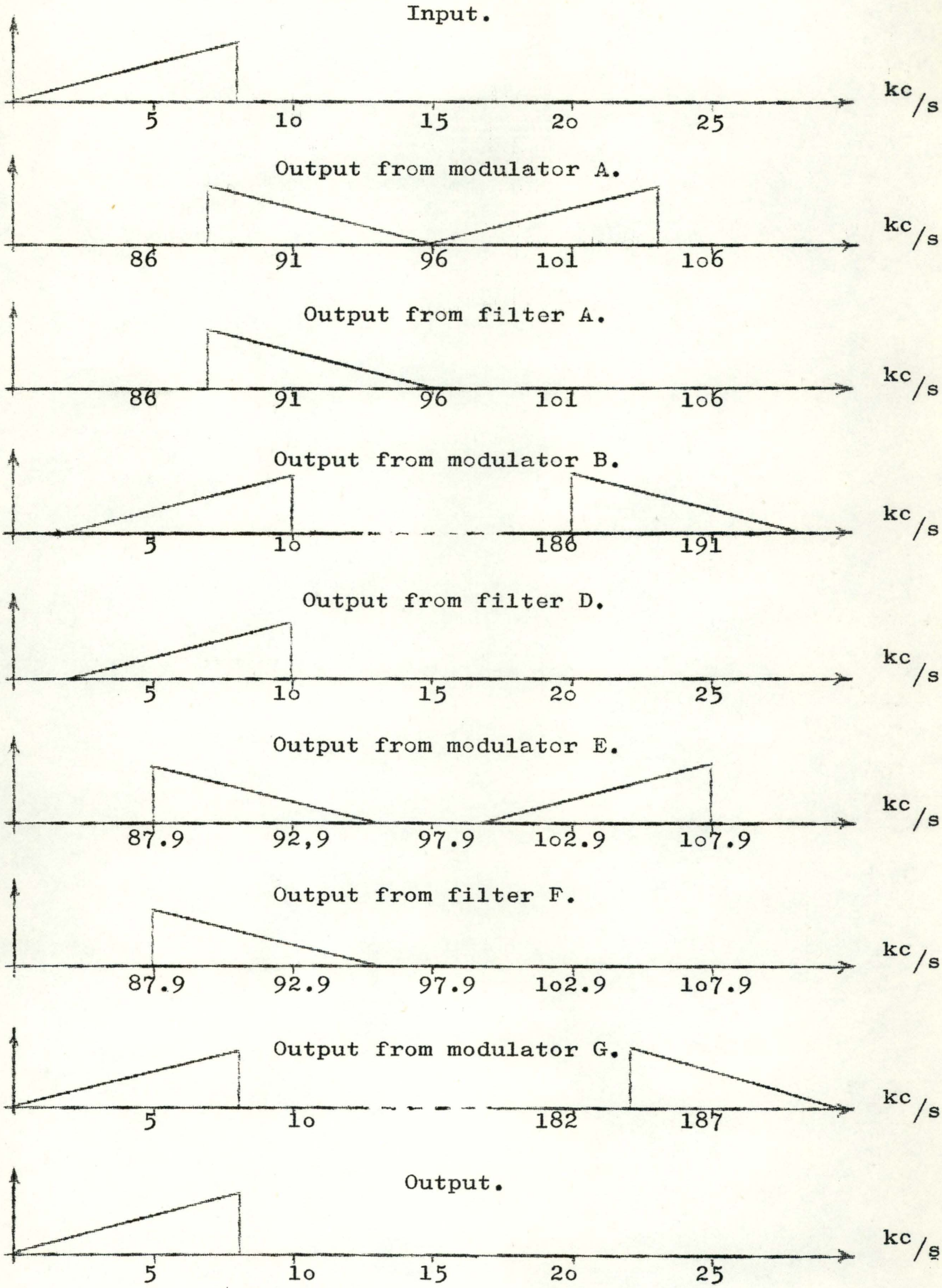


Fig. 1a. HETERODYNE SIGNAL PROCESSING
(cp. Block diagram Fig. 2)
Full frequency range

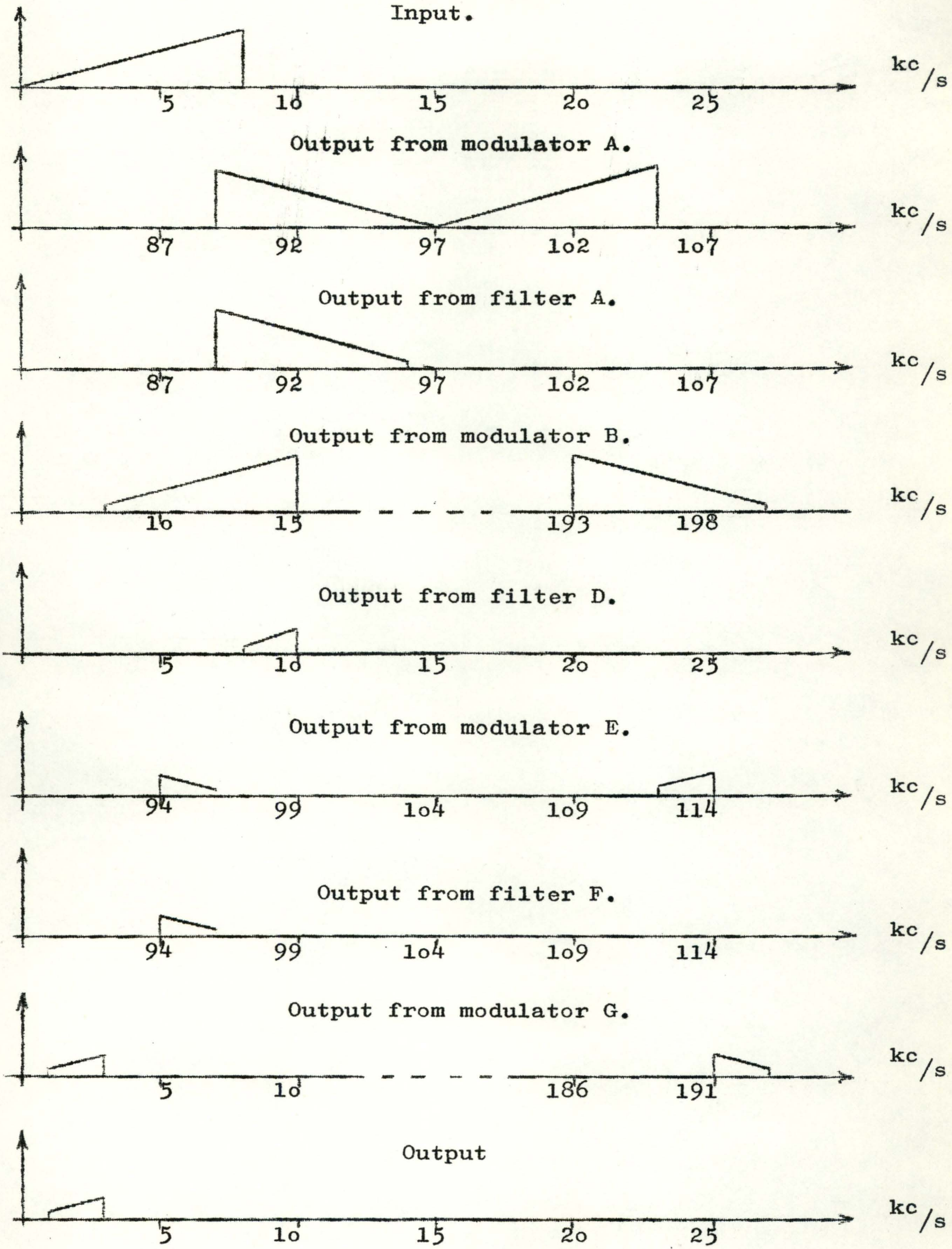


Fig. 1b. HETERODYNE SIGNAL PROCESSING
(cp. Block diagram Fig. 2)
Bandpass 1 kcps - 3 kcps

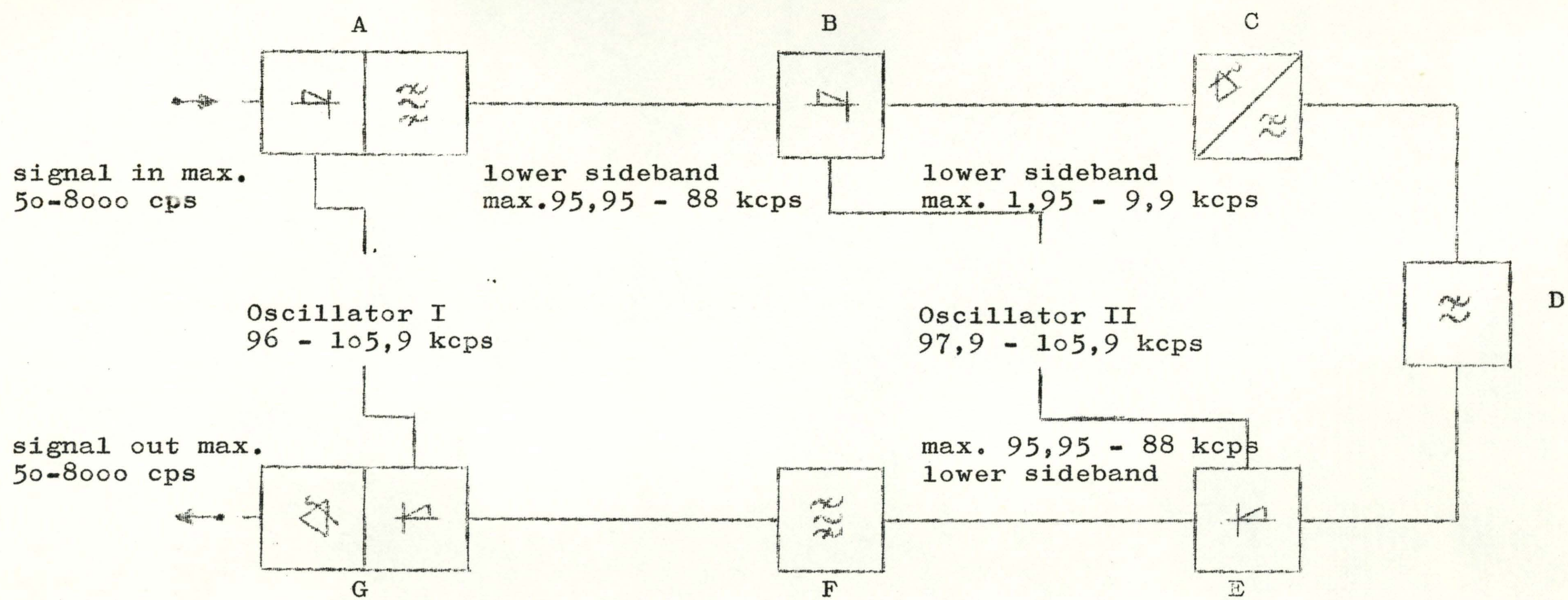
from $f_{osc}+f_{min}$ to $f_{osc}+f_{max}$, and the lower sideband reaching downwards from $f_{osc}-f_{min}$ to $f_{osc}-f_{max}$. Since the entire signal information is contained in each sideband, it is possible to filter out either the upper or the lower sideband and to transmit that one alone through the system, the original signal being restored at the output. In our setup the lower sideband (oscillator frequency minus signal frequency) is used throughout. This entails that at some points in the system the frequency spectrum is inverted, i.e. turned upside down ($f_{osc}-f_{min}$ being a higher frequency than $f_{osc}-f_{max}$). One consequence of this is that the highpass filtering of the signal is in fact performed by means of a lowpass filter.

2. The actual setup.

Fig. 2 gives a block diagram of our heterodyne filter in its present form.

The idea governing this project was to use standard components as building blocks to the greatest possible extent. As reported previously, the Danish Post and Telegraph Department has generously put a good deal of equipment for carrier telephony and radio telephony at our disposal. This equipment included a number of modulators and fixed bandpass filters. Various arrangements of these have been tried out, but since the filtering process implies that the carrier frequency oscillators are tuned off the frequencies for which the units are designed, it has proved difficult to obtain a satisfactory performance at all cutoff frequencies. Whistle interference among the many frequencies generated in the modulators is a very disturbing phenomenon likely to occur unless the utmost care is taken in the choice of frequency bands.

The solution we prefer at present is to perform the highpass filtering (which is physically performed as a lowpass filtering) at a high frequency, and to perform the lowpass filtering at a low frequency. For the former purpose a very sharp filter with a cutoff frequency of 95,95 kcps (block F in Fig. 2) is at our disposal. For the latter purpose we employ a filter (block C-D in Fig. 2) with a sharp cutoff at 9,8 kcps (which of course dictates the upper limit of the entire frequency range of the signal). This filter was designed and built by us at the Institute, whereas all



A Siemens modulator "1" with filter.
Rel 14R 168 a1

B Siemens modulator "2"
Rel 15C 710 b

C Equalizing network for filter

D 9,9 kcps lowpass filter

E Siemens modulator "3"
Rel 15C 706b

F Siemens filter
Rel str 14R 181 a dt/en

G Siemens modulator "4" with 10 kcps
lowpass filter
Rel 14R 174 a1

Fig. 2. HETERODYNE FILTER

other sections (including the 95,95 kcps filter) are commercial equipment.

The maximum **usable** bandwidth of the system is at present 8 kcps. (We intend to replace modulators 2 and 3 by a different type of modulator, which will extend the **usable** frequency range, cp. p. 32.) It is possible, for example, to transmit the frequency range 5o cps - 8 kcps and to suppress frequencies outside this range, cp. Fig. 1a. The carrier frequency of the first modulator is then chosen to be 96 kcps. When audio frequencies are applied to the modulator and converted to sidebands around the carrier the lower of these sidebands (which is inverted) will pass through the 95,95 kcps lowpass filter except for the 5o cps range closest to the carrier frequency, i.e. the sideband will be limited in a way corresponding to a 5o cps highpass filtering of the original audio signal (5o cps being changed to 96-o,o5 = 95,95 kcps, 8 kcps being changed to 96-8 = 88 kcps, and so on). The upper sideband is eliminated. If now the carrier frequency of the second modulator is chosen to be 97,8 kcps, the sideband extending from 95,95 to 88 kcps is moved back to a low frequency location, and turned upside down again. The 9,8 kcps lowpass filter situated after the second modulator limits the sideband in a way corresponding to an 8 kcps lowpass filtering of the original audio signal (95,95 kcps being changed to 97,8o-95,95 = 1,85 kcps, 88 kcps being changed to 97,8o - 88 = 9,8o kcps, and so on). Modulators 3 and 4 reverse this procedure, so that we end up with the original audio signal, only limited to the range 5o cps - 8 kcps. A lowpass filter after modulator 3 (block F in the diagram, Fig. 2) cuts at 95,95 kcps (like the one after modulator 1) and removes the unwanted sideband and distortion products at this stage, and a lowpass filter in the output amplifier ensures that only audio frequency components reach the output. It is seen that at least four filters are needed in the heterodyne system.

If the carrier frequency applied to modulators 1 and 4 ($f_{osc\ 1}$) is raised, obviously a greater portion of the audio spectrum is cut off, since a greater portion of the lower sideband generated in the modulator will be raised above the cutoff frequency of the 95,95 kcps lowpass filter. This corresponds to highpass filtering with a higher cutoff frequency. E.g., a carrier **frequency** of 96,95 kcps conditions an effective highpass filtering at 96,95-95,95 = 1 kcps, and a carrier frequency of 1oo,95 kcps conditions a highpass filtering at 1oo,95-95,95 = 5 kcps.

If the carrier frequency applied to modulators 2 and 3 ($f_{osc\ 2}$) is raised, it is also clear that an increasing portion of the audio spectrum is cut off, since a greater portion of the lower sideband generated in this modulator will be raised above the cutoff frequency of the 9,8 kcps lowpass filter. This corresponds to lowpass filtering with a lower cutoff frequency. E.g., provided that $f_{osc\ 1}$ is kept at 96 kcps (lower sideband 95,95 to 88 kcps), and $f_{osc\ 2}$ is raised to 1oo,8 kcps, we get an effective lowpass filtering at 5 kcps, since an audio signal of 5 kcps is first moved to 96-5 = 91 kcps, and afterwards to 1oo,8-91 = 9,8 kcps so that it just passes through the 9,8 kcps lowpass filter, whereas higher frequencies would be suppressed.

It is, of course, not the absolute frequency of $f_{osc\ 2}$, but the increment of this frequency above $f_{osc\ 1}$ that determinesthe lowpass cutoff. Thus with the carrier of the first modulator set at 1oo,95 kcps (highpass 5 kcps) the carrier of the second modulator must be raised above 1oo,95+5 = 1o5,95 kcps in order to get any signal through the system, 1o5,95 kcps being in this case the proper setting for lowpass cutoff at 5 kcps.

It is seen that the setting of the oscillators demands a bit of calculation work, although an exact adjustment of the carrier frequencies can be carried out more easily by measuring the response of the system while turning the knobs of the oscillators. It is a definite drawback of the system that it is not possible to calibrate the oscillator dials in true cutoff frequency, this making the heterodyne filter less expedient for purposes requiring a fast change of cutoff freguencies. In return, the heterodyne filter offers the possibility of continuously varying the cutoff frequencies and to place the frequency cuts at exactly the frequencies desired, if only the user takes the time to adjust it properly.

3. Performance of the heterodyne filter: frequency response.

Filters are generally characterized as devices that pass certain frequencies but suppress others. It is, however, to some extent a matter of definition whether a particular frequency is "passed" or "suppressed" by the filter. Nevertheless, in order to obtain a condensed characterization of the filter response it is often necessary to distinguish categorically between frequencies that are passed and frequencies that are not passed.

With filters characterized by monotonically increasing attenuation beyond a certain frequency it is customary to define the passband as the frequency range in which the attenuation exhibits values between 0 and 3 dB, i.e., the cutoff frequency equals the "-3 dB point". The performance in the "attenuation band" may then be specified in either two ways: (a) in terms of rate of increasing attenuation (expressed in decibels per octave) or (b) in terms of an absolute (minimum) value of attenuation required to suppress unwanted frequencies: "A_{min}". With the latter specification the attenuation band is split up into two: a transition region (immediately adjacent to the passband) with attenuation values between 3 dB and "A_{min}", and a stopband with attenuation values exceeding "A_{min}". When very sharp filtering is required the latter specification is definitely more meaningful. It might be desired, for example, to remove F1 of a vowel while keeping F2 and higher formants intact in order to study the effect of this removal on perception. It might then be estimated that all frequencies within the effective range of F1 should be attenuated at least 40 dB (or possibly more), whereas the envelope of F2 should be minimally distorted, i.e. the entire transition region between cutoff and A_{min} (= 40 dB attenuation) should fall in the valley between the formants.

Sharp filtering obviously requires that the transition region between cutoff and the A_{min} point be as narrow as possible. In filtering of speech it is often reasonable to employ a filter whose sharpness of cutoff is proportional to the cutoff frequency (A_{min} being reached, for example, at one third of an octave from the - 3 dB point), and this will normally be true of stepwise variable filters (including the LC highpass filter of the Institute of Phonetics). The heterodyne filter , however, operates with fixed filters, i.e. the sharpness of cutoff as expressed in cps is constant irrespective of the effective cutoff frequency (determined by the carrier oscillator). This means that the fixed filters must be extremely frequency selective in order that the cutoff be sufficiently sharp at low frequency settings. For our purpose it is sometimes interesting to "cut" between two harmonics of a vowel sound (particularly in the low frequency end of the spectrum), i.e. the transition region between passband and stopband should not - at least for the highpass section - be more than some 100 cps wide.

Another important consideration is the required minimum attenuation in the stop band: A_{min}. For purposes involving perceptual tests a difference in level between the effective passband and the effective stopband of some 40-50 dB is often necessary. Obviously, if a very high value of A_{min} is required, it is more difficult to obtain a narrow region between the passband and the stopband.

In the heterodyne system the situation is particularly complicated, since the filters perform a double function: (a) to highpass and/or lowpass filter the audio signal, (b) to remove unwanted frequencies generated in the modulators, i.e. to eliminate nonlinear distortion. The latter frequencies are partly situated rather far from the useful frequency bands, but on the other hand they must be very effectively suppressed. The filters employed are well suited for this double purpose, since in addition to a sharp cutoff they exhibit a rising attenuation throughout part of the stopband, the final attenuation exceeding 60 dB. For the user, however, the behaviour near cutoff is most interesting. Therefore, in the specifications given below, we have arbitrarily defined A_{min} as 40 dB attenuation, and described the sharpness of cutoff with reference to this value of stopband attenuation.

Finally, the behaviour of the filters in their passbands may deserve special attention. Above, the passband was defined as the frequency band with attenuation between 0 and 3 dB. However, the filter system exhibits a certain amplitude ripple in the passband, and with this type of behaviour the above measure does not characterize the passband response well. It goes without saying that an amplitude ripple of the magnitude of 3 dB throughout the passband is not desirable, although it would not violate the definition of passband. It is therefore essential to specify also the permissible ripple or amplitude fluctuation in the passband, and accordingly to specify a "linear portion" of the passband, in which the response should keep within predescribed limits. There is no general standard as to the permissible ripple. Considering the irregularities found with other pieces of equipment used in acoustic-phonetic studies (in particular, microphones and loudspeakers), a ripple of 2 dB or even more might perhaps be tolerated. However, we have strived to reduce the ripple to one dB, i.e. to keep the amplitude response in the passband within $\pm$ 0,5 dB of the average value. Accordingly, if the maximum value is denoted "0 dB", we define the cutoff frequency of the linear portion of the passband as the point at which the attenuation exceeds 1 dB (cp. reference (3)). By designing the filter in such a way that the -1 dB point is quite close in frequency to the -3dB point, a nearly ideal response is obtained: with a strictly linear passband and a stopband, separated by a very narrow region of transition.

By introducing the -1 dB point as an additional measure of cutoff we get a filter response that is advantageous when filtered speech is used for tests on perception, since for such a purpose a faithful reproduction of frequencies near cutoff is just as interesting as the reproduction of frequencies further off in the passband. Filters with a slowly increasing attenuation from 0 to 3 dB in the passband introduce an error which may not be entirely negligible.

Fig. 3 shows the frequency response of the whole heterodyne system with some different choices of cutoff frequencies. It is seen that the passband as well as the cutoff slopes have a very regular form, except that in cases of narrowband filtering the two cutoff slopes are not quite equal, the highpass section having a considerably sharper cutoff than the lowpass section. On the highpass side the -3 dB point is reached some 15 cps from the -1 dB point, and the -40 dB point is reached some 40 cps from the -3 dB point. On the lowpass side the -3 dB point is reached some 20 cps from the -1 dB point, and the -40 dB point is reached only some 240 cps from the -3 dB point. This difference is in many cases of no consequence, since the highpass filtering is most likely to be used at rather low frequencies, and the lowpass filtering at somewhat higher frequencies, so that the relative sharpness of filtering will often be approximately the same.

The reason for the difference between the two filter sections is that we have built the lowpass section with ordinary coils and condensers and altogether without the use of very advanced (and expensive) techniques, whereas - as stated above - the highpass section includes a commercial high quality crystal filter designed for single sideband transmission. In order to be able to compete with this filter we designed the lowpass filter section for the lowest cutoff frequency compatible with the desire to obtain a sufficient maximum bandwidth, and this is the major reason for modulating down to a low frequency range in modulator "2" (see Fig. 1). Since, everything else being equal, the "sharpness" of a filter will be proportional to the cutoff frequency, the absolute distance in cps between passband and stopband can be reduced by designing the filter for a low cutoff frequency. However, even in the 10 kcps range it is quite a problem to design a sufficiently sharp filter. A distance of 260 cps between the -1 and-40 dB points entails that A_{min} be obtained at a frequency that is only approximately 1.026 times the cutoff frequency (defined as the -1 dB point). We have obtained this only in a roundabout way.

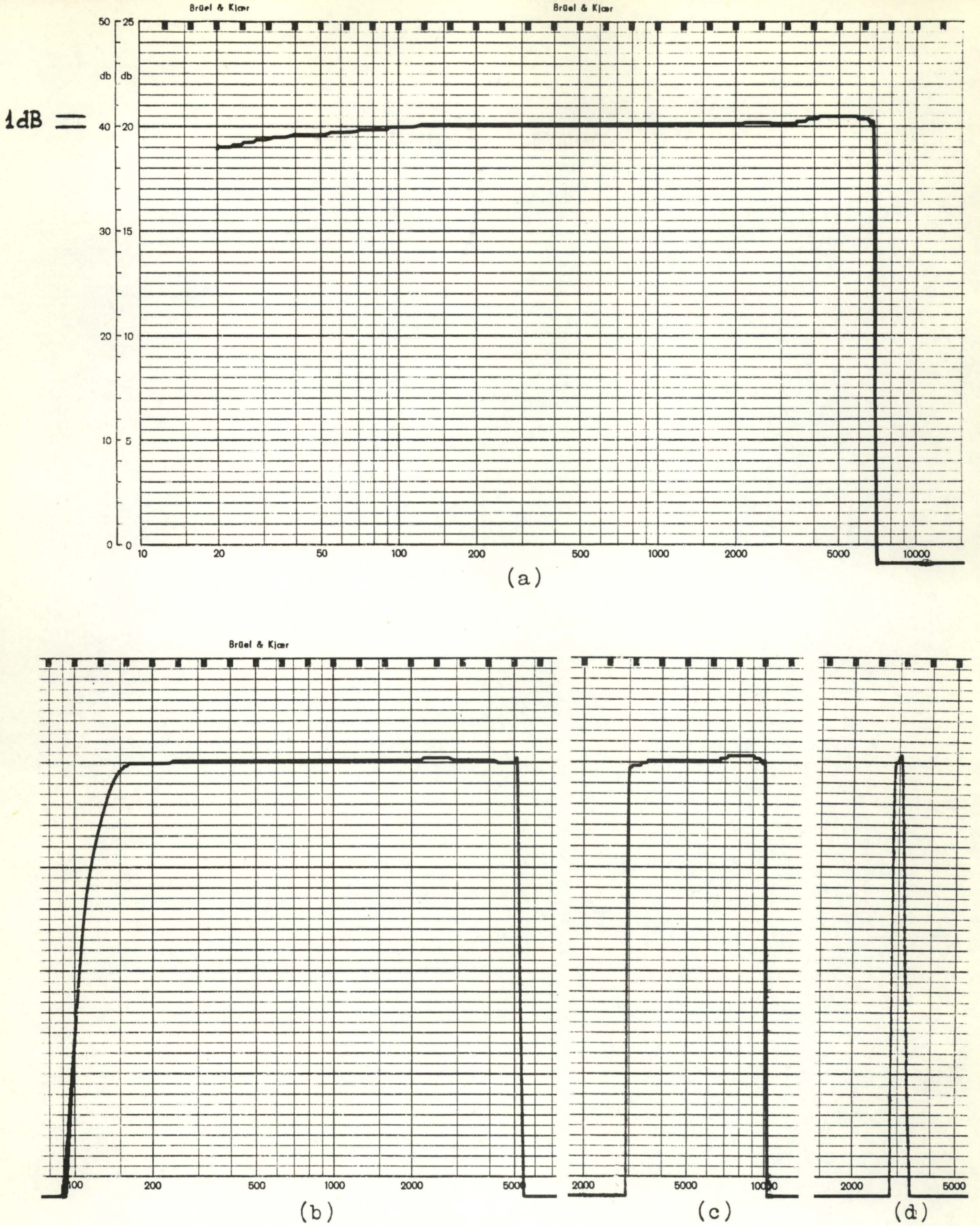


Fig. 3.

Frequency response of heterodyne filter at various settings. Passband amplitude ripple adjusted to be less than ±0,5 dB; paper range 50 dB.

(a) wide band response (-1 dB points at 90 cps and approx. 7000 cps; notice that the amplitude is only 3 dB down at 20 cps).

(b) & (c) various bandpass settings (-1 dB points in (b) at approx. 150 cps and 5000 cps, in (c) at 3000 and 10000 cps).

(d) narrow band response (-1 dB points at 2900 and 3100 cps; this corresponds to a bandwidth of some 235 cps between the -3 dB points. Notice that the bandwidth at the 40 dB attenuation level is only some 500 cps).

The lowpass filter consists of two sections. The first of these (Fig. 2 : C) is a pre-shaping unit, whereas the second (Fig. 2 : D) is a passive LC filter of the conventional constant-k type (image parameter filter) with seven sections, which gives a very high attenuation in the stop band. This filter was designed for a nominal cutoff frequency of 1o kcps. In view of the great number of sections a certain attenuation at cutoff might be expected (due to lossy components). However, the actual attenuation at cutoff is as high as 12 dB and thus exceeds what might be predicted, possibly because there are alignment difficulties with a filter of this complexity. In order to improve the behaviour of the filter near cutoff we have added the amplitude correction equalizer (pre-shaping network) before the filter. This network, which was constructed by a cut-and-try procedure, includes an LC pole circuit resonating at a frequency just below the nominal cutoff, an LC zero circuit resonating at a frequency just above the nominal cutoff, and some RC circuits improving the response at frequencies well below cutoff. When carefully adjusted the equalizer provides a satisfactory overall response for the combined system. (It is possible here to trade some of the linearity in the passband for better sharpness of cutoff, the maximum sharpness obtainable being -4o dB at a distance of 195 cps from the -3 dB point; this gives a passband ripple of slightly more than 2 dB.)

4. Performance of the heterodyne filter: noise, nonlinear distortion, and phase distortion.

The signal-to-noise ratio of the heterodyne filter is very good. With sinewave testing it is possible to obtain a maximum output (before clipping) of o,32 volts RMS with o,44 volts RMS applied to the input terminals. The background noise at the output terminals is more than 65 dB below maximum signal output. However, with maximum input voltage the distortion is relatively high, and in order to minimize distortion it is preferable to reduce the input voltage to max. 1oo millivolts. Even so, the signal-to-noise ratio exceeds 5o dB, which competes well with most recording equipment.

Whistle interference among some of the several frequencies generated in the modulators, is apt to be a disturbing factor. In our first setup we used a lower carrier frequency for modulators 2 and 3, the lowpass filtering being performed (on the inverted sideband, i.e., physically as a highpass filtering) by a Siemens bandpass filter with cutoff at 6okcps. However, it appeared that audible interference tones were generated at certain settings of the carrier oscillators (the interference being between higher harmonics of the two carriers). With the present choice of frequencies we have more or less eliminated this problem, although the transposition of the signal to a low frequency range in modulator 2 (see Fig. 1) creates problems, too. It has proved impossible to utilize the entire passband of the 9,8 kcps lowpass filter, because the occurrence of very low frequencies at this place in the system creates strange distortion products. We do not know at present to what extent this is due to the design of the modulators (which are intended for operation at higher frequencies), or to other factors as well. For better exploitation of the passband modulators 2 and 3 must probably be replaced by other units.

As stated above the suppression of signals outside the passband is very effective. E.g., a o,44 volts signal at 11 kcps does not give any detectable contribution to the output voltage, i.e. the attenuation of the lowpass section at this frequency is (considerably) better than 6o dB. The performance of the highpass section is similarly high, however, at first sight the suppression of the signals below the highpass cutoff frequency does not seem to exceed some 48 dB. This is due to the presence of distortion products. When a sine wave is applied to the heterodyne filter, the inevitable nonlinearity of the modulators conditions the presence of a (weak) third harmonic of the signal, which eventually passes through even if the signal itself is suppressed.

When complex signals are filtered, both harmonic distortion and intermodulation among the components of the signal will occur. The presence of these types of distortion is an inherent limitation of the heterodyne system, although their effect can be reduced by careful design and operation of the filter. In our present setup the distortion due to nonlinearity is typically of the order of 1 to 2%. This is not a shocking figure compared to the performance of medium quality tape recorders and amplifiers, but it should be kept in mind that the generation of distortion products in the stopbands may seriously blur the picture of the filtering function. This consideration also motivates the modest requirement on minimum attenuation in the stopbands which was formulated in the preceding section.

In this context it may be added that according to our experience filtered signals should be recorded and manipulated with the utmost care. It goes without saying that a lowgrade tape recorder may more or less spoil the result. For example, if the fundamental of a harmonic spectrum (e.g., a vowel) is cut off by highpass filtering, intermodulation among the higher harmonics may regenerate the fundamental, the result being an apparently imperfect filtering. This kind of distortion may pass unnoticed under normal circumstances, i.e. with unfiltered speech, since the frequencies generated by intermodulation are already there, so that the contribution of the distortion products to the intensity level is negligible.

Finally, the response of the filter in the time domain should be considered. For analyses of transient sounds (consonants) it is important that the filter does not "smear out" the signal, i.e. the ringing effect must be within narrow limits. With quasi-periodic sounds including vowels the effect of ringing may be less harmful, since the ear seems to be rather insensitive to phase. However, visual inspection of the waveform reveals the distortion. It is well-known that a sharp cutoff brings with it an extremely nonconstant delay of different frequencies within the passband, i.e. a sharp filter inevitably distorts the waveshape of the signal, especially because the phase shift of the system changes rapidly near cutoff.

The phase characteristics of the heterodyne filter are on the whole poor. The rate of phase shift of the lowpass section is rather uniform, and square wave testing shows a rather good reproduction of the waveform, but the highpass section exhibits an enormous shift from one end of the passband to the other. Thus the complete heterodyne system badly distorts the waveshape of signals, and it must be concluded that it is not suitable for analyses involving a closer inspection of phenomena in the <u>time</u> domain, although it is well suited for slowly changing phenomena like vowel stimuli used in perception tests. In spite of the very outstanding frequency discriminating properties of the highpass filter section, a modification of the heterodyne system might profitably start with a redesign of this section, possible with a shift of the operating range to a much lower frequency, where a simpler filter might do the same job with less phase shift. However, it remains to be decided to what extent high fidelity in the time domain can be important for our experiments with filtered speech.

9. Acknowledgements.

Our thanks are due to the Post and Telegraph Department, which has put most of the equipment used in the filter at our disposal, to the State Institute of Speech Defects, which has put a wave analyzer used for control of the system at our disposal, and to the Danish Council for Scientific and Industrial Research (Danmarks teknisk-videnskabelige Forskningsråd), whose financial support has made the project possible.

References:

(1) Jørgen Rischel, "Heterodyne Filter", ARIPUC 1/1966, pp.13-14.

(2) Gunnar M. Fant, The Heterodyne Filter (Stockholm, 1952).

(3) A.I. Zverev, Handbook of Filter Synthesis (New York, London, Sydney, 1967), pp. 63 ff.

SPEECH SYNTHESIZER

Jørgen Rischel and Svend Erik Lystlund

In the past year we have been experimenting with a method to generate control voltages (varying as a function of time) for the formant circuits of a terminal analog synthesizer. Also a modulator system for the synthesizer is being tried out. The work is still in a too early phase for presentation. An account of it will probably appear in the next report of our Institute.

Acknowledgements.

The work is supported by a grant from the Danish Council for Scientific and Industrial Research.

PHONETIC ANALYSIS OF BREATHY (MURMURED) VOWELS IN GUJARATI *)

Eli Fischer-Jørgensen

1. INTRODUCTION.

The Indian language Gujarati, spoken in a district north of Bombay, has a contrast between clear and breathy vowels. This contrast has been treated both from a phonemic and a phonetic point of view by P.B. Pandit (1), who called the breathy vowels "murmured", a terminology which will be used in the following. We also follow his transcription according to which underlining indicates murmur.

Gujarati has eight clear vowels / i e ɛ a ɔ o u ə/ and the corresponding murmured vowels / i̲ e̲ ɛ̲ a̲ ɔ̲ o̲ u̲ ə̲/. The nasalized vowels / ĩ ɛ̃ ã ɔ̃ ũ ə̃ / and nasalized murmured vowels are not included in the present investigation with the exception of the example /pɔ̃ci - pɔ̲̃ci/. Nasalized murmured vowels are rare.

Historically the murmured vowels have developed from (i) vowel + h + vowel, (ii) h + vowel, and (iii) from a fusion of the aspiration of a final voiced aspirated stop with the preceding vowel (2).

Phonemically murmured vowels may be interpreted as vowel + h, but phonetically they form one segment. Pandit (1) gives the following description: "Murmur is voiced breath, low pitched and simultaneous with the vowel" (p. 169) - "sotto voce, with voicing and slight lowering of pitch" (p. 170). He also published a spectrogram and drew attention to "the presence of random distribution of energy, more noticeable at higher frequencies" (p. 172).

A spectrographic investigation of the formant frequencies of murmured vowels compared to clear vowels has been undertaken by Radhekant Dave (2).

The purpose of the present investigation is to give a more all round phonetic description of these vowels in respect of airflow, duration, fundamental frequency, overall intensity, formant frequen-

----

*) This is a summary of a paper to be published in <u>Indian Linguistics</u> 1968. The numbers of tables and figures have been retained, although some have been left out here.

cies, and distribution of spectral energy, and including a restricted number of auditory tests. **)

2. THE INFORMANTS.

Seven informants have been used: RD, RT, SK, DD, GU, PPB, PvB (PvB is female, the others male). DD had a high-pitched falsetto voice, and only a small part of his recordings have been used. All informants have murmured vowels in their natural speech. In RD's and PBP's speech murmur is optional. RT tended to pronounce an [h] when reading aloud, and a good deal of his examples had to be removed. RD, SK and RT have no consistent distinction between close /e o/ and open /ɛ ɔ/. Only [e o] have been used in the transcriptions of their speech.

3. THE TEXTS.

The texts consisted of series of isolated words and short sentences, spoken several times by each speaker. It appears from the tables and graphs which words were used and how often they were repeated. Series of words, particularly when repeated, may lead to a sort of rhythmical singsong which may reduce the differences between the two types of words. Nevertheless only averages of the word series are given in the tables and have been statistically treated. This procedure has been chosen, partly because more examples had been recorded of the word series, partly because the sentences were spoken in pairs by some of the speakers, and this gave rise to contrastive intonations. RT's tape recording forms an exception because he also spoke the sentences five times and not in pairs. In this case words and sentences have been combined in the averages. - For the other informants the sentence examples are used for comparison.

**) The investigations have been carried out in the laboratory of our institute, and I am grateful to several of the staff members and students for help.

The experiments with synthetic and filtered vowels have been carried out in cooperation with Jørgen Rischel, who has been responsible for the technical part of these experiments.

I am particularly indebted to Radhekant Dave, who has found most of the informants, set up the word lists, controlled the recordings, and listened to the tests.

4. INSTRUMENTAL SET-UP.

4.1. A tape recording of PBP was made in Dehradun in 1957, a recording of PvB in the Institute of Phonetics in Amsterdam in 1964, tape recordings of the other informants in Copenhagen 1966-67.

4.2. A selected number of words containing the vowels [i e ɛ a o ə] from the tape recordings of six of the informants (DD was left out) have been used for spectrographic analysis by means of the Kay Electric Sonagraph. Narrow band spectrograms and narrow band sections (both with "High Shaping" and "Flat 2") were taken of all the words selected, and some spectrograms were taken with wide band (see Figs. 1-3). Moreover some spectrograms were taken with high compression, and some with half speed of the tape in order to examine the presence of noise in the higher frequency regions.

The spectrograms were used for measurements of formant frequency, formant amplitude and for the measurement of the amplitude of F_o and of the second and third harmonics H_2 and H_3. (Generally F_3 and F_4 could not be measured for [o].) Amplitude measurements were made on the basis of the sections. For the formants above c. 1000 cps the high shaping section had to be used, but on the basis of the high shaping curve of the instrument the measurements were corrected by subtracting the appropriate number of decibels. As a measurement of the amplitude the root mean square value has been chosen, which means that the harmonics within the range of the formant have been summed up according to a pre-established table of dB-values. Two close formants make difficulties, but nevertheless this method was found to give more reliable results than the measurement of the envelope peak, which can be difficult to place exactly. (3).

Some reservations must be taken as to the exactitude of the measurements, because the new model of the Kay Electric Sonagraph (A 6061) has been found to produce difference tones by intermodulation. However, the formant peaks of normal vowel spectrograms do not seem to be distorted (spectrograms taken with high shaping, with "Flat 1", "Flat 2" and with low pass filtering did not show any difference in amplitude relations), but the valleys in between the formants are hardly reliable.

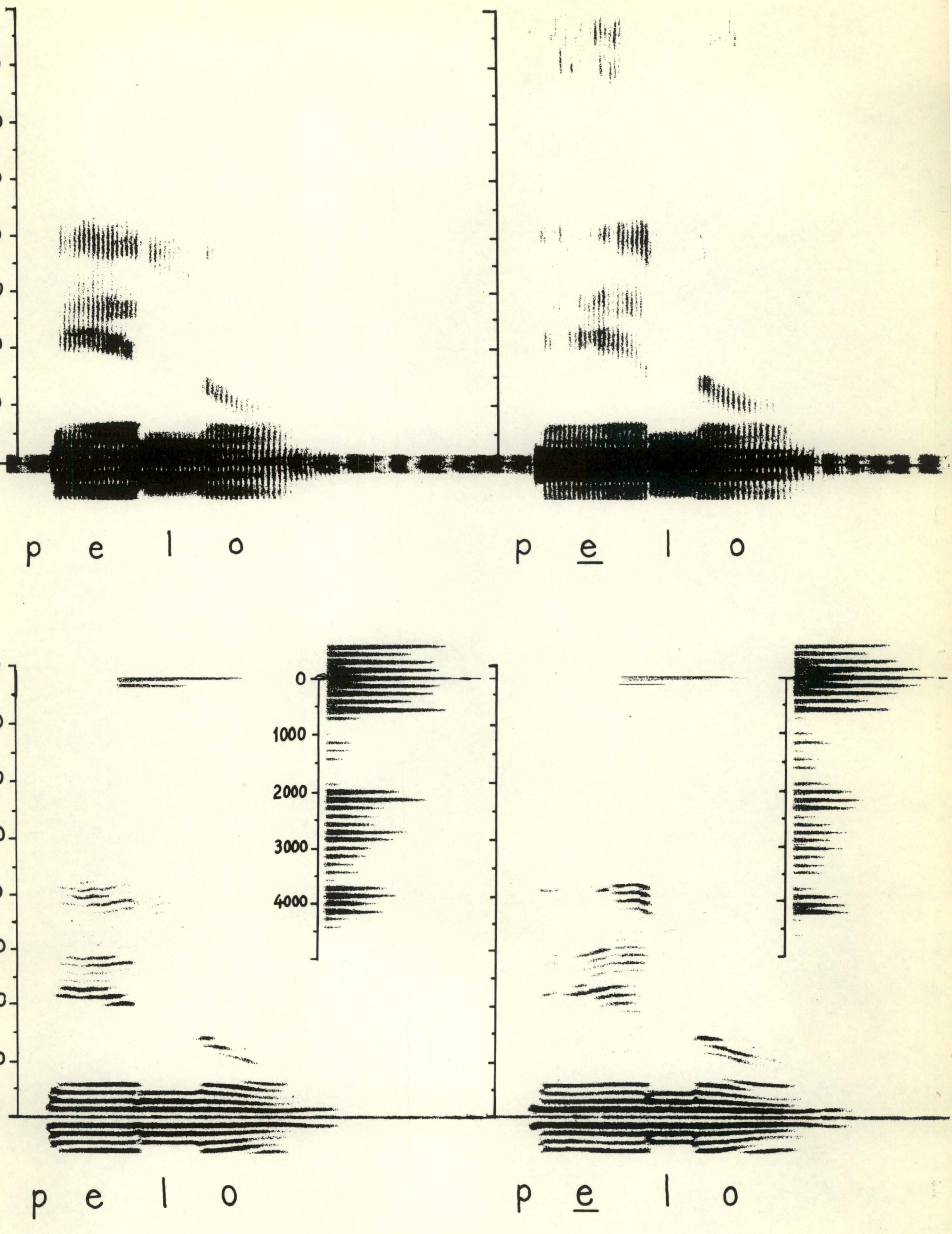


Fig. 1. Spectrograms of pelo and p<u>e</u>lo (RD).
Wide band and narrow band spectrograms and narrow band section.

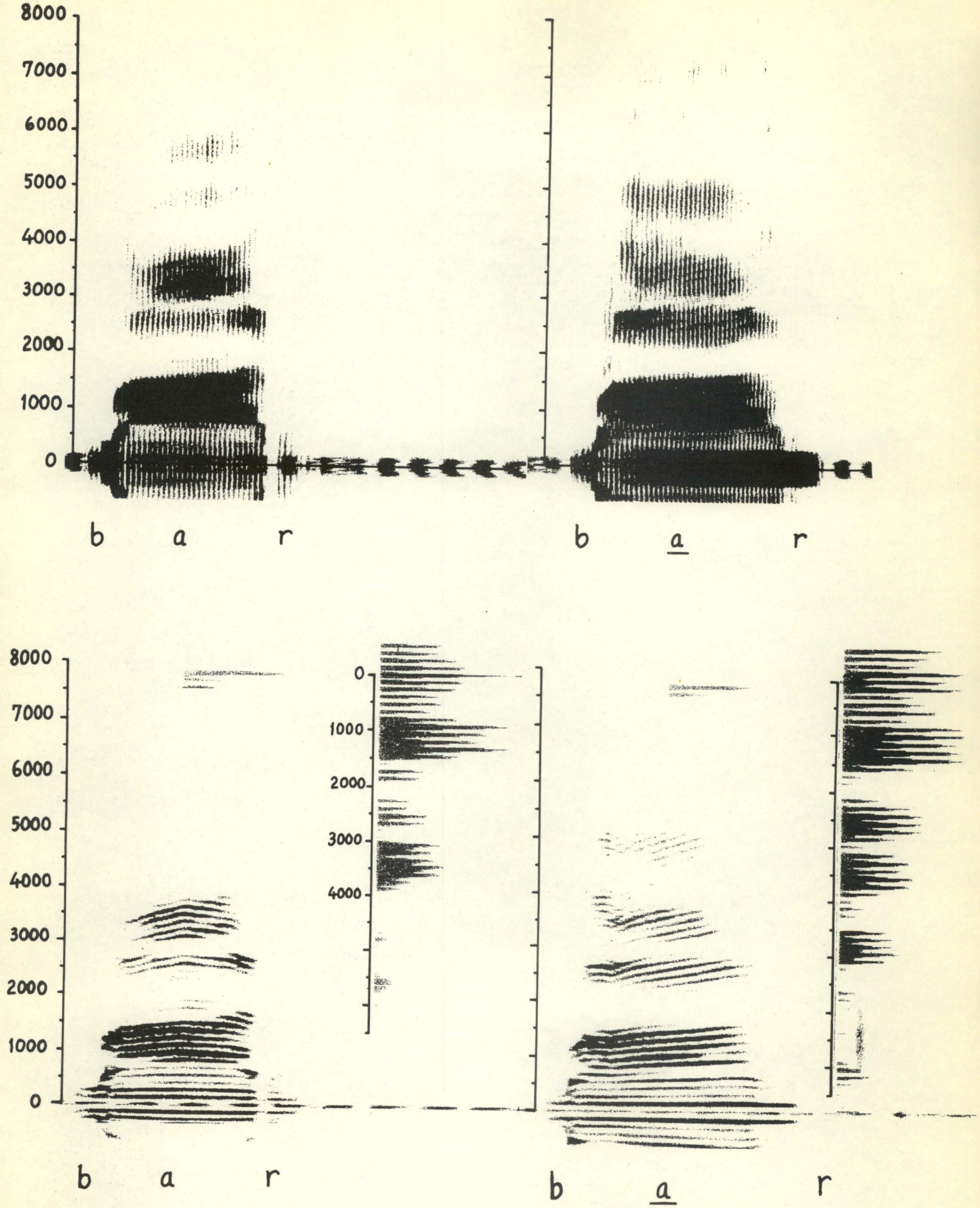


Fig. 2. Spectrograms of bar and ba̲r̲ (GU).
Wide band and narrow band spectrograms and narrow band section.

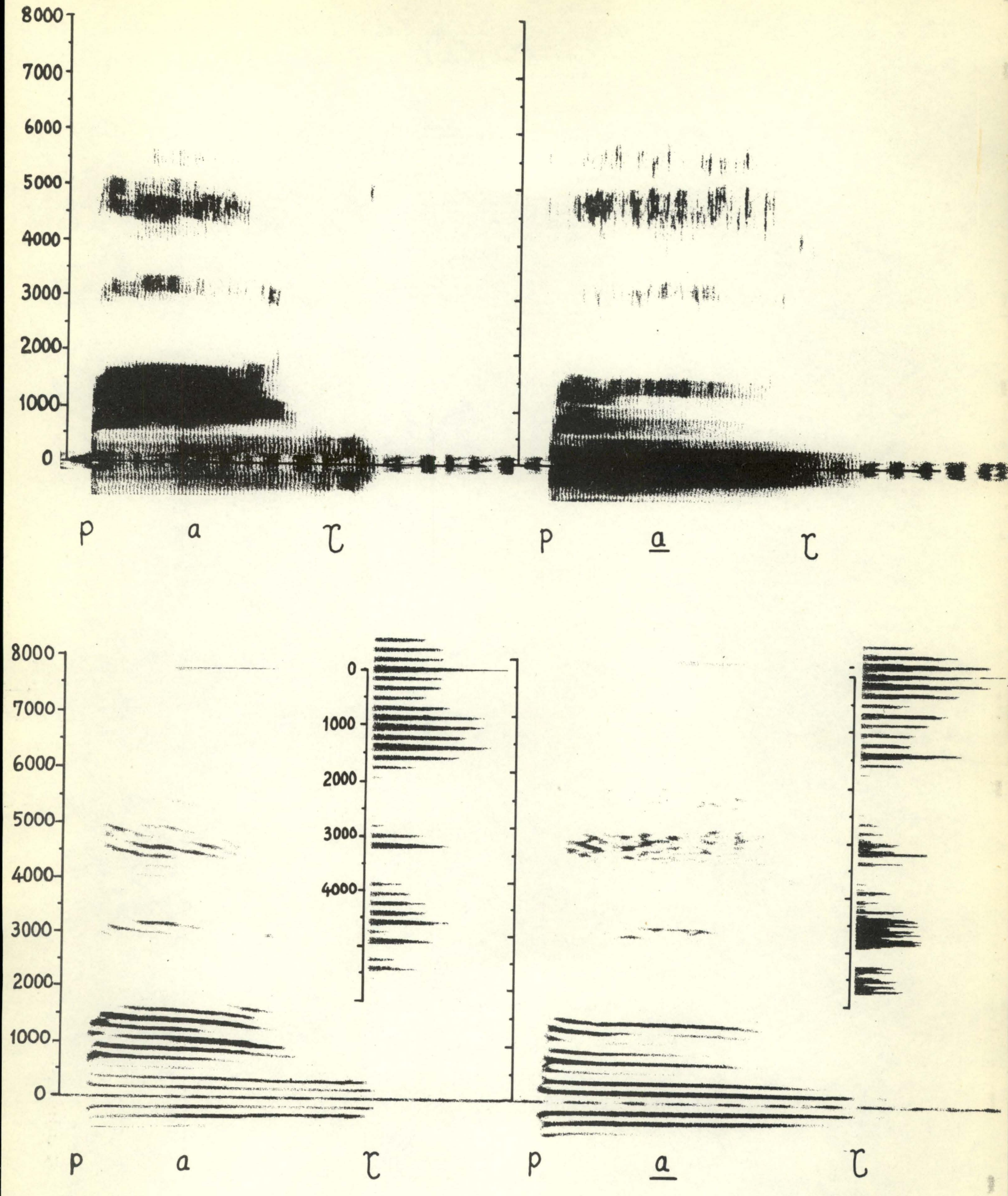


Fig. 3. Spectrograms of paɽ and pa̲ɽ (PvB).
Wide band and narrow band spectrograms and narrow band section.

4.3. All of the tape recordings were used for analysis of fundamental frequency and overall intensity be means of the trans-pitchmeter and intensity-meter connected with a four channel Elema Mingograph. The following curves were taken: (1) a duplex oscillogram, (2) a sharply highpass-filtered logarithmic intensity curve with an integration time of 5 ms (2.5 ms for the female subject), (3) an intensity curve with flat frequency response and an integration time of 10 ms (5 for the female speaker), and (4) a fundamental frequency curve. The highpass-filtering was made by means of an external passive filter (insertion loss filter), with a nominal cutoff frequency of 315 cps. The 3 dB point was at 340 cps, and the attenuation was more than 45 dB at 270 cps (see Fig. 4).

4.4. The selected texts used for spectrographic recording were also used for an oscillogram recorded on the Mingograph with double speed of the paper (20 cs) and with the tape played back at one quarter of the speed. In this way the upper frequency limit of the mingograph recordings was raised from 800 to about 3200 cps, and the distance from one period to the next was eight times the distance of the other mingograms taken with a paper speed of 10 cm per second. The purpose was to see details of the temporal change of the vowel.

4.5. Three of the subjects (RD, RT and GU) spoke the list of words and sentences into the Aerometer used for airflow measurements. Curves of fundamental frequency and overall intensity were taken simultaneously from a throat microphone. Only one intensity curve was taken (in some cases with and in some cases without highpass filtering). The words were spoken in groups of four and five with varying order (see Figs. 6, 7 and 8).

4.6. A few recordings comprising five isolated vowels and five words, spoken each six times by RD, were made by means of the Fabre glottograph, combined with curves of fundamental frequency and airflow (see Fig. 5).

4.7. Attempts were made to synthesize sounds with the spectral characteristics of clear and murmured vowels. The synthesizer has been described by Jørgen Rischel (4).

4.8. Filtering. Some filtering experiments were made with highpass, lowpass and a combination of highpass and lowpass filtering of clear and murmured vowels with the purpose of observing the consequences for the perception of murmured vowels. For the highpass

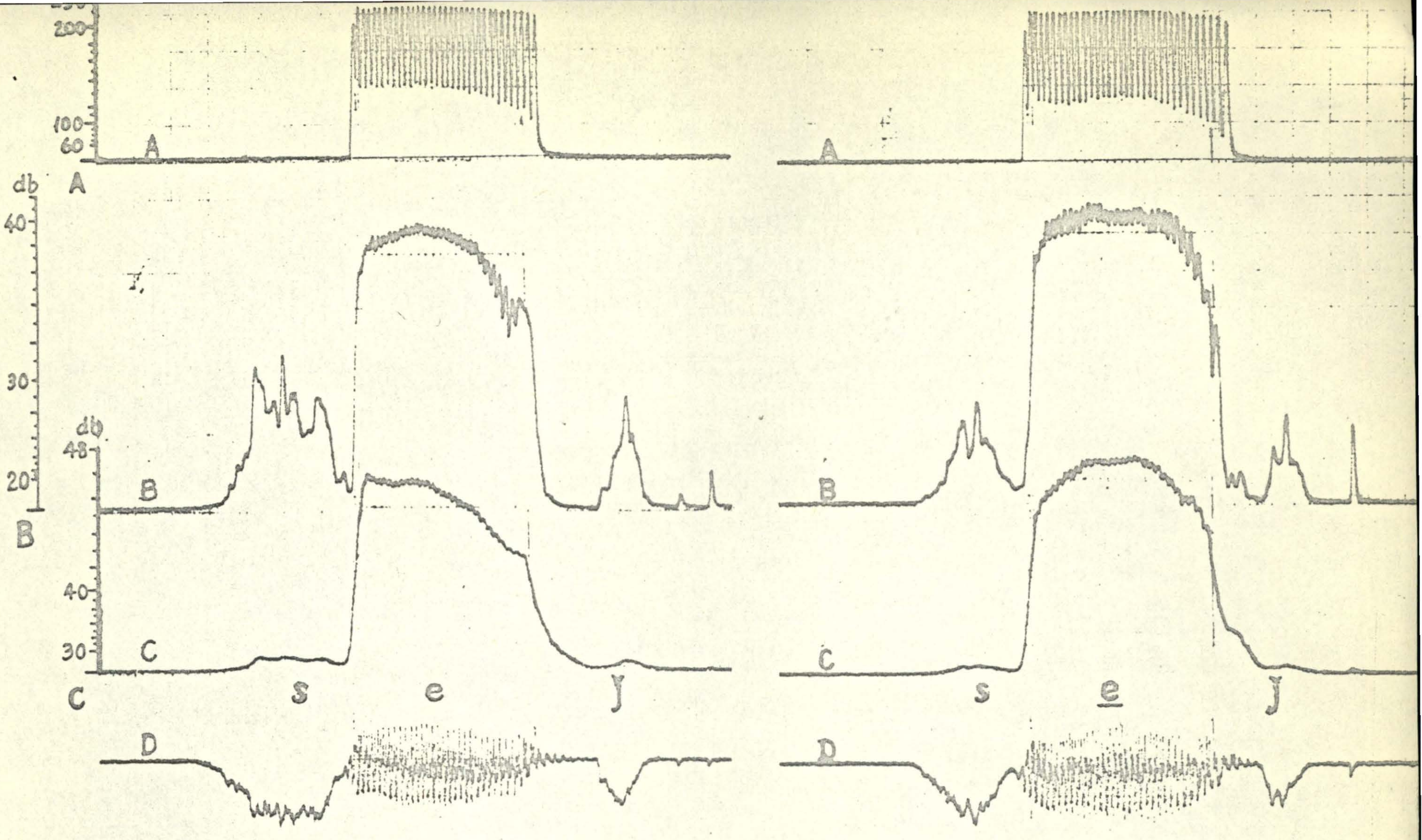


Fig. 4. Mingogram of seɟ and seɟ (RD).

A. Fundamental frequency curve.

B. Highpass filtered logarithmic intensity curve.

C. Linear intensity curve without filtering.

D. Duplex oscillogram.

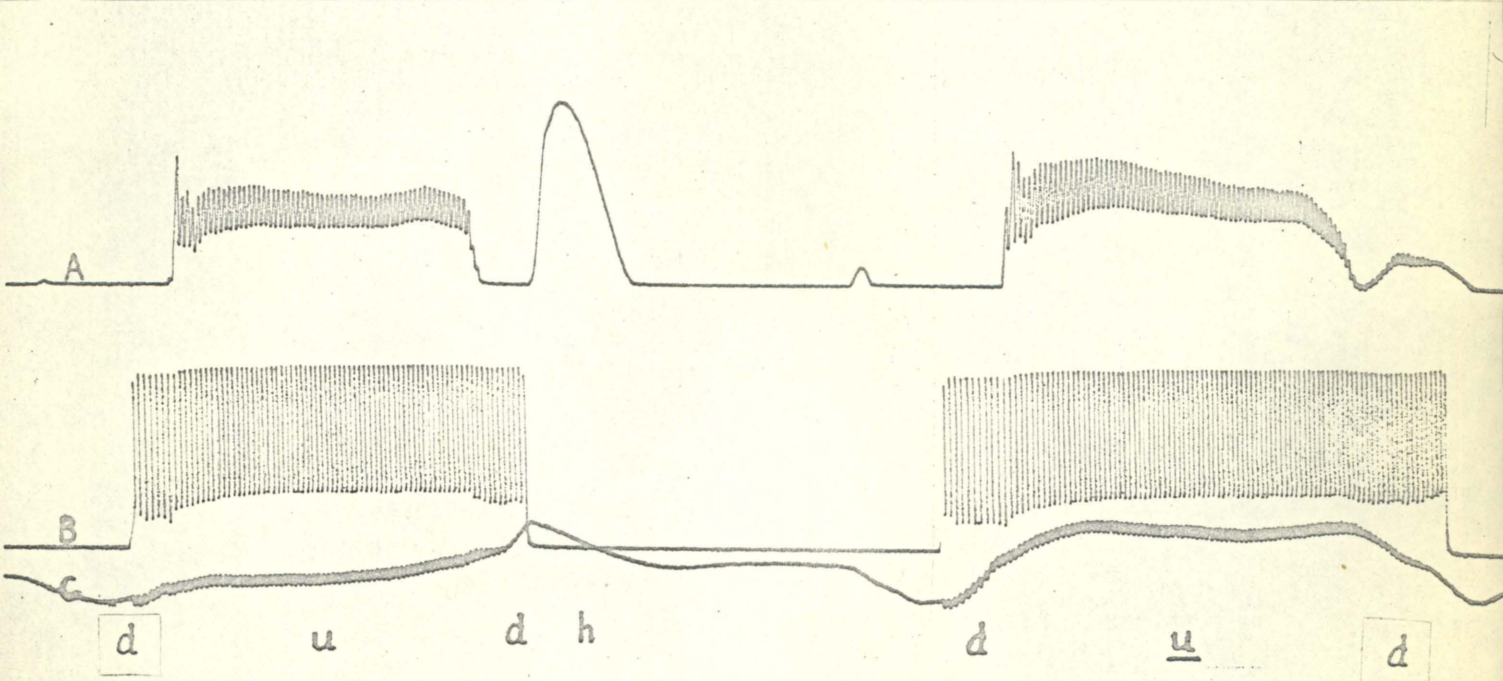


Fig. 5. Mingogram of dudh and dud (RD).

A. Airflow curve.

B. Fundamental frequency curve.

C. Glottogram.

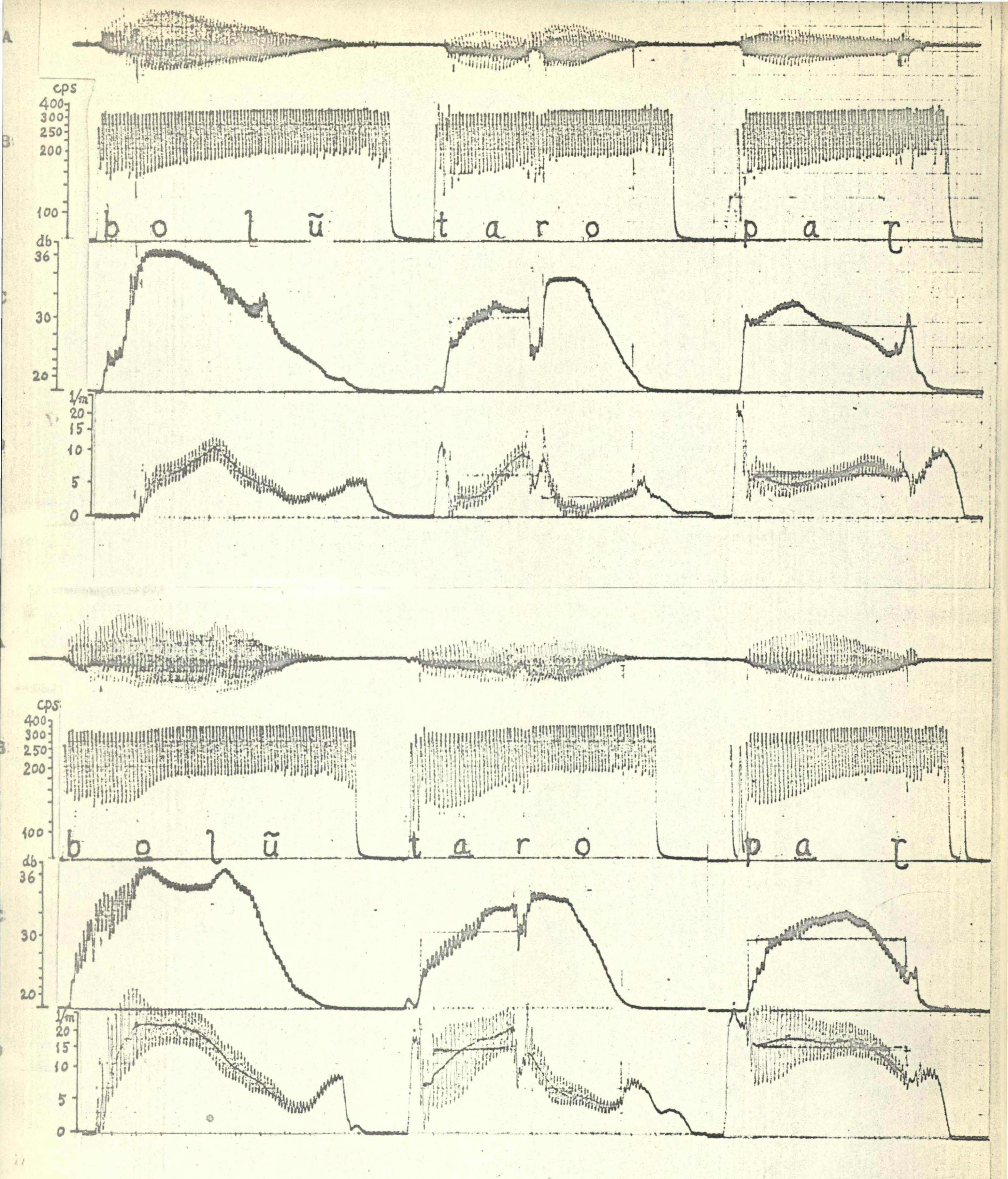


Fig. 6. Mingograms of boɭũ, taro, paɽ and bo̲ɭũ, ta̲ro, pa̲ɽ (RD).

A. Oscillogram.

B. Fundamental frequency curve.

C. Logarithmic intensity curve, highpass filtered 315 cps.

D. Airflow curve.

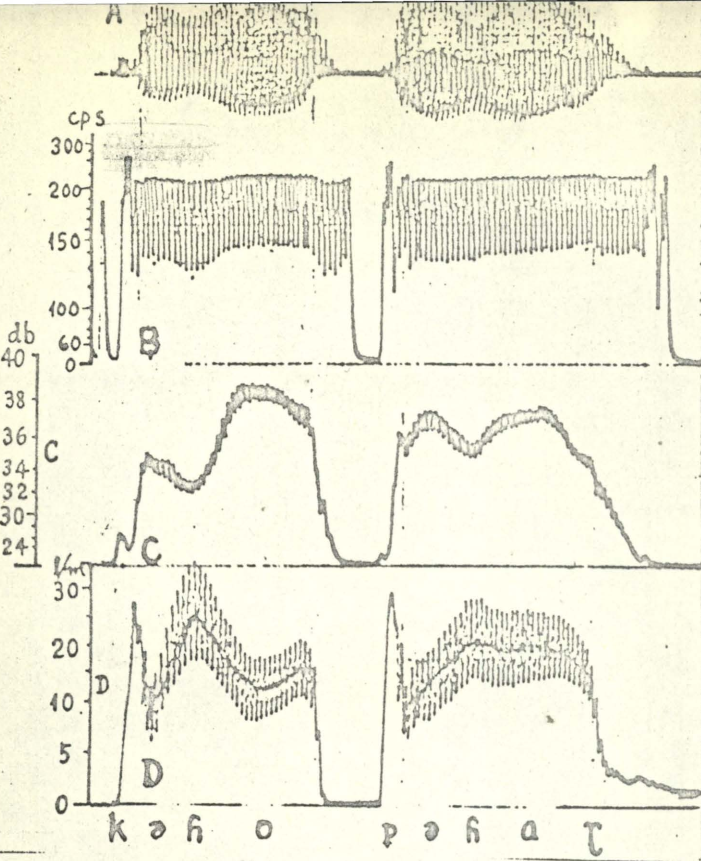


Fig. 7a.

Mingogram of intervocalic h in ko, paʈ (RT).

A. Oscillogram.
B. Fundamental frequency curve.
C. Intensity curve.
D. Airflow curve.

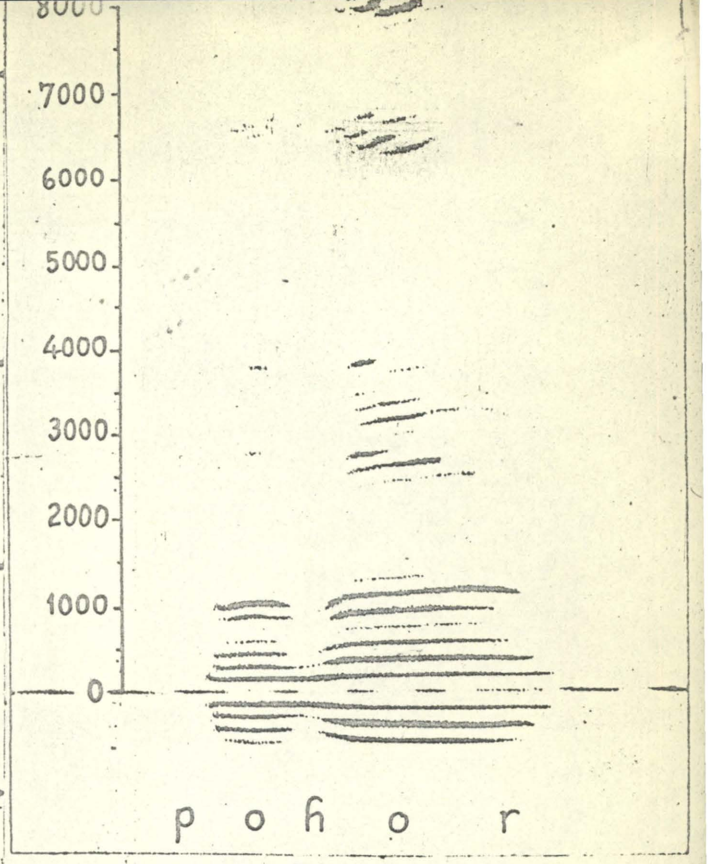


Fig. 7b.

Narrow band spectrogram of intervocalic h in por (RT).

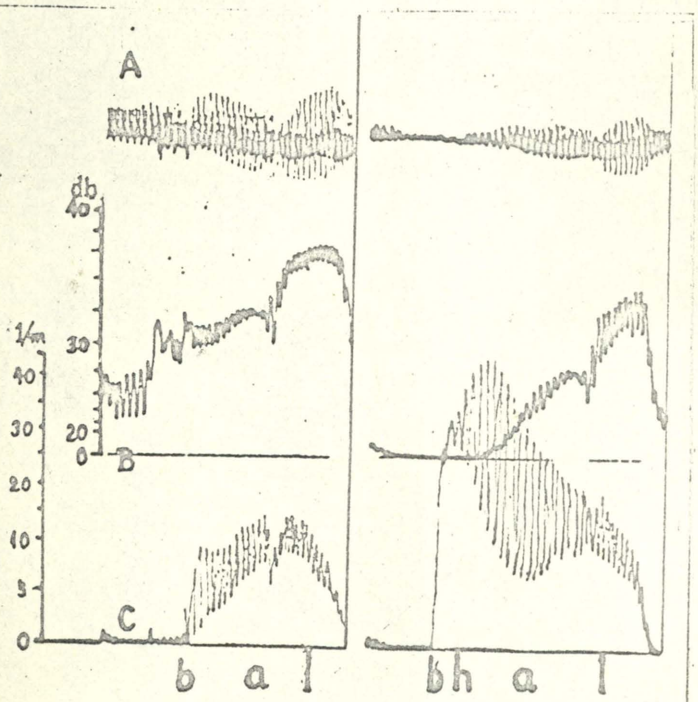


Fig. 8a. Mingogram of b and bh (RD).

A. Oscillogram.
B. Fundamental frequency curve.
C. Airflow curve.

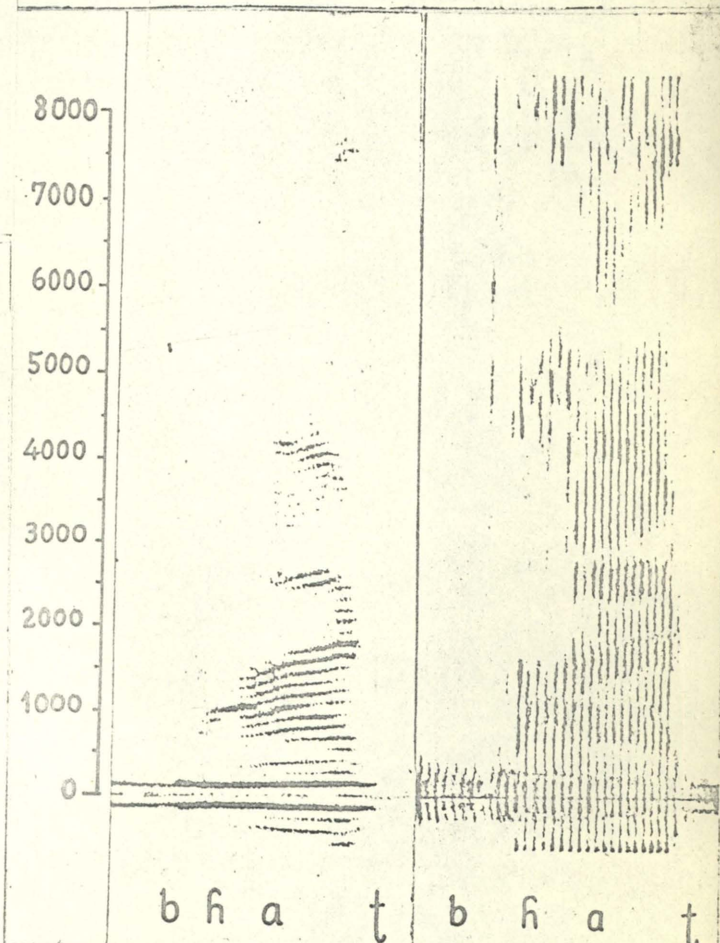


Fig. 8b.

Narrow and wide band spectrograms of bh in bɦaʈ (RD).

and lowpass filtering the heterodyne filter was used (5), and the cutoff frequencies (here defined as the points of 3 dB attenuation) were set to 23o cps and 31oo (in one case 32oo) cps respectively. This filter was also used, in connection with another filter, to exclude frequencies between 2oo and 5oo cps. Two passive filters were used for a similar purpose with the frequency limits 16o and 5oo cps.

The filters were very steep and the attenuation should be more than 4o dB some 5o cps from the cutoff frequency, except for the lowpass filtering at 31oo and 32oo cps, which was performed with a somewhat less sharp cutoff. Control measurements of the tape recordings showed, however, that the attenuation of harmonics outside the passband(s) amounted to only some 25 dB. This is clearly a distortion effect, which may be ascribable to the signal level being too high during the recording. *)

5. RESULTS.

5.1. Physiological Analysis.

5.1.1. Airflow. According to the subjective impression, particularly the impression of the speaker, the most obvious characteristic of murmured vowels is an increase of airflow.

The airflow has been measured by means of the aerometer (see section 4.5), and the curves indicate the amount of air passing per time unit. The measurements comprise the maximum airflow, the average airflow (the scale l/m indicates litre per minute), and the distance (in cs) from the beginning of the vowel to the maximum. The time constant of the instrument is sufficiently small to show also the voice ripple. A mid-line was drawn through the ripple by hand. The maximum was measured as the highest point on this line. The average airflow of the vowel has been found by a graphical approximation by means of a horizontal line drawn by hand. This could be done with an exactitude of somewhat below 1 l/m.

The airflow has been measured for three subjects (RD, RT and GU), the total material comprising 1442 words. Examples of the curves are given in Fig. 6.

*) In a later filtering experiment with an improved setup we succeeded in keeping the out-of-passband distortion products (measured at the loudspeaker) some 4o dB below the level of the strongest harmonics. However, a listening session comprising a restricted number of words made under these improved conditions gave only slightly deviating results from those presented in § 8.3.3 below.

The main results are given in Table 1.*)

Table 1 AIRFLOW

	Maximum Airflow			Average Airflow		
Speaker	RD	RT	GU	RD	RT	GU
Number of words	552	676	214	5o4	644	214
l/m murm	18.6	17.2	36.6	15.1	12.1	3o.5
l/m clear	8.9	11.4	25.7	7.2	8.3	19.3
difference	9.7***	5.8***	1o.9***	7.9***	3.8***	11.2***
increase in %	1o9	55	46	11o	46	6o

RD and RT have also spoken the word pairs in sentences. For RD the difference between clear and murmured vowels was slightly larger in the sentences, for RT considerably larger.

The averages of the different word pairs in the series of isolated words are given in graphical form in Fig. 9. **) The stability of the difference appears clearly from the graph. There is no single exception. For single word pairs (words spoken under the same conditions in the same reading) the difference is also 1oo per cent stable in the speech of RD and GU, 96-97 per cent in the speech of RT.

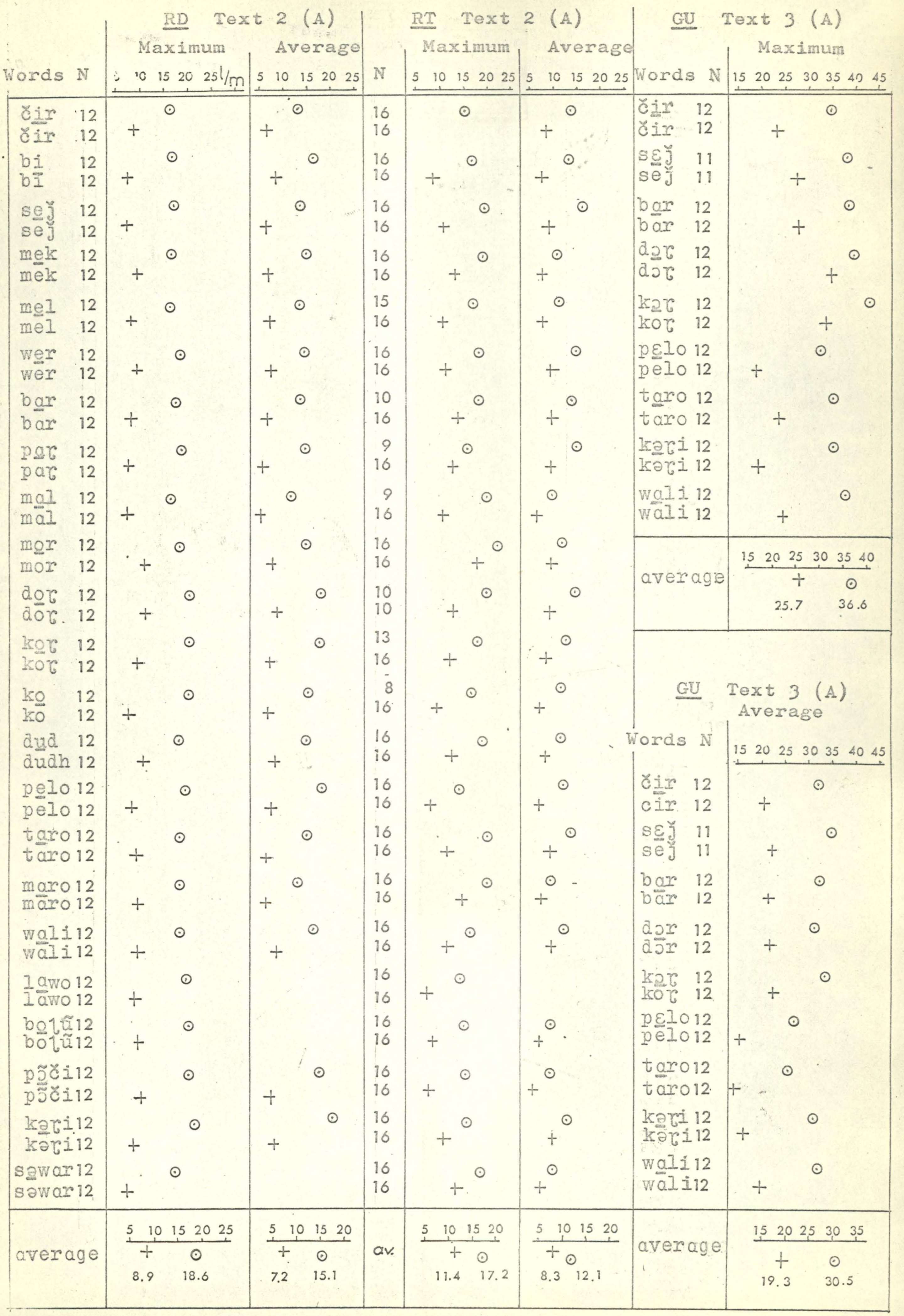
----

*) In this and the following tables the level of significance is indicated by means of asterisks, one asterisk indicating a significance level of 5 per cent, two asterisks a level of 1 per cent, and three a level of o.1 per cent. The significance has been calculated by means of a pair test (see Croxton, Elementary Statistics, Dover 1959, p. 241) generally based on the word averages or in the cases where the number of different pairs were too small (e.g. PvB, DD and sometimes GU), on a comparison between single pairs within the same reading.

The difference always indicates murmured vowel minus clear vowel, i.e. the number is positive if the murmured vowel has a higher value than the corresponding clear vowel.

**) In this and the following graphs clear vowels are indicated by crosses, murmured vowels by rings.

Fig. 9. AIRFLOW (⊙ = murmured + = clear)



5.1.2. In dissyllabic words the characteristic difference in airflow is spread out over both syllables. Only in RD's speech is the difference in the second vowel considerably smaller than the difference in the first vowel (see Fig. 6). RT has even a greater difference in the second vowel. On the other hand RD and, partly GU, have more airflow in an initial unvoiced consonant before murmured vowel than before clear vowel (see Fig. 6).

5.1.3. The airflow in the beginning and end of the vowels is strongly influenced by the surrounding consonants. It is often particularly low after voiced stops, which, in Gujarati, may be more or less implosive, particularly often so in PvB's and GU's speech.

5.1.4. A very short series of whispered words were spoken by RD, There is also in whispered speech an obvious difference in airflow between murmured and clear vowels, but the difference is smaller than in normal speech.

5.2. Vocal Cord Opening

The increased airflow of murmured vowels can be explained either by a wider glottis, or by an increased activity of the expiratory muscles or by both.

It has only been possible to investigate the glottal opening of one subject (RD). By means of a normal laryngoscope combined with stroboscopic light his glottis has been observed by a medical doctor and two phoneticians while he pronounced clear and murmured [a] and [ɛ]. All have noticed a wider opening in the rear part of the glottis for murmured vowels.

Moreover, recordings were made with the Fabre glottograph (see section 4.6) of a text (containing 4 word pairs and four pairs of isolated vowels) spoken 6 times by RD. - No calibration was possible, and the zero-line was not stable, but the average level of the curve of murmured vowels was in all cases higher than that of neighbouring clear vowels. The differences were most pronounced for [u-u̲] (see Fig. 5). This higher level indicates a higher resistance, which may be due to the opening of the glottis. An attempt was made to take glottograms of clear and murmured vowels with the new photo-electric glottograph of the Institute of Phonetics, but without success. It proved very difficult to place the tube with the photocell in the appropriate position because of Dave's narrow throat and strong reflexes.

5.3. Hypothesis about Activity of Expiratory Muscles.

Electromyographic investigations have not been undertaken. But as the measurements of overall intensity do not show any consistent difference between clear and murmured vowels, it is highly probable that the loss of intensity which one might have expected in murmured vowels because of the leaking glottis, is compensated for by a stronger activity of the expiratory muscles, keeping the subglottal pressure on the same level. This assumption is supported by Dave's subjective impression of a stronger "stress" in murmured vowels.

It is thus probable that the stronger airflow of murmured vowels is due both to glottal and expiratory conditions.

6. RESULTS OF THE ACOUSTIC ANALYSIS

6.1. General Remarks.

The measurements of duration, fundamental frequency and intensity have been based partly on the tape recordings (below indicated by T), partly on the aerometer recordings (below indicated by A). The spectral analysis has been based on tape recordings only.

6.2. Duration

6.2.1. The material included in the general averages (i.e. all word lists and RT's sentences)comprises 2oo7 words in all. The main results are given in Table 2, and the averages of single words in Figs. 1o and 11. The duration is measured in centiseconds, and the measurement was made with an exactitude of one cs.

It appears from the table that murmured vowels are longer than clear vowels and that the difference is significant for all speakers. But both the absolute and the relative difference is small. The sentences show somewhat larger differences (15-47 %), particularly for RT (45 %).

6.2.2. On the whole dissyllables show a somewhat greater relative difference than monosyllables, which increases the difference between the words with murmured and clear vowels(see Fig. 1).

6.2.3. The well known general tendency to lengthen the vowel before [r] and to shorten the vowel before stops is also seen in this

Table 2 DURATION (Word Series)

Monosyllables

Speaker	RD(A)	RD(T)	RT(A)	RT(T)	GU(A)	GU(T)	SK	PBP	PvB	DD
Number of Words	336	132	308	72	120	100	142	60	46	8
cs. murm.	29.3	27.9	22.1	23.4	32.4	24.3	18.0	32.6	35.3	19.0
cs. clear	28.1	24.0	19.3	21.2	30.2	21.4	16.7	30.5	33.2	15.8
Difference	1.2	3.9	2.8	2.2	2.2	2.9	1.3	2.1	2.1	2.2
Increase in %	4	17	15	11	7	14	9	10	6	14

Dissyllables

Speaker	RD(A)	RD(T)	RT(A)	RT(T)	GU(A)	GU(T)	SK	PBP	PvB	DD
Number of Words	120	48	160	87	96	60	65	12	8	34
cs. murm.	17.9	16.9	14.5	16.8	15.3	13.7	12.1	16.5	19.5	16.0
cs. clear	15.5	13.7	12.1	12.5	14.3	12.2	10.5	14.4	16.8	13.5
Difference	2.4	3.2	2.4	4.3	1.0	1.5	1.6	2.1	2.7	2.5
Increase in %	16	24	22	35	8	15	17	18	16	23

Monosyllables + dissyllables

Speaker	RD(A)	RD(T)	RT(A)	RT(T)	GU(A)	GU(T)	SK	PBP	PvB	DD
Number of Words	456	180	468	159	216	160	207	72	54	42
cs. murm.	26.3	25.0	20.1	19.7	24.8	20.3	16.3	29.9	33.1	16.6
cs. clear	24.8	21.3	17.4	16.4	23.2	17.9	14.9	27.8	30.9	13.9
Difference	1.5***	3.7***	2.9***	3.3***	1.6***	2.4***	1.4***	2.1**	2.2**	2.5***
Increase in %	6	17	18	24	8	14	11	10	7	21

Fig. 1o. DURATION (⊙ = murmured + = clear)

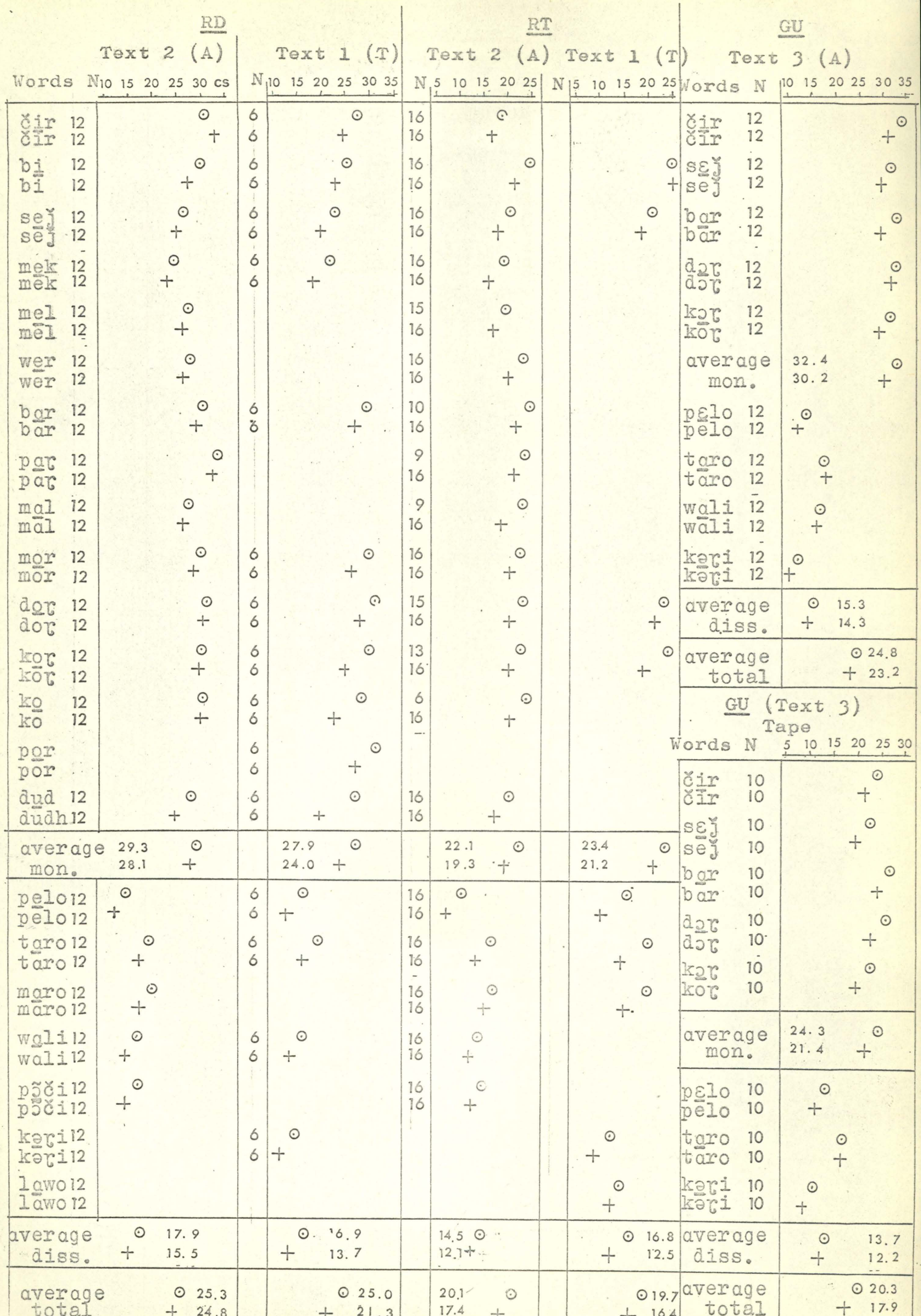


Fig. 11. DURATION (⊙ = murmured += clear)

PvB (Text 5)

Words	N	Tape (15 20 25 30 35 40 cs)	
ba̱r	3	⊙	
bar	3	+	
pa̱ɽ	4	⊙	
paɽ	4	+	
ka̱n	4	⊙	
kan	4	+	
da̱m	4	⊙	
dam	4	+	
ko̱ɽ	4	⊙	
koɽ	4	+	
mo̱ɽ	4	⊙	
moɽ	4	+	
average mon.		⊙	35.3
		+	33.2
wa̱li	4	⊙	
wali	4	+	
average total		⊙	33.1
		+	30.9

PBP (Text 4)

Words	N	Tape (10 15 20 25 30 35 40 cs)	
bi̱k	3	⊙	
bik	3	+	
ba̱r	3	⊙	
bar	3	+	
pa̱ɽ	3	⊙	
paɽ	3	+	
kɔ̱	3	⊙	
kɔ	3	+	
ko̱ɽ	3	⊙	
koɽ	3	+	
dɔ̱ɽ	3	⊙	
dɔɽ	3	+	
du̱d	3	⊙	
kud	3	+	
ũ̱	3	⊙	
ũ	3	+	
lo̱i	3	⊙	
loi	3	+	
nə̱i	3	⊙	
ləi	3	+	
average mon.		⊙	32.6
		+	30.5
pɛ̱lo	3	⊙	
pelo	3	+	
wa̱li	3	⊙	
wali	3	+	
average diss.		⊙	16.5
		+	14.4
average total		⊙	29.9
		+	27.8

SK (Text 1)

Words	N	Tape (0 5 10 15 20 25 cs)	
či̱r	6	⊙	
čir	6	+	
bi̱	3	⊙	
bi	3	+	
se̱ǰ	9	⊙	
seǰ	9	+	
me̱k	9	⊙	
mek	9	+	
ba̱r	9	⊙	
bar	9	+	
ko̱ɽ	9	⊙	
koɽ	9	+	
do̱ɽ	5	⊙	
doɽ	5	+	
po̱r	5	⊙	
por	5	+	
mo̱r	8	⊙	
mor	8	+	
ko̱	8	⊙	
ko	8	+	
average mon.		⊙	18.6
		+	16.7
pe̱lo	8	⊙	
pelo	9	+	
ma̱ro	8	⊙	
maro	9	+	
kə̱ɽi	8	⊙	
kəɽi	9	+	
wa̱li	7	⊙	
wali	7	+	
average diss.		⊙	12.1
		+	10.5
average total		⊙	16.3
		+	14.9

DD (Text 1)

Words	N	Tape (0 5 10 15 20 25 30)	
se̱ǰ	4	⊙	
seǰ	4	+	
pe̱lo	5	⊙	
pelo	5	+	
kə̱ɽi	3	⊙	
kəɽi	3	+	
ma̱ro	3	⊙	
maro	3	+	
wa̱li	2	⊙	
wali	2	+	
average diss.		⊙	16.0
		+	13.5
average total		⊙	16.6
		+	13.9

material.

6.2.4. RD has also spoken five examples of long isolated [a] and four of [a]. In this case the relation between murmured and clear vowels was reversed ([a] 54 cs, [a] 82 cs) without any overlapping. The reason is probably that in such long vowels the great amount of air required for the murmured vowels sets relatively narrow limits to their duration.

The reason for the longer duration of murmur vowels in normal speech is thus to be sought in the historical origin of murmured vowels as a fusion between one or two vowels and /h/.

6.3. Fundamental Frequency.

6.3.1. According to P.B. Pandit's description (1), murmured vowels are characterized by a slight lowering of pitch. As the disappearance of /h/ before or after the vowels in other Indian languages has been accompanied by a development of tonal differences (6), it is of particular interest to investigate this feature.

6.3.2. As word tones are always modified by sentence intonation (and even isolated words or word series will have a definite sentence intonation), it is important to restrict the comparison to words found under the same rhythmical conditions. In RD's A-recording and in both RT's recordings the words are spoken with rising or rising-level intonation. In RD's T-recording some words were spoken in pairs, and in this case the first word had rising intonation, the second rising-falling, with dominating fall. In the words spoken in series he has rising intonation in one reading, and rising-falling in another. The words of his T-recording have therefore been distributed on two categories marked as / and /\. GU has rising-falling intonation, PBP variable intonation (only the frequency at the start has been measured). SK's curves could not be measured because of hoarseness.

The sentence intonation varies in the same way for clear and murmured vowels, so that the general trend of the movement is the same, and differences must be looked for in smaller details.

6.3.3. The frequency was measured at some selected points of the frequency curve, namely the beginning and end of the vowel, as well as maxima and minima. It was measured with an exactitude of five cps. Moreover the distance from the beginning of the vowel to the frequency maximum was measured (in cs).

As it appears from Table 4, comprising the averages for each recording, the words with murmured vowel have a lower minimum than the words with clear vowel (for the words with rising-falling tone this minimum is the first one). Although the difference is very small, it is relatively constant, and is statistically significant for all recordings except RT (A). PvB's curves could not be measured because the base line of the calibration curve was uncertain, but she shows the same tendency to have a lower start of murmured vowels.

6.3.4. A significant difference has also been found for the distance from the beginning of the vowel to the frequency peak (Table 5), but when the difference is measured in percentage of the vowel, it is extremely small or non-existent.

6.3.5. Figs. 13 and 16 show the differences in minimum, maximum, rise and distance from beginning to maximum for the averages of each word pair for some of the subjects.

6.3.6. The graphical display of Figs. 13 and 16 is simplified in the sense that the minimum has been identified with the start of the vowel. In reality the vowel may start with a small fall, and this is particularly often the case in murmured vowels (especially for RD and GU). The form of the curve is therefore often more complex in murmured vowels ( ) than in clear vowels ( ), (see Fig. 6).

6.3.7. The sentences show the same features, but with some greater variation.

6.4. Intensity.

6.4.1. As already mentioned (5.3.) there is no consistent difference in overall intensity between clear and murmured vowels. The averages for the different speakers and recordings are given in graphical display in Fig. 17a.

6.4.2. More consistency is found in the shape of the curve. All speakers have, on the average, a longer distance from the beginning of the vowel to the peak of the curve in murmured vowels than in clear vowels, but the difference is not significant in GU's and PvB's recordings. The general averages are given in Table 7, and the averages of the individual words are shown graphically in Fig. 18 and Fig. 19 (see also Fig. 6).

Table 4. FUNDAMENTAL FREQUENCY

Minimum (Beginning)(cps).

Speakers	RD(A)	RD(T)		RT(A)	RT(T)	GU(A)	GU(T)	PBP
Number of Words	552	89	91	348	159	1oo	147	72
cps murm.	137	114	12o	133	133	131	11o	129
cps clear	148	119	126	134	136	137	117	14o
Difference	÷11***	-5***	÷6***	÷1	÷3**	-6***	-7***	÷11***
Lowering in %	8	4	4	1	2	5	6	7

Maximum (cps).

Difference	o	o	o	+2	+3	-4	-4	
Rise								
cps murm.	36	21	12	18	38	11	13	
cps clear	25	16	6	16	32	9	1o	
Difference	11***	5***	6***	2**	6***	2***	3***	

TABLE 5. FUNDAMENTAL FREQUENCY

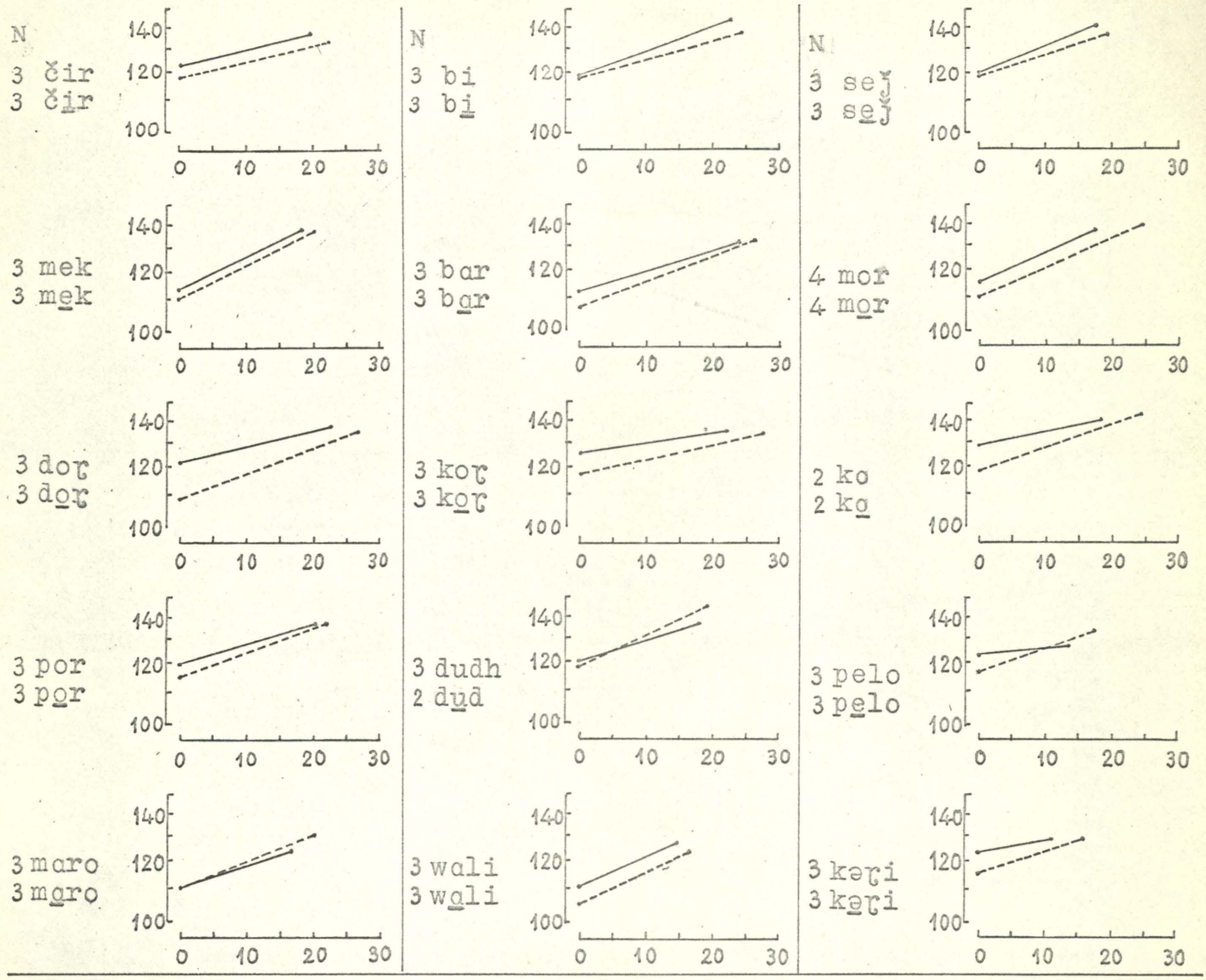
Distance beginning-maximum (cs)

Speaker	RD(A)	RD(T)		RT(A)	RT(T)	GU(A)	GU(T)
Monosyllables							
Number of Words	336	/ 66	^ 66	206	72	78	96
cs murm.	19.8	23.4	12.0	18.7	20.4	19.6	16.5
cs clear	19.2	20.3	8.7	16.9	18.0	15.6	12.2
Difference	0.6	3.1	3.3	1.8	2.4	4.0	4.3
% of duration murm.	68	85	42	85	87	60	68
% of duration clear	68	84	36	88	85	52	57
Difference	0	1	6	÷3	2	8	11
Dissyllables							
Number of Words	120	24	24	143	67	22	32
cs murm.	16.8	16.5	16.6	12.7	17.3	14.8	13.7
cs clear	14.2	13.9	13.7	9.5	12.4	13.2	12.2
Difference	2.6	2.6	3.0	3.2	4.9	1.6	1.6
% of duration murm.	94	95	100	88	100	97	100
% of duration clear	92	100	100	78	100	92	100
Difference	2	-5	0	10	0	5	0
Monosyllables + Dissyllables							
Number of Words	456	90	90	348	139	100	148
cs murm.	19.0	21.5	13.3	16.8	18.8	18.2	15.4
cs clear	17.9	18.6	10.1	14.5	15.2	15.0	12.2
Difference	1.1***	2.9***	3.2***	2.3***	3.6***	3.1***	3.2***
% of duration murm.	72	87	52	84	95	76	76
% of duration clear	72	87	46	83	93	65	68
Difference	0	0	6	1	2	11	8

Fig. 13. FUNDAMENTAL FREQUENCY

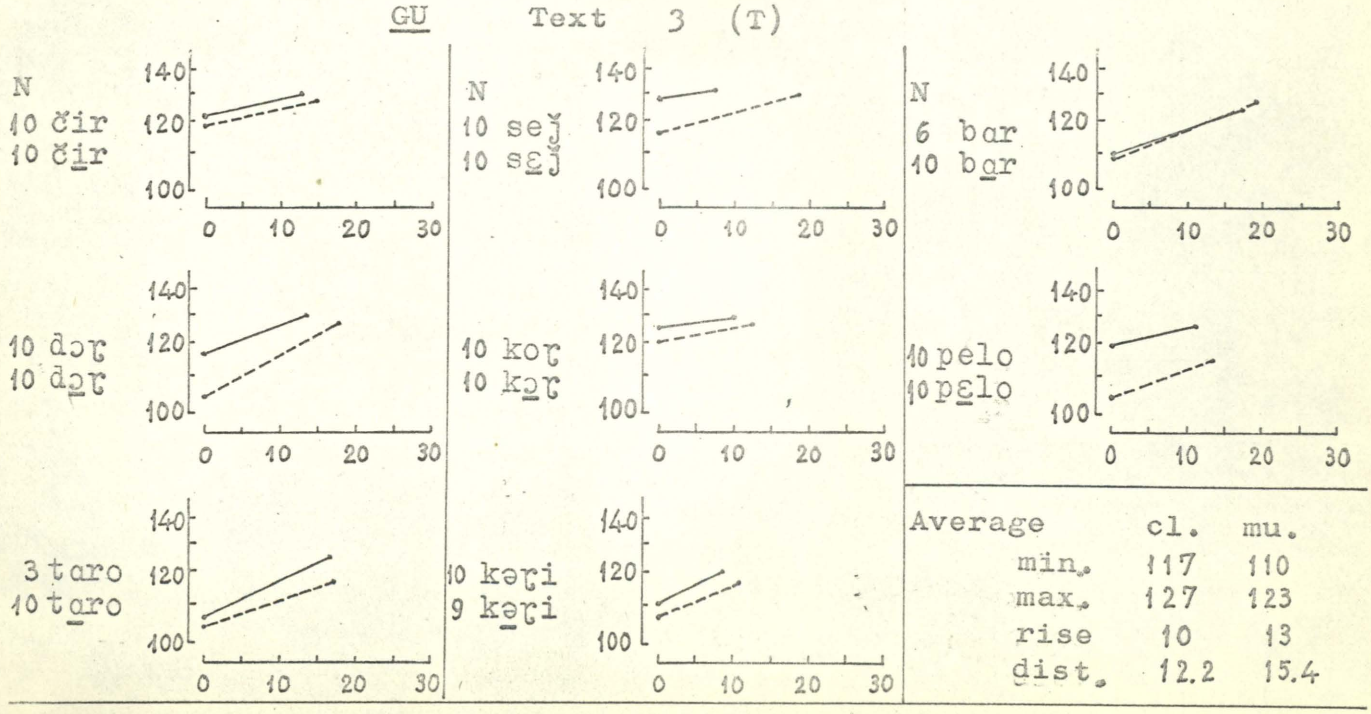
MINIMUM (cps), MAXIMUM (cps), DISTANCE MIN.-MAX. (cs)

RD Text 1 (T) (/) — =clear ---- =murmured



Average		min.	max.	rise	dist.
	cl.	119	135	16	18.6
	mu.	114	135	21	21.5

GU Text 3 (T)

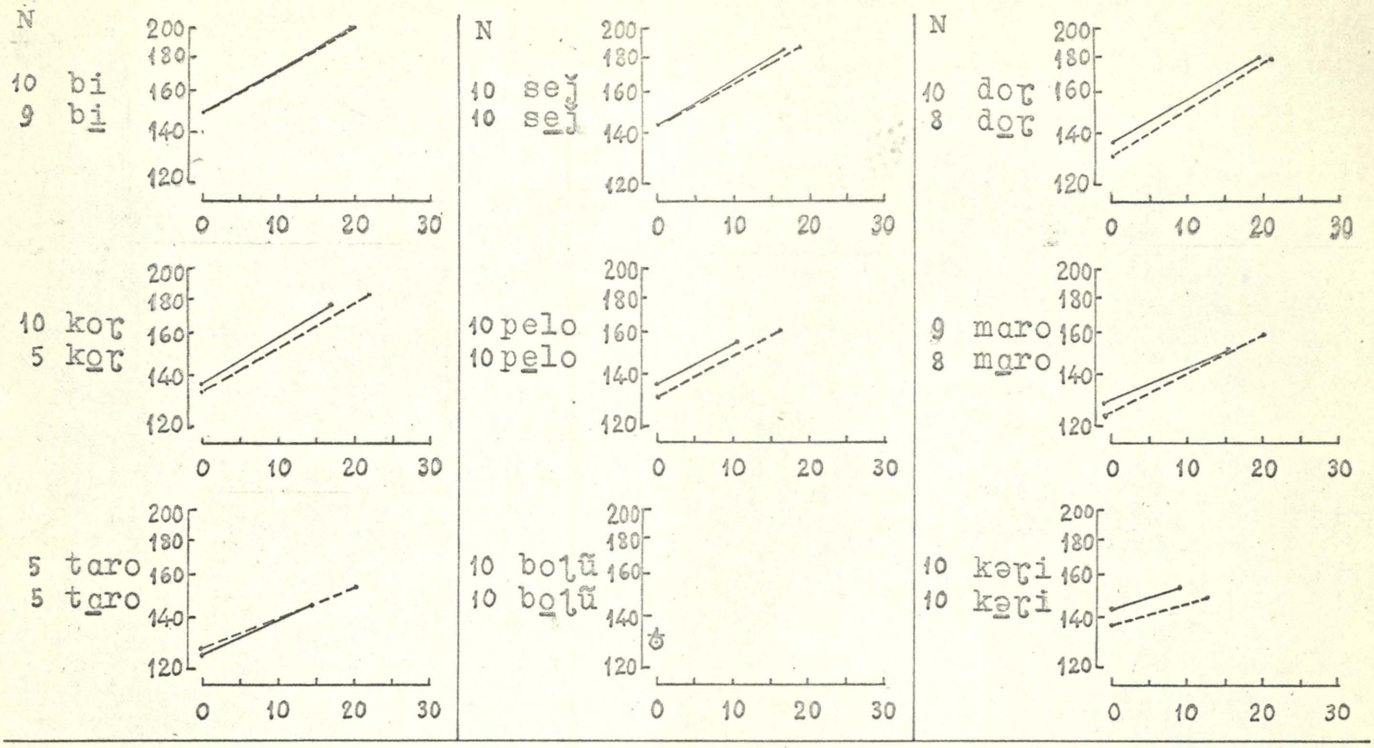


Average	cl.	mu.
min.	117	110
max.	127	123
rise	10	13
dist.	12.2	15.4

Fig. 16. FUNDAMENTAL FREQUENCY

MINIMUM (cps), MAXIMUM (cps), DISTANCE MIN.-MAX. (cs)

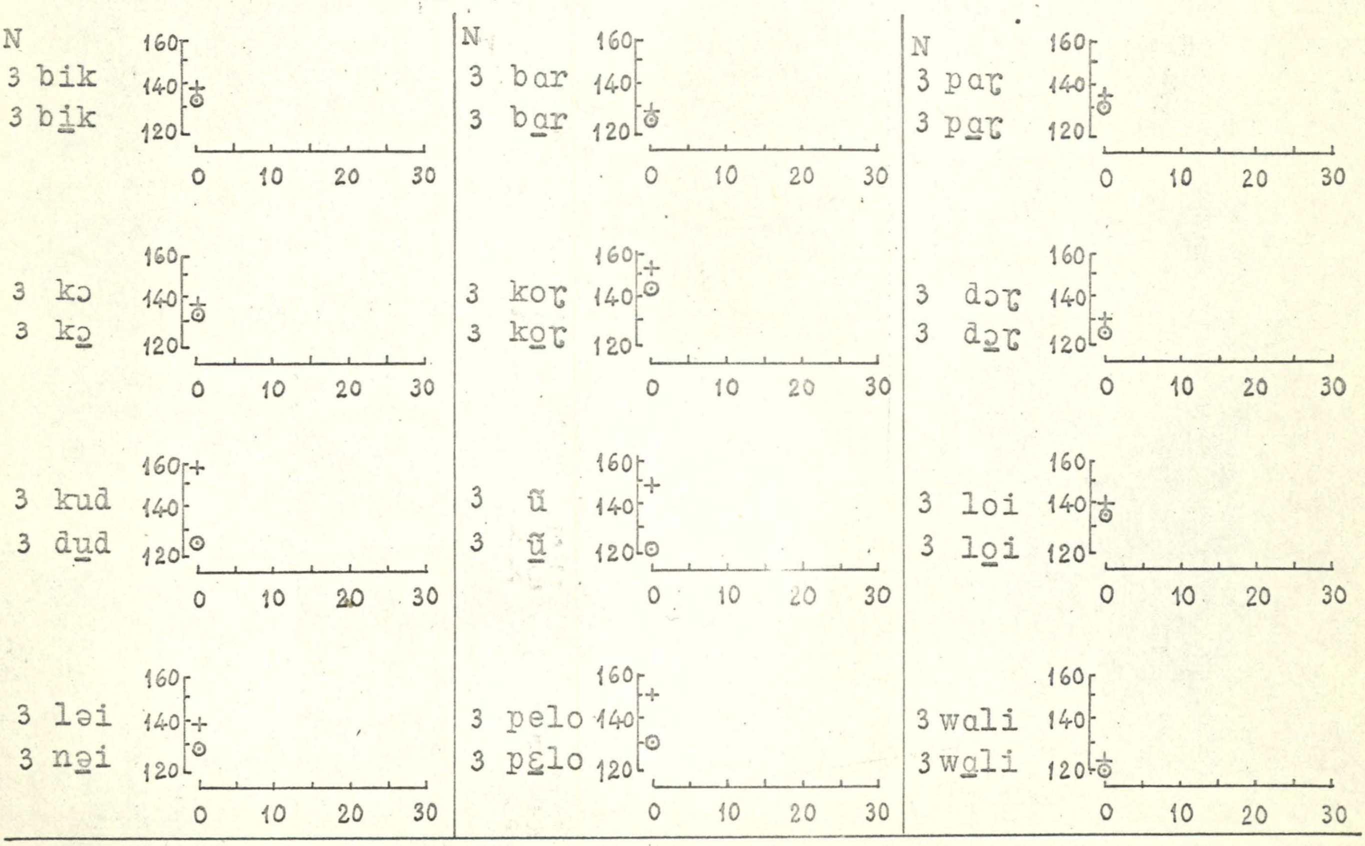
RT Text 2 (T)



Average

	min.	max.	rise	dist.
cl.	136	168	32	15.2
mu.	133	171	38	18.8

PBP Text 4



Average

	min.
cl.	140
mu.	129

Fig. 17. INTENSITY

ATTENUATION IN HIGHPASS FILTERED CURVE (⊙ = murmured + = clear)

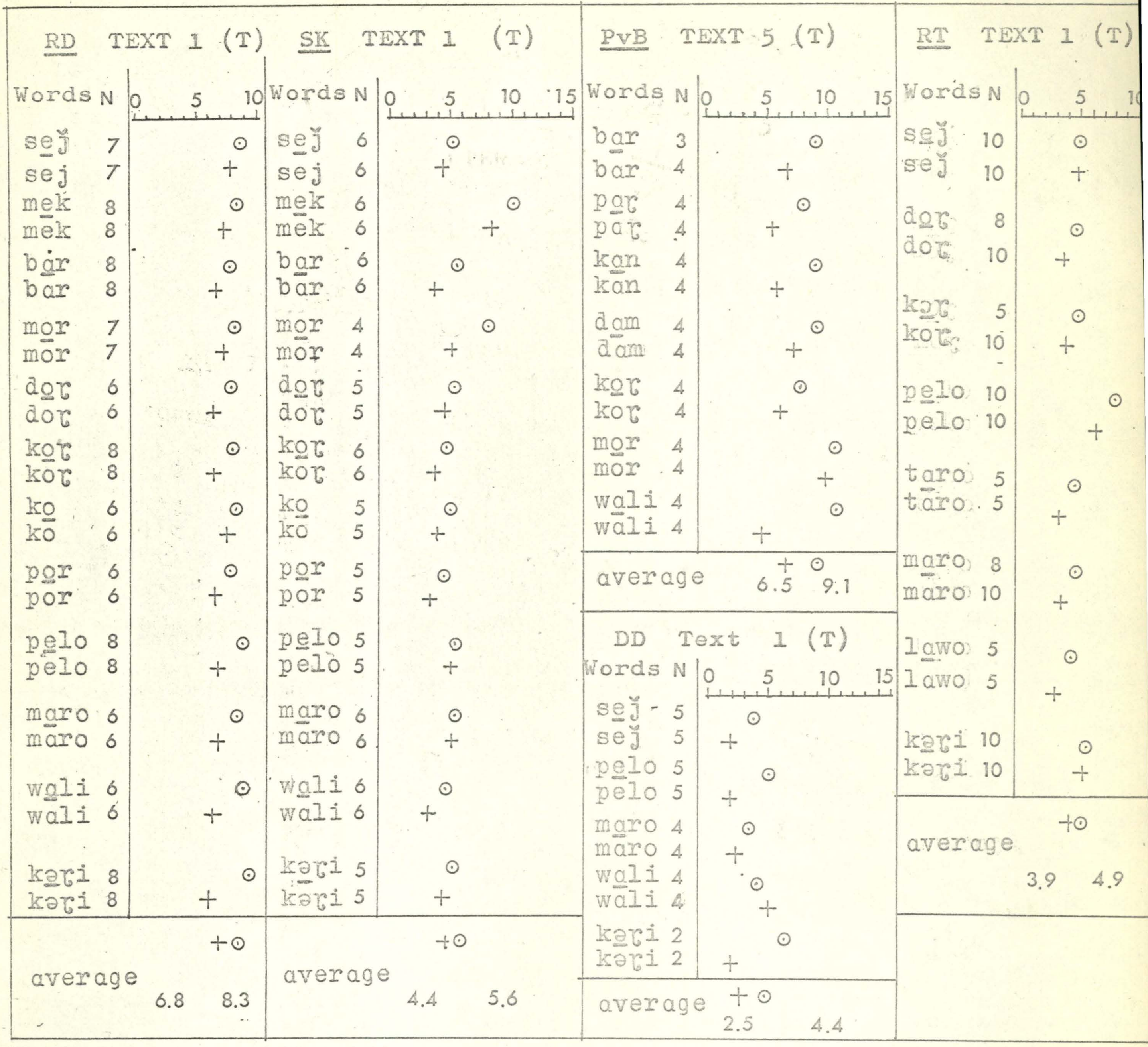


Fig. 17 a. AVERAGE INTENSITY

DIFFERENCE BETWEEN MURMURED AND CLEAR VOWELS

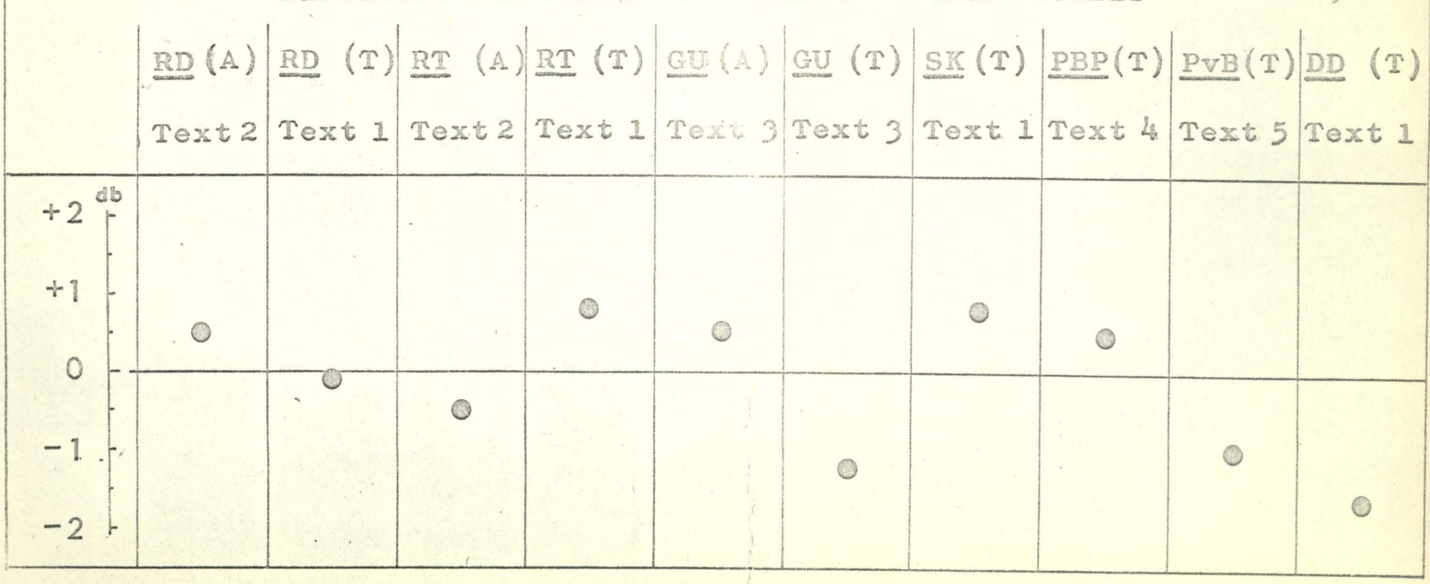


Fig. 18. INTENSITY

DISTANCE FROM BEGINNING TO MAXIMUM (CS) (⊙ = murmured + = clear)

	RD		RT	
words	Text 2 (A) N	Text 1 (T) N	Text 2 (A) N	Text 2 (T) N
čir	12	8	8	
čir	12	8	8	
bi	12	8	8	9
bi	12	8	8	10
sej	12	7	8	10
sej	12	7	8	10
mek	12	8	8	
mek	12	8	8	
mel	12		7	
mel	12		8	
wer	12		8	
wer	12		8	
bar	12	8	6	
bar	12	8	8	
paɽ	12		5	
paɽ	12		8	
mal	12		8	
mal	12		8	
mor	12	7	8	
mor	12	7	8	
doɽ	12	6	8	8
doɽ	12	6	8	10
koɽ	12	8	6	5
koɽ	12	8	8	10
ko	12	6	2	
ko	12	6	8	
por		6		
por		6		
dud	12	7	8	
dudh	12	8	8	
pelo	12	8	8	10
pelo	12	8	8	10
taro	12		8	5
taro	12		8	5
maro	12	6	8	8
maro	12	6	8	10
wali	12	6	8	
wali	12	6	8	
lawo	12		8	5
lawo	12		8	5
boɽũ	12		8	10
boɽũ	12		8	10
pɔ̃či	12		8	
pɔ̃či	12		8	
kəɽi	12	8	8	10
kəɽi	12	8	8	10
average (+ ⊙)	8.2 12.5	8.4 12.1	9.6 11.3	9.1 12.7

GU Text 3 (A)	
words	N
čir	6
čir	6
sɛj	5
sej	5
bar	6
bar	6
dɔɽ	6
dɔɽ	6
kɔɽ	6
koɽ	6
pɛlo	6
pelo	6
taro	6
taro	6
kəɽi	6
kəɽi	6
wali	4
wali	4
average (+ ⊙)	13.2 14.4

GU Text 3 (T)	
words	N
čir	10
čir	10
sɛj	10
sej	10
bar	10
bar	10
dɔɽ	10
dɔɽ	10
kɔɽ	10
koɽ	10
pɛlo	10
pelo	10
taro	10
taro	10
kəɽi	10
kəɽi	10
average (+ ⊙)	10.2 10.5

Fig. 19. INTENSITY

DISTANCE FROM BEGINNING TO MAXIMUM (CS) (⊙ = murmured + = clear)

SK Text 1 (T)

Words	N	0 5 10 15 20
čir	6	⊙
čir	6	+
bi	3	⊙
bi	3	+
sej̃	6	⊙
sej	6	+
mek	9	⊙
mek	9	+
bar	6	⊙
bar	6	+
koɽ	5	⊙
koɽ	5	+
doɽ	5	⊙
doɽ	5	+
por	6	⊙
por	6	+
mor	4	⊙
mor	4	+
ko	5	⊙
ko	5	+
dud	9	⊙
dudh	9	+
pelo	7	⊙
pelo	7	+
maro	7	⊙
maro	7	+
wali	7	⊙
wali	7	+
kəɽi	8	⊙
kəɽi	8	+
average		+ 7.4 ⊙ 8.7

PvB Text 5 (T)

Words	N	0 5 10 15 20
bar	3	⊙
bar	3	+
paɽ	4	⊙
paɽ	4	+
kan	4	⊙
kan	4	+
dam	4	⊙
dam	4	+
koɽ	4	⊙
koɽ	4	+
mor	4	⊙
mor	4	+
wali	4	⊙
wali	4	+
average		+ 4.9 ⊙ 6.3

DD Text 1 (T)

Words	N	0 5 10 15 20
sej̃	5	⊙
sej	5	+
pelo	5	⊙
pelo	5	+
maro	4	⊙
maro	4	+
wali	2	⊙
wali	2	+
kəɽi	4	⊙
kəɽi	4	+
average		+ 3.8 ⊙ 8.6

PBP Text 4 (T)

Words	N	0 5 10 15 20 25
bik	3	⊙
bik	3	+
bar	3	⊙
bar	3	+
paɽ	3	⊙
paɽ	3	+
koɽ	3	⊙
koɽ	3	+
ko	3	⊙
ko	3	+
dɔɽ	3	⊙
dɔɽ	3	+
dud	3	⊙
kud	3	+
ũ	3	⊙
ũ	3	+
loi	3	⊙
loi	3	+
nəi	3	⊙
ləi	3	+
pɛlo	3	⊙
pelo	3	+
wali	3	⊙
wali	3	+
average		+ 8.6 ⊙ 11.5

Table 7. INTENSITY

Distance from beginning to maximum (cs).

	RD(A)	RD(T)	RT(A)	RT(T)	GU(A)	GU(T)	SK	PBP	PvB	DD
Number of words	528	215	338	70	102	80	186	72	54	40
murm.	12.5	12.1	11.3	12.7	14.4	10.5	8.7	11.5	6.3	8.6
clear	8.2	8.4	9.6	9.1	13.2	10.2	7.4	8.6	4.9	3.8
diff.	4.3***	3.7***	1.7***	3.6***	1.2	0.3	1.3*	2.8**	1.4	4.8**

The words in sentences show dimilar differences. Both murmured and clear vowels have a somewhat longer distance to the intensity peak when the tone is rising than when it is rising-falling.

6.4.3. Some of the subjects (RD, PBP, partly RT) have not only a longer distance to the peak, but also a lower start and a greater rise of the intensity curve (see Fig. 6 [taro] and [paṛ]).

6.5 Spectral Structure.

6.5.1. General Remarks.

The analysis of spectral structure is based on a restricted number of words (324 in all) comprising the vowels [a a̤ i i̤ e e̤ ɛ o o̤ ɔ ɔ̤]. The only example of [u -ṳ] (dudh-dṳd) was left out as somewhat dubious, and the examples with ə-ə̤ had very short vowels and were difficult to measure.

The results are given in graphical form in Figs. 20-24. In these charts frequency is given horizontally and amplitude vertically. The scales are logarithmic. Besides the formants (F_1, F_2, F_3, and F_4) the fundamental (F_o) is given and also the second and third harmonics (H_2 and H_3) in the cases where they do not enter into F_1. Two separate peaks in the region of F_4 have been indicated as F_{4a} and F_{4b}. N indicates the number of word pairs, (in contradistinction to the other graphs, where it indicates the number of single words). The transcription with [ɛ] and [ɔ] is only used where the speaker distinguishes between the phonemes /e/, /o/ and the phonemes /ɛ/, /ɔ/. Thus /ɔ/ need not indicate a more open vowel than [o]. Close and open e and o-sounds are treated together in the following sections as [e] and [o]. Words containing the same vowels have been combined in one figure, if they did not show any appreciable difference in formant structure.

Fig. 2o. SPECTRAL STRUCTURE

Frequency and Amplitude of the Fundamental (F_o), the Second and Third Harmonics (H_2 and H_3), and F_1 - F_4

Frequency ⟶ (cps), Amplitude ↑ (db). a - a.

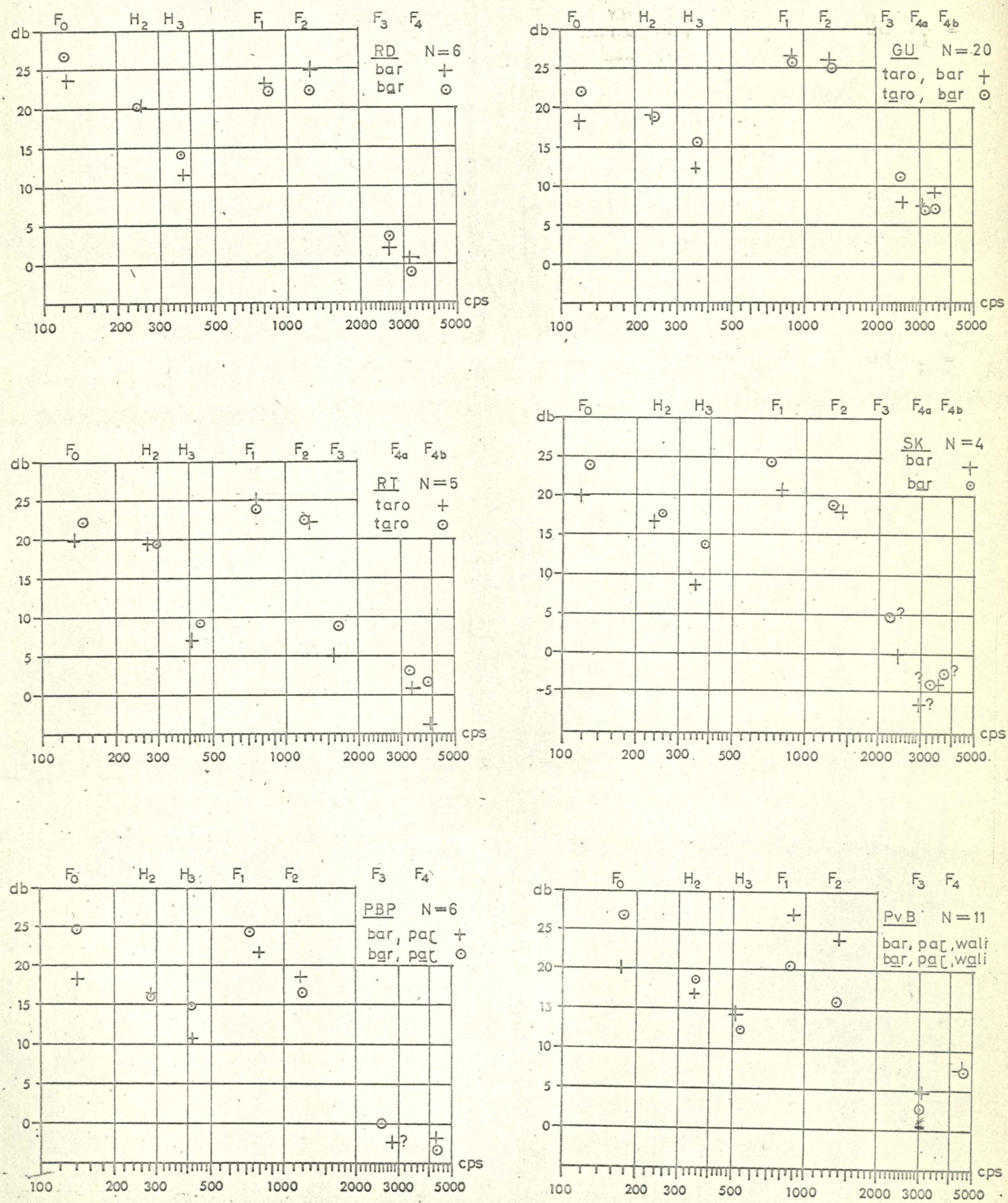
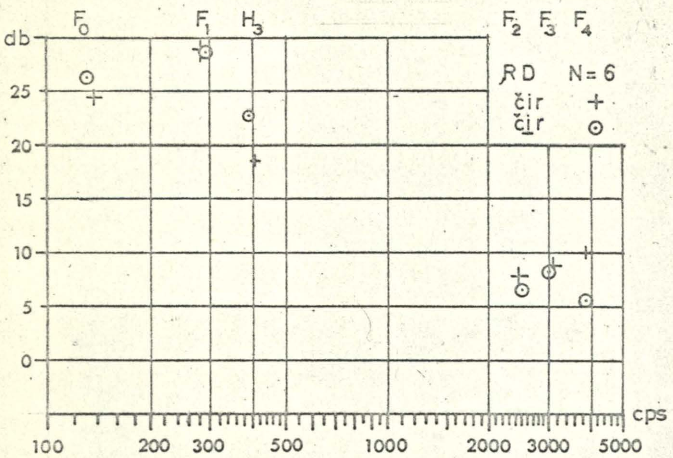
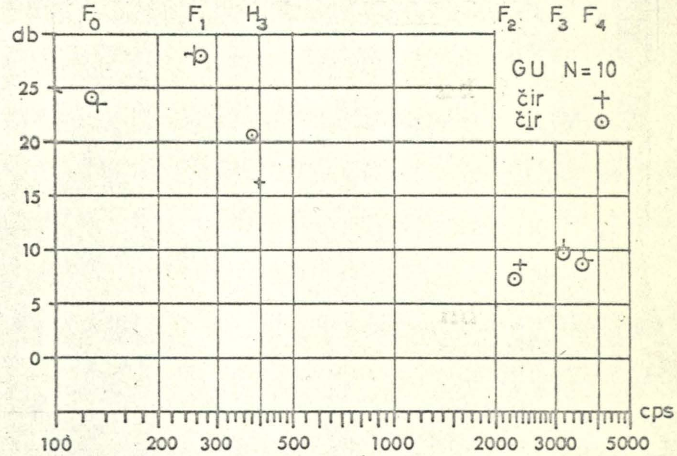


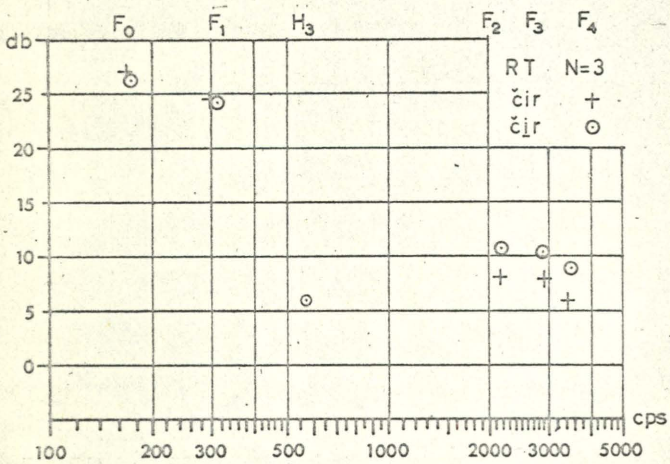
Fig. 21. SPECTRAL STRUCTURE

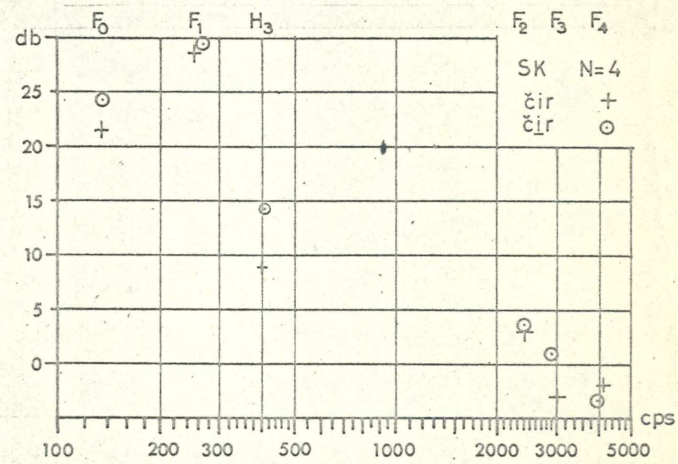
Frequency and Amplitude of the Fundamental (F_o) and F_1 - F_4

Frequency ⟶ (cps), ↑ Amplitude (db). i - i̲ .









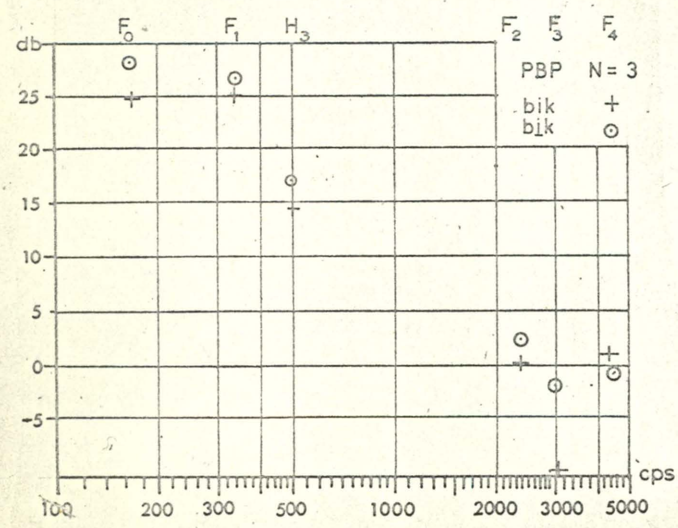
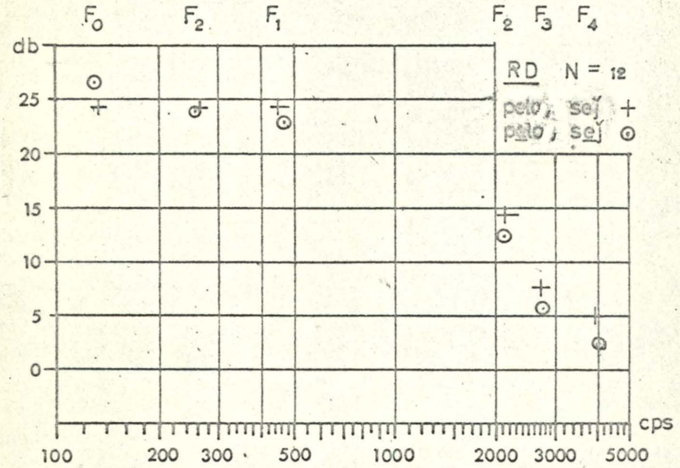
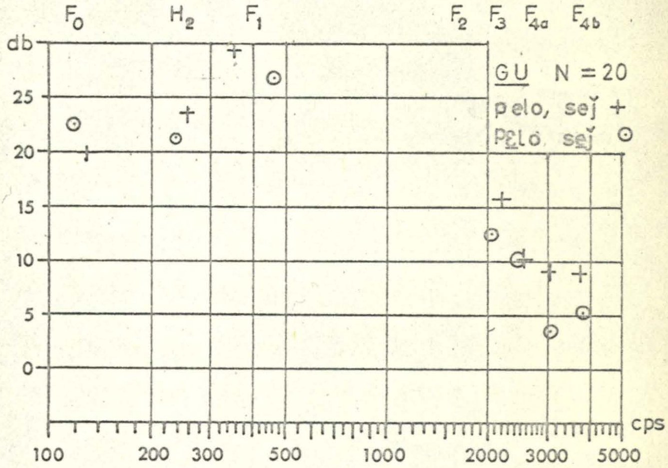


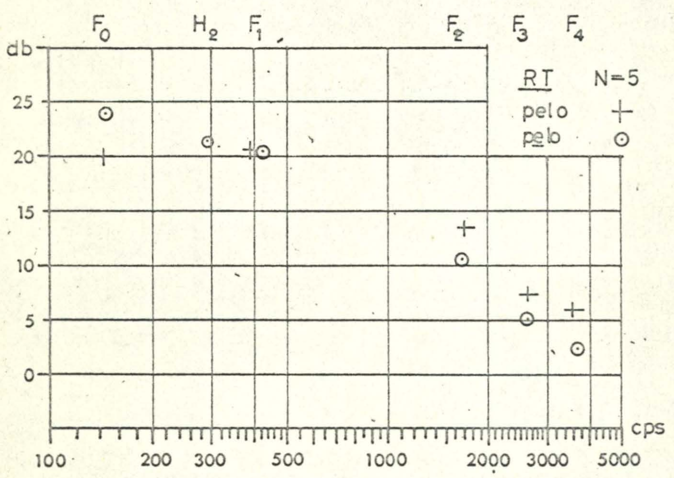
Fig. 22. SPECTRAL STRUCTURE

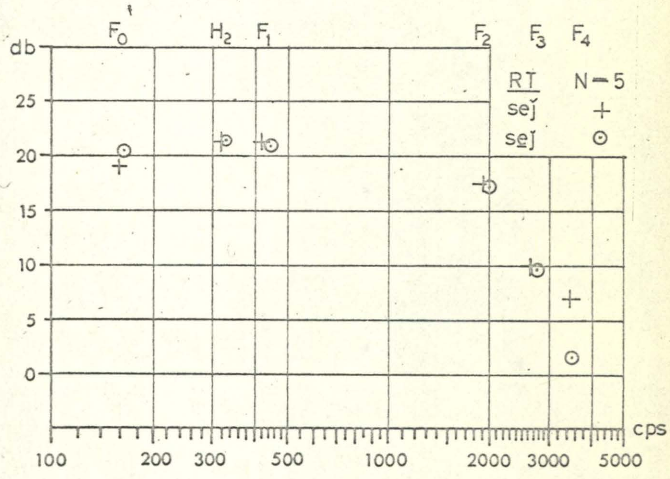
Frequency and Amplitude of the Fundamental (F_o), the Second Harmonic (H_2), and F_1 - F_4

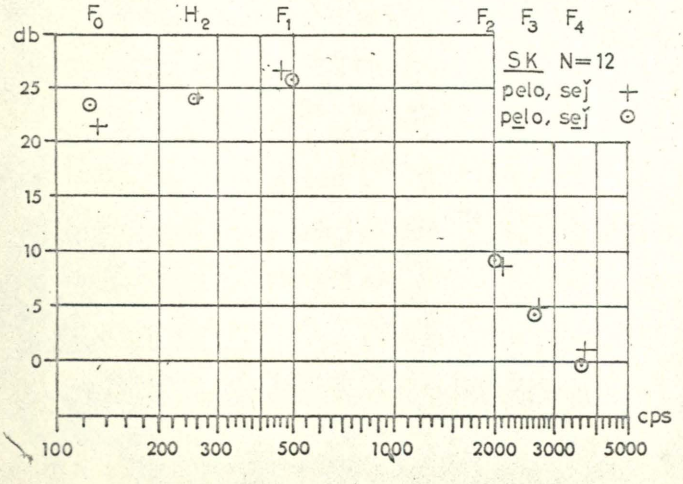
Frequency ⟶ (cps), ↑ Amplitude (db) . e - e̱.











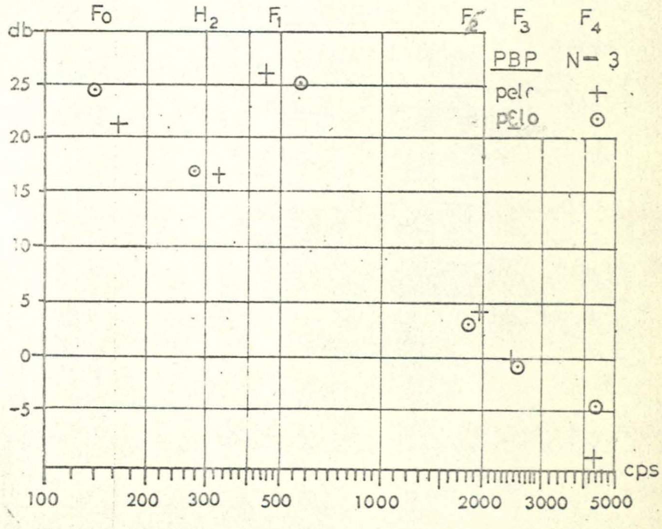
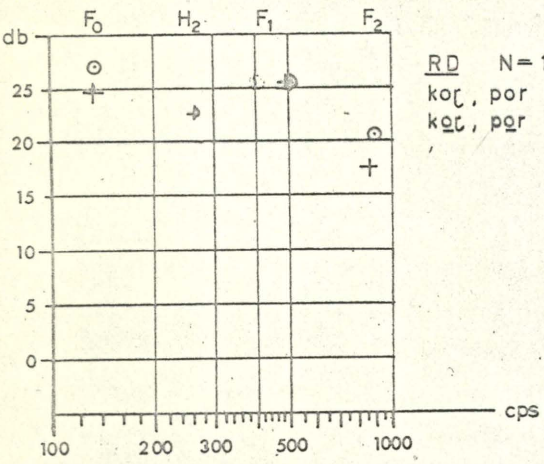
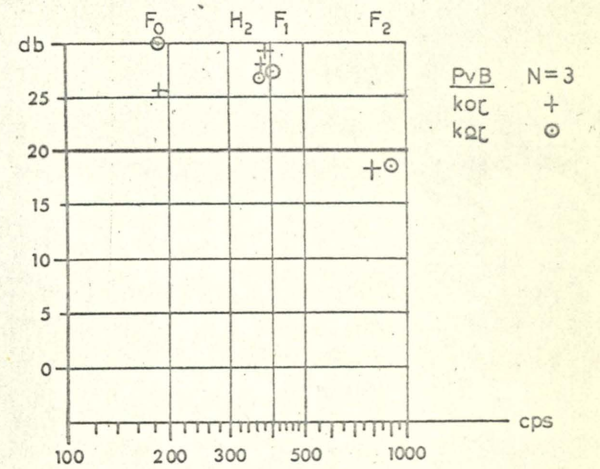


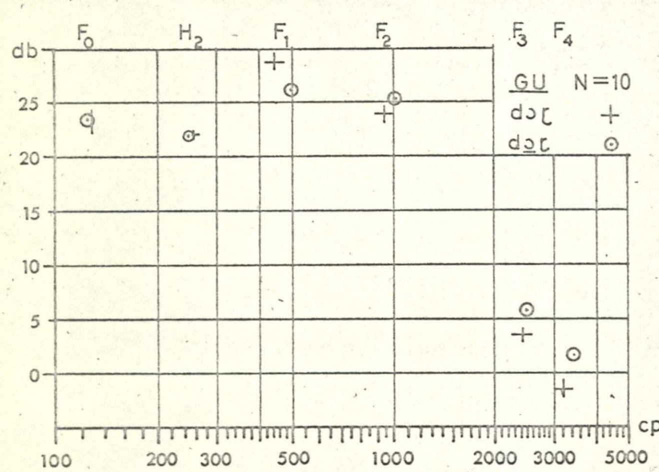
Fig. 23. SPECTRAL STRUCTURE

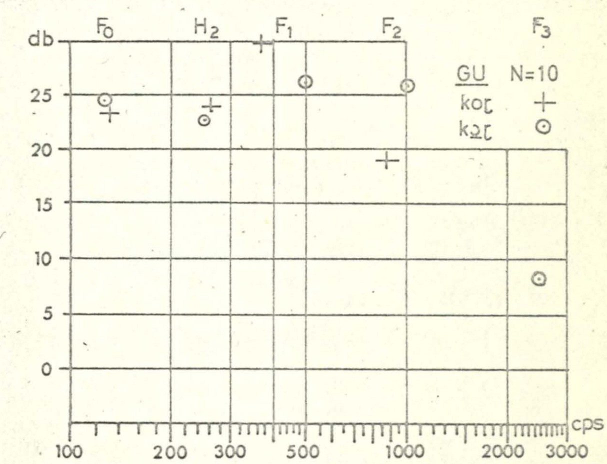
Frequency and Amplitude of the Fundamental (F_o), the Second Harmonic (H_2), and F_1 - F_2 (- F_4)

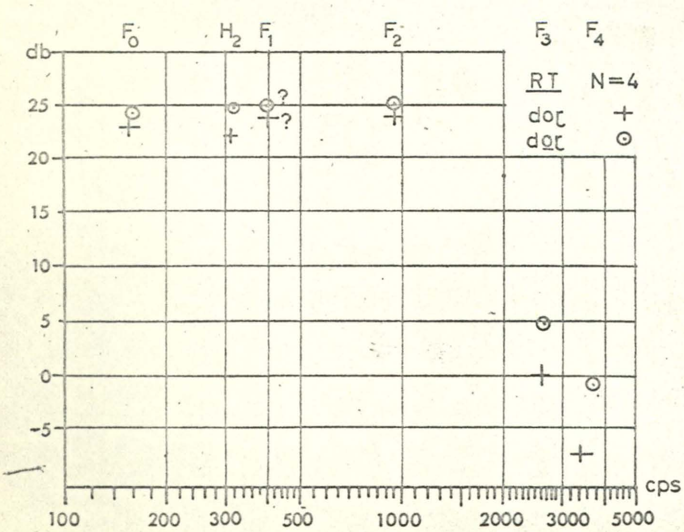
Frequency ⟶ (cps), ↑ Amplitude (db). o -o̲ .











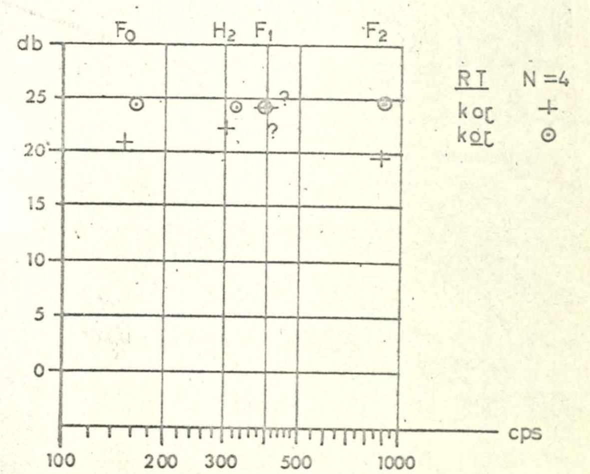
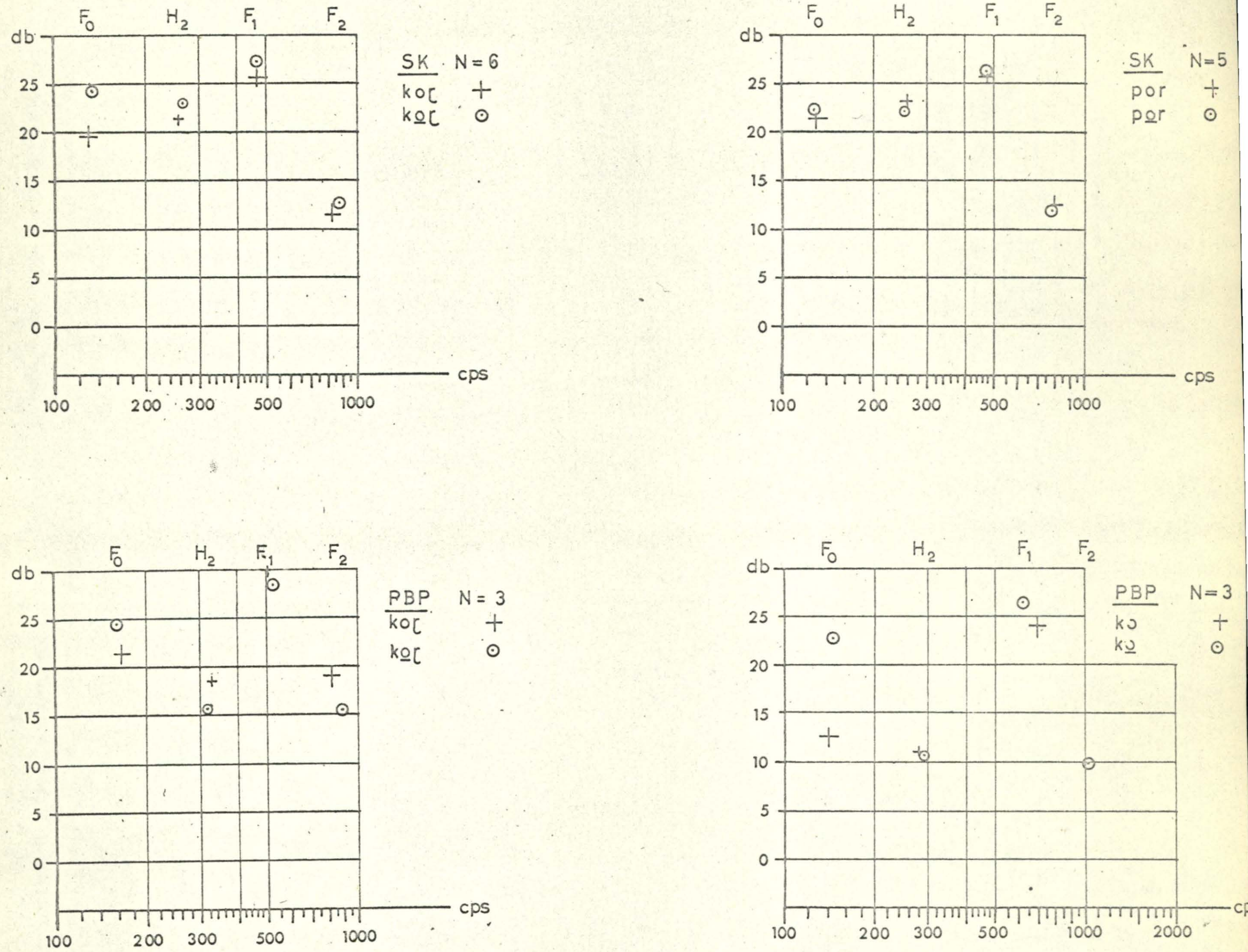


Fig. 24. SPECTRAL STRUCTURE

Frequency and Amplitude of the Fundamental (F_o), the Second Harmonic (H_2), and F_1 - F_2

Frequency ⟶ (cps), ↑ Amplitude (db) . o - o̲ .



6.5.2. Formant Frequency.

Dave did not find any consistent differences between the formant frequencies of clear and murmured vowels (2). My measurements have confirmed this result. The only difference common to all speakers is a somewhat higher F1 in [e̲] than in [e]. Four of the six speakers have also a higher F1 in [o̲], four have a lower F2 in [e̲] and five a higher F2 in [o̲]. This means that there is a tendency for [e̲] and [o̲] to be somewhat more open than clear [e] and [o]. The tendency to a shift up of F1 is what should be expected as a result of a larger coupling to the trachea (7), but it must be noted that the tendency is not found in [a-a̲], that it is very slight in [i-i̲] and also very slight for RD, SK and RT in [e-e̲] and [o-o̲]. The speakers GU, PBP, and PvB have a much more pronounced difference, but these speakers distinguish between close and open /e/ and /o/, and in most cases it would be more correct to say that they have the phonemes/e/ and /o/ in words with clear vowel and the phonemes /ɛ/ and /ɔ/ in words with murmured vowel. The examples /dɔɽ/ (GU) and /ko̲ɽ/ (PBP) show that the degree of opening is not a mechanical consequence of the murmur. In the cases where the difference is pronounced it should thus not be considered as a synchronic phonetic difference, accompanying the difference between clear and murmured vowels, but as a result of a diachronic development. It is, however, interesting that a former intervocalic /h/ (which is the historical source of the murmur) has also been considered as one of the sourcesof the open /ɛ/, e.g. in /pɛ̲lo/ (8).

6.5.3. Distribution of Spectral Energy.

The most obvious spectral characteristic of murmured vowels is the relatively high level of the fundamental compared to F_1-F_4.

6.5.3.a. A rough quantitative picture of this difference can be obtained from a comparison between the highpass filtered intensity curve, used for delimitation of the mingograms, and the intensity curve without highpass filtering.

By subtracting the attenuation caused by highpass filtering of oral vowels from that of murmured vowels, one gets a measure of the greater predominance of the lowest harmonics in the latter. Table 8 gives the general averages, and Fig. 17 the averages of individual word pairs in graphical form.

TABLE 8. REDUCTION IN LEVEL BY HIGHPASS FILTERING
Difference between murmured and clear vowels. (db).

Speakers	RD(T)	RT(T)	SK	PvB	DD
Number of words	174	131	130	55	40
Difference murm.-clear	1.5***	1.0***	1.2***	2.6***	1.9***

Only the T-recordings, which comprise the two types of intensity curves for the same words, have been utilized. The close vowels /i-i/ and /u-u/ have been left out, because in these vowels most of the first formant was removed too. GU's and PBP's highpass filtered curves could not be used because of a technical mistake.

6.5.3.b. The spectrograms have been analysed in more detail. The results are given in graphical form in Fig. 20-24 (vertical dimension), and Table 9 gives the differences between murmured and clear vowels as regards the absolute level of F_o (L_o) and the relative level of F_o compared to the levels of the second harmonic and the level of F_1-F_4, i.e. L_o-$L(H_2)$ and L_o-L_1, L_o-L_2, L_o-L_3, and L_o-L_4.

It appears from Table 9 and from the graphs Figs. 20-24 (see also 1-3) that the absolute level of F_o is higher in murmured than in clear vowels, and that L_o-L_1, L_o-L_2, and L_o-L_4 is higher in the murmured vowels [a, e] and [i] than in the corresponding clear vowels, whereas for [o] only L_o-L_1 is higher. L_o-L_3 is variable except for [e]. Some older spectrograms of [u-u] show the same tendency as found here for [o-o].

The relative drop of intensity above F_o is evident already in the second harmonic; L^o-$L(H_2)$ is higher in murmured vowels than in clear vowels. The spectra of [a-a] and [i-i] show that this is not true of the third harmonic (in e- and o-sounds the problem is more complicated because H_3 is part of F_1).

This fact may be interpreted in different ways. (i) One might imagine a larynx spectrum with an increase of intensity in the odd harmonics. This assumption, however, was not corroborated by an inspection of the sections. (ii) It might be due to a broadening of the first formant of i and of the subformant of a, and the increase of F_o might be part of this broadening. But a number of extra recordings of bar-bar spoken by RD and GU, partly with rising-falling fundamental, partly on a very low fundamental showed that at low

Table 9. SPECTRAL AMPLITUDE.

Difference between murmured and clear vowels.

(Murmured - clear)(dB)

[a̱-a]

Speakers	RD	RT	GU	SK	PBP	PvB
Words	ba̱r/bar	ta̱ro/taro	ba̱r/bar ta̱ro/taro	ba̱r/bar	ba̱r/bar pa̱ɽ/paɽ	ba̱r/bar pa̱ɽ/paɽ wa̱li/wali
N	12	10	40	8	12	22
L_o	3.0	2.5	3.7	4.1	6.2	6.7
$L_o-L(H_2)$	3.2	2.6	3.9	3.1	6.7	5.0
L_o-L_1	4.1	3.8	4.6	0.3	3.6	13.7
L_o-L_2	5.7	2.3	4.9	3.4	8.4	14.6
L_o-L_3	1.5	-1.3	0.5	-0.5	4.0	8.6
L_o-L_4	4.9	0.2	5.9	1.4	7.4	3.3

[i̱-i]

Speakers	RD	RT	GU	SK	PBP
Words	či̱r/cir	či̱r/cir	či̱r/cir	či̱r/cir	bi̱k/bik
N	12	6	20	8	6
L_o	1.9	-0.9	0.6	2.8	3.5
$L_o-L(H_2)$	3.4	-0.9	0.9	2.4	1.7
L_o-L_1	2.0	-0.5	0.4	1.9	1.8
L_o-L_2	3.1	-3.6	1.9	2.2	1.3
L_o-L_3	2.5	-3.4	1.0	-1.2	-4.3
L_o-L_4	6.3	-3.8	1.3	4.3	5.3

[e̱-e]

Speakers	RD	RT		GU	SK	PBP
Words	se̱ǰ/seǰ pe̱lo/pelo	se̱ǰ/seǰ	pe̱lo/pelo	se̱ǰ/seǰ pe̱lo/pelo	se̱ǰ/seǰ pe̱lo/pelo	pɛ̱lo/pelo
N	24	6	6	40	24	6
L_o	2.1	1.4	3.7	2.4	1.9	3.0
$L_o-L(H_2)$	2.4	1.2	3.5	4.8	1.8	3.0
L_o-L_1	3.5	1.6	3.9	5.1	2.7	3.9
L_o-L_2	3.9	1.7	6.8	5.7	1.4	4.1
L_o-L_3	3.3	1.9	5.9	2.6	2.4	3.7
L_o-L_4	4.7	6.8	7.4	8.1	3.5	-1.5

[o̱-o]

Speakers	RD	RT		GU		SK		PBP		PvB
Words	ko̱ɽ/koɽ po̱r/por	ko̱ɽ/ koɽ	do̱ɽ/ doɽ	kɔ̱ɽ/ koɽ	dɔ̱ɽ/ doɽ	ko̱ɽ/ koɽ	po̱r/ por	ko̱ɽ/ koɽ	kɔ̱/ kɔ	ko̱ɽ/ koɽ
N	24	8	8	20	20	12	10	6	6	8
L_o	2.4	3.4	1.3	1.1	0.2	4.6	0.9	3.0	10.0	4.3
$L_o-L(H_2)$	2.3	1.5	-1.2	2.5	0.5	3.0	2.0	6.0	10.3	5.5
L_o-L_1	2.5	3.2	0.2	4.7	2.8	3.0	0.3	4.5	7.5	5.5
L_o-L_2	-0.5	-1.8	0.1	-5.7	-1.1	3.5	1.5	7.5	?	4.1
L_o-L_3			-3.3	-8.8	-2.0					
L_o-L_4			-5.0	-8.7	-2.9					

fundamental frequencies (100-110 cps) there is a valley between the fundamental and the subformant at 250-300 cps, and the difference between clear and murmured vowels lies, in these cases, in the fundamental.

(iii). It is thus more probable that the increase in the level of F_o is due to a change in the voice source spectrum, caused by a relaxation of the glottis, and that the increase in the level of H_3 compared to H_2 is due to an increase in the frequency of the subformant in[a] and of F_1 in[i](which is thus probably slightly higher than indicated in the table and graphs). This would also fit with the assumption of a rise in F_1 in breathy vowels (cp. 6.5.1)

It has only been possible to apply significance tests to a restricted part of the material because of the small number of pairs. But the differences have been found to be significant for the absolute level of F_o (L_o), for L_o-$L(H_2)$, and L_o-L_1 for o-o and [e-e](GU and RD) and for[a-a](GU and PvB), and moreover for the difference L_o-L_2 and L_o-L_4 for [e-e](GU and RD) and [a-a](GU and PvB).

6.5.4. Other Spectral Characteristics.

(a) It would be natural to expect more noise in murmured vowels because of the stronger airstream. Pandit has also found some random distribution of energy particularly at higher frequencies in [pɔr] compared to [pɔr] (1,p. 172). In the present material noise at higher frequencies was clearly seen in PvB's murmured vowels but in spite of the fact that special curves were taken with this purpose it was rarely found in PBP's and RD's vowels and hardly ever in those of the other informants.

(b) It has been assumed that breathy vowels should have broader formants (3). A certain broadening can sometimes be seen in the sections, but by no means always. The oscillograms taken with 1/4 speed of the tape (4.4) show more damping of the single periods in [a] than in [a] for four of the informants, but very little for other vowels.

(c) The oscillograms show a tendency to asymmetry in the murmured vowels of some informants, the amplitude of the first positive deflection of each period being stronger than the corresponding negative deflection. This may have something to do with the damping or perhaps, according to a suggestion made by Gunnar Fant, to a slower closing of the vocal cords.

(d) The fundamental frequency curve is often more smooth for murmured vowels (see Fig. 6). This is particularly evident for informants with hoarse voice, and is probably due to a more relaxed glottis.

(e) A later start or a momentary weakening of higher formants (F_3 and F_4, and sometimes F_2) is often found in RD's murmured vowels (see Fig. 1), and sometimes in the vowels of other speakers.

(f) A less sharp delimitation between vowel and [r ɽ l ɭ] pointing to a less precise articulation is a relatively frequent feature (see Fig. 6 taro-taro).

7. COMPARISON BETWEEN MURMURED VOWEL AND [h].

As murmured vowels have come into existence through a fusion of vowel and [h], a comparison between murmur and [h] may be of interest. Voiced [h] is found in Gujarati after voiced stops. Figs. 8a and 8a contain mingograms and spectrograms of [bh-]. Fig. 8a shows that [h] has a very strong F_o, weak higher formants and some noise at higher frequencies. Figs. 7a and b contain curves of intervocalic [h] spoken by RT instead of murmured vowels, the spectrogram (7b) shows the same strong intensity of F_o and weakness of higher formants as 8a, Fig. 7a an increase of airflow combined with a decrease of intensity and of fundamental frequency compared to the surrounding vowels. (The [h] of [kəho] is, in all respects, stronger than that of [pəhaɽ]). The strong airflow, low frequency, and the relatively strong intensity of F_o have all been found as characteristics of murmured vowels. The cases with a drop of frequency and low intensity in the beginning often found in RD's curves of murmured vowels are signs of an incomplete fusion of [h] with the vowel, so that the murmur-element is stronger in the beginning (see Fig. 6).

Ilse Lehiste (9) has measured the airflow of voiceless [h] and found it to be very strong (this has also been found for /h/ in other languages, it is e.g. obvious in Danish for both voiced and voiceless [h]). Moreover she found a pronounced weakening of F_1, and a rise in frequency, in the cases where F_1 was visible (this has also been seen in spectrograms of Danish [h]).

The difference between murmured and clear vowels in Gujarati is neutralized after aspirated consonants. The vowel found in this position is considered as clear, but curves of vowels preceded by aspirated consonants spoken by PBP, PvB, and RD show a certain assimilation of the beginning of the vowel to [h] : the fundamental is stronger, the airflow stronger, and there may be some noise at higher frequencies (the latter phenomenon is particularly obvious in PvB's curves). - For the stronger fundamental see Fig. 8b.

8. THE RELATIVE IMPORTANCE OF THE ACOUSTIC CUES.

8.1 Stability.

As one of the possible criteria of stability we have chosen the percentage of individual word pairs characterized by the difference in question; as pairs are considered the words of the same reading standing under the same conditions. The percentages are listed in Table 10.

Table 10. STABILITY

Percentage of word pairs characterized by the different acoustic cues.

Speaker	RD(A)	RD(T)	RT(A)	RT(T)	GU(A)	GU(T)	SK	PBP	PvB	DD
1. Duration	68	90	86	85	75	80	74	63	77	(67)
2. Frequency		´ ˆ								
a Minimum	89	71 67	35	54	81	76		71	48	
b Rise	83	69 71	51	72	48	56		73	67	
c Distance to peak	66	77 84	81	85	79	76				
Intensity										
Distance to peak	74	72	60	81	75	63	66	66	55	(67)
Distribution of spectral energy										
L_o		92		77		79	78	94	100	
$L_o - L(H_2)$		89		64		83	74	89	93	
$L_o - L_1$		94		69		90	78	78	100	
$L_o - L_2$		67		55		69	74	83	100	
		(87)		(57)		(91)	(69)	(75)	(100)	
$L_o - L_4$		91		65		82	78	83	91	

The number of words on which the percentages are based can be seen in tables 2-9. The numbers in parentheses under L_o - L_2 for RD and GU are the percentages for [a-a, e-e, i-i] without [o-o]). The percentages above 70 have been underlined. The percentage for L_o - L_4 is based on [a-a, e-e, i-i] alone.

Besides the cues listed RD and RT have a relatively high stability for the extent of the rise of the intensity curve, the rise being more extensive in murmured vowels. The percentages are: RD(A) 77, RD(T) 81, RT(T) 71, but RT(A) 41.

The sentences have somewhat higher percentages, particularly for RD(A) and RT(A).

8.2. Independency.

The question must be raised whether the acoustic cues found are mutually independent. If not, it is problematic to consider them as separate cues. However, no constant correlations have been found. It is only necessary to make some reservations for the distance to the frequency peak. It is in principle not dependent on duration, but the size and significance of the difference may be partly due to differences of duration, because the duration simply sets a limit to the place of the peak. This is certainly true of most of the dissyllables and also of the monosyllables with rising intonation in RD's T-recording and in both of RT's recordings.

8.3 Influence on Perception.

It has only been possible to investigate this very important aspect of the question in a very preliminary and insufficient way, in the first place because our synthesizer is under construction, and its possibilities are restricted, and secondly because we had only two Gujarati listeners (RD and RT) at our disposal.

8.3.1 Synthesis. The present state of our synthesizer allowed us to synthesize isolated vowels containing up to five formants which could be varied independently in frequency, bandwidth and amplitude, and a low frequency channel which permitted the variation of the amplitude of the fundamental and of the second harmonic. The vowels were composed of four normal formants and a subformant below F_1. The frequency contour of the fundamental could be given a fixed inflection (either level or slightly rising) or it could be produced manually, but of course not with precision. - Rischel succeeded without difficulty in synthesizing a clear [a] matching the spectral characteristics of Dave's [a] closely, which Dave recognized as a good Gujarati [a]. It sounded even very much like Dave's voice. But it proved impossible to produce a vowel which he would recognize as murmured, although it was possible to simulate all the amplitude relations found in the previous analysis. Only when these characteristics were exaggerated, he found that the vowel approached a murmured vowel very slightly. This seems to indicate that the specific spectral structure mentioned in 5.3.4 is not sufficient for identification. If F_1 was weakened too much, the vowel sounded nasal.

Addition of noise in higher formants helped a little, but the vowel was still not recognized as murmured. - It was also attempted to mix an [h] spoken by Dave (in [aha]) with the best of the synthesized vowels, but without result (noise and vowel did not fuse completely). The same was tried with a natural [a], but with the same negative result. The synthetic vowels had slightly rising tone. Some rather primitive attempts at making a frequency or an intensity dip in the beginning did not give any improvement.

8.3.2 Listening to unchanged words in random order.

a. Each speaker had listened to his own tape recording and found it satisfactory. Moreover Dave had listened carefully to all recordings, and the words which he did not find quite murmured (this was a small number) or which he perceived as containing vowel + h + vowel (a good deal of RT's words) were discarded before the acoustic analysis. The material should thus be satisfactory. But the perceptual differences were small, even to a trained ear, and it was therefore decided to try whether the Gujarati speakers themselves could distinguish the words when they were given in random order. A test was prepared containing a number of words with different vowels from RD's, RT's, GU's, PBP's, and PvB's recordings. It was attempted to include words with different predominant cues. The test contained 45 different words spoken by RD, 36 by GU, 25 by RT, 16 by PBP, and 16 by PvB. The words spoken by different speakers were kept apart in separate groups. Each word was repeated twice with one second's interval and there was a pause of 5 seconds between the different words. A number was spoken by me before each word. The order of the words was quasi-randomized. Words belonging to the same word pair were never given in succession, and words with the same vowel were only rarely brought in succession. There was thus, generally, at least two other words between e.g. [bar] and [bar]. RD (Dave) and RT acted as listeners. Dave was asked to indicate on a sheet whether the vowel of the word in question was murmured or clear, RT (who is not a linguist) was asked to identify the word as one of two words written in Gujarati orthography on the sheet after each number. He was asked to underline the word he heard.

b. Table 13 contains the number of correctly identified words.

Table 13. IDENTIFICATION OF UNCHANGED WORDS BY DIFFERENT SPEAKERS AND LISTENERS.

Speaker		RD			RT			GU			PBP			PvB		
		mu.	cl.	tot.	mu.	cl.	tot.	mu.	cl.	tot.	mu.	cl.	tot.	mu.	cl.	tot.
Number of words presented		25	20	45	14	11	25	21	15	36	8	8	16	9	7	16
Number correctly identified	RD	24	18	42	6	9	15	12	14	26	4	8	12	8	7	15
	RT	17	16	33	8	10	18	14	11	25	6	8	14	6	7	13
Percentage correctly identified	RD	96	90	93	43	82	60	57	93	72	50	100	75	89	100	94
	RT	68	80	75	57	91	72	67	73	69	75	100	88	67	100	81

The most astonishing result is the low percentage of murmured vowels identified. It is the more astonishing as the listening conditions should be good. High quality tape recorders and head phones were used, and the level was chosen so as to be convenient for the listeners, and RD knew all the speakers (except PvB) and had worked with his own recordings and those of RT. Moreover, the words not identified by RD and RT are often different, so that the percentage of murmured vowels identified by both is still lower. For the different speakers the percentage is: RD 65, RT 29, GU 33, PBP 25, PvB 67. One of the reasons is that the two listeners have difficulty with different vowels. In table 14 the results are grouped according to different vowels instead of different speakers.

Table 14. IDENTIFICATION OF UNCHANGED WORDS FOR DIFFERENT VOWELS AND LISTENERS.

Vowel		$\underline{i}$	i	tot.	$\underline{e}$	e	tot.	$\underline{a}$	a	tot.	$\underline{o}$	o	tot.
Number of words presented		18	12	30	14	13	27	25	19	44	20	17	37
Number correctly identified	RD	16	11	27	13	9	22	15	19	34	10	17	27
	RT	6	10	16	12	7	19	20	18	38	13	17	30
Percentage correctly identified	RD	89	92	90	93	69	81	60	100	77	50	100	73
	RT	33	84	53	86	54	73	80	95	86	65	100	81

It appears from the tables and graphs that clear vowels are generally much better identified than murmured vowels, i.e. they are rarely heard as murmured. The opposite difference in RD's perception of his own vowels is too small to be of any significance. The only real exception is that there is a tendency to perceive clear [e] as murmured (and murmured [$\underline{e}$] is perceived correctly more often than other vowels.

Generally, however, it seems as if murmured vowels are conceived as the marked member of the opposition. One might have expected the opposite, at any rate in the case of RT, since many Gujarati speakers have free variation between clear and murmured vowel in e.g. /b$\underline{a}$r/, but only clear vowel in /bar/, so that a perceived [b$\underline{a}$r] can only be interpreted as /b$\underline{a}$r/, whereas a perceived [bar] could be interpreted as both /bar/ and /b$\underline{a}$r/. But both listeners seem to have identified a vowel as murmured only in the cases where they perceived some positive indication of murmur. This means probably that a significance test should not be based on the assumption that the two answers "clear" and "murmured" have the same probability. (No significance test has been applied to this material because of the relatively small number of answers.)

c. In view of the small number of words and listeners one should not draw too many conclusions from details of the test, and since the decision was often difficult, there are probably a good deal of answers which are due to chance.

In spite of this uncertainty a detailed comparison in respect of duration, fundamental frequency, intensity and spectrum of the words which were correctly identified with those which were misheard has given some results.

Duration seems to be of very little importance for the choice, but this does not exclude the possibility that duration may play a role in connected speech.

As to fundamental frequency, there is some evidence that words with purely falling tone (which also have a relatively high start) tend to be heard as clear. Only a few words with murmured vowel and falling tone have been identified correctly, and they have a particularly typical spectral structure. A low start in combination with a tone dip is in some cases seen to be more important than spectral structure and intensity.

As for the distance to intensity and frequency peak no conclusions can be drawn from the material.

The distribution of spectral energy on the other hand seems to be decisive in many cases, particularly when the spectral structure of the vowels of speaker and listener are compared.

In RD's perception of [a] there is only one case where the spectral structure seems to be irrelevant. Also RD's perception of [o] and [e] and RT's perception of [a, e] and [o] are influenced by spectral structure, but not to the same degree. RT has only identified six of eighteen examples of [i]. This may perhaps also be explained by spectral structure in the sense that he has himself no difference, and may therefore not be prepared to listen for this feature.

It may thus be concluded that in this listening test the most important cue seems to have been distribution of spectral energy, but that fundamental frequency is also of importance.

The listening test was carried out after the completion of the acoustic analysis. One might ask whether the stability of the cues would not have been better if only words identified correctly in the test had been used. The analysis of the mistakes show, however, that there would hardly have been much change, except for the distribution of spectral energy. Here the stability would probably have been raised, but not to a hundred percent.

8.3.3. Listening to filtered vowels.

a. Since the relative prominence of the fundamental had proved to be very constant in murmured vowels, it was found of interest to suppress it by filtering and to examine the result for perception.

The restricted text used for spectrographic analysis was used for this purpose; only SK was left out because of his hoarseness.

The texts were highpass filtered by means of a heterodyne filter (5) with a cutoff frequency of 230 cps (3 dB point), the words with [a-a̲] also with a cutoff frequency of 500 cps. Also lowpass filtering with a cutoff frequency of 3200 cps was undertaken. In a preliminary listening session Dave found the murmured vowels with highpass filtering slightly weakened, those with lowpass filtering unchanged except his own [o̲]. As highpass filtered murmured [a̲] was somewhat more difficult to perceive with a cutoff frequency of 500 cps than with a cutoff frequency of 230 cps, it was assumed that the subformant around 250-300 cps might be of some importance, and therefore a band stop filtering was made with the cutoff frequencies 200-500 (and for the speakers with low F_o 160-500) by which the subformant was weakened. But Dave did not find that the bandstop filtering had any influence. The loudness level seemed to be more important. Murmured vowels require a good deal of loudness for their identification.

A shorter test was then made with a selected number of words in random order. The test contained 91 highpass filtered and 101 band-stop filtered words spoken by RD, RT, GU, PBP, and PvB, and 17 lowpass filtered words spoken by RD. The recording and the listening conditions were as in the test with unchanged words. Only Dave acted as a listener.

There were 26 highpass filtered, 38 band-stop filtered and 4 lowpass filtered clear vowels. They were almost all heard correctly (96,95, and 100% respectively). -

The lowpass filtered murmured vowels were also heard correctly.

The results for the highpass filtered and band-stop filtered murmured vowels are given in Tables 15 and 16, grouped according to speakers and vowels respectively. Vowels heard correctly and incorrectly by Dave when unfiltered have been separated in the tables and designated by + and - respectively.

Table 15. IDENTIFICATION OF FILTERED MURMURED VOWELS GROUPED ACCORDING TO SPEAKER (LISTENER RD).

1. BS(16o(2oo)-5oo)

Speaker	RD		RT		GU		PBP		PvB		tot.	
	unfilt.		unfilt.		unfilt.		unfilt.		unfilt.		unfilt.	
	+	-	+	-	+	-	+	-	+	-	+	-
Number presented	12	3	7	6	1o	8	4	4	8	1	41	22
Number correctly identified	9	o	5	1	7	3	4	3	7	o	32	7
% correctly identified	75	o	71	17	7o	38	1oo	75	88	o	78	32

2. HP 23o

	RD		RT		GU		PBP		PvB		tot.	
Number presented	12	2	6	2	11	5	4	4	8	1	41	14
Number correctly identified	1o	o	2	o	9	o	4	1	6	o	31	1
% correctly identified	83	o	33	o	82	o	1oo	25	75	o	76	7

Table 16. IDENTIFICATION OF FILTERED MURMURED VOWELS GROUPED ACCORDING TO VOWELS (LISTENER RD).

1. BS (16o(2oo)-5oo)									2. HP 23o							
Vowels	i		e		a		o		i		e		a		o	
	unfilt.		unfilt.		unfilt.		unfilt.		unfilt.		unfilt.		unfilt.		unfilt.	
	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-
Number presented	12	2	9	1	12	1o	8	9	12	o	8	o	12	6	9	8
Number correctly identified	11	o	8	o	11	4	2	3	11	o	3	o	12	1	5	o
% correctly identified	92	o	89	o	92	4o	25	33	92	o	38	o	1oo	17	56	o

The high percentage of correctly identified vowels for the band-stop filtered recording is not astonishing. There was no reason to expect that the murmured character of the vowels should be damaged as long as the fundamental was preserved. Of the 41 vowels heard correctly, when unfiltered, 32 (or 78%) were heard correctly when band-stop filtered. Moreover of the 22 vowels heard incorrectly when unfiltered, 7 (or 32%) were heard correctly when filtered.

Most of the mistakes (six out of nine) concern the vowel [o̲]. This can be explained by the fast that in [o̲] the whole spectral difference lies in the amplitude relation between the fundamental and the second harmonic and (particularly) the first formant, and these are cut out (or weakened) by the filtering.

It is more astonishing that 76% of the <u>highpass filtered murmured</u> vowels have been heard correctly (the only difference from the band-stop filtered vowels is that one vowel has been improved). As mentioned in section 4, the fundamental has only been attenuated about 25 dB, but even this attenuation must reverse the relations between the level of the fundamental and the formants completely, and in the test with unchanged vowels this relation had been found to be of importance for perception. It should, however, be kept in mind in the first place that distribution of spectral energy is not the only cue, and in the second place that the difference in distribution of spectral energy is not limited to the relation between fundamental and higher formants. Almost all mistakes concern [e̲] and [o̲] , whereas [a̲] and [i̲] are heard correctly, and in the latter vowels there is also a clear difference in the relation between the level of F_3 and F_4 (and partly F_2) and a difference in the level of H_3, i.e. the subformant of [a̲] and the first formant of [i̲] are somewhat raised and broadened. Murmured [i̲] sounds more open or "lax" than clear [i], also when highpass filtered. Moreover six of the 12 words with [a̲] were spoken by PvB and had strong noise.

<u>8.3.4.</u> <u>Perception of words with murmured vowels after tape-cutting.</u>

As some of the acoustic cues found were concentrated to the beginning of the vowel, it might be of interest to examine the effect of removing the beginning or the end of the vowel. The investigation undertaken with this purpose is very restricted and purely preliminary, and it was only possible to find one listener at the time (Dave).

Two isolated examples of [a̲] spoken by RD, and 12 words spoken by RD, GU, and PvB were used. Cuts were made at steps of approximately three cs both from the beginning and from the end of the vowel until about 1o-12 cs were left. The initial and final consonants were cut off at the first step from the beginning and end respectively. The cuts were made by hand and controlled by means of spectrograms. After each cut the words were played over to a second tape recorder. Dave listened a couple of times to the word series in the order recorded, i.e. with diminishing length, and tried to decide when the murmured character of the word became dubious, and when the vowel became clear. It was not so that he found it increasingly difficult to decide whether the vowel was clear or murmured with diminishing length; at a certain point the vowel was heard as clear. This, again, shows that murmur is perceived as a marked feature.

The results of the listening showed, in the first place, that a vowel must have a certain duration in order to be perceived as murmured. Below about 12 cs almost all vowels are heard as clear. This may partly be due to the accompanying reduction of loudness.

In the second place, it was obvious that cutting from the beginning is more detrimental than cutting from the end. When about 11 cs are cut off (from 5 to 16 cs for the different vowels) the vowel is heard as clear (with one exception), irrespectively of the length of the remaining part (which was from 11 to 45 cs). Eleven of the twelve words had also been represented in the test with unchanged vowels, and all identified as murmured, and seven of the words had been presented in the test with highpass filtering and all, except one, had been identified correctly in this test, too. This means that shortening of the vowel, particularly from the beginning, is more detrimental than removing the fundamental. And as an inspection of the spectrograms has shown that the specific spectral structure is kept throughout the vowel, it is obvious that the reasons for the importance of the beginning of the vowel must be sought in other types of cues. The cues concentrated in the beginning are low start of the tone movement, tone dip, and, sometimes, weak intensity. Almost all the vowels in this test had the small tone dip. Dave has this tone dip very often in his speech, and it seems to be of importance for his perception also. It was unfortunate that RT, who does not have this tone dip, and for whom it did not seem to play any role in the test with unchanged vowels, could not take part in this test.

One of the words was heard as clear as soon as the consonant was cut off. This word did not have the characteristic spectral distribution. -

One word was not influenced at all by the tape cutting, namely PvB's /pa̱ɽ/. This word had strong noise.

8.3.5. Moreover, a detailed examination of the answers to the test with unchanged vowels has shown that none of the acoustic cues are necessary, and none is alone sufficient. For all the cues cases can be found where a correct identification has taken place, although the cue was lacking, and for all the cues cases can be found where the cue is present in a strong degree, and nevertheless the vowel has not been identified as murmured. Generally a certain number of cues are combined, and the lack of one may be compensated for by a stronger degree of one of the others.

9. GENERAL CONCLUSIONS.

The phonetic investigation of murmured vowels in Gujarati presented on the preceding pages has brought the following results:

1. On the physiological level murmured vowels are characterized by a strong airflow. This is a very stable feature. It seems to be due to the presence of a small opening in the rear part of the glottis. Since murmured vowels have, in spite of this opening, the same physical intensity as clear vowels, a stronger activity of the expiratory muscles may be assumed.

2. Acoustically murmured vowels are characterized by one or more of the following features: longer duration, lower start of the fundamental frequency, sometimes a small tone dip in the beginning, higher rise, longer distance to the tonal peak, sometimes lower intensity at the start, longer distance to the intensity peak, relatively strong level of the fundamental compared to the second harmonics, and to formant 1, 2, and 4, but not to the third harmonic and to formant 3, sometimes a momentary weakening of formants 2-4 in the beginning of the vowel accompanied by slight noise,

a slight rise in the frequency of formant 1, more damping, asymmetry of the oscillogram, noise in the higher formants.

All these features may be explained by the opening of the glottis with the exception of the longer duration, which is probably due to the historical origin of the murmured vowels as a fusion of vowel with [h] (compare also that isolated murmured vowels seem to be shorter than clear vowels).

The most stable of these cues are duration, distribution of spectral energy, and distance to the tonal peak (the stability of the latter may, however, partly be a consequence of the differences of duration). The least stable features are lower intensity at the start, momentary weakening of higher formants, formant frequency, damping, and asymmetry.

3. The cues which are most important for perception seem to be distribution of spectral energy, low frequency start, and a dip of frequency at the beginning.

None of these cues are necessary, and none is alone sufficient. We are thus faced with a situation where a large number of instable acoustic cues correspond to a simple physiological difference. And murmured vowels may be quoted as an example in favour of the motor theory of speech perception (9), according to which the motor center is involved in the identification of incoming signals. It may well be that a murmured vowel is identified by the speaker as a vowel which he would produce with strong airflow.

The general interest of the investigation lies in this point.

Postscriptum:

For an additional investigation made after the completion of the paper, see p.84.

Inverse filtering analysis.

By inverse filtering the ripples of the formants can be removed from the oscillogram, so that only the glottal curve is left. Such a filtering of one example of the word pair /bar - b<u>a</u>r/, spoken by RD, has been undertaken in the Speech Transmission Laboratory of the Royal Technical High School in Stockholm by J. Lindquist. The result is seen below (Fig. 27). It is obvious that the shape of the curve is more sinusoidal and the closure phase is shorter in the murmured vowel than in the clear vowel. This points to a relatively stronger fundamental (as it was also found in the spectral analysis). The murmured vowel has also a somewhat slower fall of the curve corresponding to a slower closing movement of the vocal cords (cp. the suggestion made by Gunnar Fant concerning the asymmetry of the oscillogram 6.5.4.c).

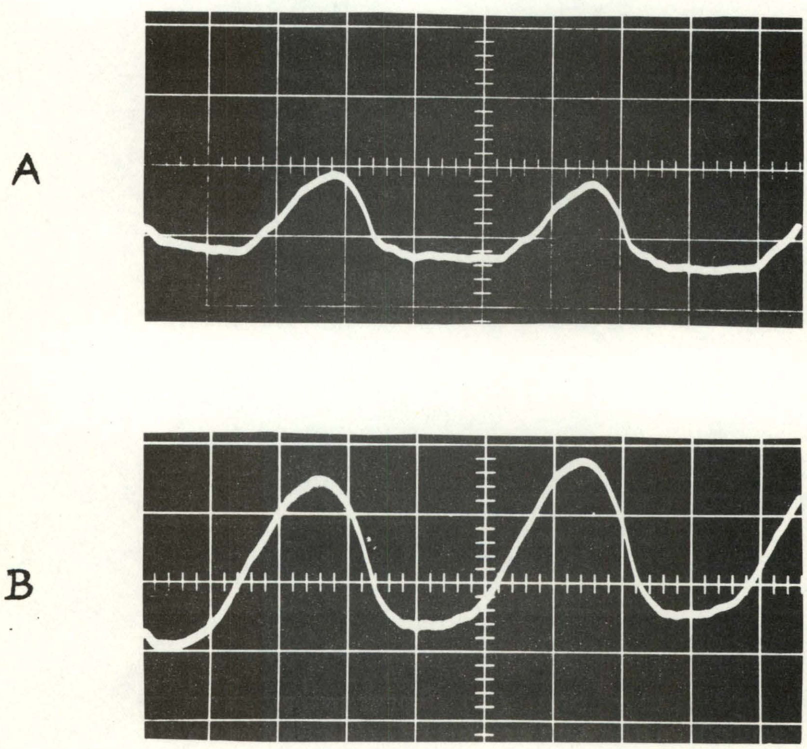


Fig. 27. Inverse filtering curve of the beginning of the vowel in A. /bar/, B. /b<u>a</u>r/, spoken by RD.

REFERENCES

1. P.B. Pandit, "Nasalisation, Aspiration and Murmur in Gujarati", Indian Linguistics 17 (1955-56) 1957, pp. 165-172.

2. R. Dave, "A Formant Analysis of the Clear, Nasalized and Murmured Vowels in Gujarati", this report (ARIPUC 2/1967), pp. 119-132, to be published in full in Indian Linguistics.

3. G. Fant, K. Fintoft, J. Liljencrantz, B. Lindblom and J. Mártony, "Formant Amplitude Measurements", JASA 35 (1963), pp. 1753-62.

4. Jørgen Rischel, "Instrumentation for Vowel Synthesis", ARIPUC I/1966, pp. 15-21.

5. Jørgen Rischel, "Heterodyne Filter", ARIPUC 1 (1966), pp. 13-14, and "The Heterodyne Bandpass Filter of the Institute of Phonetics", this report, (ARIPUC 2/1967), pp. 20-34.

6. K. Ch. Bahl, "Tones in Punjabi", Indian Linguistics 17 (1955-56) 1957, pp. 139-147, and Ved Kumari Ghai, "Word Tones in Dogri", this report (ARIPUC 2/1967), pp.133-157, to be published in full in Indian Linguistics.

7. K. Fintoft, B. Lindblom, and J. Mártony, "Measurements of Formant Level in Human Speech", QPSR 2/1962, Speech Transmission Laboratory, Stockholm, pp. 9-17 (p. 16).

8. P.B. Pandit, "E and O in Gujarati", Indian Linguistics XV, 1955-56 reprint edition, pp. 617-55 (p. 628).

9. Ilse Lehiste, "Acoustical Characteristics of Selected English Consonants", International Journal of American Linguistics 3o, Part IV, 1964, 197 pp., see particularly Chapter 3, pp. 141-18o: "Acoustic Characteristics of /h/ and Whispered Speech".

1o. Alvin Liberman, "Some Results of Research on Speech Perception", JASA 29 (1957), pp. 117-22, and A.M. Liberman, F.S. Cooper, Katherine S. Harris, and P.G. MacNeilage, "A Motor Theory of Speech Perception", Proceedings of the Speech Communication Seminar, Stockholm 1962, Volume 2, D3, and the Proceedings of the Seminar on Speech Production and Perception Leningrad 1966, to be published in Zeitschrift für Phonetik und allgemeine Sprachwissenschaft. -

IS DANISH 'sj-' ONE OR TWO SOUNDS?

Hans Basbøll

The opinions held by phoneticians about the phonetic nature of the sound (here called 'sj') which in genuine Danish words corresponds to the letters sj- (as in sjælden 'seldom') differ widely. In the two official textbooks for the course in general and Danish phonetics (Andersen (1), Fischer-Jørgensen (4); also Hansen (5)) it is stated that the sound in question is in fact a palatalized s followed by a j (which is normally described as unvoiced). Uldall (11), Arnholtz & Reinhold (2), and Martinet (8) (1937) call the Danish 'sj' one sound, a sort of ʃ. Jespersen (6), Diderichsen (3), Koefoed (7), and Spore (9) only say that Danish 'sj' may be pronounced as one or two sounds. The opinion of Martinet (10) (1949) is not clear.

All the above-mentioned authors seem to base their opinions only on auditory observations; it might therefore be interesting to make an acoustic investigation on this point (but it was not intended to give a general acoustic description of Danish 'sj').

As there is never in Danish commutation between s+j and ʃ (but there is between s and 'sj'), the manifestation of /sj-/ (corresponding to sj in the orthography) can be rather varying. But one manifestation is generally considered to be the normal one in standard Danish.

A series of 25 Danish words (with initial sj, s, fj, f, tj, and a few others) were recorded from 12 persons.

Four persons having a professional knowledge of the norm of the standard language judged whether these recordings were normative, particularly concerning the pronunciation of 'sj'. (Andersen identified most of the 'sj's as the sound he had transcribed [s_jj̥] in his book (1).)

No attempt was made to give an absolute definition of 'one sound' (-segment). The method was one of comparison of spectrograms, in order to see whether Danish 'sj' clearly resembles segments which are always described as one sound (e.g. English ʃ in 'she') or segments which are always described as two sounds (e.g. Russian s_jj in 'Cья').

The procedure was as follows: spectrograms with Russian [s_jæ] and [$s_j$jæ], with French [aʃe] and [aʃje], with British [ʃu:] and [sju:], with German [ʃ, ç, s] and [ʃç, sç], and with Danish [fɛlˀ] and [fjɛlˀ] (and a few others) were compared in order to determine which acoustic characteristics correspond to the auditory impression of a j after an unvoiced fricative; or, more exactly, to find some acoustic differences between fricative + j + vowel and fricative + vowel. (All the above-mentioned examples are phonemically different). Four such criteria (which are not constant) were found:

(a). A shift in intensity in the unvoiced segment (s being more intense than j̥). This is not an important criterion.

(b). A shift in the distribution of spectral energy in the unvoiced segment. This is the most important criterion, as Danish j̥ has another distribution of energy than s, s_j, and ʃ.

(c). A voiced segment (after the unvoiced one) with less intensity than the vowel and a low F1.

(d). The vowel transitions may be important, but as the transitions of F2 and F3 from ʃ to the vowel e.g. in [ʃæ] can be similar to the initial bendings of these formants in [jæ], this criterion is not so important as (b) and (c), but more important than (a).

The length of the various segments and transitions must be taken into account.

Then all the spectrograms of Danish 'sj' were examined in order to determine whether any of the acoustic criteria (abcd) for j were present. The main conclusion is that the persons whose pronunciation of 'sj' was found completely normative by all the 'language norm judges' only had one 'sound segment' before the vowel, i.e. they had none of the acoustic criteria (abcd) for the presence of j (cp. Figs. 1 & 2). But in the material as a whole all degrees between clearly one and clearly two segments were found, with a significant majority for one segment.

As Danish 'sj' has been described as ʃ, s_j, and s_jj, the following test was made. English shock, Russian Свя and Ся, and Danish sjal were recorded on the segmentator. Then still greater parts of the initial consonant + the vowel transitions of each of these four words were cut off, and by each new cutting four well-trained phoneticians described the initial sound auditorily

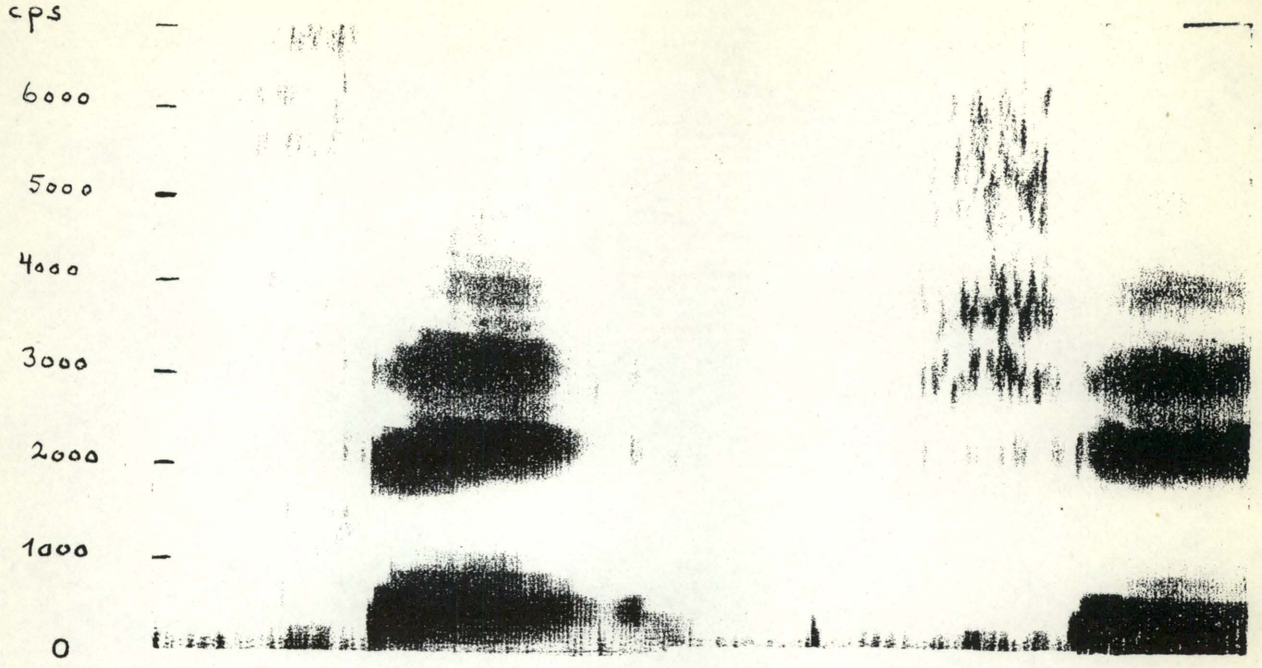


Fig. 1. s æ ʔ l ʃ æ ʔ (l)

('sal') ('sjal')

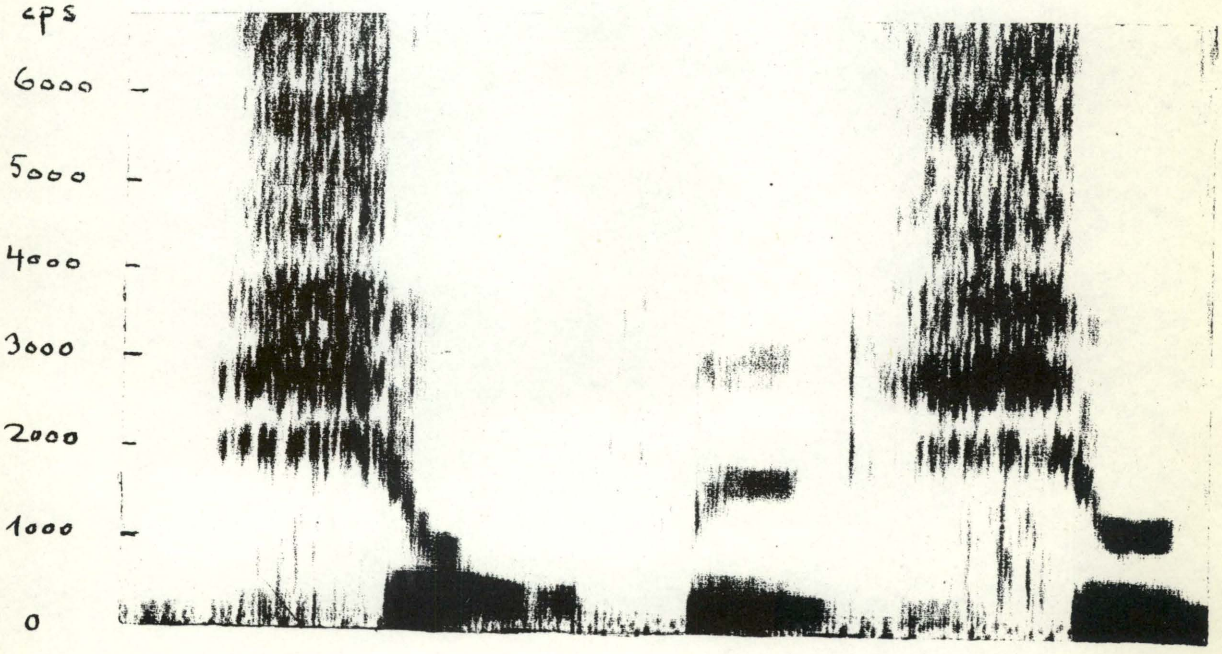


Fig 2. ʃ o ʔ f l̩ ʃ u (sg̊ə)

('sjofel') ('sjuske')

in as great detail as possible. Everything was recorded on tape and on an oscillograph. It should be determined whether the 'cutting evolution' of Danish sjal clearly resembled the 'cutting evolution' of one of the three foreign words more than that of the others. Particularly it was to be tested whether (by some stop in the cutting) a j could be heard in some but not all of the four words, and if this were the case, how Danish 'sj' would be in this respect.

No clear results were found although when the word was cut just when the vocal vibrations start, there was a tendency to hear a diphthong [i̯æ] in Russian Свя, but a (dental) stop + a vowel in the three other words). Nevertheless the method might prove useful with greater material and a more elaborate procedure (e.g. involving spectrographical recordings of the test words in order to find some acoustic 'fix-points' where you could cut, so that the cut test words can be more easily comparable).

References:

(1) Poul Andersen, Dansk fonetik, Chapter XV of Nordisk Lærebog for Talepædagoger (1954), p. 341.

(2) Arnholtz & Reinhold, Einführung in das dänische Lautsystem mit Schallplatten (1936), p. 49.

(3) Paul Diderichsen, Danish Pronunciation (Key to Gramophone Record)[2] (1953), pp. 16-17.

(4) Eli Fischer-Jørgensen, Almen fonetik [3] (1962), pp. 68-69.

(5) Aage Hansen, Udtalen i moderne dansk (1956), p. 83.

(6) Otto Jespersen, Modersmålets fonetik[2] (1934), pp. 38-39.

(7) H. A. Koefoed, Teach Yourself Danish (1958), § 27.

(8) André Martinet, La phonologie du mot en danois (1937), p. 39.

(9) Palle Spore, La langue danoise (1965), p. 42.

(1o) Ingeborg Stemann, Manuel de la langue danoise (1949), p.39 (the phonetic section is by André Martinet).

(11) H.J.Uldall, A Danish Phonetic Reader (1933), p. IX.

THE WORD TONES OF SERBO-CROATIAN AN INSTRUMENTAL STUDY.

Per Jacobsen

This study describes the results of an instrumental investigation made at the Institute of Phonetics of the University of Copenhagen in the spring of 1967. The purpose of the investigation was to examine the phonetic character of the Serbo-Croatian word tones, i.e. to measure vowel quantities, fundamental frequency, and intensity.

By means of the commutation test, 4 contrasting types of accent are established. These 4 types are traditionally called: long falling (usually symbolized with (⌒)) [V̀:], short falling (\\) [V̀], long rising (/) [V́:], and short rising (\) [V́]. In syllables after the ictus only the contrast long:short is phonemic, ([V:]:[V]). Although the results of this investigation show that the traditional names of the word tones do not always correspond to the phonetic reality, they are used here to label the four types of word tones. The distribution of the four accents is limited: Falling accents occur only on the first syllable. Rising accents may occur on any syllable except on the last. Monosyllabics have falling accents.

In his book <u>Osnovi Fonetike Srpskog Jezika</u> Branko Miletić describes the Serbo-Croatian word tones as follows:

<u>Melody</u>. Accent (\\) [V̀] occurs in two main types: 1) One falling, in the central dialects (Bosnia, Hercegovina, Mačva, Dubrovnik, and other regions) and 2) one level, in Belgrade and other eastern regions. The first type (the falling) is more typical and probably older.

Accent (⌒) [V̀:] is <u>falling</u> in all dialects, but in such a way that the tone ordinarily forms an arch: it rises in the beginning and then falls steeply towards the end; this fall is greater or smaller, depending on the dialect, but it is a constant feature.

Accent (\) [V́] is rising in all dialects; it is either <u>rising</u> from the beginning to the end or <u>partly</u> rising, so that part of the tone is level.

Accent (/) [V́:] is the most stable tone in all dialects: it is always <u>rising</u>. There are, however, dialect differences: in certain dialects the tone is more rising than in others.

INTENSITY. The falling accents have parallel tone and intensity movements: The intensity rises in the beginning or stays constant for some time and then falls. On the contrary there is no such complete parallelism between tone and intensity in the rising accents: the intensity rises only in the beginning of the vowel and then falls towards the end.

QUANTITY. Accentuated long vowels are twice as long as accentuated short vowels, and accentuated vowels are usually longer than unaccentuated ones. Although ̏ [V̀] and ̀ [V́] are short accents and ̑ [V̀:] and ́ [V́:] are long, there are some differences between them: ́ [V́:] is usually somewhat longer than ̑ [V̀:], while the typical ̏ [V̀] is longer than ̀ [V́], often considerably longer. -]

This description covers the traditional conception of the phonetic nature of the Serbo-Croatian word tones.

Examples of tone and quantity contrasts:

sȇdī	['sɛ̀:di:]	'he paints grey'
sédi	['sɛ́:di]	'grey'
sédī	['sɛ́:di:]	'his hair is getting grey'
sèdī	['sɛ́di:]	'he sits down'
sèdi	['sɛ́di]	'sit down!'
kȕka	['kȕka]	'a hook'
kȕkā	['kȕka:]	'he laments'
kùka	['kúka]	'of the hip'
Rósa	['rɔ́:sa]	'the girl's name Rosa'
rȍsa	['rɔ́sa]	'dew'

Material, informants, and recording.

A number of monosyllables and disyllables were selected for the investigation. As the physical duration usually depends on vowel quality, the material is arranged so that long a is compared with short a, long e with short e, etc. In the selection of the examples various factors, which might influence the duration or tone of the accentuated vowels, have been considered. Thus the following consonant might play a role for vowel duration. Therefore, the vowels have been analysed when preceding different consonants: nasal/liquid, voiced and unvoiced fricatives, and plosives. Vowel duration might be influenced also by the quality of the following syllable.

For this reason, the disyllabic material consists of both long and short second syllables.

The material can be arranged as follows:

Table 1 a. Monosyllables

Following consonant:	Nasal/ liquid	Fricative		Plosive	
		unvoiced	voiced	unvoiced	voiced
long falling	sȃm	pȃs	pȃž	sȃt	sȃd
short falling	sȁn	pȁs	sȁv	čȁk	sȁd

Table 1 b. Disyllables

Intervocalic consonant:		Liquid/ nasal		Fricative			
				unvoiced		voiced	
Tone label:		falling	rising	falling	rising	falling	rising
1st syllable long	2nd syllable short:	čȃri	táman	spȃsa	kásu	kȃžu	fáza
	2nd syllable long:	pȃrā	sámīm	pȃsā	pásē	kȃžē	fázā
1st syllable short	2nd syllable short:	pȁra	pàra	pȁša	pàša	pȁzi	tàvan
	2nd syllable long:	pȁmēt	tàmān	pȁšē	pàsūlj	pȁzī	fàzān

Table 1 b. Disyllables - continued

Intervocalic consonant		Plosive			
		unvoiced		voiced	
Tone label:		falling	rising	falling	rising
1st syllable long	2nd syllable short:	kȃpi	pápa	tȃbor	sábor
	2nd syllable long:	skȃčē	pácōv	sȃdīm	nádā
1st syllable short	2nd syllable short:	šȁpat	šàtor	Šȁbac	sàda
	2nd syllable long:	pȁkōst	kàpūt	pȁdā	kàdēt

Similarly for each of the syllabic sounds a,e,i,o,u,r.

The tone usually starts at a lower point when the initial consonant is voiced than when it is unvoiced. In order to get as uniform a material as possible the majority of the examples have an unvoiced initial consonant, whereas a smaller number, for comparison, have a voiced initial consonant.

At last the structure of the stressed syllable is considered. All disyllables are of the type sada or Šabac, i.e. with open first syllable. *

A number of the words investigated (16 monosyllables and 132 disyllables) form minimal pairs in which tone and/or length determine the meaning.

The words were written on cards, one word on each card, and arranged in quasi-random order, the only restriction being that minimal pairs did not occur adjacent to each other. The words were placed in a frame sentence: piše...na karti ('...is written on the card'). By this method a complete uniformity in the placement of the words in a sentence context is obtained, and the sentence intonation does not, at the place in question, disturb the word tone too much. The utterances were recorded in a silent room on a professional tape recorder at the Institute of Phonetics.

Four informants participated in the investigation:
MA (female) was born in 1938 in Sremski Karlovci in Vojvodina. She has a university degree in Serbo-Croatian language and literature. Her material consists of 47 monosyllables and 242 disyllables, 289 utterances in all.

CJ (female), born in 193o in Valjevo in Serbia. Her material consists of 43 monosyllables and 2o5 disyllables, 248 utterances in all.

SM (male), born in Šabac in Serbia in 1941. Material: 48 monosyllables and 196 disyllables, 244 utterances in all.

DA (male) born in Zlarina near Šibenik in Dalmatia and grown up in Bosnia and Vojvodina. His material consists of 37 monosyllables and 164 disyllables, 2o1 utterances in all.

The complete material consists of 982 utterances.

----
* The material was intended to consist of 6o mono- and 24o disyllables, a number I did not reach. OJ, MA, and SM have spoken more than 24o disyllables, but this is because there are doublets of several examples.

In order to try whether long and short vowels were consistently distinguished by the speakers the minimally contrastive words were rerecorded on a special tape where they were placed immediately after one another. The informants were then asked to identify the utterances. This listening test showed that all 4 persons distinguished between the four types of accent, easily and without making errors. On the other hand they did not in all cases distinguish long and short unaccented vowels. The only exception was the short a of nom. sing. fem. in contradistinction to the long a of gen.plur. but here, too, there was some vacillation.

Therefore, I have desisted from any conclusion concerning the influence of the subsequent syllable on the quantity of the stressed syllable. In the calculations on vowel quantities I have combined disyllables assumed to have long and short second syllables into one group. If no phonemic distinction is found between long and short unstressed vowels the possible differences in duration are too accidental to be considered in the calculations.

Recording of acoustic curves.

The utterances were subjected to acoustic analysis by means of an intensity meter (KTH type) and Frøkjær-Jensen's pitch meter and recorded on the mingograph. Four synchronous traces were recorded on the ink writer: an intonation curve, a logarithmic intensity curve with an integration of 2,5 ms for female voices and 5 ms for male voices and a high-pass filtering (-3dB at 5oo cps) for all voices. The high-pass filtering in connection with the logarithmic scale conditions that the consonants stand out rather distinctly. The third curve is an intensity curve (linear scale) with an integration of 5 and 1o ms, respectively, and with linear frequency response. The fourth curve is a 'duplex oscillogram', which is a combination of an ordinary oscillogram and a high-pass filtered intensity curve.

The logarithmic intensity curve and the duplex oscillogram were used to delimitate the sounds before measurements were made on the intonation curve and the linear intensity curve. The speed of the paper was 1oo mm/sec, (one mm corresponding to 1/1oo sec).

The mingograph recordings were made by cand.art. Hans Peter Jørgensen of the Institute.

Table 2. Vowel duration in cs.

2 a. Monosyllables

	â [à:]	ê [ɛ̀:]	î [ì:]	ô [ɔ̀:]	û [ù:]	r̂ [r̀:]
MA	20,7	20,4	19,2	19,8	19,4	20,0
OJ	21,0	20,4	17,1	19,4	21,3	17,3
SM	24,8	21,4	20,2	23,8	16,6	17,0
DA	23,2	18,1	19,4	19,2	18,7	-
average	22,4	20,1	19,0	20,6	19,0	18,1

	ȁ [à]	ȅ [ɛ̀]	ȉ [ì]	ȍ [ɔ̀]	ȕ [ù]	ȑ [r̀]
MA	13,3	14,9	10,1	12,3	12,5	12,0
OJ	13,0	11,0	10,0	11,1	10,6	9,8
SM	12,6	11,7	8,7	10,9	9,6	7,7
DA	14,5	12,5	10,3	11,2	11,3	11,0
average	13,4	12,5	9,8	11,4	11,0	10,1

2 b. Disyllables; (duration of stressed vowels):

	â [à:]	á [á:]	ê [ɛ̀:]	é [ɛ́:]	î [ì:]	í [í:]	ô [ɔ̀:]	ó [ɔ́:]
MA	22,4	21,9	20,8	21,4	18,8	18,8	20,9	20,6
OJ	21,5	21,7	17,5	20,3	16,0	17,1	18,6	20,0
SM	24,4	20,9	21,0	21,1	18,2	18,5	19,8	19,5
DA	22,0	19,8	18,0	21,9	17,4	17,2	21,7	21,4
average	22,6	21,1	19,3	21,2	17,6	17,9	20,3	20,4

	û [ù:]	ú [ú:]	r̂ [r̀:]	ŕ [ŕ:]	ȁ [à]	à [á]	ȅ [ɛ̀]	è [ɛ́]
MA	20,3	20,1	20,4	20,4	14,2	15,0	15,1	14,9
OJ	18,0	16,8	16,3	18,7	12,7	11,8	11,0	11,6
SM	20,5	19,3	21,3	18,8	11,6	10,6	10,9	11,1
DA	19,8	19,8	16,5	19,5	13,2	12,8	12,0	11,4
average	19,7	19,0	18,6	19,3	13,1	12,5	12,3	12,3

	ȉ [ì]	ì [í]	ȍ [ɔ̀]	ò [ɔ́]	ȕ [ù]	ù [ú]	ȑ [r̀]	r̀ [ŕ]
MA	11,3	10,1	14,4	14,8	11,3	12,1	12,0	11,5
OJ	9,2	7,3	11,6	11,5	9,2	9,0	10,1	10,3
SM	9,4	7,7	11,5	9,8	9,8	10,8	9,5	11,6
DA	9,0	9,3	11,2	12,3	9,9	9,1	9,3	9,0
average	9,8	8,6	12,2	12,1	10,0	10,2	10,2	10,6

Average vowel duration in cs for four informants:

Table 3.

monosyll. ⌒[v̀:] 19,9	monosyll. ˵[v̀] 11,4	⌒[v̀:]* 19,8	˵[v̀]* 11,3
disyll. ⌒[v̀:] 19,7	disyll. ˵[v̀] 11,2	´[v́:] 19,8	`[v́] 11,1

Significance testing is not necessary to show that the differences between monosyllables and disyllables are <u>not</u> significant.

The influence of the following consonant on vowel length is seen in the following table:

Table 4.

	nasal/liquid	unvoiced fricative	voiced fricative	unvoiced plosive	voiced plosive
long vowel	2o,4	2o,2	2o,2	18,6	19,7
short vowel	11,o	12,5	11,9	9,9	1o,6

As seen in table 3 there is no significant difference between monosyllables and disyllables; nor is there a significant difference between rising and falling wordtones in this respect. This result disagrees with the statement in Miletic's <u>Osnovi fonetike</u>.

The following consonant influences vowel length, vowels followed by plosives being somewhat shorter than vowels followed by fricatives, nasals or liquids, and vowels followed by unvoiced plosives being shorter than vowels followed by voiced plosives.

The material is clearly divided into long and short vowels. The overlapping is minimal. A diagram of the dispersion of the vowel <u>i</u> is given as an example (Fig. 1.).

Altogether the durations of the individual vowels are as follows:

----

* Monosyllables and disyllables are taken together since the difference in duration is not significant

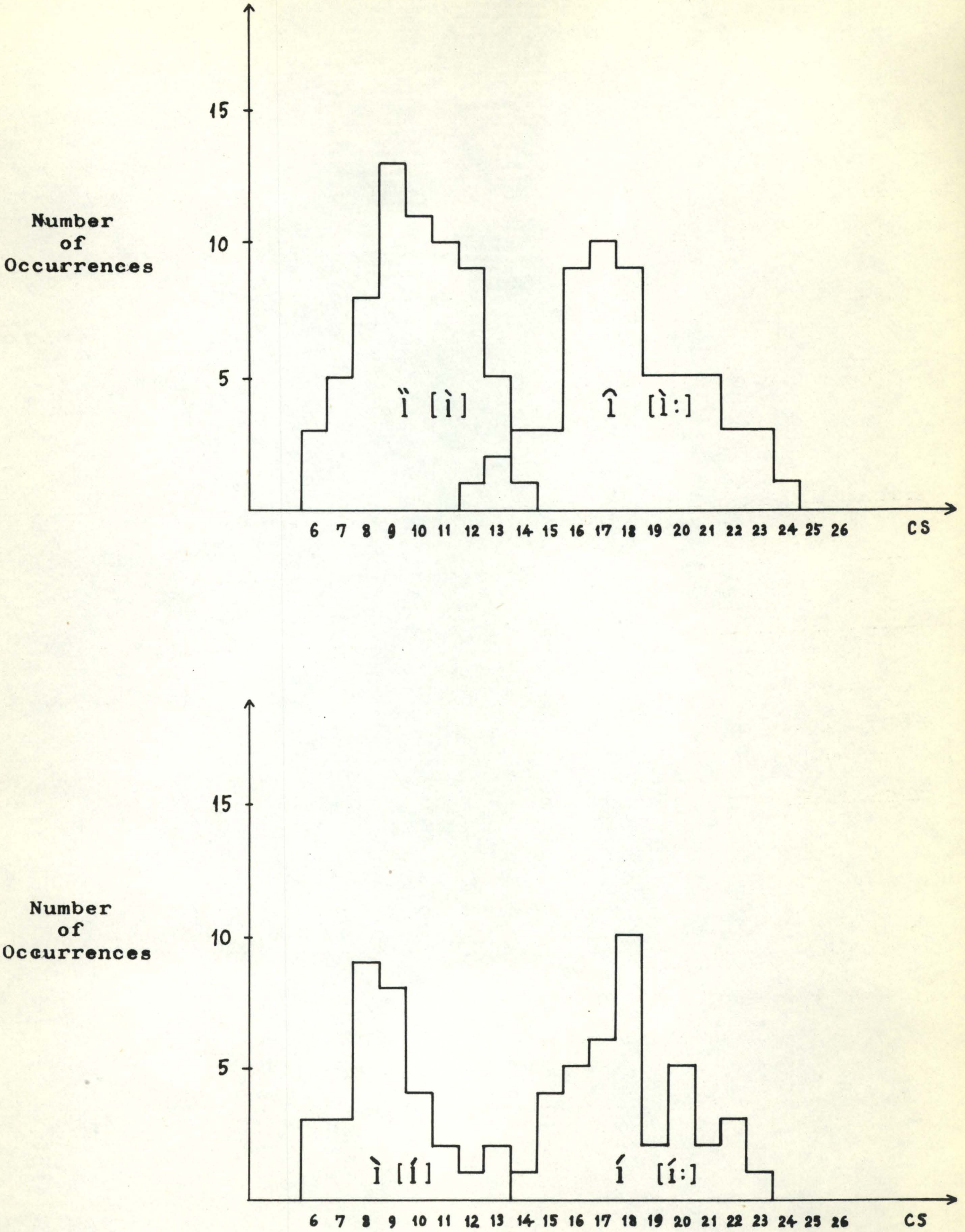

Number of Occurrences
15
10
5
ȉ [ì]
î [ì:]
6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
CS
Number of Occurrences
15
10
5
ì [í]
í [í:]
6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
CS

Fig. 1

Table 5. Vowel durations.

long vowels:	i: 18,2	r: 18,7	u: 19,2
	e: 2o,3		o: 2o,4
		a: 22,1	
short vowels:	i 9,4	r 1o,3	u 1o,4
	e 12,4		o 11,9
		a 13,o	

- which was to be expected, the closed vowels being physically shorter than the more open ones.

Table 6.

Average fundamental frequencies (beginning, peak, and end)-of the syllabic sounds. The distance of the fundamental frequency peak (if present) from the beginning of the syllabic sound is given as a percentage of the total duration.

Monosyll./⌒/ [V̀:]	with unvoiced initial consonant			with voiced initial consonant		
	1st syllable	%	2nd syllable	1st syllable	%	2nd syllable
MA	237-259-143	3o		212-27o-155	33	
OJ	222-289-149	24		213-262-154	26	
SM	148-165- 93	17		112-128- 8o	35	
DA	165-191-124	37		147-178-117	3o	
Disyll./⌒/ [V̀:]						
MA	215-249-169	38	159- 14o	198-239-169	48	164- -142
OJ	221-282-191	31	173- -128	218-273-2o7	5o	171- -122
SM	145-171- 9o	29	9o- - 81	121-145-1o1	43	92- - 82
DA	177-192-132	34	125- -117	173-198-13o	37	128- -118
Monosyll./\\/ [V̀]						
MA	218-246-19o	45		193-232-214	46	
OJ	223-296-232	41		218-288-254	5o	
SM	134-158-122	46		113- -113		
DA	17o-188-153	51		146-183-139	6o	

(%: Place of peak in percent.)

Table 6. - continued.

Disyll./ˋˋ/ [V̀]	with unvoiced initial consonant			with voiced initial consonant		
	1st syllable	%	2nd syllable	1st syllable	%	2nd syllable
MA	205-235-208	66	175- -136	191-242-235	74	175- -138
OJ	221-291-258	55	182- -129	211-269-259	71	174- -117
SM	145-181-143	46	99- - 90	132- -144		103- - 89
DA	176-198-162	53	132- -113	164-182-173	61	149- -116
/ʹ/ [V́:] (only disyllabic)						
MA	185- -204		223- -187	177- -210		235- -198
OJ	194- -260		260- -171	182- -236		248- -162
SM	120- -143		139- - 96	103- -140		137- - 86
DA	140- -167		174- -136	131- -164		175- -139
/ˋ/ [V́] (only disyllabic)						
MA	189- -191		220- -173	183- -206		228- -190
OJ	199- -248		261- -162	191- -237		255- -178
SM	129- -126		148- -105	113- -128		141- -105
DA	147- -159		182- -142	140- -163		191- -146

(%: Place of peak in percent)

Description of frequency

Although the averages presented here show a difference in the stressed vowels between short "falling" and "rising" a large part of the examples show that the tone of the stressed vowel is not always relevant for the opposition. (Cf. illustration Fig. 4b of žȅnī [ʒɛ̀niː]*) and žèni [ʒɛ́ni]). The opposition cannot be regarded as an opposition between falling and rising tone in the stressed syllable. On the other hand it is evident and common for all four informants that the opposition between / ̏/ [ˋ] and / ̀/ [ˊ] is established by the relation between the tone of the stressed syllable and that of the following syllable. The tone in the syllable after the ictus in words with / ̀/ [ˊ] starts as high or even higher than the end of the preceding (stressed) syllable, while it is considerably lower in words with / ̏/ [ˋ].

The movement of the tone is continued in voiced consonant after short vowel.

The tone in the stressed syllables in words with / ̑/ [V̀ː] and / ́/ [V́ː] is in itself sufficient to establish the opposition.

Table 7

Intensities in dB. - Averages for all four informants

Word tone		1. syllable (vowel)			place of peak	2. syllable (vowel)			place of peak
		Beg.	Peak	End	percentage	Beg.	Peak	End	percentage
mono-syll.	/ ̑/ [V̀ː]	41	5o	34	29				
	/ ̏/ [V̏]	41	5o	38	45				
di-syll.	/ ̑/ [V̀ː]	41	51	41	39	41	-	36	-
	/ ̏/ [V̏]	45	51	41	52	43	47	38	28
	/ ́/ [V́ː]	41	5o	44	61	47	48	38	34
	/ ̀/ [V́]	41	49	43	58	46	49	38	33

----

*) or as vowel length after stressed syllable is not relevant: žȅni [ʒɛ̀ni].

Description of intensity.

The intensity usually rises suddenly in the beginning of the vowel. It may have its peak in the beginning and may then fall towards the end, or it may exhibit a break after the sudden rise, rise slightly towards the end and then suddenly fall just before the end of the vowel. When the peak is situated in the first half of the vowel the intensity is falling, and when it is in the second half the intensity is rising.

Although the table shows falling intensity with /ȃ/ [\`:] (less evident with /ȁ/ [\`]) and rising with /á/ [´:] and /à/ [´], the deviation in the individual occurrences is so large that one may speak of a tendency only.

On the other hand there seems to be a difference in the relation between the syllables. In words with /ȃ/ [\`:] and /ȁ/ [\`] the intensity of the second syllable is lower than the intensity of the first syllable, while the second syllable in words with /á/ [´:] and /à/ [´]has (almost) the same intensity as the first syllable. Regarded in this way it is correct to underline the parallelism between tone and intensity.

Conclusion

On the basis of the present material I cannot share the traditional conception of the Serbocroatian accent system. In short vowels it seems impossible to locate the distinctive characteristics of word tone within the syllable traditionally said to carry the accent. In long vowels, however, the usual assumption seems to be true (cf. Miletić).

Whether intensity plays any role for the identification is difficult to tell. Experiments with synthetic speech where the different parameters of speech signal can be arbitrarily changed, might throw light on this question.

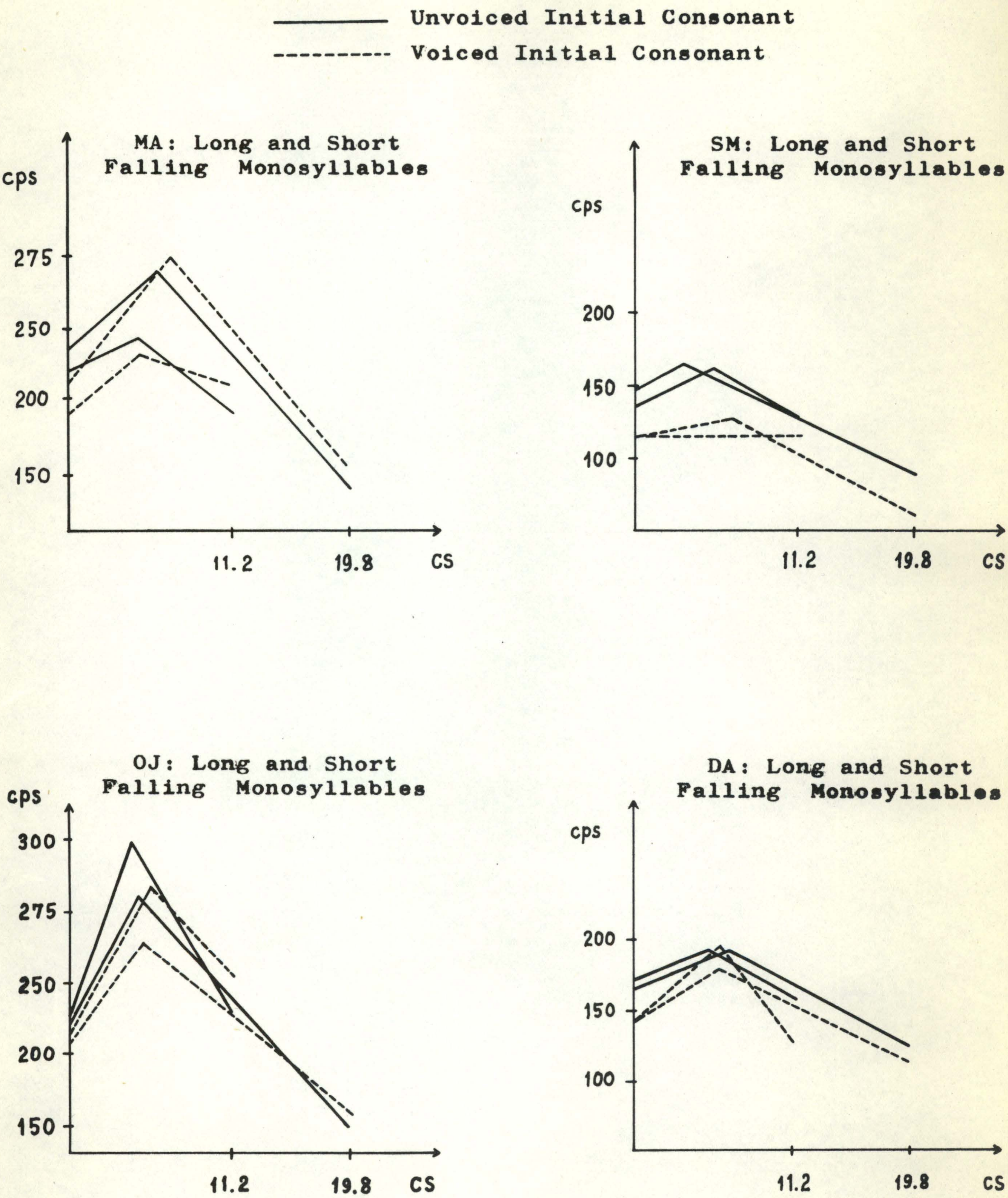
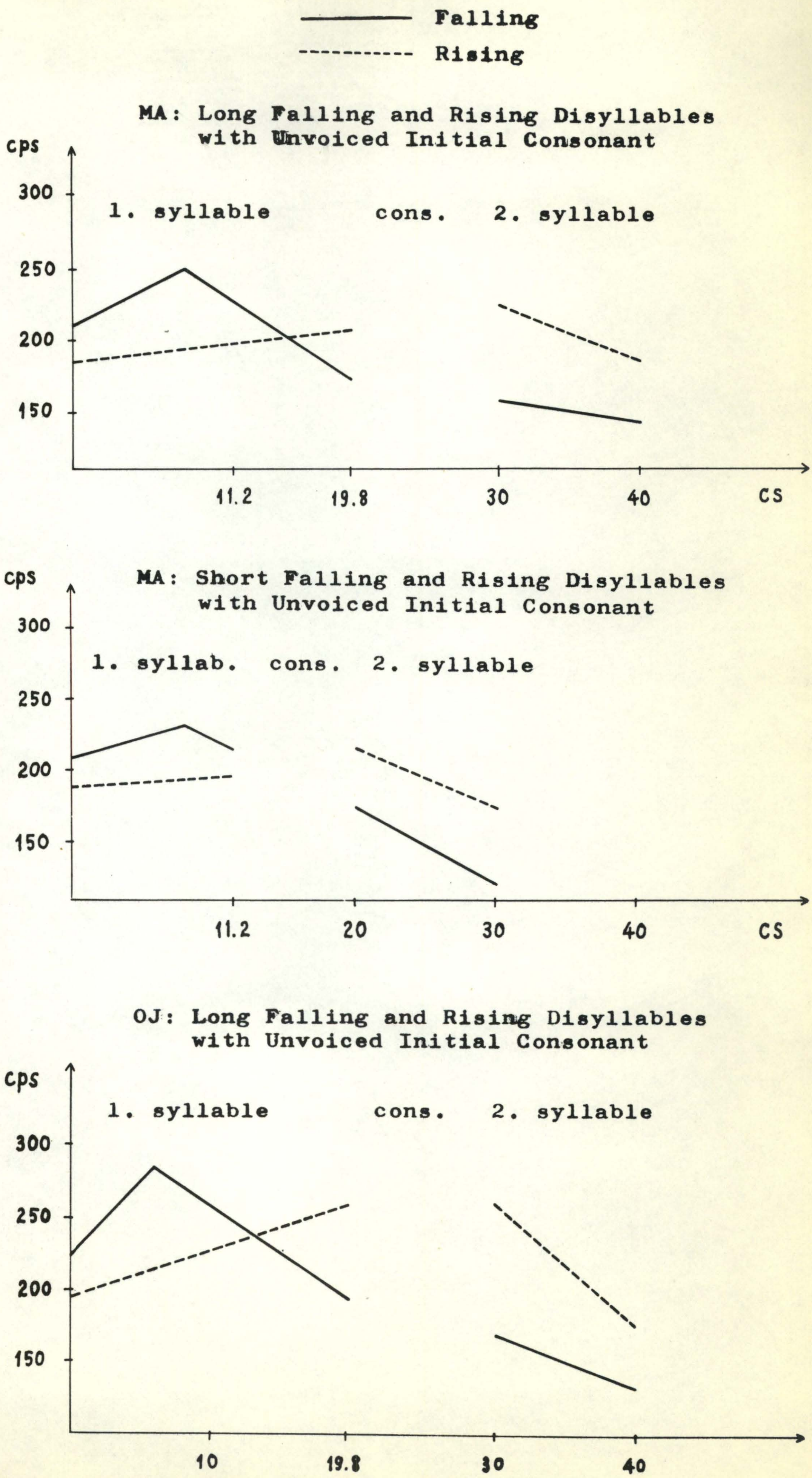


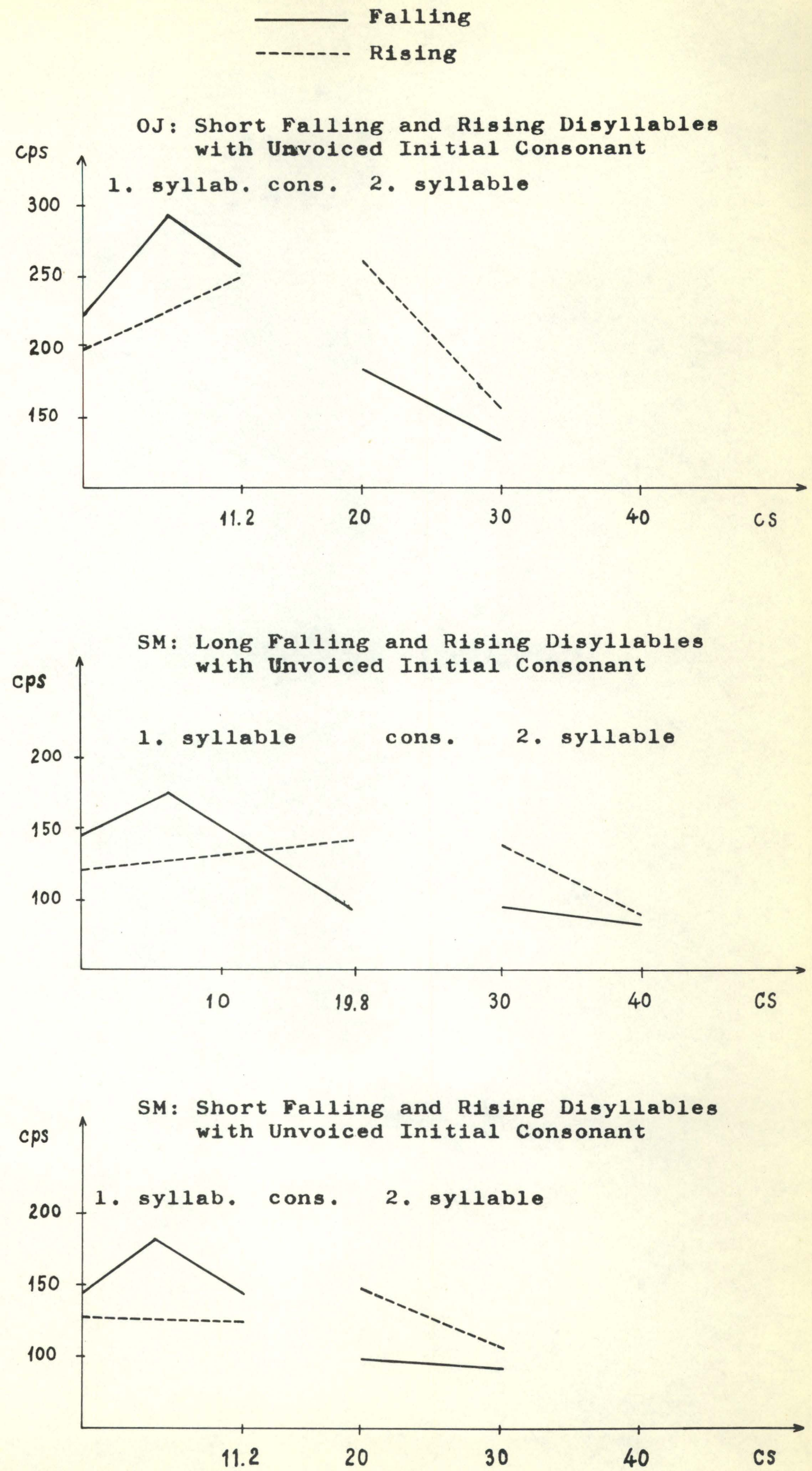
Fig. 2a.

Fundamental Frequency.



Falling
Rising
MA: Long Falling and Rising Disyllables with Unvoiced Initial Consonant
cps
300
250
200
150
1. syllable
cons.
2. syllable
11.2
19.8
30
40
cs
MA: Short Falling and Rising Disyllables with Unvoiced Initial Consonant
cps
300
250
200
150
1. syllab.
cons.
2. syllable
11.2
20
30
40
cs
OJ: Long Falling and Rising Disyllables with Unvoiced Initial Consonant
cps
300
250
200
150
1. syllable
cons.
2. syllable
10
19.8
30
40

Fig. 2b.



Falling
Rising
OJ: Short Falling and Rising Disyllables
with Unvoiced Initial Consonant
cps
1. syllab. cons. 2. syllable
300
250
200
150
11.2
20
30
40
cs
SM: Long Falling and Rising Disyllables
with Unvoiced Initial Consonant
cps
1. syllable
cons.
2. syllable
200
150
100
10
19.8
30
40
CS
SM: Short Falling and Rising Disyllables
with Unvoiced Initial Consonant
cps
1. syllab. cons. 2. syllable
200
150
100
11.2
20
30
40
CS

Fig. 2c.

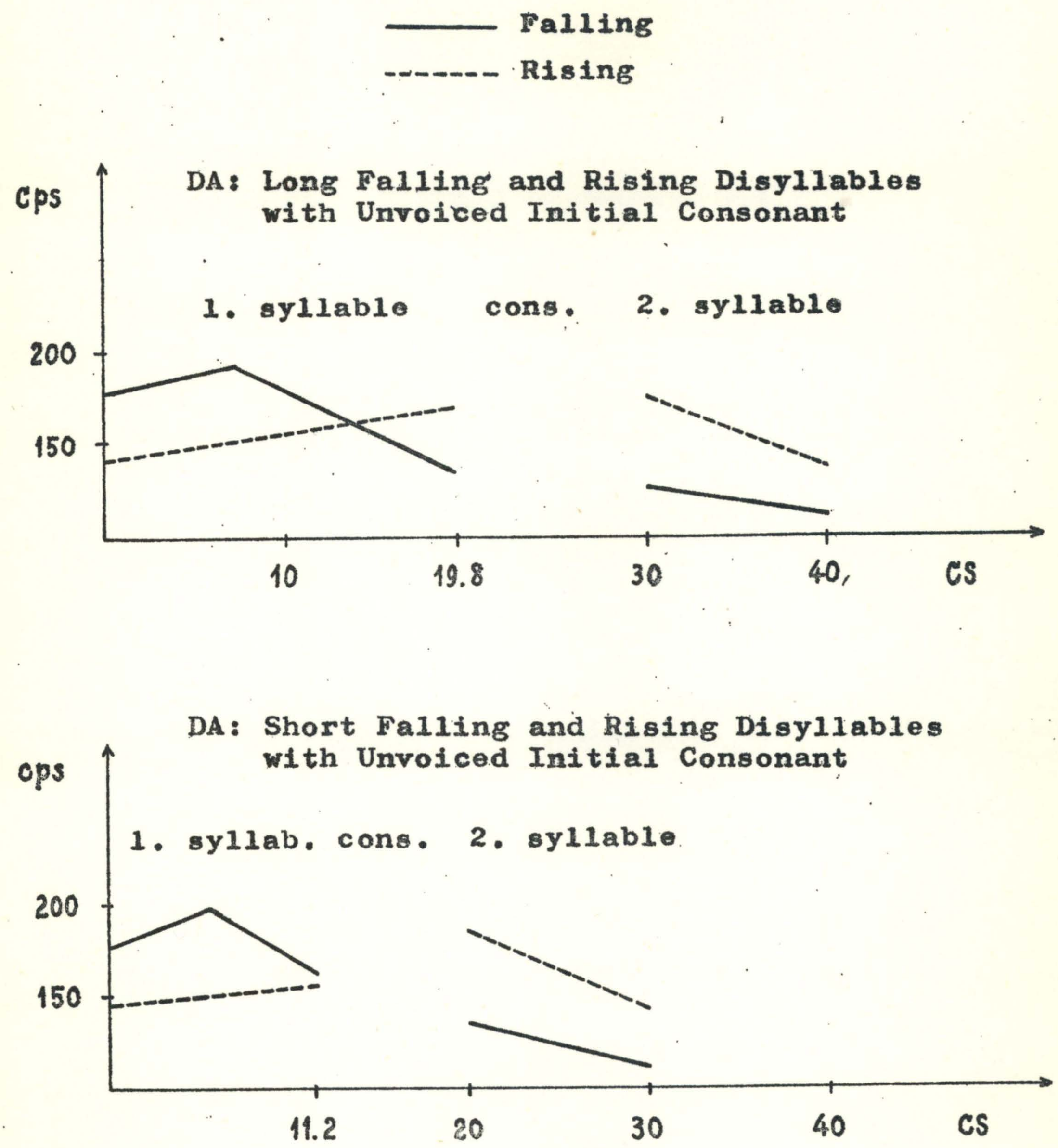
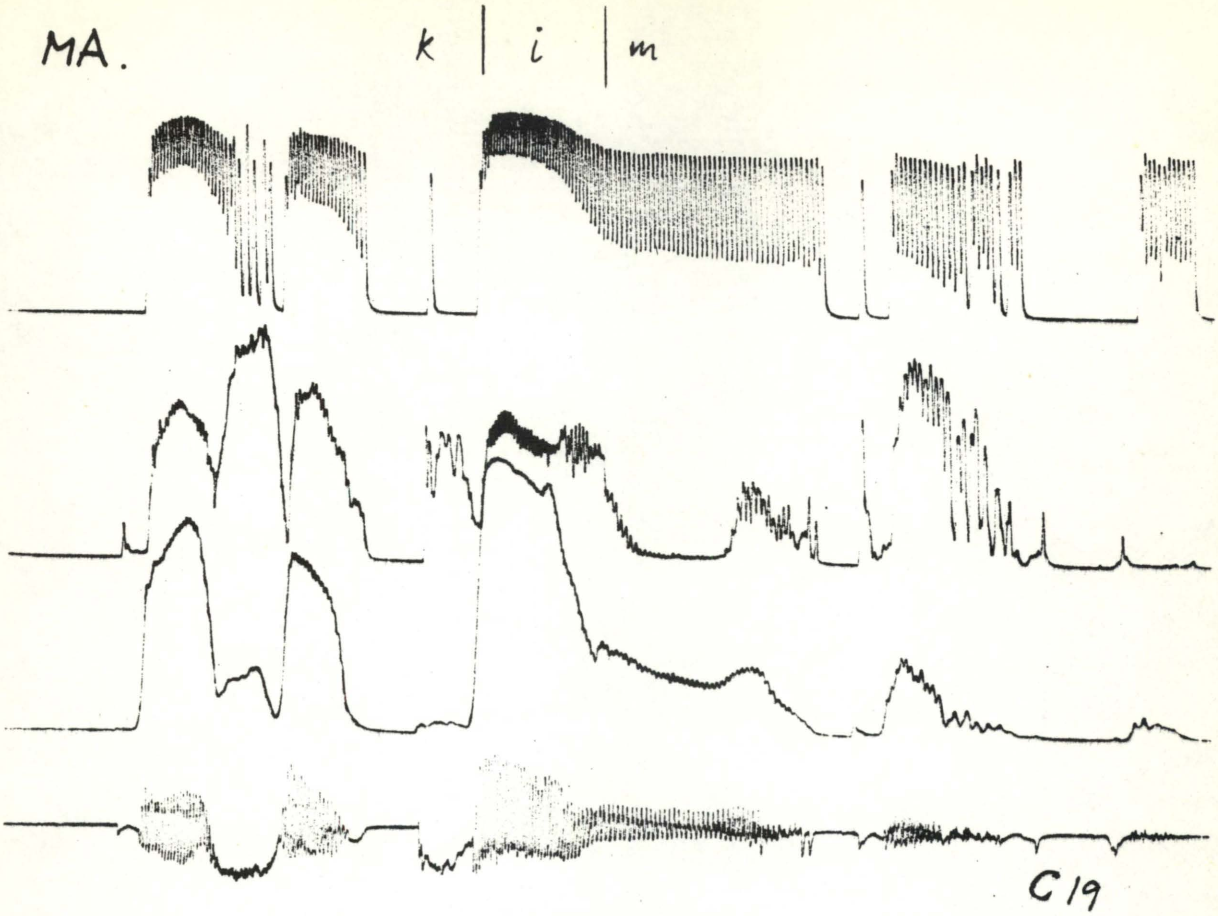


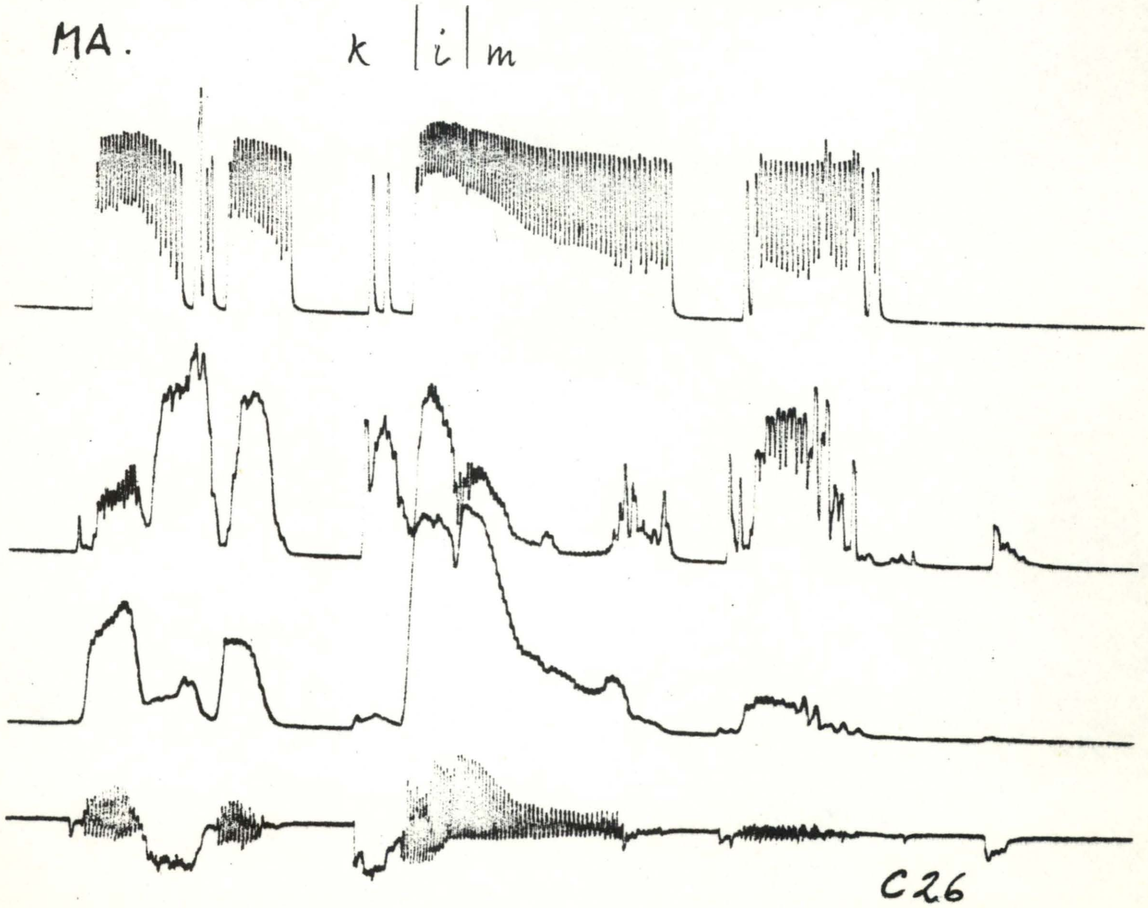
Fig. 2d

The diagrams above are graphic illustrations of the measured values given in Table 6. The short vowel duration shown here represents the average of all short vowels, i.e. 11.2 cs. The long vowel duration represents the average of all long vowels, i.e. 19.8 cs. The consonants and the second syllables which have not been measured are arbitrarily set to 10 cs each (the consonants in long disyllables to 10.2 cs).

In monosyllables ——— symbolizes long and short vowels preceded by an unvoiced initial consonant, while ----- symbolizes vowels preceded by a voiced initial consonant. In disyllables ——— symbolizes falling, ------ rising word tone.

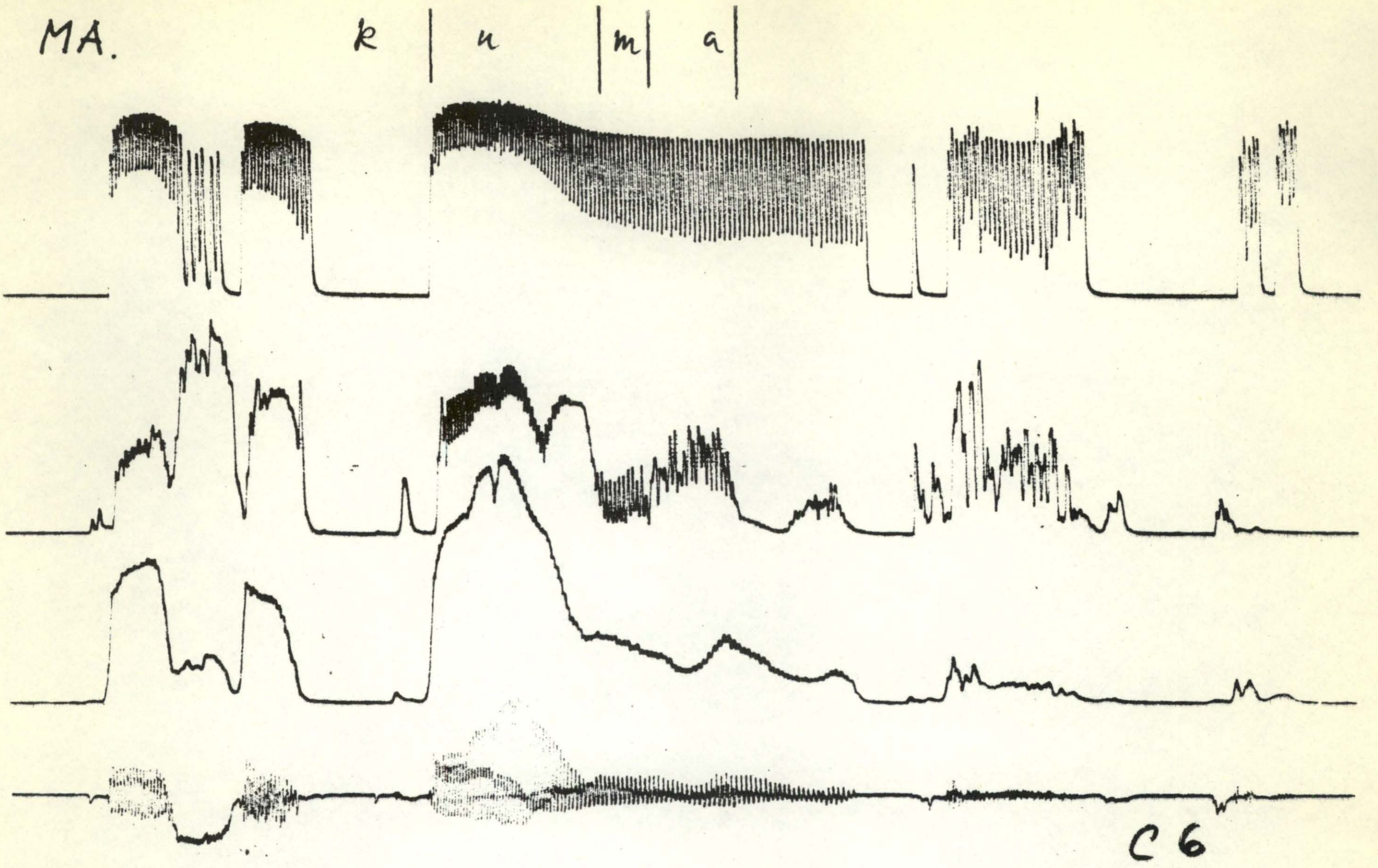


kîm 'who, instr. sg. masc./neut.' (V̀:)
(Traces from top to bottom: (1) pitch curve, (2) highpass filtered intensity curve, (3) unfiltered intensity curve, (4) "duplex oscillogram. See text for details.)

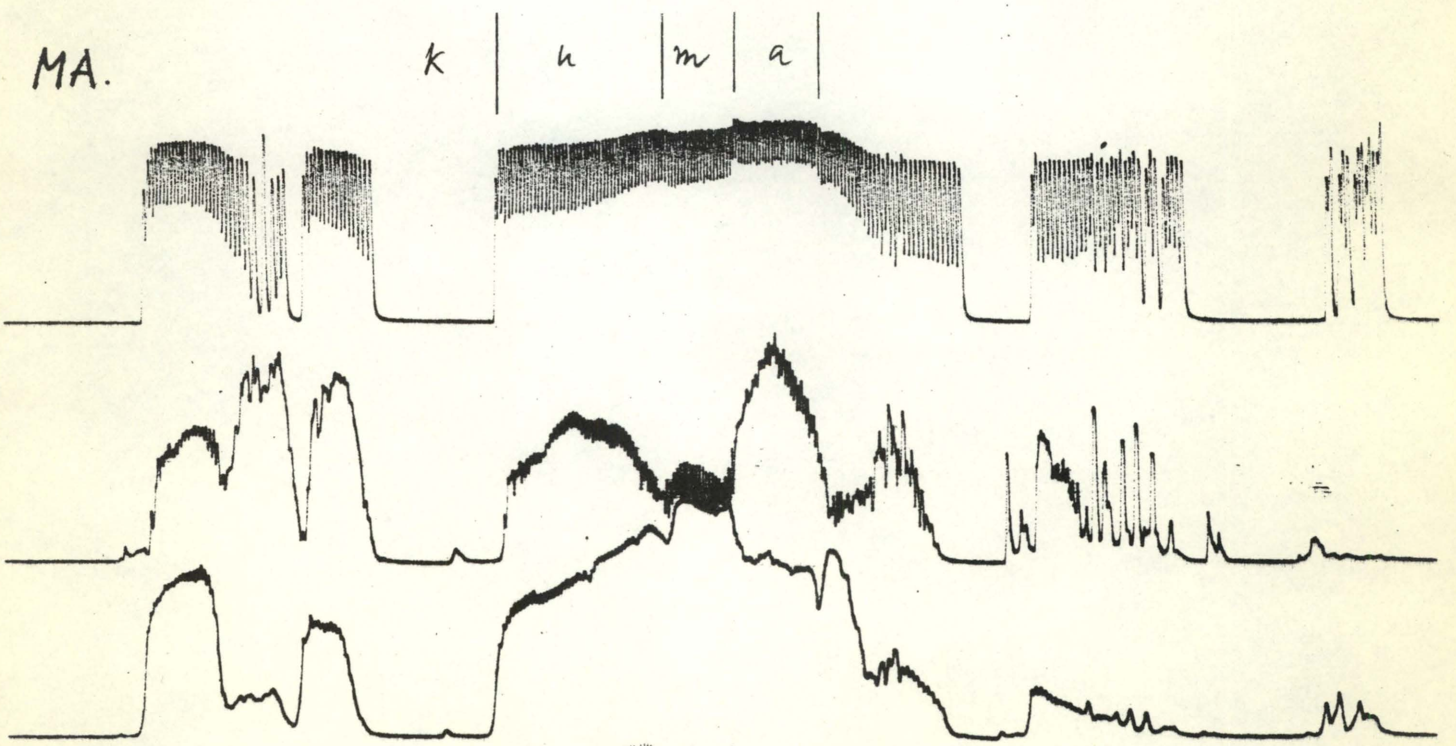


kȉm 'caraway' (V̀)

Fig. 3
Mingograms of monosyllables.

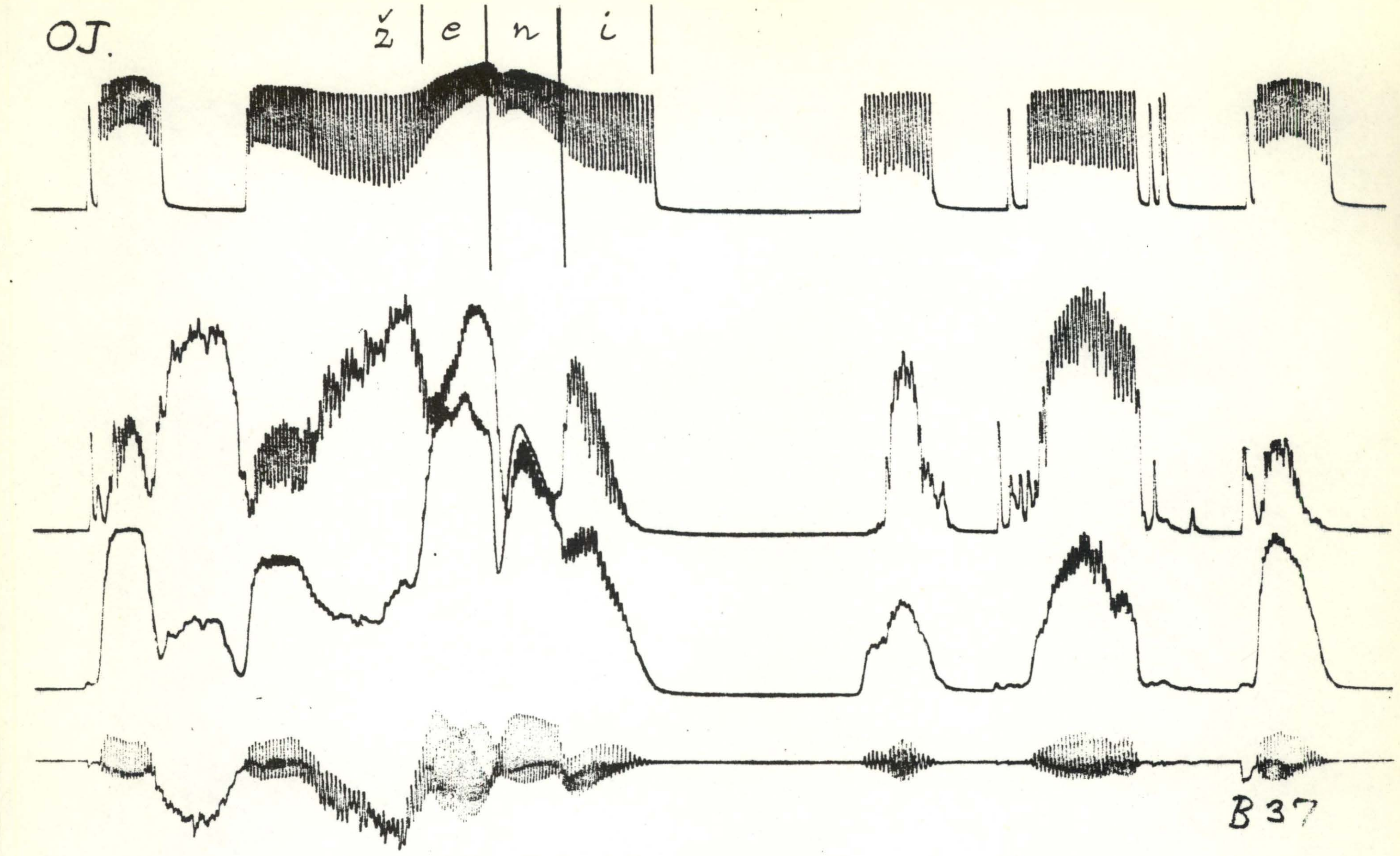


k$\hat{\text{u}}$ma 'godfather, acc./gen. sg.'　　(v̀:)

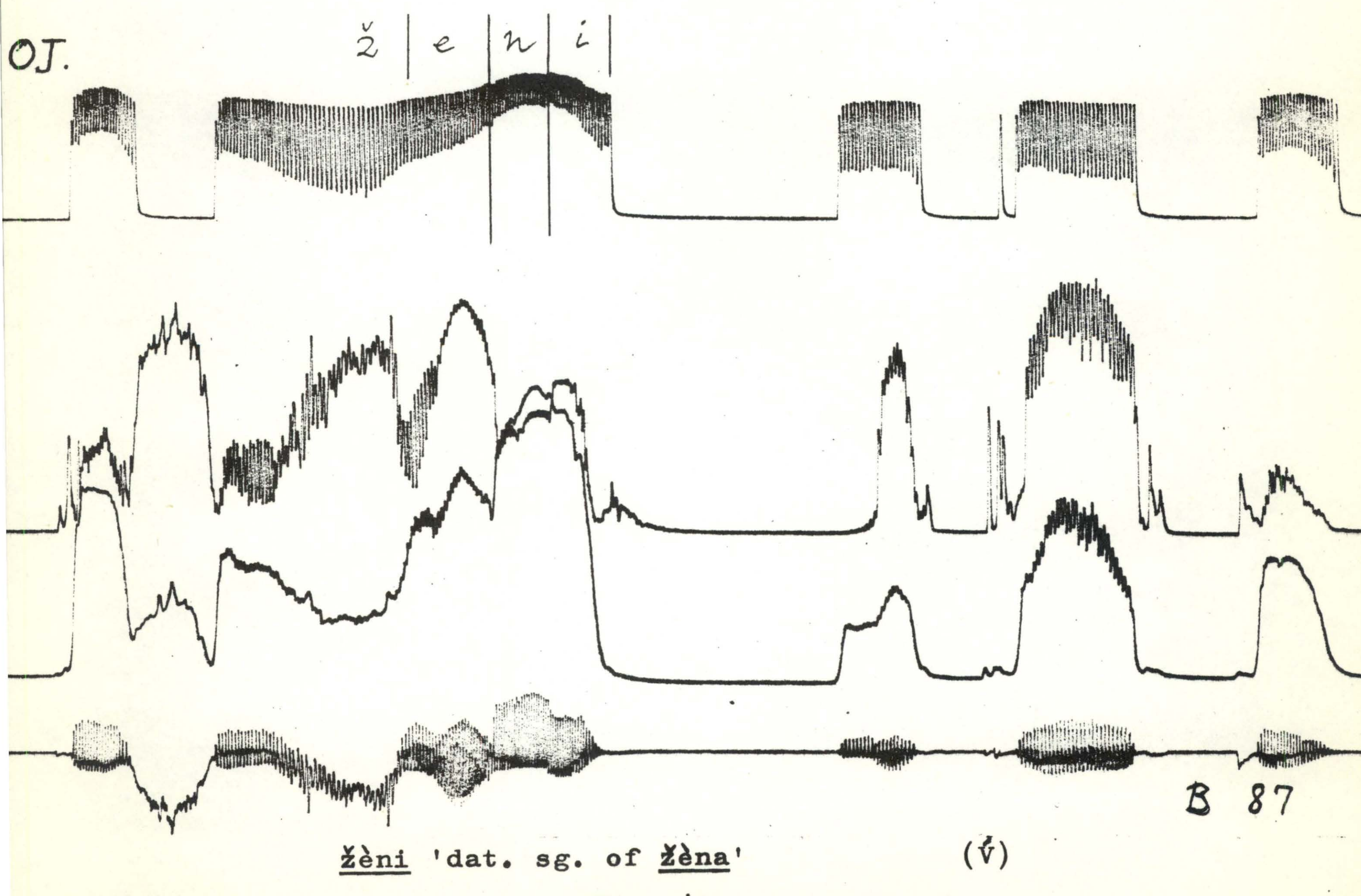


kúma 'godmother'　　(v́:)

Fig. 4a
Mingograms of disyllables.



<u>žȅnī</u> 'he gives in marriage' (v̀)



<u>žèni</u> 'dat. sg. of <u>žèna</u>' (v́)

Fig. 4b
Mingograms of disyllables.

DURATION OF FRENCH VOWELS BEFORE FRICATIVES *)

Karen Landschultz

According to the general opinion as regards vowel duration in French, only the final (stressed) vowel of a tone group can be prolonged. The traditional rule tells that in this position the prolongation applies to 1) all vowels before voiced fricatives and [r], 2) [o], [ø], [ɑ] before any pronounced consonant, and 3) the nasal vowels likewise before any pronounced consonant.

The main purpose of this work was to examine the validity of 1) and 2) - with exception of the position before [r] due to difficulties in determining the boundaries of this sound on a mingogram. - Furthermore the duration of the fricatives was examined. - Four Frenchmen (two men: AM 57 and SM 26 years old and two women: CHH 30 and ThM 17 years old) were used.

The following vowels and positions were examined (mainly based upon Straka's vowel scheme in <u>Bulletin de la faculté des lettres de Strasbourg</u> (1950), pp. 220 and 373):

stressed vowel										before
i	y	u	o	ɔ	ɛ	a	-	-	œ	f / v
i	y	u	o	-	ɛ	a	ɑ	ø	-	s / z
i	y	u	o	ɔ	ɛ	a	ɑ	-	-	ʃ / ʒ

As it might be of some interest to see whether vowel duration is influenced by a following voiced fricative when the vowel is not stressed, some examples were set up in which the vowel occurred in an unstressed open syllable, the next syllable beginning with a fricative. The following vowels were examined:

unstressed vowel					before
i	u	o	a	-	f / v
i	u	o	- **)	ɑ	s / z
i	u	o	a	ɑ	ʃ / ʒ

---

*) Thesis work for the cand.art. degree, completed in November 1966. The results will appear in a more detailed form in <u>Revue Romane</u>.

**) The examples in this position were deleted by mistake during the recordings.

The conditions are not alike in the two positions, the syllable in stressed position being closed, that in unstressed open. It was rather difficult to find a sufficiently large material in closed unstressed syllable, as it was required that the conditions should be quite alike in stressed and unstressed position.

The preceding consonants were for the most part unvoiced stops or unvoiced fricatives.

All stressed as well as all unstressed vowels appeared in two examples each recorded three times. Thus for each subject there are six measurements in every position. - The material was recorded in random order.

The examples were presented in short sentences, stressed vowels in utterances like, for instance, 'C'est un homme actif qui fait bien des choses', 'On le voit qui bouge mais on n'entend rien', unstressed vowels in, for instance, 'Elle va l'attifer et c'est très mauvais', 'Il vient de bouger malgré mes conseils'.

The rate of speech was controlled. There was no considerable variation.

A material consisting of the above-mentioned examples presented isolated in a sort of enumeration was also recorded, but it turned out to be inadequate due to uneven accentuation and varying rate of speech.

The recordings were made in the acoustical laboratory of the Technical High School (Danmarks Tekniske Højskole), and mingograms were made at the Institute of Phonetics showing pitch and intensity in linear as well as logarithmic display. A fourth trace presented a duplex oscillogram. Segmentation of the sounds were made by comparing the different traces. A control by spectrograms gave evidence that the delimitations were correct.

I. Relative vowel duration.

The following averages have been calculated for the duration of the stressed vowels (measured in centiseconds):

	(AM)						(SM)					
before:	[f]	[v]	[s]	[z]	[ʃ]	[ʒ]	[f]	[v]	[s]	[z]	[ʃ]	[ʒ]
[i]	10.9	27.9	9.6	16.3	9.1	17.4	12.2	18.0	12.5	22.0	11.1	23.6
[y]	9.8	18.8	9.9	17.0	11.3	15.3	10.4	19.2	14.2	20.8	14.8	22.2
[u]	10.4	19.8	12.0	15.8	11.1	25.2	10.8	22.9	13.8	24.1	13.9	26.4
[o]	16.1	16.7	21.2	23.2	16.8	22.0	14.8	22.2	18.2	19.8	16.3	24.6
[ɔ]	12.4	15.6	-	-	13.8	18.4	15.1	23.1	-	-	12.3	23.6
[ɛ]	10.6	21.9	10.6	22.9	11.5	18.3	12.8	16.6	12.0	21.7	12.8	17.2
[a]	17.0	19.9	13.7	20.5	14.0	18.6	17.8	23.6	13.8	22.8	15.3	26.2
[ɑ]	-	-	14.7	26.2	21.2	22.9	-	-	16.8	25.2	18.9	23.5
[ø]	-	-	18.9	24.7	-	-	-	-	-	-	-	-
[œ]	11.0	22.8	-	-	-	-	13.2	20.1	-	-	-	-
	(ThM)						(CHH)					
[i]	10.3	13.5	10.3	16.7	11.7	15.8	8.5	12.5	9.2	14.1	9.8	14.1
[y]	12.3	16.2	11.9	18.1	12.8	16.3	8.4	10.8	8.7	12.8	11.0	11.3
[u]	11.3	15.2	12.9	18.0	12.8	20.4	9.0	12.0	10.3	14.4	9.4	16.4
[o]	15.6	18.4	15.2	18.6	15.7	20.2	10.5	13.2	13.2	14.7	12.8	15.8
[ɔ]	17.1	19.6	-	-	15.2	20.9	11.6	14.8	-	-	11.6	16.6
[ɛ]	11.9	16.9	15.8	20.3	16.1	20.3	9.6	13.1	11.3	14.6	12.7	15.5
[a]	17.8	18.7	16.3	22.9	18.3	20.0	13.7	18.8	12.7	17.3	12.1	15.5
[ɑ]	-	-	17.5	22.5	18.7	23.0	-	-	11.3	17.5	14.3	16.0
[ø]	-	-	15.8	19.9	-	-	-	-	12.3	18.3	-	-
[œ]	14.7	18.0	-	-	-	-	12.4	12.1	-	-	-	-

A rather stable tendency toward shortest duration for the closed [i y u] and toward longest duration for the open [a ɑ] is seen. The half-closed [o ø] are rather long for AM and SM before unvoiced fricative. The half-open [ɛ œ ɔ] fluctuate between the durations of the shortest and the longest vowels. ([ɛ] is throughout rather short.) (Fig. 1 illustrates the relations for ThM.) The durations and differences are throughout smaller for ThM and CHH than for AM and SM. - In spite of the exceptions which can be found in the table, it seems reasonable to state that the duration of the vowels - except for [o ø] - tends to be proportional to the degree of opening.

```
i   uεy           œ o      ɔ a                          [f]
      12      14      16      18      20      22

i    y      u      o  εøa     α                         [s]
      12      14      16      18      20      22

     i   y,u        ɔoε         aα                      [ʃ]
      12      14      16      18      20      22

            i       u   y ε    œoa  ɔ                   [v]
      12      14      16      18      20      22

                          i   uyo     øε        αa      [z]
      12      14      16      18      20      22

                      iy             a͡oεuɔ        α     [ʒ]
      12      14      16      18      20      22
```

Fig. 1

The relative vowel duration. ThM.

```
 f  s ʃ            v      z,ʒ                           [i]
       10       12       14       16       18

fs         vʃʒ        z                                 [y]
       10       12       14       16       18

   fʃ     s       v         z        ʒ                  [u]
       10       12       14       16       18

          f          ʃ v,s    z   ʒ                     [o]
       10       12       14       16       18

                f,ʃ           v        ʒ                [ɔ]
       10       12       14       16       18

      f     s        ʃ v     z    ʒ                     [ε]
       10       12       14       16       18

                   ʃ s   f        ʒ       z       v     [a]
       10       12       14       16       18

           s               ʃ        ʒ   z               [ɑ]
       10       12       14       16       18

                  s                           z         [ø]
       10       12       14       16       18

                 vf                                     [œ]
       10       12       14       16       18
```

Fig. 2

Influence from the place of articulation of the following consonant and from voicing. CHH.

The following averages have been calculated for the duration of the unstressed vowels (measured in centiseconds):

before:	(AM) [f]	[v]	[s]	[z]	[ʃ]	[ʒ]	(SM) [f]	[v]	[s]	[z]	[ʃ]	[ʒ]
[i]	7.3	9.8	8.7	12.8	8.3	12.9	6.3	8.4	6.1	9.9	5.9	10.9
[u]	9.2	9.5	9.3	11.6	8.7	14.3	6.3	8.1	7.5	8.3	7.4	11.8
[o]	9.7	13.2	11.9	17.4	13.2	18.6	8.9	9.6	9.0	10.4	9.7	11.9
[a]	11.4	13.8	-	-	11.3	13.6	8.3	11.2	-	-	10.2	13.1
[ɑ]	-	-	10.8	16.2	13.1	18.1	-	-	8.5	12.0	9.8	12.7

	(ThM)						(CHH)					
[i]	5.7	6.4	5.4	6.1	5.3	6.7	4.8	6.1	6.8	7.0	5.3	8.2
[u]	5.8	5.2	5.3	8.4	5.8	10.8	7.3	6.0	5.6	5.9	5.6	8.7
[o]	6.4	7.1	7.8	9.4	7.7	11.2	7.2	7.5	7.9	9.3	7.9	11.0
[a]	8.8	9.7	-	-	9.9	11.2	9.2	9.4	-	-	7.8	10.9
[ɑ]	-	-	7.8	12.5	10.8	12.2	-	-	8.3	12.2	11.0	12.7

In this position the degree of opening is easily seen to be a determinating factor for duration except for [o]. The differences are small, but rather stable. These facts should not surprise: the unstressed vowels examined here are - except for [o] - articulatorily in extreme positions and therefore the overlapping is naturally smaller. Furthermore the unstressed syllable can be considered to be less exposed to influences from (greater or smaller) changes in stress, rate of speech, and intonation. The latitude of variation is thus smaller for the unstressed than for the stressed vowels.

II. Influence from the place of articulation of the following consonant.

An examination of the influence due to place of articulation of the following fricative (illustration fig. 2, CHH) reveals in stressed syllable a tendency for the vowel length to be shortest before [f / v], in unstressed syllable only before [f] . The relations are throughout rather unstable for the remaining fricatives, - however, a certain tendency can be seen in stressed syllable toward longest duration before [ʃ/ʒ], in unstressed only before [ʒ].

The theory according to which a larger movement conditions a longer duration can explain the relative vowel duration as well as the influence due to place of articulation of the following consonant. As consonants are characterized by stop or fricative articulation, closed vowels are more similar and open vowels less similar to the surrounding consonants. Open vowels consequently need longer duration than closed vowels. - The half-closed [o ø] form an exception. Their long duration might be explained as diachronic compensatory prolongation. - According to the theory back vowels should be longer than the corresponding front vowels, the back of the tongue articulating more slowly than the front of the tongue. Furthermore rounded vowels should be longer than the corresponding unrounded, the rounding being a rather complicated articulation. This material does not confirm these two points. - The movement of articulation from vowel to [f/v] does not imply any change in tongue position. The movement consequently has no prolongating effect. The relatively complicated movement of articulation of [ʃ/ʒ] might explain the few tendencies toward longest vowel duration before these fricatives.

III. Influence of voicing.

As regards the influence from the voiced counterparts of the six fricatives upon vowel duration, a prolongation is quite clearly seen in stressed position; it is less evident but still relatively stable in unstressed position. (See fig. 2, CHH, stressed position). ThM and CHH pronounced practically all [v z ʒ] (partly) unvoiced. The preceding vowels did, however, show a significantly longer duration than the vowels followed by [f s ʃ]. The average prolongation was in stressed position for AM 7.1 cs, for SM 8.1 cs, for ThM 4.3 cs, and for CHH 3.6 cs. In unstressed position the prolongation was for AM 3.8 cs, for SM 2.6 cs, for ThM 1.9 cs, and for CHH 1.6 cs. The differences are thus smallest for ThM and CHH.

A calculation of the percentual average prolongation in stressed position gave the following results:

	[i]	[y]	[u]	[o]	[ɔ]	[ɛ]	[a]	[ɑ]	[ø]	[œ]	
AM:	156.4	90.7	90.5	3.6	25.5	107.1	17.1	-	-	106.8	[v/f]
	69.6	71.7	31.9	9.5	-	116.5	50.1	78.4	30.7	-	[z/s]
	91.7	35.6	127.1	31.3	33.3	59.4	32.9	8.3	-	-	[ʒ/ʃ]
SM:	48.0	84.1	111.5	50.0	53.0	29.7	32.8	-	-	52.6	[v/f]
	76.0	47.1	75.1	9.2	-	80.5	65.1	49.5	-	-	[z/s]
	112.6	50.0	89.9	51.3	92.5	34.4	70.6	24.3	-	-	[ʒ/ʃ]
ThM:	31.7	31.1	33.8	18.2	14.6	42.0	4,7	-	-	22.8	[v/f]
	62.5	40.2	39.5	22.6	-	28.4	41.0	28.6	25.8	-	[z/s]
	35.8	28.1	59.1	28.7	37.9	26.4	9.6	23.3	-	-	[ʒ/ʃ]
CHH:	47.1	28.8	33.3	25.3	28.1	36.5	37.8	-	-	neg.	[v/f]
	53.7	47.1	39.5	11.4	-	29.6	36.3	54.5	48.7	-	[z/s]
	43.2	3.0	74.4	22.8	43.2	22.4	28.3	12.3	-	-	[ʒ/ʃ]

The table reveals that the percentual prolongation differs according to the degree of opening: The closed vowels tend to show a larger percentual prolongation than the open vowels, [o], and [ø]. The half-open vowels, however, show no quite clear tendencies. Except for these vowels, the percentual prolongation is thus inversely proportional to the relations found for the relative vowel duration. The fact that the "long" vowels are not prolonged to the same extent as the "short" vowels, can be considered as a certain tendency to limit the sound duration.

IV. Duration of the fricatives.

Finally as well as initially (for instance in 'acti<u>f</u>'/'acti<u>v</u>e' 'atti<u>f</u>er'/'acti<u>v</u>er') the unvoiced fricatives of AM and SM showed a longer duration than their voiced counterparts. (Figs. 3 and 4 show this relation for SM). [v z ʒ] of ThM and CHH, too, which were pronounced (partly) without voicing showed a significantly smaller duration than the corresponding [f s ʃ]. Thus the duration of the fricatives is inversely proportional to the duration of the preceding vowel: a short vowel is followed by a long fricative, and vice versa. A calculation proved that the group vowel + unvoiced fricative and

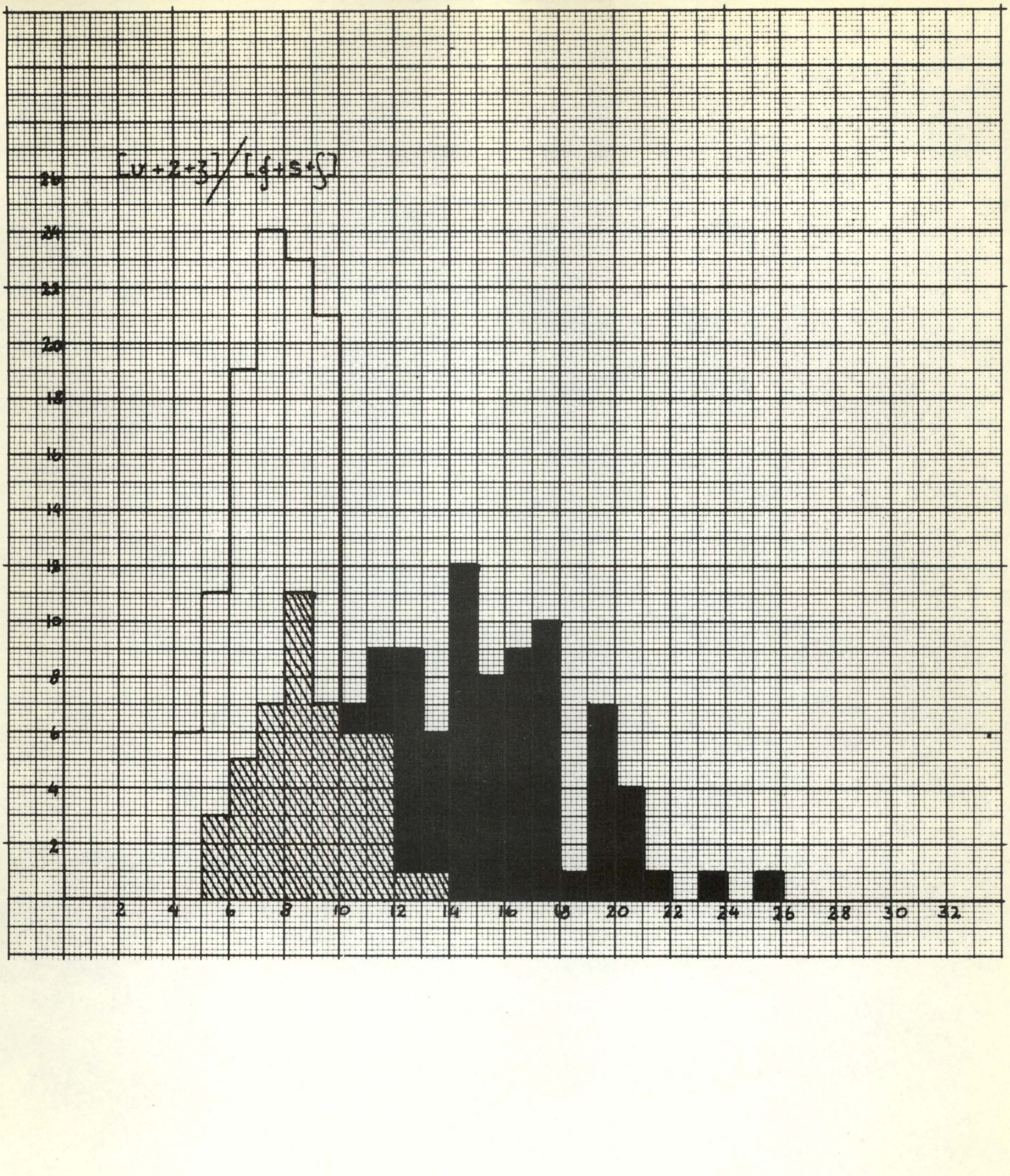


Fig. 3

Duration of the final stressed fricatives. SM.

Average (centiseconds): v+z+ʒ : 7,8
f+s+ʃ : 13,0

White areas = voiced fricatives
Black areas = unvoiced fricatives
Hatched areas = overlapping voiced and unvoiced fricatives.

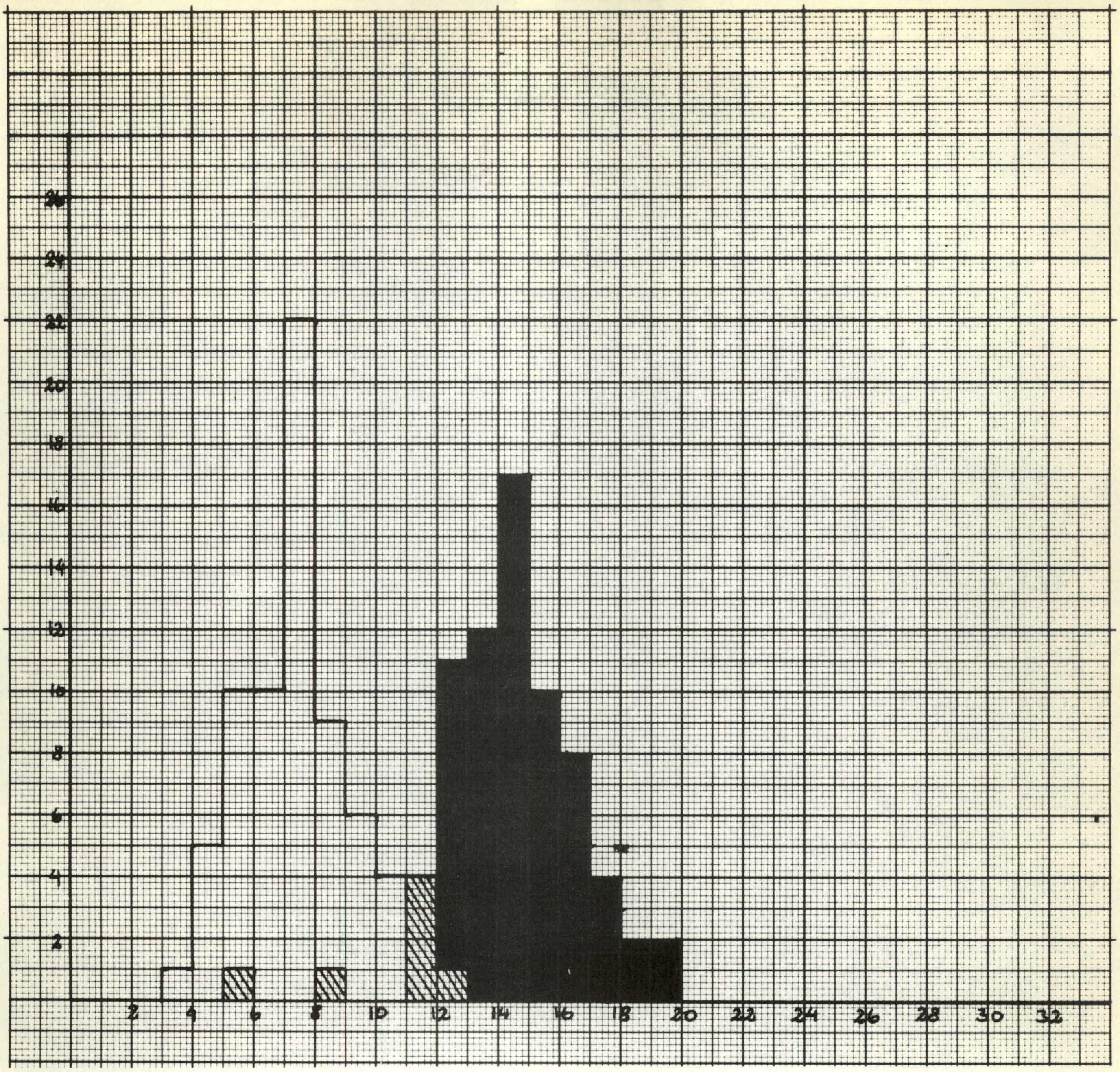


Fig. 4

Duration of the initial stressed fricatives. SM.

Average (centiseconds): v+z+ʒ : 7,3
f+s+ʃ : 14,2

White areas = voiced fricatives
Black areas = unvoiced fricatives
Hatched areas = overlapping voiced and unvoiced fricatives.

the group vowel + voiced fricative tend to be quantitatively alike. - Furthermore the final fricative showed a considerably larger latitude of variation than the initial fricative. -

The prolongation before voiced fricatives and the relatively large percentual prolongation of "short" vowels may both be explained as a tendency to maintain a certain rhythm. Thus the varying duration is due to compensation. - It would be of great interest in this connection to examine why unvoiced fricatives are longer than their voiced counterparts.

It remains to be mentioned that a calculation of the significance of the various distributions gave satisfying results.

Reference:

E. Fischer-Jørgensen, "Sound Duration and Place of Articulation", Zeitschrift für Phonetik 17, Heft 2-4 (1964), pp. 175-2o7.

A FORMANT ANALYSIS OF THE CLEAR, NASALIZED AND MURMURED VOWELS IN GUJARATI

R. Dave

Gujarati, a language of Gujarat State of Western India, has a series of clear (normal), nasalized and murmured (breathy or aspirated) vowels, the study of which is interesting both from a phonemic and from a phonetic point of view. I undertook a comparative formant analysis of these vowels based on spectrograms in June, 1966. Two lists were spoken by three informants: SK, RD and RT. List I contained clear and nasalized vowels spoken in words and in isolation. List II contained clear and murmured vowels spoken in words, phrases and sentences. List II contained commutable pairs. Nasalized murmured vowels were taken only as specimens.

Care was taken to prepare the lists according to the principles laid down by various phoneticians about the influence of the consonants on the vowel formants. Combinations of dental (alveolar) consonants with front vowels and of labial consonants with back vowels have generally been preferred in list I. Wherever this was not possible, an [h] or a velar or [l] was selected. It did not become possible to follow this criterion in the case of list II.

Both lists were read twice by each informant. RT read list II five times. Spectrograms were taken of one reading. Thus, in list I one example of each word has been measured for each person. There were 41 samples of the clear vowels and 44 of the nasalized vowels. List II contained 4-5 examples of each word, altogether 126 samples in pairs. RT had a few more. Three kinds of spectrograms were taken of each sample: wide band, narrow band, and narrow band section. In all, approximately 2270 spectrograms have been taken and measured.

F_o (the fundamental) was measured on the section by dividing the distance between the first and the eleventh harmonic by ten. Generally the section was taken in the middle of the vowel. The frequencies of the formants were determined by a comparison between the three kinds of spectrograms. As the wide band spectrograms gave somewhat higher values, and seemed to be less reliable than the narrow band and section, the measurements were mainly based on the two latter types. The peak of the envelope on the section was considered to be the centre frequency of the formant.

Generally there was no difficulty in measuring F_o. It was difficult to locate the centre frequency for weak and split formants and for two very close formants. Nasal consonants or [h] preceding a vowel also gave some difficulties.

The Kay Electric Sonagraph purchased in 1953 was used for analysis, the new Sonagraph purchased in 1967 was used for some control spectrograms.

The charts give a graphical display of F1 versus F2. There is overlapping between the areas of [a] and [ã] for all informants. Moreover, the areas of [a] and [ə̃] are very close to each other. The same is true of [u] and [ũ]. In RT a slight overlapping between [u] and [ũ] is seen. In SK [a], [ã], [ə̃] (and also [ə]) are very close to each other, while in RT [ɔ̃] is closer to [a]. In SK and RD [a] and [ə̃] are both found in the same area as [a].

There is almost a complete overlapping of the F1-F2 areas of the clear and murmured vowels (see charts).

Tables of averages of the clear and nasalized vowels, and the clear and murmured vowels have been given on pages 121-124.

Some of the charts have been given on pages 128 ff.

<u>Clear and nasalized vowels. The major characteristic features of nasalized vowels.</u>

The following features of the nasalized vowels have been observed:

(i) There is no constant difference in F_o. In many cases F_o is higher for the nasalized vowels, but not always. The average of F_o for [ĩ], [ũ] and [ã] of all informants is higher than in the corresponding oral vowels. Although the differences are small and not found in all vowels, it should be noted that they are constant in the case of the close or narrow vowels. More investigations are necessary.

(ii) The open nasalized vowels ([ɛ̃ ã ə̃ ɔ̃]) have a higher F1 than the corresponding clear vowels, whereas [ũ] has a lower F1 than [u]. In SK and RT the nasalized [ĩ] has a somewhat lower F1 than [i].

(iii) F1 for the open nasalized vowels is weakened. This has been said to be the major cue for nasality by various scholars, (1-5) and it is quite evident on the spectrograms.

Table 1. SK. averages of clear and nasalized vowels. (List I. A-B)

No.	Samples	No. of samples	F_o	Fs	F1	F2	F3	F4
1.	[i]	(9)	130		263	2358	2989	4000
2.	[ĩ]	(5)	143		255	2758	3250	4033
3.	[e]	(6)	131	(225)(1)	420	2200	2783	3820
4.	[ɛ̃]	(6)	132	237	671	2221	2825	3683
5.	[a]	(8)	118	(204)( )	747	1409	2212	(2983?)
6.	[ã]	(6)	126	238	779	1375	2137	(3650?)(1)
7.	[ə]	(5)	128		603	1330	2455	(3800?)(1)
8.	[ə̃]	(4)	130	239	706	1250		3683
9.	[o]	(5)	127	(222)	493	770		
10.	[ɔ̃]	(6)	126	242	630	905	(2800?)(1)	
11.	[u]	(5)	128		313	721		
12.	[ũ]	(6)	142		253	625	(2300?)(1)	2862

Table 2. RD. averages of clear and nasalized vowels. (List I. A-B)

No.	Samples	No. of samples	F_o	Fs	F1	F2	F3	F4
1.	[i]	(6)	135		274	2358	2900	3804
2.	[ĩ]	(5)	139		278	2620	3203	3940
3.	[e]	(5)	132	(205)	427	2166	2660	3890
4.	[ɛ̃]	(6)	133	283	627	2190	2775	3927
5.	[a]	(9)	124	(204)	849	1315	2525	3211
6.	[ã]	(6)	130	290	932	1314	2379	3442
7.	[ə]	(5)	131	(210)	595	1450	2475	3794
8.	[ə̃]	(4)	130	236	785	1360	2462	(3600?)(1)
9.	[o]	(5)	131	(200)	474	912	(2450?)(2)	3450
10.	[ɔ̃]	(6)	134	256	617	949	2528	3542
11.	[u]	(5)	131		330	816	(2500?)(1)	
12.	[ũ]	(6)	133		268	678	2612	

Table 3. RT. averages of clear and nasalized vowels. (List I. A-B)

No.	Samples	No. of samples	F_o	Fs	F1	F2	F3	F4
1.	[i]	(6)	169		277	2104	2558	3312
2.	[ĩ]	(5)	177		255	2430	3060	3350?
3.	[e]	(6)	147		357	1896	2530	3454
4.	[ɛ̃]	(6)	169	222	748	1971	2667	3450
5.	[a]	(8)	142		781	1196	2537	3200
6.	[ã]	(6)	165	(205??)	795	1183	(2723?)	3212
7.	[ə]	(5)	160		595	1225	2500	3256
8.	[ə̃]	(4)	157	224	701	1175	2644	3382
9.	[o]	(5)	164		396	880	2606	3235
10.	[ɔ̃]	(6)	161	256	677	925	2487??	3370
11.	[u]	(5)	149		276	816	2490	(3250?)(2)
12.	[ũ]	(6)	186		253	700	2228	(??)

Table 4. SK. averages of clear and murmured vowels. (List II-C.)

No.	Samples	No. of samples	F1	F2	F3	F4
1.	[bi, či:r]	11	270	2409	3027	3954
2.	[bi, či:r]	11	272	2394	2959	3802
3.	[pelo, se:ǰ]	8	452	2059	2639	3583
4.	[pelo, se:ǰ]	8	478	1962	2631	3607
5.	[mɛ:k]	5	542	2215	3050	3535
6.	[mɛ:k]	5	555	2205	2925	3585
7.	[ba:r, maro, wali]	12	729	1344	2298	3647
8.	[ba:r. maro. wali]	9	712	1333	2278	3525
9.	[kəɽi]	5	562	1397	2208	3422
10.	[kəɽi]	5	545	1378	2238	3675(2)
11.	[mo:r, ko:ɽ, po:r, do:ɽ. ko]	20	493	830	2413	3656
12.	[mo:r, ko:ɽ, po:r, do:ɽ. ko]	18	494	867	2437	3623
13.	[du:dh]	4	280	942	2575	3825
14.	[du:d]	4	282	908	2550	3687

Table 5. RD. averages of clear and murmured vowels. (List II-C.)

No.	Samples.	No. of samples	F1	F2	F3	F4
1.	[bi, či:r]	8	253	2438	3150	3803
2.	[bi̱, či̱:r]	8	249	2487	3028	3831
3.	[pelo, se:ǰ, me:k]	12	463	2101	2689	3686
4.	[pe̱lo, se̱:ǰ, me̱:k]	12	466	2101	2760	3931
5.	[ba:r, maro, wali]	11	852	1229	2595	3288
6.	[ba̱:r, ma̱ro, wa̱li]	11	852	1229	2595	3288
7.	[kəɽi]	4	564	1363	2298	3725
8.	[kə̱ɽi]	4	577	1375	2419	3800
9.	[mo:r, ko:ɽ, po:r, do:ɽ, ko]	19	482	872	2505	3348
10.	[mo̱:r, ko̱:ɽ, po̱:r, do̱:ɽ, ko̱]	19	504	911	2607	3273
11.	[du:dh]	4	304	931	2442	3650?(1)
12.	[du̱:d]	4	350	944	2410	3460

Table 6. RT. averages of clear and murmured vowels. (List II-C and D.)

No.	Samples.	No. of samples	F1	F2	F3	F4
1.	[bi, či:r]	8	259	2169	2824	3467
2.	[bi̱, či̱:r]	7	259	2121	2764	3468
3.	[pelo, se:ǯ, me:l]	12	392	1817	2604	3446
4.	[pe̱lo, se̱:ǯ, me̱:l]	6	378	1751	2662	3508
5.	[mɛ:k]	4	485	1887	2700	3419
6.	[mɛ̱:k]	2	550	1862	2725	3550
7.	[ba:r, maro, wali, pa:ɽ, taro, ma:l, lawo]	22	759	1172	2688	3698
8.	[ba̱:r, ma̱ro, wa̱li, pa̱:ɽ, ta̱ro, ma̱:l, la̱wo]	10	759	1161	2704	3761
9.	[kəɽi]	4	514	1415	2279	3438
10.	[kə̱ɽi]	4	507	1300	2402	3444
11.	[mo:r, ko:ɽ, po:r, do:ɽ, ko, boɭũ]	22	388	867	2608	3583
12.	[mo̱:r, ko̱:ɽ, po̱:r, do̱:ɽ, ko̱, bo̱ɭũ	13	390	882	2650	3633
13.	[du:dh]	4	290	962	2525	3062
14.	[du̱:d]	3	282	932	2525	2983

I have noted weakening of F2 for the nasalized [ũ] in SK. Fant has referred to such a possibility. (1)

(iv) Higher formants are weakened. F3 is more affected than F4, and is sometimes eliminated. Both F3 and F4 tend to be raised.

The exact location of F3 and F4 sometimes makes difficulties because the formants may be split, and there may be extra resonances.

This observation is also generally in accordance with the observations made by other phoneticians.

(v) A sub-formant (a nasal formant) in the open nasalized vowels appears between 200-300 cps. This F_s is said to be a secondary cue for nasality. However, it seems to be a very important cue. F_s of nasalized vowels is higher in frequency than a sub-formant occurring in clear vowels. It is also stronger in intensity.

(vi) There seems to be a zero in the open nasalized vowels between 400-500 cps. This is also mentioned as a characteristic feature of nasalization of vowels. The valley in this frequency area seems to be due to an anti-resonance and not simply due to the raising of the formants in the process of nasalization.

Weakening of the third harmonic and reinforcement of the second is worth noting.

(vii) Small weak peaks are found at irregular frequencies. Extra formants are found, but they differ according to vowels and informants. However, the frequency region pointed out by Hattori, Fujimura and Kajiyama still seems to be the main region of such formants (between 1000 and 2500 cps).

(viii) I have noticed an increase in the bandwidths of formants, especially of F1, in some spectrograms. However, I have not observed this in F1 of [ĩ] and [ũ]. Probably F2 of [ũ] sometimes has more bandwidth than F2 of [u]. I have not undertaken any measurements of formant bandwidth.

(ix) Gujarati vowels seem to be uniformly nasalized throughout. In this respect they differ from the French nasalized vowels (cp. Hattori (2). Also see Kongsdal in ARIPUC(1966) (3)). The nasalized vowels in French Canadian, as investigated by Jean Gondron, seem to possess similar features. Kongsdal has, however, found a constant degree of nasalization in isolated French vowels.(3)

Clear and murmured vowels.

(i) Lowering of pitch is said to be an important characteristic feature of murmured vowels.(6) My measurements do not support this assumption. It must, however, be remembered that the measurements of F_o for different vowels were made approximately in the middle of the vowel, and not at exactly the same place.

(iia) There are no constant and regular differences between the formant frequencies of clear and murmured vowels. But many times the formants of the murmured vowels are higher than those of the clear vowels. The difference is, however, so small that one cannot take this into account. The overlapping is obvious.

The second harmonic of the murmured vowels often seems to be weaker in energy than the corresponding harmonic of the clear vowels. In nasalized vowels this harmonic is stronger than that of the clear vowels.

(iib) The formants of clear vowels are more clearly visible than those of nasalized vowels. A comparison between the nasalized and murmured vowels gives the same result. This does not mean that the formants of murmured vowels are quite normal and regular. Many irregularities are seen in the spectrograms, varying for different persons and word samples, a few being common to all informants. F1 of open murmured vowels is not always affected by the process of murmurization or aspiration. If it is affected, it is weakened. The following characteristics have been noticed: (1) Weakening of F1. (2) Weak and split higher formants. (3) A hole in the spectrogram in the higher frequency region, or sometimes lower. This may be found in the spectrograms of clear vowels, too. (4) F4 seems to be weaker than F3 in contradistinction to the nasalized vowels. (5) Only a few times noise has been noticed at very high frequencies. I did not attempt to measure formants higher than F4.

(iii) Whereas F_s of nasalized vowels is raised a little, that of the murmured vowels seems to be slightly lower.

(iv) Extra formants seem to occur in the same frequency regions as have been mentioned in the discussion of the nasalized vowels.

(v) Sometimes an extra stress is heard. It is not clear whether the murmured vowels are normally perceived as more strongly stressed than the clear vowels.

(vi) No constant change in the length of the murmured vowels has been observed, but exact measurements have not been undertaken.

In bisyllabic words the consonant following the murmured vowel is relatively shorter than the consonant following the clear vowel. In some cases the second vowel is also of shorter duration.

References:

(1) G. Fant, Acoustic Theory of Speech Production, pp. 148-149, and 159.

(2) S. Hattori, K. Yamamoto, and O. Fujimura, "Nasalization of vowels in relation to nasals", JASA 30 (1958), pp. 267-274.

(3) Ole Kongsdal Jensen, Træk af de franske nasalvokalers akustiske og fysiologiske struktur, pp. 12-23. Also "Features of the acoustical and physiological structure of the French nasal vowels", ARIPUC 1 (1966), pp. 59-66.

(4) Svend Smith, "Vocalization and added nasal resonance", Folia Phoniatrica (1951), p. 167.

(5) House and Stevens, "Analog studies of nasalization", Journal of Speech and Hearing Disorders 21 (1956), pp. 218-232.

(6) P. B. Pandit, "Nasalization, aspiration, and murmur in Gujarati", Indian Linguistics (1957).

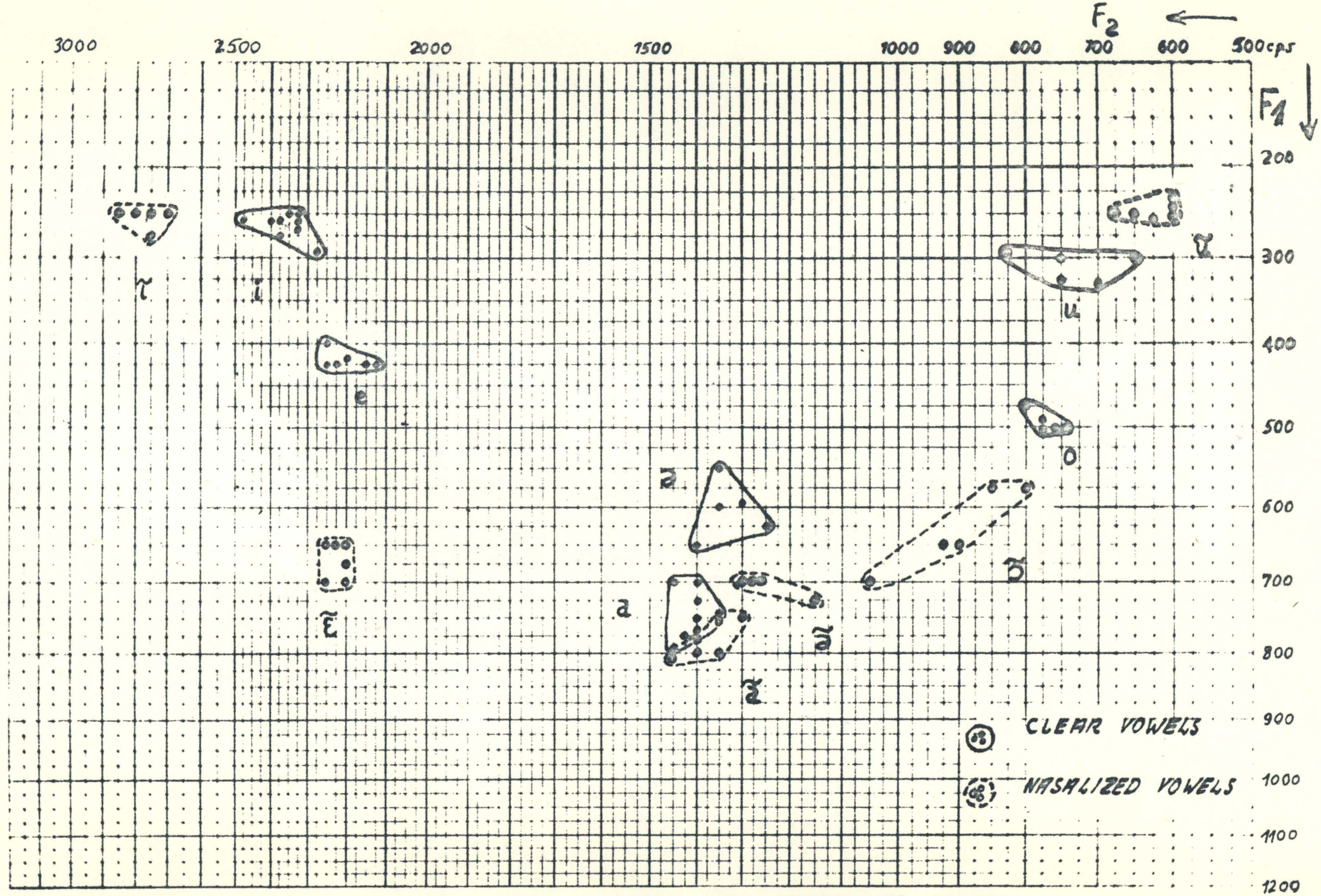


Chart 7 SK CLEAR AND NASALIZED VOWELS

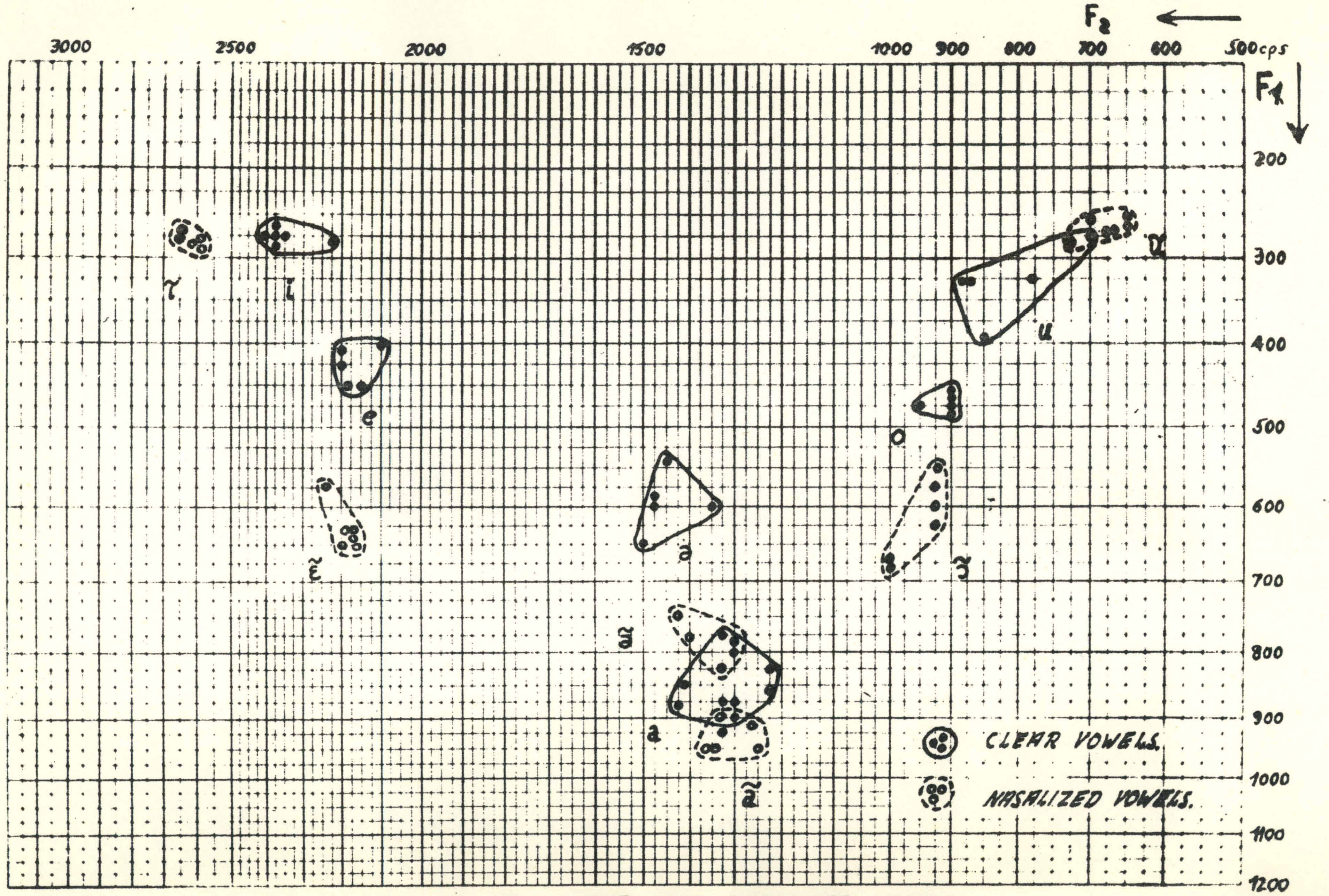


Chart 8 RD CLEAR AND NASALIZED VOWELS

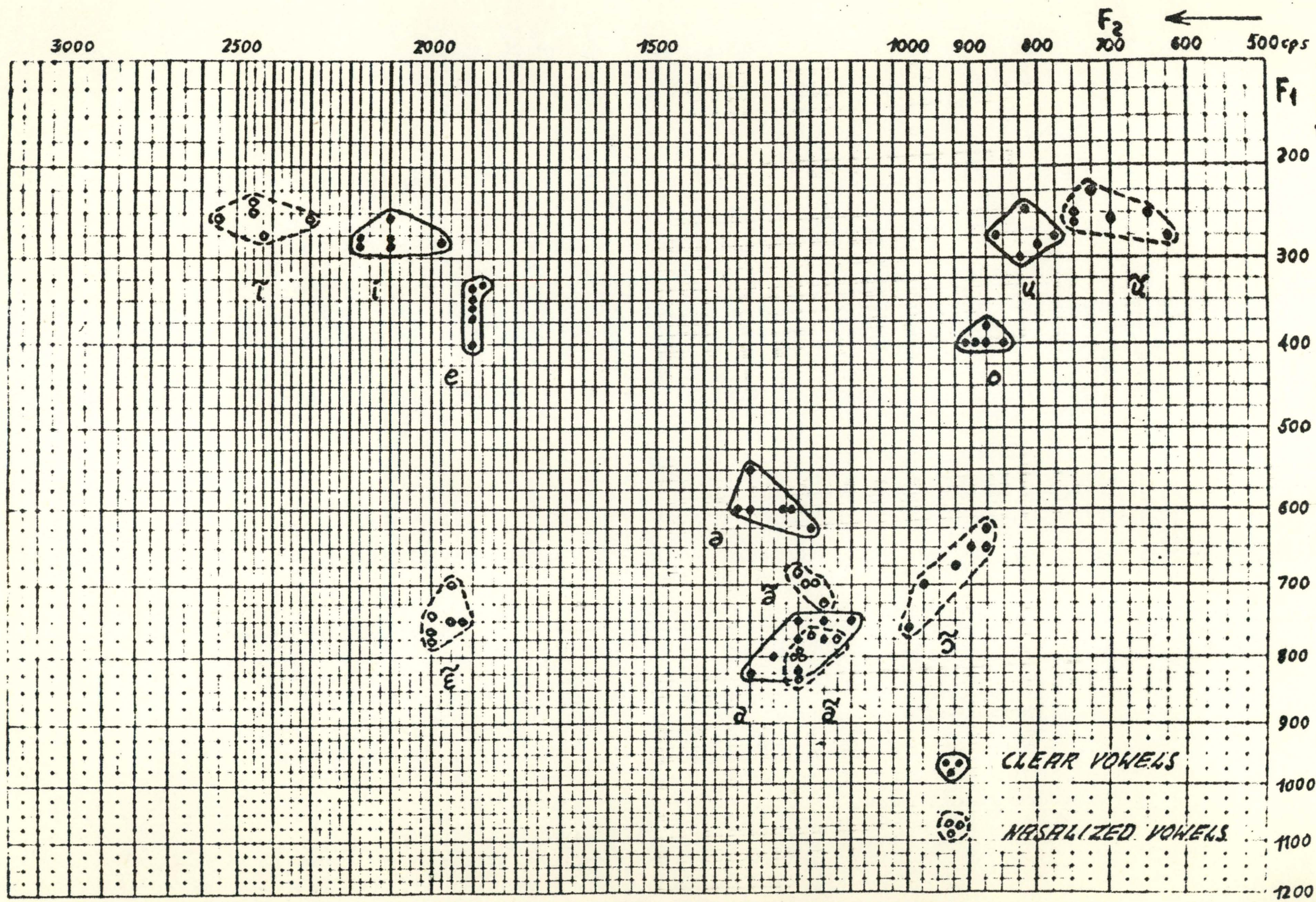


Chart 9 RT CLEAR AND NASALIZED VOWELS

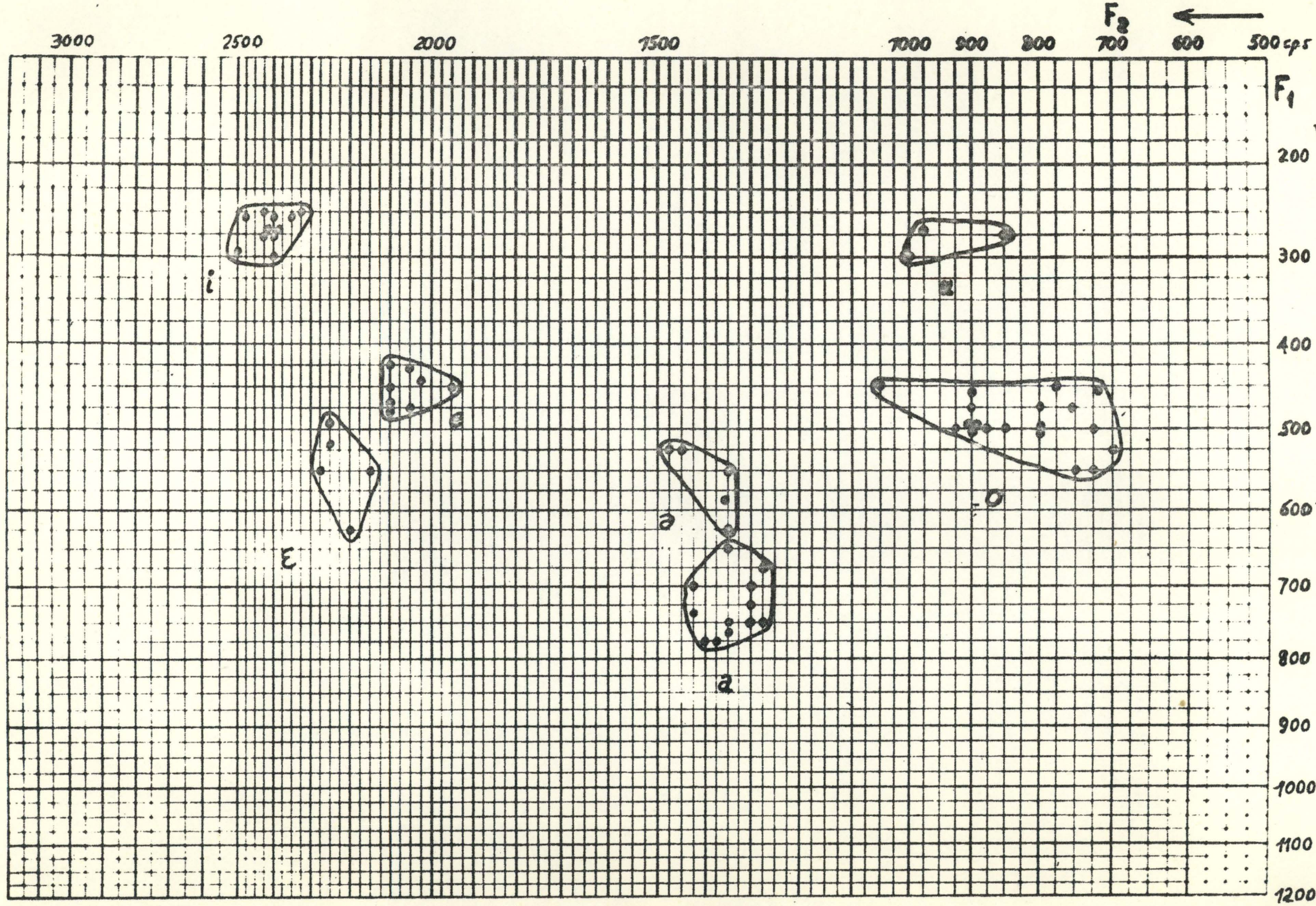


Chart 10 SK CLEAR VOWELS

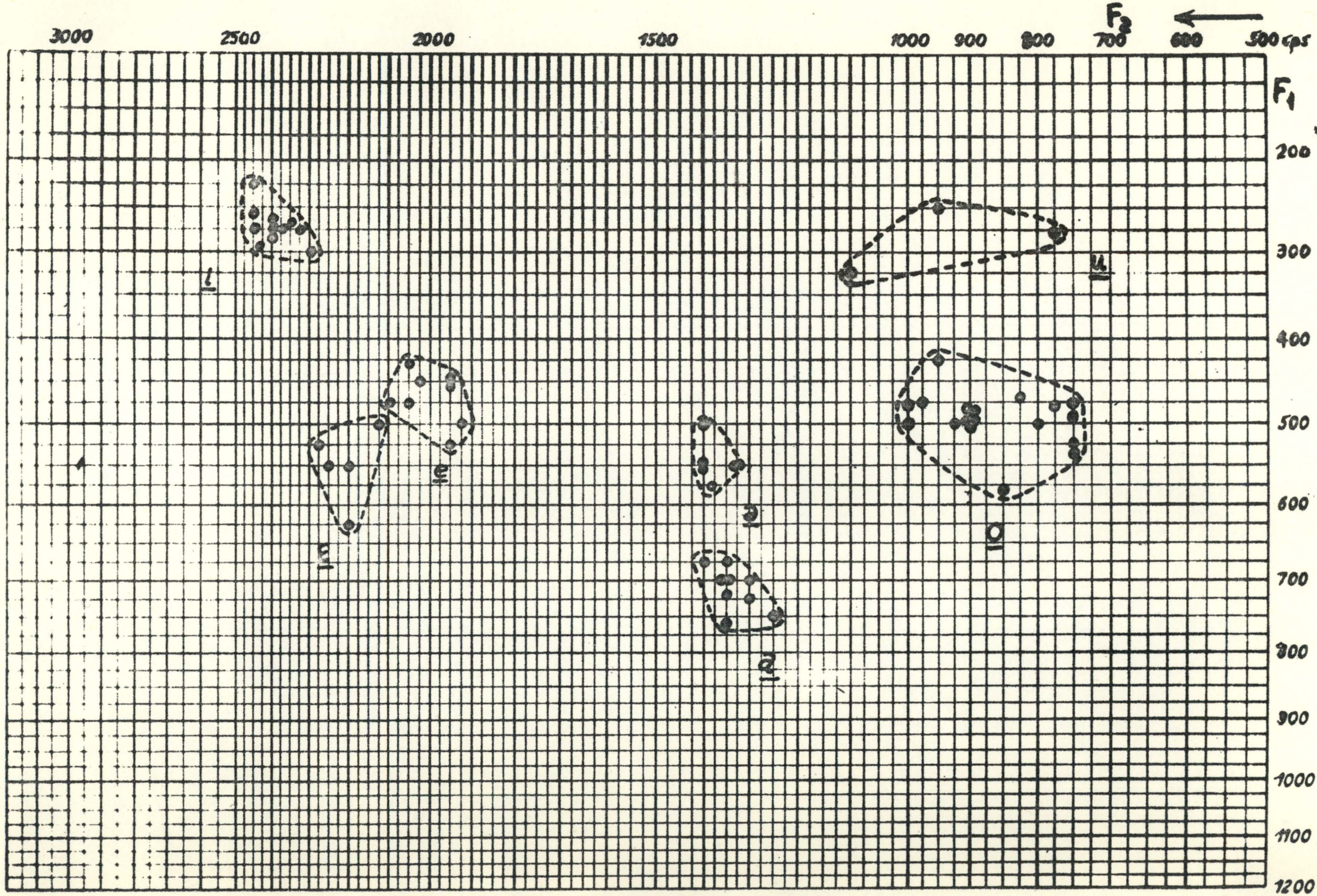


Chart 11 SK MURMURED VOWELS.

WORD TONES IN DOGRI

Ved Kumari Ghai

1. Introduction.

This paper aims at giving a brief historical, phonological and phonetic description of the tone types in Dogri as spoken in Jammu city - the winter capital of J&K State of India. According to the census of India 1961, Dogri is spoken by 879,748 people. Grierson, in his Linguistic Survey of India, has grouped Dogri with Panjabi, but Dogri possesses various characteristics which are different from those of Panjabi, and it might be interesting to study the relationship of Dogri with Panjabi on one hand and with various dialects of Western Pahari on the other.

Of the three prosodic features - stress, pitch and vowel quantity - the last is independently phonemic in Dogri as there is contrast between some long and short vowels in non-final position. As regards stress and pitch, both are involved in the tone system of the language.

Dogri, like its neighbouring language Eastern Panjabi, possesses three significant tones which may be described as neutral (tone I), falling (tone 2), and rising (tone 3). Since 1913 when professor T. G. Baily noticed the existence of significant tones in Panjabi these tones have been described by many writers, but instrumental methods have not been used by any of them. As regards tones in Dogri no study, traditional or experimental, is available as yet.

2. Phonological (phonemic) aspect.

These three tones have to be regarded as pitch phonemes (or tonemes), because they are the only distinctive features in such sets of words as k$\overset{1}{\text{a}}$:r 'work', k$\overset{2}{\text{a}}$:r 'house', k$\overset{3}{\text{a}}$:r 'line'. The meaning differentiated by these tones is mostly lexical but in some rare cases it is grammatical. The verb paɽ occurs in non causal form with tone 3 and in causal form with tone 2. All these tones are word tones in the sense that only one significant tone occurs on a simple word. They can, however, be described as syllabic tones in the sense that the nucleus of a tone occurs on any one of the syllables in a polysyllabic word while other syllables are adjusted to the starting point and the end point of the tone-bearing syllable. It is thus not only tone but also the position of the tone which is significant.

o p$\overset{2}{a}$ra: rd$\overset{1}{a}$: h$\overset{1}{a}$: 'he was filling (something)'

o p$\overset{2}{a}$ra: rd$\overset{1}{a}$: h$\overset{1}{a}$: 'he was getting (something) filled'

The syllable containing the nucleus of tone has also stronger stress (including a certain lengthening and a more precise quality of the vowel).

3. Historical aspect.

Tones are by themselves, like stress and quantity, supra segmental features, but it is interesting to know that tones in Dogri and Panjabi can be traced historically to segmental features. What we perceive as a tone is a mixture of various factors, but its main determinant is the rate of laryngeal vibration which is called fundamental frequency. It is related to such segmental features as voice, aspiration, and glottalization.

The tones of Dogri have nothing to do with the musical accent of the Vedic language, the earliest stage of Old Indo-Aryan, but the stress accent of Classical Sanskrit i.e. a later stage of Old Indo-Aryan has played a role in determining their nature. Stressed syllables in later OIA and MIA are generally preserved in Dogri as syllables with stress as well as tone. If there is no aspiration in the word in the OIA or the MIA stage, the tone is mid level or neutral, i.e. tone 1, but if there is aspiration (generally voiced) in the neighbourhood of the stressed vowel, the tone is either falling, i.e. tone 2,or rising, i.e. tone 3. The aspiration of the voiced aspirated stops of OIA and MIA and of mh nh lh as well as the h sound, which either developed from aspirated stops or sibilants of OIA or existed in words borrowed from other languages, disappears in Dogri giving rise to tone 2 if the stressed vowel follows it and to tone 3 if it precedes it.

Thus intervocalic h appears as tone 2 or tone 3

OIA	'loha:	Dogri	l$\overset{3}{o}$a:	'iron'
OIA	loha'ka:ra	Dogri	lu$\overset{2}{a}$:r	'ironsmith'

Initial h and final h are replaced by tone 2 and 3 respectively.

OIA	'haṣta	Dogri	$\overset{2}{a}$tth	'hand'
Arabic	sa'la:h	Dogri	sal$\overset{3}{a}$:	'consultation'

Similar is the case with voiced aspirated stops.

OIA sva'bha:va Dogri sub$\overset{2}{a}$: 'nature'

OIA 'la:bha Dogri l$\overset{3}{a}$:b 'benefit'

The voiced aspirates lose their voice when occurring initially or when preceded by a prefix, but this devoicing may not take place if the voiced aspirated stop is preceded by a prefix which is not realized as a prefix in Dogri, or if the prefix is inconstant in rapid speech.

OIA 'bha:ra Dogri p$\overset{2}{a}$:r 'weight'

OIA pra'dha:na Dogri prad$\overset{2}{a}$:n 'chief'

OIA a'bhya:sa Dogri by$\overset{2}{a}$:s 'practice'

The aspiration of the unvoiced aspirated stops is preserved, but there is a tendency to pronounce the neighbouring stressed vowel with tone 2 or tone 3.

OIA 'khalla Dogri kh$\overset{3}{a}$ll 'down'

OIA 'kheda Dogri kh$\overset{2}{e}$d 'play'

4. Instrumental investigation.

A preliminary study of tones in Dogri was undertaken by the author under the guidance of professor Eli Fischer-Jørgensen, director of the Institute of Phonetics at the University of Copenhagen. The study is limited for the most part to monosyllabics, although disyllabics and words having vowel sequences have also been considered to some extent.

The speech of five informants - native speakers of Dogri belonging to Jammu Province - has been used. Eighty-six words of which fifty-three are monosyllabics and thirty-three are polysyllabics including examples of all three tones have been placed in one or two or three of six sentence-frames.

Frame I ʃail - ai. 'good - is'

Frame II ʃail - ai? do., interrogative intonation

Frame III - ʃail ai. ' - good is'

Frame IV isi - 'this - '

Frame V isi - te dikkh. 'this - and see' (isi is an accusative form)

Frame VI mẽ tusẽ -glā:yā: 'I (to you) - said'

Frames I, II and III have been used for substantives, frames IV and V for verbs in imperative forms and frame VI has been used for sets of words belonging to different grammatical categories.

To avoid contrastive pitch, these sentences were arranged in randomized lists which were read twice by each informant. Tape-recordings of these lists spoken by two informants were made in Copenhagen and Stockholm respectively on professional tape-recorders. The rest of the informants recorded the text in India in the studio of Radio Kashmir Jammu.

The instruments used were the Trans Pitchmeter and the Intensity Meter built by B. Frøkjær-Jensen of the Institute of Phonetics at Copenhagen University, and the mingograph type 42 made by Elema Schönander (Stockholm, Sweden).

Pitch curves, intensity curves and duplex oscillograms were made from the material recorded on tape. In the case of the pitch curves low-pass filters with cutoff frequencies of 300 (VK), 200 (RK), and 150 (SL RN DD) cps were used. Two intensity curves, one linear without filtering and the other logarithmic with a high-pass filter set at 500 cps were made. For male voices a 5 milliseconds' integration time was used in the case of the linear curve and 2,5 milliseconds in the case of the logarithmic curve. For female voices it was 10 milliseconds and 5 milliseconds respectively.

4.2. Discussion of measurements.

4.2.1. It is relative pitch and not the absolute pitch which is significant for the perception of tone. The same absolute pitch may b be perceived differently as high or low in accordance with the different voice range of different speakers and in accordance with sentence intonation. In the material used in this investigation the voice range of the informants is as follows:

TABLE 1: Voice Range of Informants

Informant	Lowest F_o in cps	Highest F_o in cps	Max. range of modulation within one vowel in cps	Mid point of the voice range in cps
VK (Female)	120	400	165	260
RK (Female)	140	350	150	245
SL (Male)	110	250	100	180
DD (Male)	100	405	150	252
RN (Male)	65	210	130	137

4.2.2. Description of tones. Tone 1 can be described as mid level tone. It starts generally at a point lower than that of tones 2 and 3 and may remain static or fall or rise in accordance with sentence intonation. It is better to describe it in negative terms, because

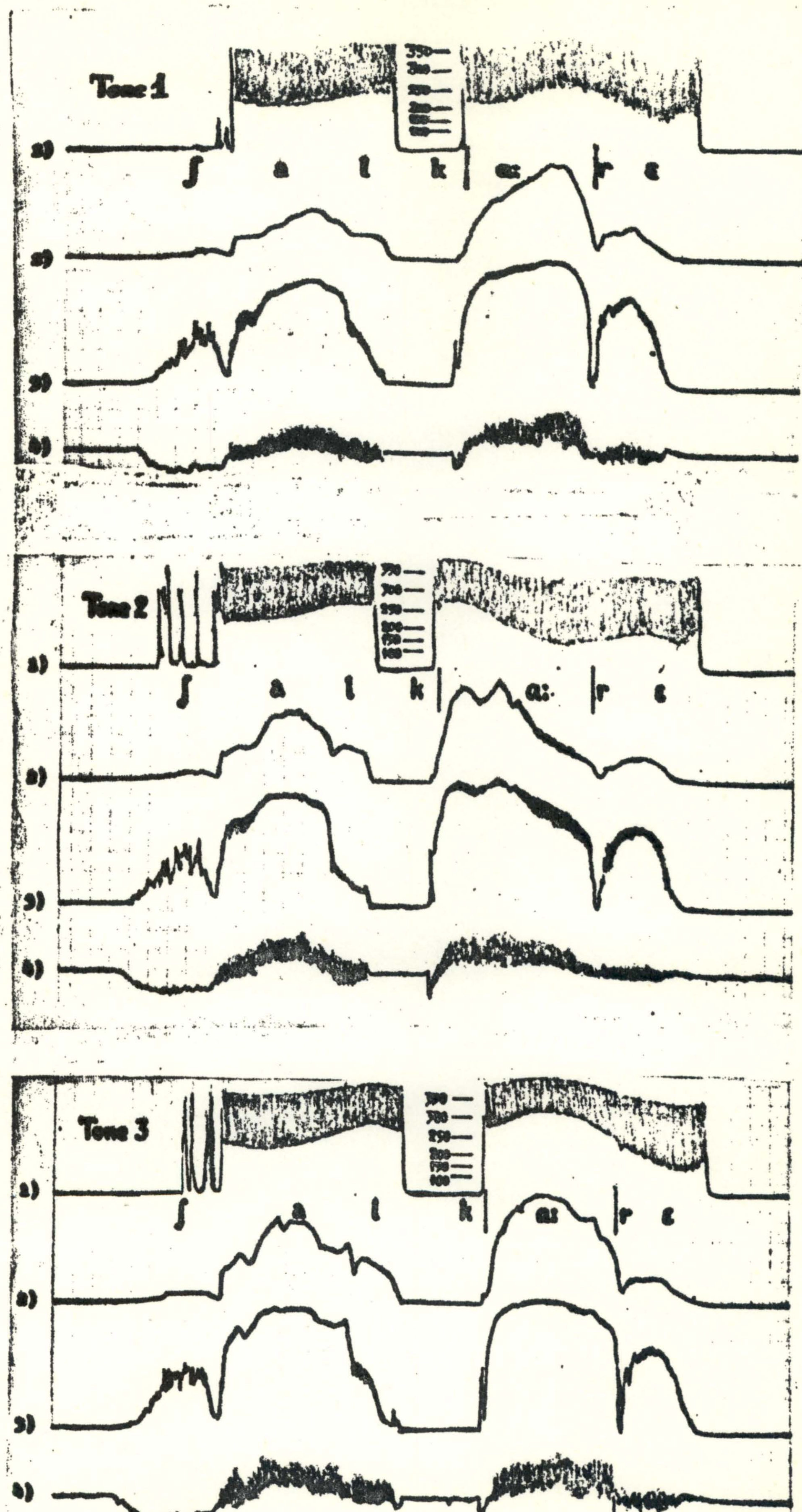


Fig. 1: Specimens of mingograms. The four traces are: 1) fundamental frequency curve; 2) intensity curve, linear, no filtering; 3) intensity curve, logarithmic, high-pass 500 cps; 4) duplex oscillogram.

its characteristic feature is absence of the features of tone 2 and tone 3.

Tone 2 is a falling tone which starts at a point generally higher than the middle of the voice range of the speaker and then falls to the lowest point. In the case of vowels of longer duration it generally rises again although this rise depends also on sentence intonation and is individually conditioned.

Tone 3 is a rising or rising-falling tone which starts at a level generally lower than the middle of the voice range and then rises to the highest level or at least to a level higher than the middle of the voice range.

Tables 2-6 contain averages of the fundamental frequency measured at a few points which seemed to be relevant - for tone 1 beginning and end of the vowel, for tone 2 beginning, minimum, and end of the vowel, for tone 3 beginning, maximum, and end. The frame has been measured at a point which seemed to be relatively constant, 1o cs after the beginning of the vowel in ʃail and the beginning of the second vowel in isi.

4.2.3. <u>Cases of similarity between tone 1 and tone 2 or 3.</u> Tone 2 shows fall and tone 3 shows rise in all the frames as uttered by all informants, but tone 1 is sometimes similar in contour to tone 2 or 3, and it is necessary to see how it is distinguished from tone 2 and tone 3.

DD's tone 1 in frame I (table 6) shows an average rise from 174 cps to 24o cps and tone 3 in the same frame shows a rise from 197 cps to 288 cps. Thus tone 3 starts at a higher pitch than does tone 1 and rises more than tone 1. Another difference is that tone 3 reaches its highest point sooner than tone 1. The maximum occurs at 23.1 cs (93.1% of the vowel duration) in the case of tone 1, and at 15.3 cs (67.1% of the vowel duration) in the case of tone 3. In the case of tone 1 the rise is generally slow in the first half of the vowel as compared with the second half. In the case of tone 3 it is quicker in the first half. The general level is also slightly higher in tone 3.

RK's tone 1 (table 3) shows a fall in frame I, and it falls even to the same level as tone 2, but as tone 2 starts higher than tone 1, the total fall in the case of tone 2 is far greater (91 cps) than in the case of tone 1 (4o cps). In frames IV and VI tone 1 shows an average rise of 28 cps and 13 cps in contrast with a corresponding rise of 99 cps and 77 cps in tone 3.

In RN's tone 1, the rise in frames IV and VI is 30 cps and 37 cps against the corresponding rise of 67 cps and 88 cps in tone 3 (table 4).

SL's tone 1 is almost static in frame III and may show a slight fall in frames IV and I. The fall however is less (11 cps in frame I and 7 cps in VII). The minimum in tone 1 in frame I occurs at 7.0 cs from the beginning of the vowel while it occurs at 17.8 cs in the case of tone 2.

VK's tone 1 shows fall or rise in frame I and rise in frames II III IV V VI, but the occasional fall is much less in tone 1 than in tone 2, and the rise is always slower in tone 1 than in tone 3.

4.2.4. <u>Tones in disyllabics.</u> The tone-bearing vowel in disyllabics shows the same contour as in monosyllabics, but part of the contour is seen on the second vowel.

4.2.5. <u>Tones in different frames.</u> Sentence intonation and the position of the test-word in the sentence affect all tones. Both frames I and III are statements, but while in I the test-word is placed in the middle, in III it is placed initially. The result is: in VK's speech tone 1 is generally slightly falling in frame I and rising in frame III, tone 2 shows a greater fall in frame I, and in tone 3 the second part (falling) is longer in frame I. In the case of DD the difference is only seen for tone 3. Similarly a comparison of frame IV and V shows that tone 1 and 3 give more rise in frame IV, and tone 2 shows more fall in frame IV.

The influence of the sentence types is seen by a comparison between frame I (statement), frame II (interrogative), and frame IV and V which are imperative sentences. Frame II brings about an overall rise of the pitch of all the tones. In frame IV all subjects show a stronger rise in tone 1 and 3 than they have in frame I, and in tone 2 all except RK show a strong rise at the end as compared to frame I.

4.2.6. <u>Individual differences.</u> DD has a strong tendency toward rising tone. In his case tone I is clearly rising and tone 2 is falling-rising in all frames. RK has no rise in tone 2 in frame I and a slight rise in only one example in frame IV.

4.2.7. Influence of surrounding sounds and the quality of vowel. As expected, the close vowel i shows a higher pitch than the open vowel a: Generally the pitch starts higher after voiced consonants. After unvoiced aspirated stops it often starts lower but shows a quick rise in the beginning. The preceding voiced consonant does not take part in the relevant tonal movement, but a following n or l generally continues the movement of the tone.

4.2.8. Vowel length. The vowel bearing tone 2 is generally longer in duration than the vowel bearing tone 1 and 3, and the vowel having tone 3 is generally shorter in duration than the vowel having tone 1 and 2. The differences are small, but relatively constant.

Illustrations.

Tables 2-6 below (= pp. 14o-15o) are followed by some examples of tone curves (Figs. II, III, V, VIII, IX, XI) drawn from mingograms (cp. Fig. 1). These generally represent the 1st reading (R I). Examples belonging to the same word types have been superposed in the same figure to give a visual impression of the variation. The averages are given below the curves ("beg." = beginning of vowel, "min." = minimum, "max." = maximum). For disyllabics the second syllable has been separated from the first by a vertical stroke.

Table 2 a.

Informant VK (fem)
Monosyllabics

	Fr.	N	Fundamental frequency (cps) Frame	Vowel of test word beg.		rise (+) fall (-)	end	rise (+) fall (-)	Vowel length (cs)	
TONE 1										
a:s,ka:r,ba:r,ʈha:r, sa:l,*ra:,(sa:ɳ), (bel),(ma:l)	I	16	205	212			197	-15	35.2	
a:s,(ka:r),ba:r,sa:l, (ʈha:r),ra:, (sa:ɳ), (bel),ma:l	II	14	202	264			300	+36	32.3	
a:s,ka:r,ba:r,ʈha:r, sa:l,*ra:,ma:l, (sa:ɳ),(bel)	III	17	188	209			263	+54	32.8	
la:,(ga:),pa:ɽ,pi:	IV	8	279	221			331	+110	37.6	
la:,ga:,pa:ɽ,pi:	V	8	234	213			235	+22	36.2	
ma:l,ta:ɽ,kol, khoɽ, pi:	VI	10		229			260	+31	28.3	
TONE 2					min.					Place of min.(cs)
a:r,ka:r,ba:r,sa:b, (ra:),(beɽ),(sa:ɳ)	I	11	210	270	149	-121	155	+6	37.5	30.9
a:r,ka:r,ba:r,sa:b, (ma:l),(ra:), (beɽ),(sa:ɳ)	II	12	206	316	213	-103	227	+14	36.8	26.6
a:r,ka:r,ba:r,sa:b, ra:,beɽ,sa:ɳ	III	11	196	238	175	-63	207	+32	34.2	23.8
la:,ba:,ca:ɽ	IV	6	278	285	162	-123	297	+35	43.7	20.8
la:, ba:, ca:ɽ	V	6	226	237	145	-92	225	+80	42.0	21.2
ma:l,ta:ɽ,kol,pi:	VI	8		282	184	-98	186	+2	30.5	26.5
TONE 3					max.					place of max.(cs)
a:r,ka:r,*sa:ɳ,tha:r, da:ɽ,*ra:,(ma:l),(bel)	I	16	206	238	314	+76	233	-81	34.0	15.0
a:r,ka:r,*sa:ɳ,tha:r, da:ɽ,*ra:,(ma:l), (bel),(sa:l)	II	18	205	260	341	+81	318	-23	33.2	18.3
a:r,ka:r,*sa:ɳ,tha:r, da:ɽ,*ra:,ma:l, (bel),(sa:l)	III	18	188	230	334	+104	304	-30	30.6	20.8
la:,ba:,ca:ɽ,pi:	IV	8	270	248	377	+129	373	-4	33.5	
la:,ba:,ca:ɽ,pi:	V	8	231	251	318	+67	208	-110	32.0	13.6
ma:l,ta:ɽ,kol,pi:, khol	VI	10		264	337	+73	327	-10	25.0	19.8

Table 2 b.

Informant VK (fem)

Words containing vowel sequences.

	Fr.	N	Fundamental frequency (cps)						Vowel length (cs)	
			Frame	Vowel of test word						
				beg.		rise (+) fall (-)	end	rise (+) fall (-)		
TONE 1										
ra:i,su:i, dua:r	I	6	203	226			192	-34	43.5	
ra:i,su:i, dua:r	II	6	183	254			304	+50	46.3	
ra:i,su:i, dua:r	III	6	192	206			281	+75	37.3	
na:i,khoi	VI	4		220			269	+49	31.5	
TONE 2					min.					place of min.(cs)
dua:r	I	2	212	262	140	-122	232	+92	51.0	39.0
dua:r	II	1	200	325	210	-115	210	+0	47.0	38.0
dua:r	III	2	190	225	175	-50	240	+65	42.0	29.5
na:i, thoi	VI	4		269	171	-98	180	+9	32.0	23.5
TONE 3					max.					place of max.(cs)
ra:i:,su:i:, ɟua:r	I	6	208	231	331	+100	210	-121	41.0	18.0
ra:i:,su:i:, ɟua:r	II	6	188	244	353	+109	328	-25	41.3	16.8
ra:i:,su:i:, ɟua:r	III	6	170	225	352	+127	290	-62	36.5	20.2
mana:i:,khoi:	VI	4		247	339	+92	289	-50	28.0	18.3

Table 2 c.

Informant VK (fem)

Disyllabics.

	Fr.	N	Fundamental frequency (cps) Frame	Vowel of test word beg.		rise (+) fall (-)	end	rise (+) fall (-)	Vowel length (cs)		Fundamental frequency second syllable beg.	end
TONE 1												
ba:ri:, na:ɽi:, pa:ɳi:, phora:	I	8	198	239			191	-48	19.4		196	192
ba:ri:, na:ɽi:, pa:ɳi:, phora:	II	8	183	283			279	-4	19.4		304	316
ba:ri:, na:ɽi:, pa:ɳi:, phora:	III	8	186	201			224	+23	23.0		247	245
phoɽa:, ma:ɽi:	VI	4		226			218	-8	19.5		239	264
TONE 2					Place of min.					Place of min.(cs)		
ba:ri:, na:ri:, ka:ɳi, thoɽi:	I	8	200	275	193	-82	193	+0	20.6	20.6	165	175
ba:ri:, na:ri:, ka:ɳi, thoɽi:	II	8	186	312	216	-84	216	+0	23.4	23.4	193	298
ba:ri:, na:ri:, ka:ɳi, thoɽi:	III	8	197	224	184	-40	184	+0	20.5	19.6	185	288
ma:ɽi, thoɽa:	VI	4		278	195	-83	195	+0	19.5	19.5	183	183
TONE 3					Place of max.					Place of max.(cs)		
da:ɽi:, ta:lli:	I	4	203	248	290	+42	290	-0	16.5	16.5	328	204
da:ɽi:, ta:lli:	II	4	187	261	331	+70	331	-0	17.0	16.2	350	338
da:ɽi:, ta:lli:	III	4	180	230	326	+96	326	-0	16.5	16.5	353	254
ma:ɽi, thoɽa:	VI			235	316	+81	316	-0	16.0	15.8	340	301

Table 3 a.

Informant RK (fem)

Monosyllabics and words containing vowel sequences.

	Fr.	N	Fundamental frequency (cps) Frame	Vowel of test word beg.		rise (+) fall (-)	end	rise (+) fall (-)	Vowel length (cs)	
TONE 1										
Monosyllabics:										
a:s, ka:r, (ba:r), tha:r, (sa:ɳ), sa:l, ra:, (bel)	I	13	242	212			172	-4o	19.o	
la:, ga:, pa:ɽ	IV	6		2o9			237	+28	21.8	
ma:l, ta:ɽ, khoɽ, (kol), pi:	VI	9		22o			233	+13	16.9	
Words containing vowel sequences:										
dua:r	I	1		25o			16o	-9o	26.o	
na:i:, khoi:	VI	4		2o9			224	+15	23.o	
TONE 2					min.					place of min. (cs)
Monosyllabics:										
a:r, ka:r, ba:r, sa:ɳ, sa:b, ra:	I	12	238	257	166	-91	166	+o	2o.1	19.8
(la:), ba:, ca:ɽ	IV	5		278	15o	-128	16o	+1o	23.o	22.o
ma:l, ta:ɽ, kol, pi:	VI	8		274	175	-99	179	+4	17.7	17.o
Words containing vowel sequences:										
dua:r	I	1		225	16o	-65	16o	+o	25.o	25.o
(na:i), thoi	VI	3		245	183	-62	183	+o	2o.o	13.o
TONE 3					max.					place of max. (cs)
Monosyllabics:										
a:r, ba:r, da:ɽ, tha:r, sa:ɳ, sa:l, ra:, (bel)	I	15	241	222	273	+51	247	-26	18.9	13.4
la:, ba:, ca:ɽ	IV	6		214	3o8	+94	3o8	-o	2o.5	17.5
ma:l, ta:ɽ, kol, pi:, khol	VI	1o		237	314	+77	3o7	-7	17.7	15.8
Words containing vowel sequences:										
mana:i	VI	2		245	3o5	+6o	26o	-45	22.5	14.5
ɟua:r	I	1		225	285	+6o	26o	-25	26.o	23.o

Table 3 b.

Informant RK (fem)

Disyllabics.

	Fr.	N	Fundamental frequency (cps) Frame	Vowel of test word beg.		rise (+) fall (-)	end	rise (+) fall (-)	Vowel length (cs)		Fundamental frequency second syllable beg.	end
TONE 1												
pa:ni:	I	2		225			175	-50	12.5		200	192
phoɽa:, ma:ɽi:	VI	4		216			193	-23	11.0		218	222
TONE 2					min.					place of min.(cs)		
ka:ɳi:	I	2		267	245	-22	245	+0	11.5	11.5		
thoɽa:, ma:ɽi:	VI	4		259	195	-64	195	+0	12.2	12.2	170	210
TONE 3					max.					place of max.(cs)		
ta:ɳi:	I	2		240	257	+17	257	-0	13.0	13.0	*)	
(thoɽa:)	VI	1		225	280	+55	280	-0	9.0	9.0	300	

*) could not be measured.

Table 4 a.

Informant RN (masc)
Monosyllabics and words containing vowel sequences

	Fr.	N	Fundamental frequency (cps) Frame	Vowel of test word beg.		rise (+) fall (-)	end	rise (+) fall (-)	Vowel length (cs)	
TONE 1										
Monosyllabics:										
a:s,ka:r,ba:r,ʈha:r, sa:ɳ, (ra:),(bel)	I	12	87	92			86	-6	25.2	
la:, ga:, pa:ɽ	IV	6		93			123	+30	27.8	
ma:l,ta:ɽ,(khoɽ), kol,pi:	VI	9		112			149	+37	21.3	
Words containing vowel sequences:										
dua:r	I	1		100			80	-20	33.0	
na:i, khoi	VI	4		93			163	+70	19.0	
TONE 2					min.					place of min.(cs)
Monosyllabics:										
a:r,ka:r,ba:r,sa:ɳ, sa:b,(ra:),(beɽ)	I	12	91	139	74	-65	82	+8	28.0	21.0
la:,ba:,ca:ɽ	IV	5		150	76	-74	144	+68	32.0	17.0
ma:l,(ta:ɽ),kol,pi:	VI	7		182	87	-95	134	+67	25.0	
Words containing vowel sequences:										
(dua:r)	I	1		145	80	-65	80	+0	29.0	29.0
na:i,thoi	VI	4		135	84	-51	149	+65	23.5	10.5
TONE 3					max.					place of max.(cs)
Monosyllabics:										
a:r,ka:r,da:ɽ,(sa:ɳ), sa:l,tha:r,(ra:), (bel)	I	13	95	95	141	+46	121	-20	24.6	17.7
ba:,ca:ɽ	IV	4		116	183	+67	158	-25	26.0	11.5
ma:l,khol,(kol),pi:	VI	7		140	228	+88	228	-0	18.6	18.6
Words containing vowel sequences:										
mana:i,khoi:	VI	4		103	204	+101	204	-0	20.5	

Table 4 b.

Informant RN (masc)

Disyllabics.

	Fr.	N	Fundamental frequency (cps)						Vowel length (cs)		Fundamental frequency second syllable	
			Frame	Vowel of test word								
				beg.		rise (+) fall (-)	end	rise (+) fall (-)			beg.	end
TONE 1												
pa:ɳi:	I	2		95			82	-13	15.o		8o	85
phoɽa:	VI	2		9o			9o	o	9.o		92	135
TONE 2					min.					place of min. (cs)		
ka:ɳi:	I	1		15o	11o	-4o	11o	+o	14.o	14.o	8o	7o
thoɽa:	VI	2		2oo	1o2	-98	1o2	+o	12.5	12.5	87	15o
TONE 3					max.					place of max. (cs)		
ʈa:ɳi:	I	2		13o	16o	+3o	16o	-o	18.5	18.5	16o	145

Table 5 a.

Informant SL (masc)
Monosyllabics and words containing vowel sequences.

	Fr.	N	Fundamental frequency (cps) Frame	Vowel of test word beg.		rise (+) fall (-)	end	rise (+) fall (-)	Vowel length (cs)	
TONE 1										
Monosyllabics:										
a:s, ka:r, ba:r, ʈha:r, sa:ɳ, sa:l, (ra:), bel	I	15	146	148			137	-11	26.3	
la:, ga:, pa:ɽ	IV	6		158			158	o	26.8	
ma:l, ta:ɽ, khoɽ, kol, pi:	VI	1o		172			165	-7	21.6	
Words containing vowel sequences:										
na:i, khoi	VI	4		164			169	+5	22.o	
dua:r	I	2		145			135	-1o	31.5	
TONE 2					min.					place of min. (cs)
Monosyllabics:										
(a:r), ka:r, ba:r, sa:ɳ, sa:b, (ra:), beɽ	I	12	148	174	122	-52	135	+13	27.7	17.8
la:, ca:ɽ, ba:	IV	6		188	131	-57	15o	+19	28.5	16.7
ma:l, ta:ɽ, kol, pi:	VI	1o		196	132	-64	138	+6	21.8	15.7
Words containing vowel sequences:										
na:i, thoi	VI	4		193	132	-61	158	+26	25.o	12.5
dua:r	I	2		162	117	-45	131	+14	33.o	23.o
TONE 3					max.					place of max. (cs)
Monosyllabics:										
a:r, ka:r, da:ɽ, tha:r, sa:ɳ, sa:l, ra:, bel	I	15	146	156	191	+35	169	-22	24.o	16.8
la:, ba:	IV	4		158	211	+53	199	-12	23.8	17.5
ma:l, (khol), (kol), pi:	VI	6		198	239	+41	239	-o	17.8	17.8(?)
Words containing vowel sequences:										
mana:i, khoi	VI	3		185	235	+5o	225	-1o	19.o	16.o
ɟua:r	I	2		147	177	+3o	167	-1o	28.o	23.o

Table 5 b.

Informant SL (masc)

Disyllabics.

	Fr.	N	Fundamental frequency (cps): Frame beg.	Vowel of test word: min./max.	rise (+) fall (-)	end	rise (+) fall (-)	Vowel length (cs)	place of min./max. (cs)	Fundamental frequency second syllable: beg.	end
TONE 1											
pa:ɳi:	I	2	157			140	-17	17.0		145	145
phoɽa:, ma:ɽi	VI	4	158			146	-12	14.0		148	157
TONE 2				min.					place of min.(cs)		
ka:ɳi:	I	2	192	127	-65	130	+3	19.0	15.0	138	143
thoɽa:, ma:ɽi:	VI	4	182	127	-55	127	+0	13.0	13.0	127	147
TONE 3				max.					place of max.(cs)		
ʈa:ɳi	I	2	175	195	+20	195	-0	16.0	16.0	205	140
thoɽa:	VI	1	190	200	+10	200	-0	12.0	12.0	200	200

Table 6 a.

Informant DD (masc)
Monosyllabics and words containing vowel sequences.

	Fr.	N	Fundamental frequency (cps) Frame	Vowel of test word beg.		rise (+) fall (-)	end	rise (+) fall (-)	Vowel length (cs)	
TONE 1										
Monosyllabics:										
*a:s,*ka:r,ma:l, sa:l,ra:,(ba:r)	I	13	200	174			240	+66	24.6	
ka:r,ba:r,ma:l, (tha:r),sa:l	III	9	196	191			261	+70	21.7	
la:,pa:ɽ,(pha:ɽ)	IV	5		188			288	+100	16.2	
la:,(ga:),pa:ɽ, (pi:),(pha:ɽ)	V	7		178			149	-29	25.7	
Words containing vowel sequences:										
ra:i,dua:r,su:i	III	6	202	197			253	+56	21.8	
TONE 2					min.					place of min.(cs)
Monosyllabics:										
a:r,ka:r,ma:l,*sa:b, pa:kh,pa:r,(ba:r)	I	14	212	222	135	-87	223	+88	25.2	10.7
ka:r,ba:r,ma:l, sa:b,(pa:kh)	III	9	188	231	171	-60	232	+61	27.8	13.4
la:,ca:ɽ,(ba:), sa:l,(sa:)	IV	8		258	178	-80	279	+101	19.4	8.4
la:,ca:ɽ,ba:,sa:l, (sa:)	V	9		262	161	-101	168	+7	29.1	10.9
TONE 3					max.					place of max.(cs)
Monosyllabics:										
a:r,(ka:r),sa:ɳ,tha:r, da:ɽ,ra:,(ga:k)	I	12	201	197	288	+91	249	-39	22.6	15.3
(ka:r),(ba:r),(tha:r), sa:ɳ,da: ,ga:k	III	9	196	207	298	+91	272	-26	20.3	17.7
la:,ca:ɽ,ba:,a:ɳ	IV	8		241	359	+118	358	-1	16.7	16.0
(la:),ca:ɽ,ba:,a:ɳ, (pi:)	V	8		212	296	+78	167	-23	22.5	10.9
Words containing vowel sequences:										
ɟua:r,ra:i,su:i		6		220	314	+94	274	-40	20.8	14.3

Table 6 b.

Informant DD (masc)

Disyllabics.

	Fr.	N	Fundamental frequency (cps) Frame	Vowel of test word beg.	min.	rise (+) fall (-)	end	rise (+) fall (-)	Vowel length (cs)	place of min. (cs)	Fundamental frequency second syllable beg.	end
TONE 1												
pa:ni:, ba:ri:, mu:rat, na:ɽi:, phoɽa:	III	10	193	209			263	+54	12.7		251	244
TONE 2												
ka:ni:, ba:ri:, (mu:rat), (thoɽa:)	III	6		231	183	-48	183	+0	12.0	9.6	192	243
TONE 3												
talli:, da:ri:, mi:na:		6		223	279	+56	280	-1	8.3	8.3	311	285

FIG. II informant VK, reading I

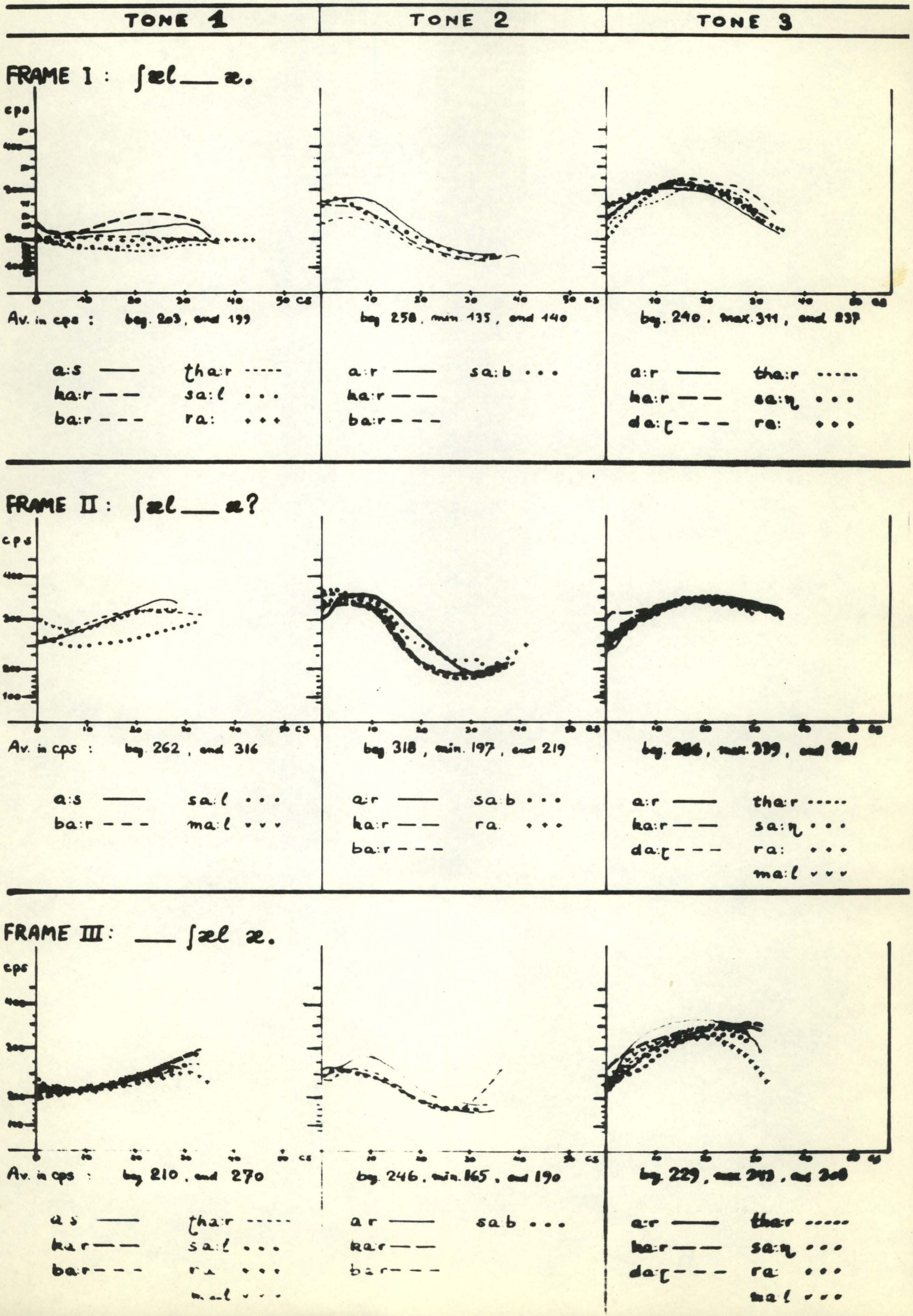


FIG. III INFORMANT VK, list 3

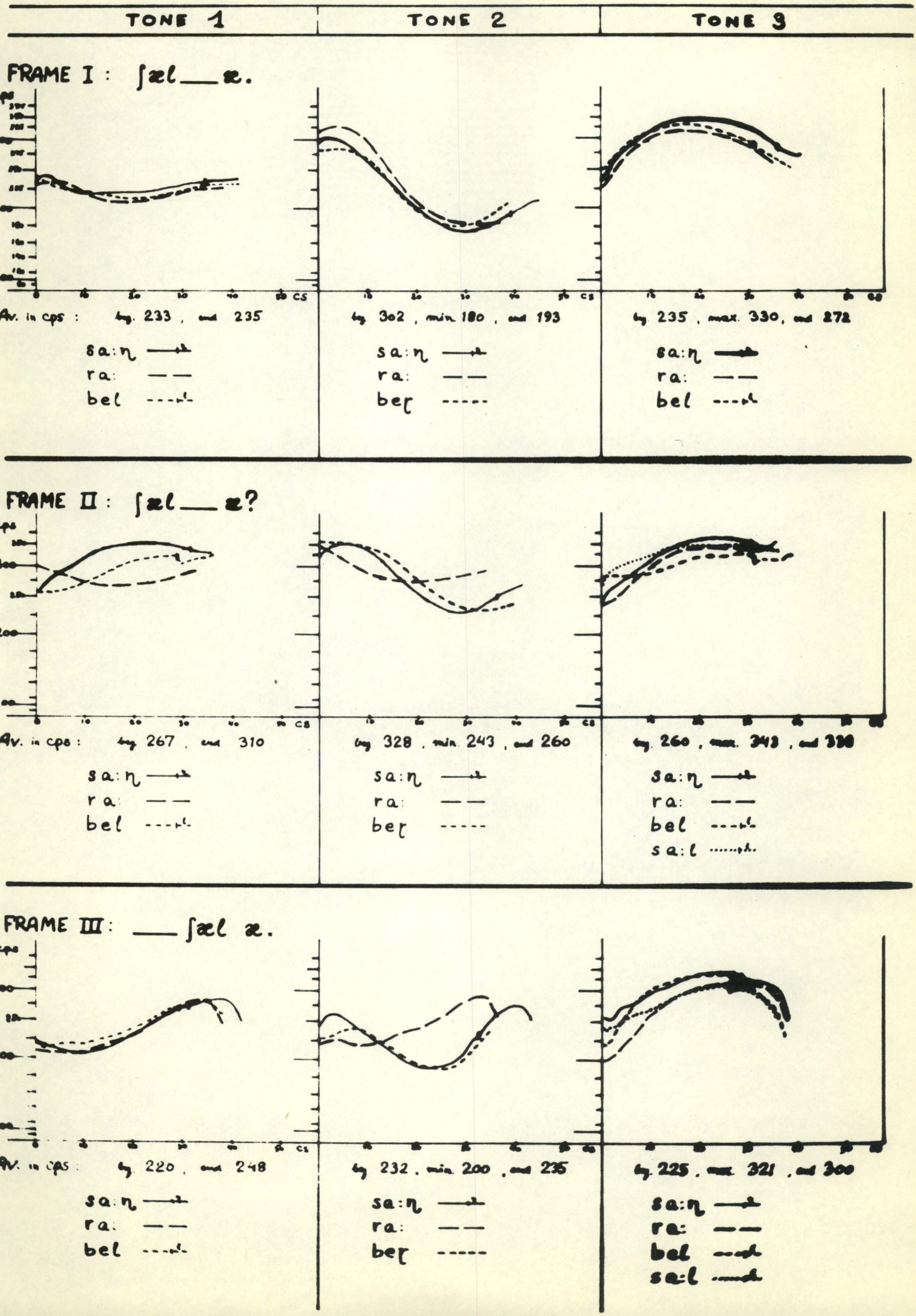


FIG. V INFORMANT VK, reading I

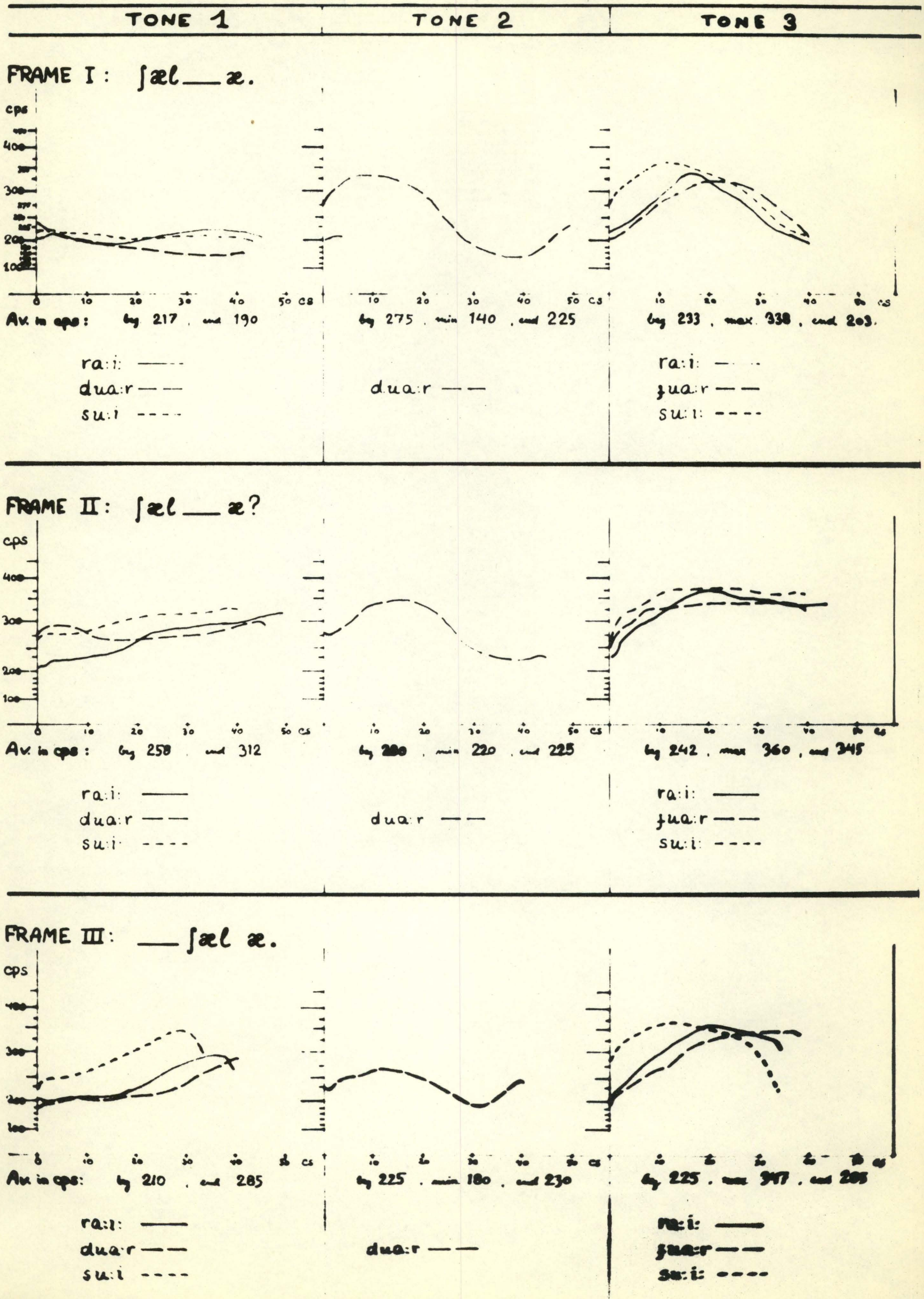


FIG. VIII INFORMANT: RK

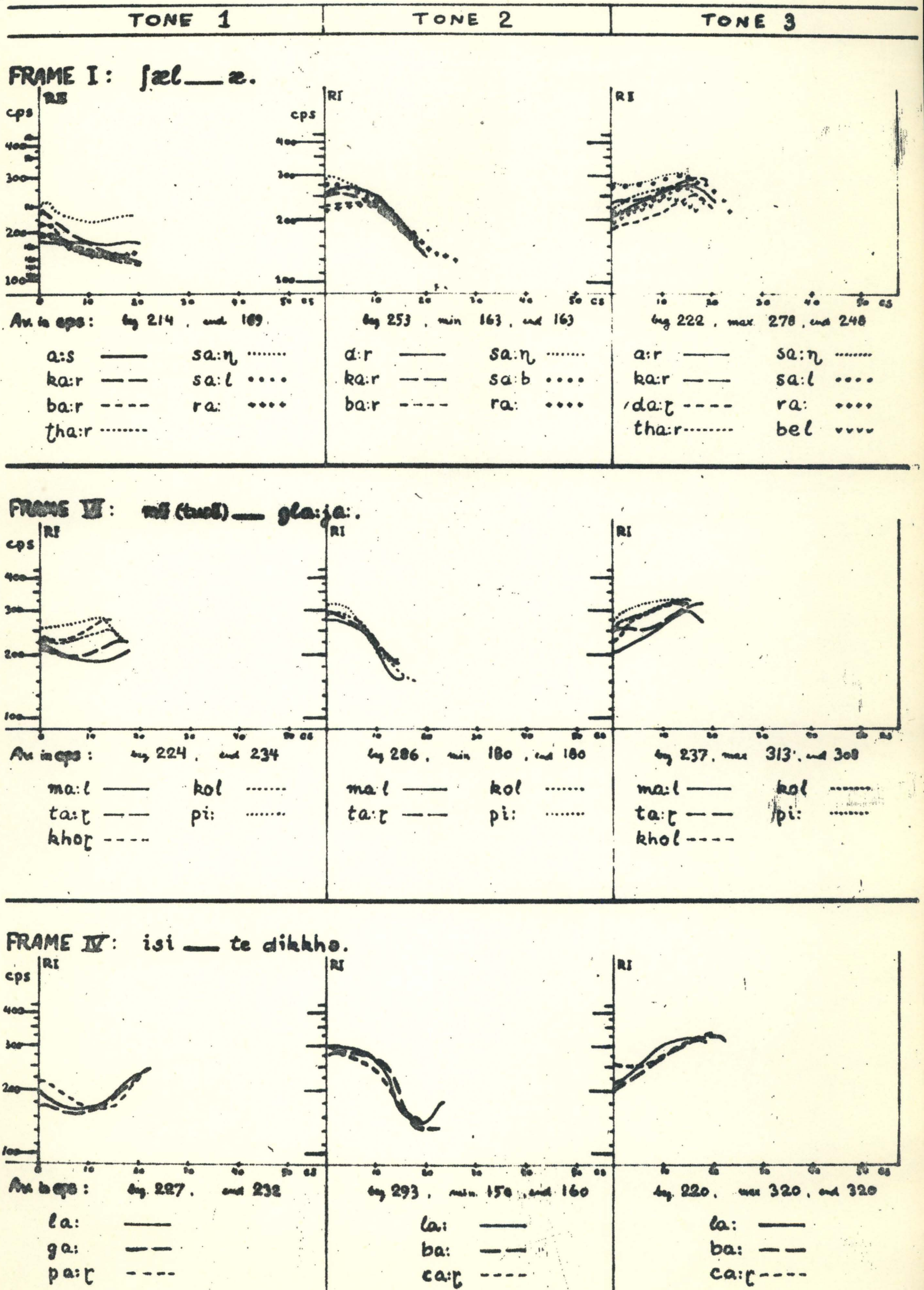


FIG. IX INFORMANT RN

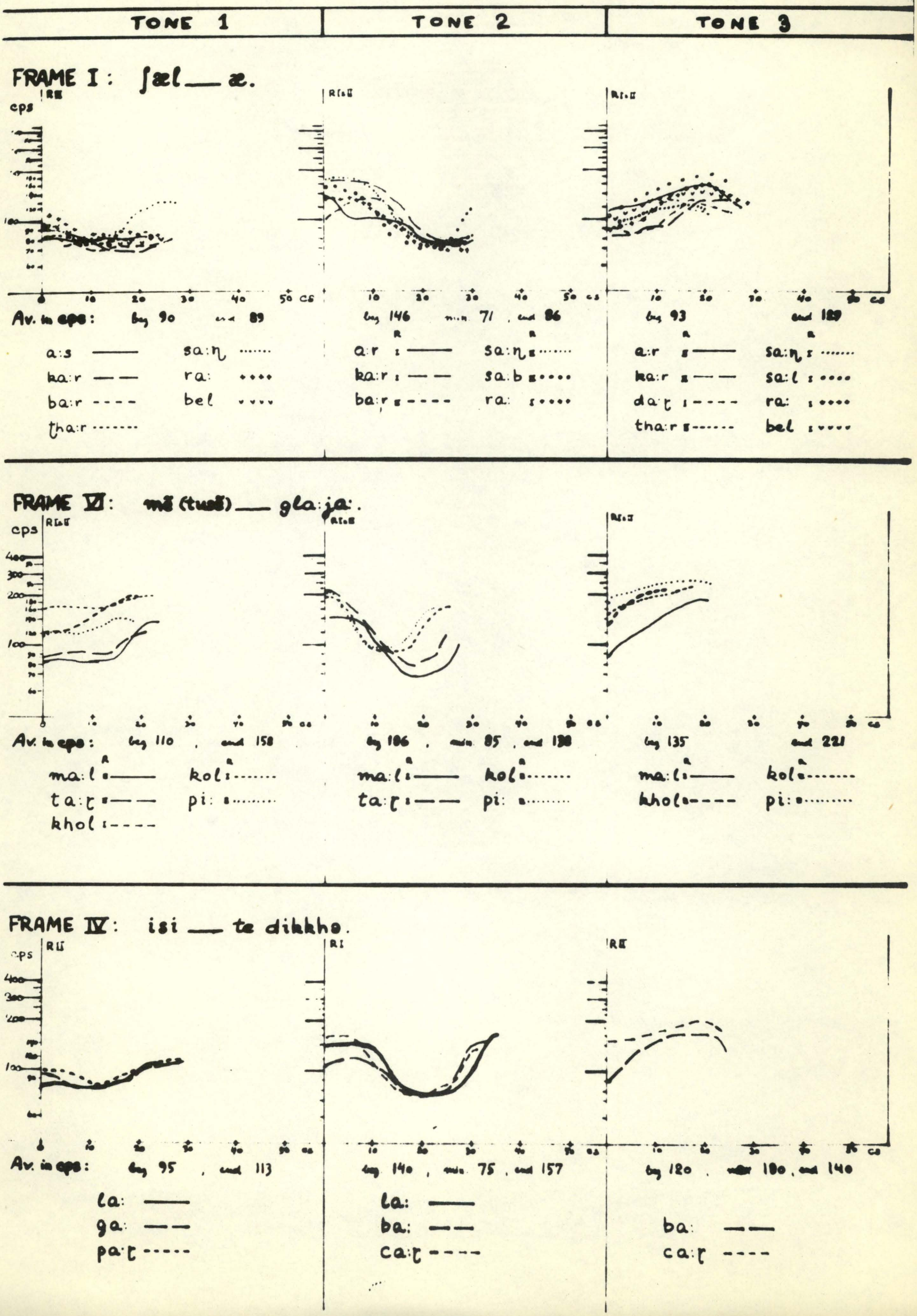
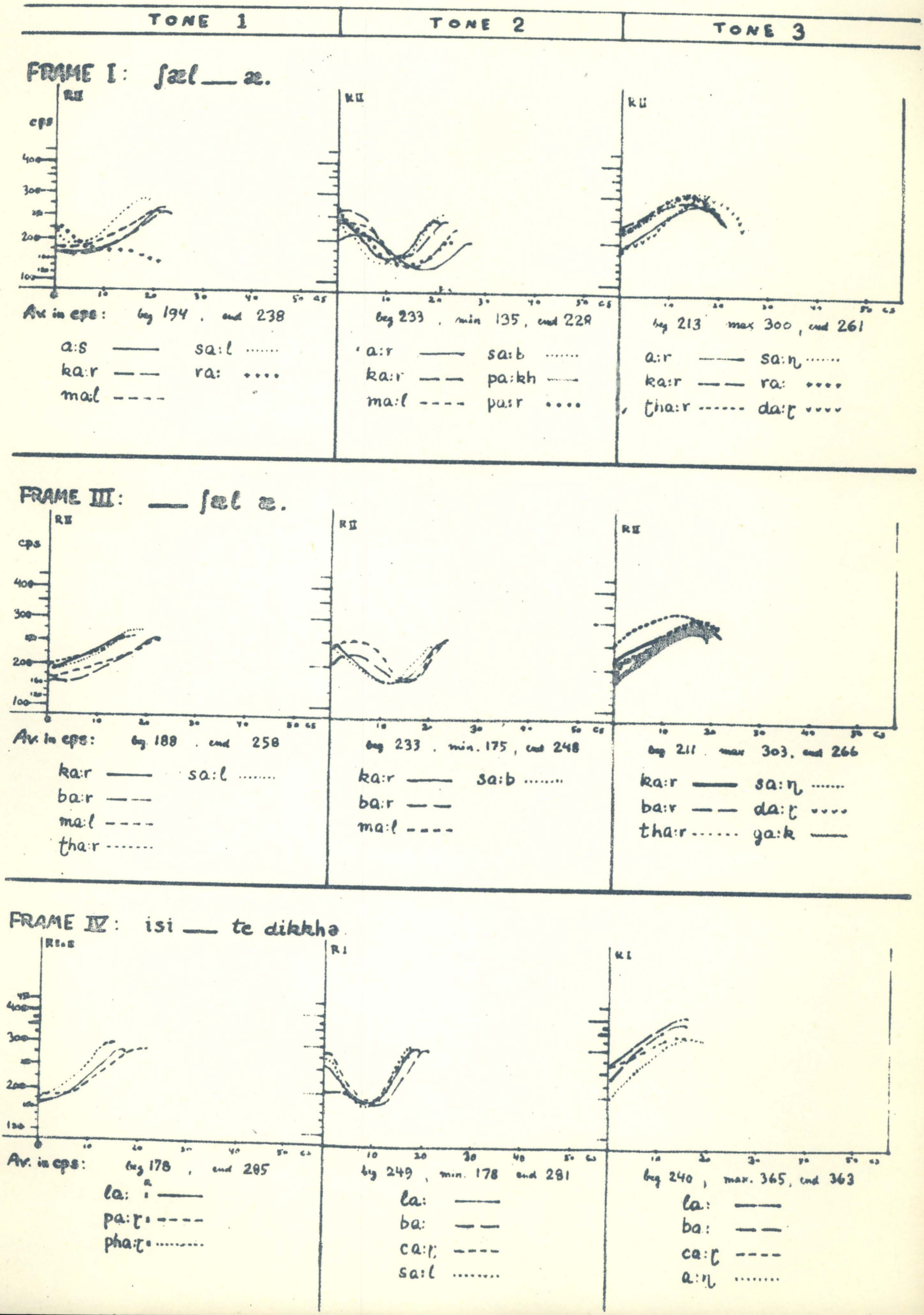


FIG. XI INFORMANT DD



STATISTIC CALCULATIONS OF FORMANT DATA.

Børge Frøkjær-Jensen

1. The Statistical Treatment of Formant Data.

(Computer Programs)

During the last few months I have been working with statistic calculations on formant data. These calculations have been carried out on an IBM 7094 calculating machine at NEUCC (Northern Europe University Computing Center) placed at the Technical University of Denmark.

For simple statistic data description of formant frequencies, formant levels, and bandwidths I have used a standard BMD program (Biomedical Computer Program) no. 01D. This program has been changed a little in order to match it better to our purposes and needs. Among other things the changed program gives both a printed and a punched output (1).

For further calculations on these outputs, and for comparisons between different groups of outputs some other programs have been worked out. This paper was primarily meant as a presentation of the output from one of these programs (called TALCOM), which is intended for comparisons between two different output groups from the changed BMD-01D program. (Each group may consist of max. 999 variables with max. 999 samples).

The program has been carefully tested. The material for one of the tests has been selected from the spectrographic material employed for a paper about the spoken Danish long vowels in our last report (2).

The material for the above mentioned test consists of 10 male, 9 female subjects, and 6 children. All these subjects have spoken the Danish long vowel phonemes once. The most common variant, the /r/-influenced combinatory variant [ɑ:] of the /a:/-phoneme is included in the material.

Illustration Fig. 1 shows the statistic calculations from the changed BMD program based on 10 male speakers. The output table contains the following columns: An identification number of the variable - each vowel/formant combination has an identification number - (VAR NO), the mean value of all the samples within a given identification number (MEAN), the standard deviation for the samples of each variable (S.D.), the standard errors of means (S.E. OF MEAN), number of samples for each mean value (SAMPLE), the maximum value of the samples (MAXIMUM), the minimum value of the samples (MINIMUM), and the dispersion range of samples (RANGE). (Reference 8.)

Fig. 2 shows the printed output based upon 9 female speakers.

Fig. 3 shows the output based upon 6 children's voices, 4 boys and 2 girls. These six children (ages between 10 and 13 years) have spoken the same text as have the adult subjects.

The output results in Fig. 1 (group A), Fig. 2 (group B), and Fig. 3 (group C) are punched by the program, and these punch cards are used as input material for the TALCOM program, the output results of which are shown in Fig. 4 and Fig. 5.

The TALCOM program calculates an output table which contains: variable numbers (I), mean values for samples in group A, which in this test equal the male formants (AMEAN), number of samples for each A-mean (K), standard errors of A-means (ASDE), standard errors of A-means in per cent (APCT), mean values for samples in group B, which in Fig. 4 equal the female formants and in Fig. 5 equal the children's formants (BMEAN), number of samples for each B-mean (L), standard errors of B-means (BSDE), standard errors of B-means in per cent (BPCT), the difference in cps between B-means and the corresponding A-means, which in Fig. 4 equals the difference between the female formants and the corresponding male formants, and in Fig. 5 equals the difference between the children's vowel formants and the corresponding male formants (DIFBA), this difference calculated in per cent (BAPCT), the absolute standard deviation of that difference (SDABS), the relative standard deviation of that difference (SDPCT), two times this standard deviation in per cent (SDPCT2), and three times this standard deviation in per cent (SDPCT3).

The job deck set-up consists of program cards and data cards. The first data card contains a text line which does not enter the computation, but is printed out as a headline immediately before the output table. The next data card contains three number codes: (a) a number (NO) indicating how many variables the computer must find in each group; (b) a number (NVAR) which tells the machine to give warnings for sample numbers which are equal to or less than this number; and (c) a code number (N) which can assume the values

BMDOID SIMPLE DATA DESCRIPTION - VERSION OF MAY 20, 1964
HEALTH SCIENCE COMPUTING FACILITY, UCLA

THE PROGRAM HAS BEEN CHANGED AND ADAPTED FOR PHONETIC ANALYSIS BY
INSTITUTE OF PHONETICS, UNIVERSITY OF COPENHAGEN, DENMARK
VERSION OF DECEMBER, 1967.

PROBLEM CARD

PROBLEM NUMBER	SPEECH	METHOD NUMBER	1
NUMBER OF CASES	10	NUMBER OF SPECIAL VALUES	0
NUMBER OF VARIABLES	44	NUMBER OF TRANSGENERATIONS	-0
NUMBER OF VARIABLES ADDED	-0	INPUT TAPE NUMBER	5

NUMBER OF VARIABLE FORMAT CARDS 1

SPECIAL VALUES CARD
0.

VARIABLE FORMAT CARD(S)
(23X,F4.0,7X,F4.0,7X,F4.0,7X,F4.0,20X)

MALE VOICES IN SPEECH, RECORD.NO. 1,1 2,1 3,1 4,1 5,1 6,1 13,1 14,1 15,1 16,1

	VAR NO	MEAN	S.D.	S.E. OF MEAN	SAMPLE	MAXIMUM	MINIMUM	RANGE
i:	1	235.	24.66	7.80	10	280.	200.	80.
	2	2119.	180.16	56.97	10	2500.	1900.	600.
	3	3013.	188.20	59.51	10	3325.	2725.	600.
	4	3354.	187.76	62.59	9	3700.	3085.	615.
e:	5	283.	30.75	9.73	10	325.	235.	90.
	6	2091.	154.86	48.97	10	2445.	1900.	545.
	7	2709.	158.27	50.05	10	2940.	2490.	450.
	8	3389.	207.76	65.70	10	3715.	3140.	575.
ɛ:	9	367.	27.99	8.85	10	400.	310.	90.
	10	1973.	130.39	41.23	10	2200.	1775.	425.
	11	2479.	187.17	59.19	10	2710.	2150.	560.
	12	3369.	192.14	60.76	10	3720.	3150.	570.
a:	13	541.	76.95	24.33	10	700.	440.	260.
	14	1708.	186.77	59.06	10	1975.	1415.	560.
	15	2289.	198.49	62.77	10	2540.	2050.	490.
	16	3423.	206.15	65.19	10	3720.	3170.	550.
ɑ:	17	698.	83.71	26.47	10	810.	585.	225.
	18	1144.	60.05	18.99	10	1240.	1060.	180.
	19	2480.	240.34	80.11	9	2845.	2035.	810.
	20	3464.	243.63	77.04	10	3800.	3040.	760.
y:	21	242.	23.93	7.57	10	280.	215.	65.
	22	1862.	83.73	26.48	10	1990.	1710.	280.
	23	2096.	104.82	33.15	10	2265.	1975.	290.
	24	3180.	201.01	67.00	9	3570.	2855.	715.
ø:	25	308.	26.89	8.50	10	370.	265.	105.
	26	1633.	85.64	27.08	10	1775.	1480.	295.
	27	2018.	111.81	35.36	10	2200.	1875.	325.
	28	3172.	201.87	67.29	9	3480.	2815.	665.
œ:	29	380.	33.08	10.46	10	425.	340.	85.
	30	1546.	80.99	25.61	10	1700.	1400.	300.
	31	2039.	99.21	31.37	10	2250.	1935.	315.
	32	3196.	186.44	58.96	10	3420.	2885.	535.
u:	33	266.	41.28	13.06	10	365.	225.	140.
	34	722.	98.90	32.97	9	875.	585.	290.
	35	2091.	122.90	50.17	6	2230.	1940.	290.
	36	3218.	266.68	119.26	5	3600.	3000.	600.
o:	37	340.	28.19	8.91	10	380.	300.	80.
	38	755.	37.08	11.72	10	800.	700.	100.
	39	2224.	182.68	69.04	7	2500.	2010.	490.
	40	3094.	235.10	88.86	7	3440.	2750.	690.
ɔ:	41	426.	31.34	9.91	10	480.	380.	100.
	42	876.	63.22	19.99	10	980.	760.	220.
	43	2204.	197.20	69.72	8	2560.	2000.	560.
	44	3173.	251.83	83.94	9	3540.	2740.	800.

FIG. 1

```
BMD01D SIMPLE DATA DESCRIPTION - VERSION OF  MAY 20, 1964
HEALTH SCIENCE COMPUTING FACILITY, UCLA

THIS PROGRAM HAS BEEN CHANGED AND ADAPTED FOR PHONETIC ANALYSIS BY
INSTITUTE OF PHONETICS, UNIVERSITY OF COPENHAGEN, DENMARK
VERSION OF DECEMBER, 1967.

PROBLEM CARD
PROBLEM NUMBER              SPEECH     METHOD NUMBER                  1
NUMBER OF CASES                  9     NUMBER OF SPECIAL VALUES       0
NUMBER OF VARIABLES             44     NUMBER OF TRANSGENERATIONS    -0
NUMBER OF VARIABLES ADDED       -0     INPUT TAPE NUMBER              5
         NUMBER OF VARIABLE FORMAT CARDS    1

SPECIAL VALUES CARD
     0.

VARIABLE FORMAT CARD(S)
(23X,F4.0,7X,F4.0,7X,F4.0,7X,F4.0,20X)

FEMALE SPEECH, RECORDINGS NO. 7,1 8,1 9,1 10,1 11,1 12,1 17,1 18,1 19,1
```

	VAR NO	MEAN	S.D.	S.E. OF MEAN	SAMPLE	MAXIMUM	MINIMUM	RANGE
i:	1	278.	25.98	8.66	9	315.	235.	80.
	2	2588.	177.36	59.12	9	2890.	2300.	590.
	3	3397.	227.03	75.68	9	3730.	3045.	685.
	4	4014.	186.41	62.14	9	4280.	3775.	505.
e:	5	334.	32.67	10.89	9	375.	265.	110.
	6	2555.	200.06	66.69	9	2800.	2125.	675.
	7	3161.	176.80	58.93	9	3455.	2890.	565.
	8	4082.	216.65	72.22	9	4455.	3815.	640.
ɛ:	9	413.	39.21	13.07	9	475.	350.	125.
	10	2416.	223.53	74.51	9	2710.	2010.	700.
	11	2984.	146.79	48.93	9	3200.	2785.	415.
	12	4132.	236.39	78.80	9	4445.	3710.	735.
a:	13	565.	108.66	36.22	9	670.	400.	270.
	14	2146.	229.45	76.48	9	2470.	1790.	680.
	15	2889.	119.12	39.71	9	3060.	2730.	330.
	16	4118.	242.39	80.80	9	4500.	3800.	700.
ɑ:	17	808.	179.50	59.83	9	1130.	525.	605.
	18	1327.	132.71	44.24	9	1550.	1130.	420.
	19	2864.	273.62	91.21	9	3200.	2240.	960.
	20	3956.	243.37	81.12	9	4285.	3450.	835.
y:	21	247.	25.63	8.54	9	290.	215.	75.
	22	2082.	136.66	45.55	9	2310.	1940.	370.
	23	2383.	210.34	74.37	8	2615.	2145.	470.
	24	3686.	207.81	69.27	9	4000.	3370.	630.
ø:	25	342.	47.96	15.99	9	390.	275.	115.
	26	1893.	123.77	41.26	9	2090.	1660.	430.
	27	2317.	178.82	59.61	9	2625.	2120.	505.
	28	3685.	236.55	78.85	9	4045.	3270.	775.
œ:	29	413.	33.63	11.21	9	460.	345.	115.
	30	1817.	150.00	50.00	9	1970.	1515.	455.
	31	2537.	92.53	30.84	9	2680.	2375.	305.
	32	3781.	160.61	56.78	8	4080.	3570.	510.
u:	33	283.	32.60	10.87	9	350.	230.	120.
	34	798.	112.62	42.56	7	970.	650.	320.
	35	3810.	0.	0.	1	3810.	3810.	0.
	36	3988.	406.59	287.50	2	4275.	3700.	575.
o:	37	360.	32.79	10.93	9	400.	320.	80.
	38	778.	31.73	11.22	8	815.	710.	105.
	39	2770.	0.	0.	1	2770.	2770.	0.
	40	3677.	377.68	168.91	5	4180.	3220.	960.
ɔ:	41	432.	49.12	16.37	9	500.	365.	135.
	42	968.	99.28	33.09	9	1135.	810.	325.
	43	2613.	3.54	2.50	2	2615.	2610.	5.
	44	3757.	281.65	114.98	6	4125.	3400.	725.

FIG. 2

BMD01D SIMPLE DATA DESCRIPTION - VERSION OF MAY 20, 1964
HEALTH SCIENCE COMPUTING FACILITY, UCLA

THE PROGRAM HAS BEEN CHANGED AND ADAPTED FOR PHONETIC ANALYSIS BY
INSTITUTE OF PHONETICS, UNIVERSITY OF COPENHAGEN, DENMARK
VERSION OF DECEMBER, 1967.

PROBLEM CARD

PROBLEM NUMBER	SPEECH	METHOD NUMBER	1
NUMBER OF CASES	6	NUMBER OF SPECIAL VALUES	0
NUMBER OF VARIABLES	44	NUMBER OF TRANSGENERATIONS	-0
NUMBER OF VARIABLES ADDED	-0	INPUT TAPE NUMBER	5

NUMBER OF VARIABLE FORMAT CARDS 1

SPECIAL VALUES CARD
0.

VARIABLE FORMAT CARD(S)
(23X,F4.0,7X,F4.0,7X,F4.0,7X,F4.0,20X)

VOWELS SPOKEN BY CHILDREN. REC. NO. 20.1 21.1 22.1 23.1 24.1 25.1

	VAR NO	MEAN	S.D.	S.E. OF MEAN	SAMPLE	MAXIMUM	MINIMUM	RANGE
i:	1	304.	35.97	14.69	6	375.	275.	100.
	2	2921.	226.90	92.63	6	3150.	2690.	460.
	3	3522.	235.21	96.02	6	3710.	3120.	590.
	4	4427.	200.22	81.74	6	4760.	4175.	585.
e:	5	407.	50.46	20.60	6	465.	350.	115.
	6	2581.	447.35	182.63	6	3025.	1985.	1040.
	7	3287.	321.80	131.38	6	3555.	2720.	835.
	8	4383.	207.50	84.71	6	4685.	4115.	570.
ε:	9	490.	33.17	13.54	6	540.	455.	85.
	10	2554.	196.45	80.20	6	2785.	2250.	535.
	11	3291.	243.20	99.28	6	3585.	2890.	695.
	12	4339.	253.80	103.61	6	4825.	4150.	675.
a:	13	537.	81.71	33.36	6	625.	400.	225.
	14	2380.	190.32	77.70	6	2660.	2165.	495.
	15	3213.	251.83	102.81	6	3520.	2795.	725.
	16	4314.	232.22	94.80	6	4780.	4160.	620.
ɑ:	17	985.	123.09	50.25	6	1165.	825.	340.
	18	1543.	77.77	31.75	6	1650.	1435.	215.
	19	2888.	265.76	108.50	6	3195.	2540.	655.
	20	3998.	211.30	86.26	6	4365.	3765.	600.
y:	21	284.	23.33	9.52	6	330.	265.	65.
	22	2181.	209.68	93.77	5	2395.	1925.	470.
	23	2568.	301.63	134.89	5	2810.	2075.	735.
	24	3863.	261.55	106.78	6	4290.	3520.	770.
ø:	25	413.	60.47	24.69	6	490.	345.	145.
	26	1873.	175.16	78.34	5	2035.	1575.	460.
	27	2599.	325.78	145.69	5	2900.	2085.	815.
	28	3932.	225.54	100.87	5	4230.	3640.	590.
œ:	29	478.	67.36	27.50	6	575.	400.	175.
	30	1815.	148.19	60.50	6	2000.	1600.	400.
	31	2849.	175.82	71.78	6	3045.	2570.	475.
	32	4018.	214.87	96.09	5	4250.	3790.	460.
u:	33	305.	31.22	13.96	5	350.	270.	80.
	34	794.	64.08	32.04	4	885.	735.	150.
	35	3553.	562.15	397.50	2	3950.	3155.	795.
	36	4270.	0.	0.	1	4270.	4270.	0.
o:	37	414.	54.17	22.11	6	485.	350.	135.
	38	876.	76.32	38.16	4	945.	782.	163.
	39	3810.	193.13	111.50	3	3980.	3600.	380.
	40	4630.	311.13	220.00	2	4850.	4410.	440.
ɔ:	41	461.	52.77	21.54	6	515.	380.	135.
	42	981.	97.49	39.80	6	1150.	875.	275.
	43	3678.	450.45	260.07	3	3975.	3160.	815.
	44	4503.	508.46	293.56	3	4950.	3950.	1000.

FIG. 3

INSTITUTE OF PHONETICS, UNIVERSITY OF COPENHAGEN, DENMARK

PROGRAM FOR COMPARISONS BETWEEN VOWEL SPECTRA GROUP A AND VOWEL SPECTRA GROUP B WITH STATISTIC CALCULATIONS.
VERSION OF DECEMBER 1967, (BFJ).

INPUT GROUP A AND INPUT GROUP B ARE TWO SETS OF VARIABLE NUMBERS,
MEAN VALUES,
STANDARD ERRORS OF MEANS, AND
NUMBER OF SAMPLES FOR EACH VARIABLE.
NUMBER OF VARIABLES MUST BE THE SAME IN GROUP A AND GROUP B.
THE MAXIMUM NUMBER OF VARIABLES IS 999.
THE MAXIMUM NUMBER OF SAMPLES IN A VARIABLE IS 999.

I = NUMBER OF A VARIABLE IN GROUP A.
AMEAN = MEANS OF SAMPLES IN GROUP A.
K = NUMBER OF SAMPLES FOR AMEAN.
ASDE = STANDARD ERRORS OF MEANS IN GROUP A.
APCT = ASDE IN PER CENT OF MEANS.
J = NUMBER OF A VARIABLE IN GROUP B.
BMEAN = MEANS OF SAMPLES IN GROUP B.
L = NUMBER OF SAMPLES FOR BMEAN.
BSDE = STANDARD ERRORS OF MEANS IN GROUP B.
BPCT = BSDE IN PER CENT OF MEANS.
DIFBA = THE DIFFERENCE BMEAN - AMEAN.
BAPCT = DIFBA IN PER CENT OF AMEAN.
SDABS = STANDARD DEVIATION FOR THE DIFFERENCE DIFBA.
SDPCT = THE RELATIVE STANDARD DEVIATION FOR DIFBA CALCULATED IN PER CENT.
SDPCT2 = TWO TIMES SDPCT.
SDPCT3 = THREE TIMES SDPCT.
NO = THE TOTAL NUMBER OF VARIABLES IN A GROUP.

COMPARISON BETWEEN FEMALE AND MALE SPOKEN VOWELS.

VAR NO	AMEAN	K	ASDE	APCT	BMEAN	L	BSDE	BPCT	DIFBA	BAPCT	SDABS	SDPCT	SDPCT2	SDPCT3	
1	235.	10	7.80	3.32	278.	9	8.66	3.12	43.	18.30	11.65	4.55	9.10	13.66	i:
2	2119.	10	56.97	2.69	2588.	9	59.12	2.28	469.	22.13	82.10	3.53	7.06	10.58	
3	3013.	10	59.51	1.98	3397.	9	75.68	2.23	384.	12.74	96.28	2.98	5.95	8.93	
4	3354.	9	62.59	1.87	4014.	9	62.14	1.55	660.	19.68	88.20	2.42	4.85	7.27	
5	283.	10	9.73	3.44	334.	9	10.89	3.26	51.	18.02	14.60	4.74	9.48	14.21	e:
6	2091.	10	48.97	2.34	2555.	9	66.69	2.61	464.	22.19	82.74	3.51	7.01	10.52	
7	2709.	10	50.05	1.85	3161.	9	58.93	1.86	452.	16.69	77.32	2.62	5.25	7.87	
8	3389.	10	65.70	1.94	4082.	9	72.22	1.77	693.	20.45	97.63	2.62	5.25	7.87	
9	367.	10	8.85	2.41	413.	9	13.07	3.16	46.	12.53	15.78	3.98	7.96	11.94	ɛ:
10	1973.	10	41.23	2.09	2416.	9	74.51	3.08	443.	22.45	85.16	3.73	7.45	11.18	
11	2479.	10	59.19	2.39	2984.	9	48.93	1.64	505.	20.37	76.80	2.90	5.79	8.69	
12	3369.	10	60.76	1.80	4132.	9	78.80	1.91	763.	22.65	99.50	2.62	5.25	7.87	
13	541.	10	24.33	4.50	565.	9	36.22	6.41	24.	4.44	43.63	7.83	15.66	23.49	a:
14	1708.	10	59.06	3.46	2146.	9	76.48	3.56	438.	25.64	96.63	4.97	9.93	14.90	
15	2289.	10	62.77	2.74	2889.	9	39.71	1.37	600.	26.21	74.28	3.07	6.13	9.20	
16	3423.	10	65.19	1.90	4118.	9	80.80	1.96	695.	20.30	103.82	2.73	5.47	8.20	
17	698.	10	26.47	3.79	808.	9	59.83	7.40	110.	15.76	65.42	8.32	16.64	24.96	ɑ:
18	1144.	10	18.99	1.66	1327.	9	44.24	3.33	183.	16.00	48.14	3.72	7.45	11.17	
19	2480.	9	80.11	3.23	2864.	9	91.21	3.18	384.	15.48	121.40	4.54	9.07	13.61	
20	3464.	10	77.04	2.22	3956.	9	81.12	2.05	492.	14.20	111.87	3.03	6.05	9.08	
21	242.	10	7.57	3.13	247.	9	8.54	3.46	5.	2.07	11.41	4.66	9.33	13.99	y:
22	1862.	10	26.48	1.42	2082.	9	45.55	2.19	220.	11.82	52.69	2.61	5.22	7.83	
23	2096.	10	33.15	1.58	2383.	8	74.37	3.12	287.	13.69	81.42	3.50	7.00	10.50	
24	3180.	9	67.00	2.11	3686.	9	69.27	1.88	506.	15.91	96.37	2.82	5.65	8.47	
25	308.	10	8.50	2.76	342.	9	15.99	4.68	34.	11.04	18.11	5.43	10.86	16.29	ø:
26	1633.	10	27.08	1.66	1893.	9	41.26	2.18	260.	15.92	49.35	2.74	5.48	8.22	
27	2018.	10	35.36	1.75	2317.	9	59.61	2.57	299.	14.82	69.31	3.11	6.23	9.34	
28	3172.	9	67.29	2.12	3685.	9	78.85	2.14	513.	16.17	103.66	3.01	6.03	9.04	
29	380.	10	10.46	2.75	413.	9	11.21	2.71	33.	8.68	15.33	3.87	7.73	11.60	œ:
30	1546.	10	25.61	1.66	1817.	9	50.00	2.75	271.	17.53	56.18	3.21	6.42	9.64	
31	2039.	10	31.37	1.54	2537.	9	30.84	1.22	498.	24.42	43.99	1.96	3.92	5.88	
32	3196.	10	56.96	1.84	3781.	8	56.78	1.50	585.	18.30	81.86	2.38	4.76	7.14	
33	266.	10	13.06	4.91	283.	9	10.87	3.84	17.	6.39	16.99	6.23	12.47	18.70	u:
34	722.	9	32.97	4.57	798.	7	42.56	5.33	76.	10.53	53.84	7.02	14.04	21.06	
35	2091.	6	50.17	2.40	3810.	1	0.	0.	1719.	82.21	50.17	2.40	4.80	7.20	
	WARNING CNLY, - NUMBER OF SAMPLES (L) FOR BMEAN IS LESS THAN OR EQUALS 3														
36	3218.	5	119.26	3.71	3988.	2	287.50	7.21	770.	23.93	311.25	8.11	16.21	24.32	
	WARNING CNLY, - NUMBER OF SAMPLES (L) FOR BMEAN IS LESS THAN OR EQUALS 3														
37	340.	10	8.91	2.62	360.	9	10.93	3.04	20.	5.88	14.10	4.01	8.02	12.03	o:
38	755.	10	11.72	1.55	778.	8	11.22	1.44	23.	3.05	16.22	2.12	4.24	6.36	
39	2224.	7	69.04	3.10	2770.	1	0.	0.	546.	24.55	69.04	3.10	6.21	9.31	
	WARNING CNLY, - NUMBER OF SAMPLES (L) FOR BMEAN IS LESS THAN OR EQUALS 3														
40	3094.	7	88.86	2.87	3677.	5	168.91	4.59	583.	18.84	190.86	5.42	10.84	16.25	
41	426.	10	9.91	2.33	432.	9	16.37	3.79	6.	1.41	19.14	4.45	8.89	13.34	ɔ:
42	876.	10	19.99	2.28	968.	9	33.09	3.42	92.	10.50	38.66	4.11	8.22	12.33	
43	2204.	8	69.72	3.16	2613.	2	2.50	0.10	409.	18.56	69.76	3.16	6.33	9.49	
	WARNING ONLY, - NUMBER OF SAMPLES (L) FOR BMEAN IS LESS THAN OR EQUALS 3														
44	3173.	9	83.94	2.65	3757.	6	114.98	3.06	584.	18.41	142.36	4.05	8.09	12.14	

THE AVERAGE PERCENTAGE OF FORMANT 1 IN GROUP B IN RELATION TO GROUP A IS 9.50 PER CENT,

THE AVERAGE PERCENTAGE OF FORMANT 2 IN GROUP B IN RELATION TO GROUP A IS 16.16 PER CENT,

THE AVERAGE PERCENTAGE OF FORMANT 3 IN GROUP B IN RELATION TO GROUP A IS 24.52 PER CENT,

THE AVERAGE PERCENTAGE OF FORMANT 4 IN GROUP B IN RELATION TO GROUP A IS 18.99 PER CENT.

THE TOTAL AVERAGE PERCENTAGE OF FORMANTS IN GROUP A IN RELATION TO FORMANTS IN GROUP B IS 17.29 PER CENT.

CALCULATIONS HAVE BEEN COMPLETED.

STOP

FIG. 4

INSTITUTE OF PHONETICS, UNIVERSITY OF COPENHAGEN, DENMARK

PROGRAM FOR COMPARISONS BETWEEN VOWEL SPECTRA GROUP A AND VOWEL SPECTRA GROUP B WITH STATISTIC CALCULATIONS.
VERSION OF DECEMBER 1967, (BFJ).

INPUT GROUP A AND INPUT GROUP B ARE TWO SETS OF VARIABLE NUMBERS,
MEAN VALUES,
STANDARD ERRORS OF MEANS, AND
NUMBER OF SAMPLES FOR EACH VARIABLE.
NUMBER OF VARIABLES MUST BE THE SAME IN GROUP A AND GROUP B.
THE MAXIMUM NUMBER OF VARIABLES IS 999.
THE MAXIMUM NUMBER OF SAMPLES IN A VARIABLE IS 999.

I = NUMBER OF A VARIABLE IN GROUP A.
AMEAN = MEANS OF SAMPLES IN GROUP A.
K = NUMBER OF SAMPLES FOR AMEAN.
ASDE = STANDARD ERRORS OF MEANS IN GROUP A.
APCT = ASDE IN PER CENT OF MEANS.
J = NUMBER OF A VARIABLE IN GROUP B.
BMEAN = MEANS OF SAMPLES IN GROUP B.
L = NUMBER OF SAMPLES FOR BMEAN.
BSDE = STANDARD ERRORS OF MEANS IN GROUP B.
BPCT = BSDE IN PER CENT OF MEANS.
DIFBA = THE DIFFERENCE BMEAN - AMEAN.
BAPCT = DIFBA IN PER CENT OF AMEAN.
SDABS = STANDARD DEVIATION FOR THE DIFFERENCE DIFBA.
SDPCT = THE RELATIVE STANDARD DEVIATION FOR DIFBA CALCULATED IN PER CENT.
SDPCT2 = TWO TIMES SDPCT.
SDPCT3 = THREE TIMES SDPCT.
NO = THE TOTAL NUMBER OF VARIABLES IN A GROUP.

COMPARISON BETWEEN SPOKEN MALE VOWELS AND SPOKEN CHILDREN'S VOWELS.

VAR NO	AMEAN	K	ASDE	APCT	BMEAN	L	BSDE	BPCT	DIFBA	BAPCT	SDABS	SDPCT	SDPCT2	SDPCT3	
1	235.	10	7.80	3.32	304.	6	14.69	4.83	69.	29.36	16.63	5.86	11.72	17.59	
2	2119.	10	56.97	2.69	2921.	6	92.63	3.17	802.	37.85	108.75	4.16	8.31	12.47	i:
3	3013.	10	59.51	1.98	3522.	6	96.02	2.73	509.	16.89	112.97	3.37	6.73	10.10	
4	3354.	9	62.59	1.87	4427.	6	81.74	1.85	1073.	31.99	102.95	2.63	5.25	7.88	
5	283.	10	9.70	3.43	407.	6	20.60	5.06	124.	43.82	22.77	6.11	12.23	18.34	
6	2091.	10	48.97	2.34	2581.	6	182.63	7.08	490.	23.43	189.08	7.45	14.91	22.36	e:
7	2709.	10	50.05	1.85	3287.	6	131.38	4.00	578.	21.34	140.59	4.40	8.81	13.21	
8	3389.	10	65.70	1.94	4383.	6	84.71	1.93	994.	29.33	107.20	2.74	5.47	8.21	
9	367.	10	8.85	2.41	490.	6	13.54	2.76	123.	33.51	16.18	3.67	7.34	11.00	
10	1973.	10	41.03	2.08	2554.	6	80.20	3.14	581.	29.45	90.09	3.77	7.53	11.30	ɛ:
11	2479.	10	59.19	2.39	3291.	6	99.28	3.02	812.	32.76	115.59	3.85	7.69	11.54	
12	3369.	10	60.76	1.80	4339.	6	103.61	2.39	970.	28.79	120.11	2.99	5.98	8.98	
13	541.	10	24.33	4.50	537.	6	33.36	6.21	-4.	-0.74	41.29	7.67	15.34	23.01	
14	1708.	10	59.06	3.46	2380.	6	77.70	3.26	672.	39.34	97.60	4.76	9.51	14.27	a:
15	2289.	10	62.77	2.74	3213.	6	102.81	3.20	924.	40.37	120.46	4.21	8.43	12.64	
16	3423.	10	65.19	1.90	4314.	6	94.80	2.20	891.	26.03	115.05	2.91	5.82	8.72	
17	698.	10	26.47	3.79	985.	6	50.25	5.10	287.	41.12	56.80	6.36	12.71	19.07	
18	1144.	10	18.99	1.66	1543.	6	31.75	2.06	399.	34.88	37.00	2.64	5.29	7.93	ɑ:
19	2480.	9	80.11	3.23	2888.	6	108.50	3.76	408.	16.45	134.87	4.95	9.91	14.86	
20	3464.	10	77.04	2.22	3998.	6	86.26	2.16	534.	15.42	115.65	3.10	6.20	9.30	
21	242.	10	7.57	3.13	284.	6	9.52	3.35	42.	17.36	12.16	4.58	9.17	13.75	
22	1862.	10	26.48	1.42	2181.	5	93.77	4.30	319.	17.13	97.44	4.53	9.06	13.59	y:
23	2096.	10	33.15	1.58	2568.	5	134.89	5.26	472.	22.52	138.90	5.49	10.97	16.46	
24	3180.	9	67.00	2.11	3863.	6	106.78	2.76	683.	21.48	126.06	3.48	6.95	10.43	
25	308.	10	8.50	2.76	413.	6	24.69	5.98	105.	34.09	26.11	6.58	13.17	19.75	
26	1633.	10	27.08	1.66	1873.	5	78.34	4.18	240.	14.70	82.89	4.50	9.00	13.50	ø:
27	2018.	10	35.36	1.75	2599.	5	145.69	5.61	581.	28.79	149.92	5.87	11.75	17.62	
28	3172.	9	67.29	2.12	3932.	5	100.87	2.57	760.	23.96	121.25	3.33	6.66	9.99	
29	380.	10	10.46	2.75	478.	6	27.50	5.75	98.	25.79	29.42	6.38	12.76	19.13	
30	1546.	10	25.61	1.68	1815.	6	60.50	3.33	269.	17.40	65.70	3.72	7.44	11.17	œ:
31	2039.	10	31.37	1.54	2849.	6	71.78	2.52	810.	39.73	78.34	2.95	5.90	8.86	
32	3196.	10	58.96	1.84	4018.	5	96.09	2.39	822.	25.72	112.74	3.02	6.04	9.06	
33	266.	10	13.06	4.91	305.	5	13.96	4.58	39.	14.66	19.12	6.71	13.42	20.14	
34	722.	9	32.97	4.57	794.	4	32.04	4.04	72.	9.97	45.97	6.09	12.19	18.28	
35	2091.	6	50.17	2.40	3553.	2	397.50	11.19	1462.	69.92	400.65	11.44	22.88	34.33	u:
	WARNING ONLY, - NUMBER OF SAMPLES (L) FOR BMEAN IS LESS THAN OR EQUALS 3														
36	3218.	5	119.26	3.71	4270.	1	0.	C.	1052.	32.69	119.26	3.71	7.41	11.12	
	WARNING ONLY, - NUMBER OF SAMPLES (L) FOR BMEAN IS LESS THAN OR EQUALS 3														
37	340.	10	8.91	2.62	414.	6	22.11	5.34	74.	21.76	23.84	5.95	11.90	17.85	
38	755.	10	11.72	1.55	876.	4	38.16	4.36	121.	16.03	39.92	4.62	9.25	13.87	
39	2224.	7	69.04	3.10	3810.	3	111.50	2.93	1586.	71.31	131.14	4.27	8.53	12.80	o:
	WARNING ONLY, - NUMBER OF SAMPLES (L) FOR BMEAN IS LESS THAN OR EQUALS 3														
40	3094.	7	88.86	2.87	4630.	2	220.00	4.75	1536.	49.64	237.27	5.55	11.10	16.66	
	WARNING ONLY, - NUMBER OF SAMPLES (L) FOR BMEAN IS LESS THAN OR EQUALS 3														
41	426.	10	9.91	2.33	461.	6	21.54	4.67	35.	8.22	23.71	5.22	10.44	15.66	
42	876.	10	19.99	2.28	981.	6	39.80	4.06	105.	11.99	44.54	4.65	9.31	13.96	
43	2204.	8	69.72	3.16	3678.	3	260.07	7.07	1474.	66.88	269.25	7.75	15.49	23.24	ɔ:
	WARNING ONLY, - NUMBER OF SAMPLES (L) FOR BMEAN IS LESS THAN OR EQUALS 3														
44	3173.	9	83.94	2.65	4503.	3	293.56	6.52	1330.	41.92	305.33	7.04	14.07	21.11	
	WARNING ONLY, - NUMBER OF SAMPLES (L) FOR BMEAN IS LESS THAN OR EQUALS 3														

THE AVERAGE PERCENTAGE OF FORMANT 1 IN GROUP B IN RELATION TO GROUP A IS 24.45 PER CENT,

THE AVERAGE PERCENTAGE OF FORMANT 2 IN GROUP B IN RELATION TO GROUP A IS 22.92 PER CENT,

THE AVERAGE PERCENTAGE OF FORMANT 3 IN GROUP B IN RELATION TO GROUP A IS 38.81 PER CENT,

THE AVERAGE PERCENTAGE OF FORMANT 4 IN GROUP B IN RELATION TO GROUP A IS 29.72 PER CENT.

THE TOTAL AVERAGE PERCENTAGE OF FORMANTS IN GROUP A IN RELATION TO FORMANTS IN GROUP B IS 28.98 PER CENT.

CALCULATIONS HAVE BEEN COMPLETED.

STOP

FIG. 5

0, 1, and 2. If the code is "0" only the above mentioned output table will be calculated and printed out. If the code number is "1", the computer will add a table of the formant frequency ratios between group B and group A, as shown in Figs. 4 and 5. If the code number is "2", the output will consist of the output table with means and standard deviations, formant frequency ratios, and finally a table of formant level ratios. (Reference 9.)

In order to avoid mistakes some typical defects or errors in the input material will cause error messages in the output. These error messages have been tested by means of deliberately incorrectly punched input cards.

2. The Relation between Female and Male Formant Frequencies.

As a preliminary example of the phonetic application of the two statistic programs described above I shall now proceed to discuss the output tables given in Figs. 4 and 5.

Illustration Fig. 6 shows a diagram based on the female/male formant frequency calculations in Fig. 4, and Fig. 7 shows a diagram based on the children/male formant frequency calculations presented in Fig. 5.

The scale factor k (called BAPCT in the TALCOM program), which equals the differences in per cent between the female and the corresponding male formant frequencies for each formant in the 11 vowels, is indicated by means of small circles.

The vertical lines extending from each k-circle indicate an interval of $\pm$ twice the standard deviation of k. Provided that the distribution is normal more than 95 % of all possible female/male formant ratios must fall inside this interval.

k_1, k_2, k_3, and k_4 are placed over each other in the diagram. In this way it is possible to compare the k scale factor variations directly in the four formant regions. The vowels are plotted along the horizontal axis, and the variations of k in per cent along the vertical axis.

How do these Data correlate with Investigations from other Authors?

As for the female/male and children/male formant ratios the first investigation on this subject was carried out in Japan in

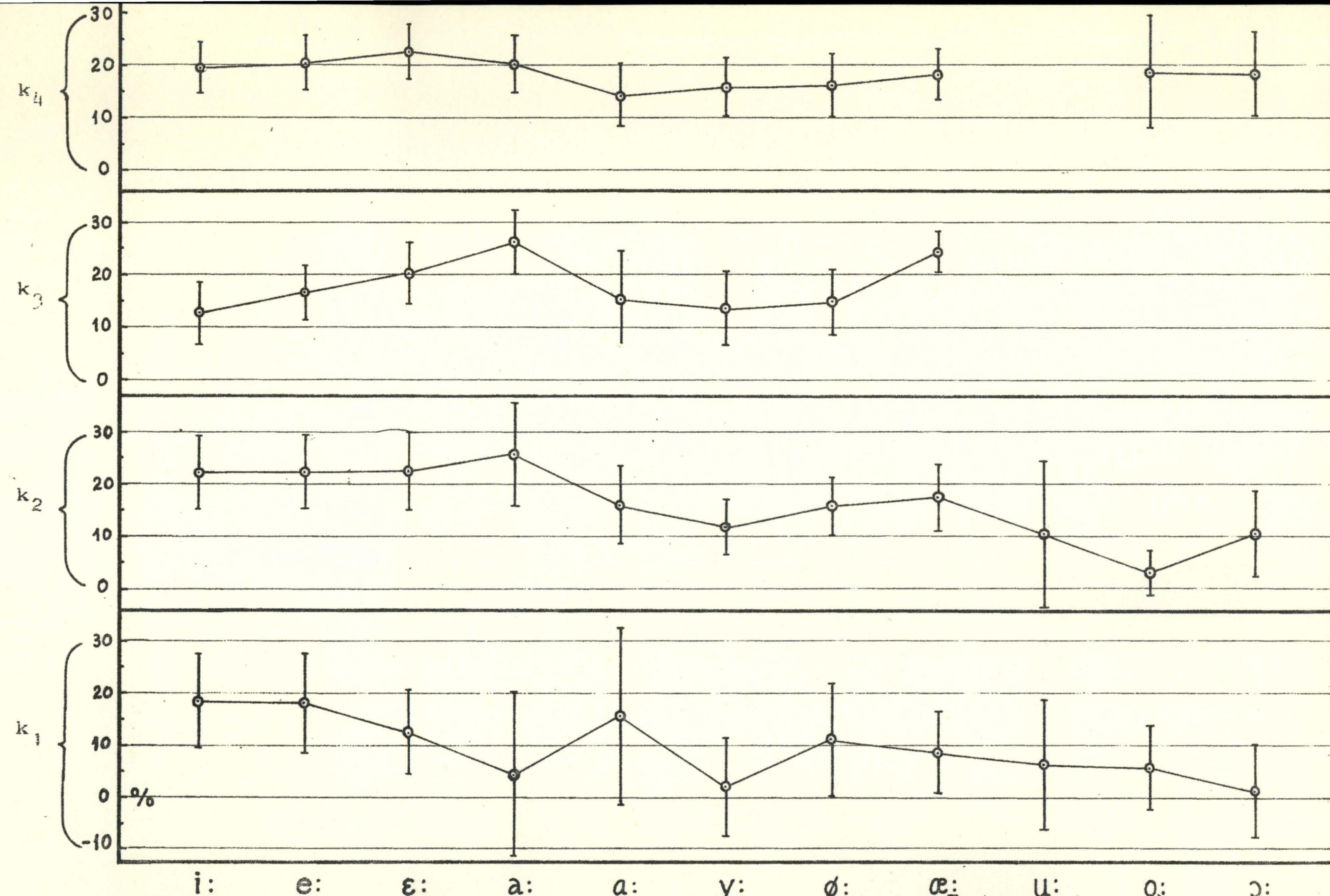
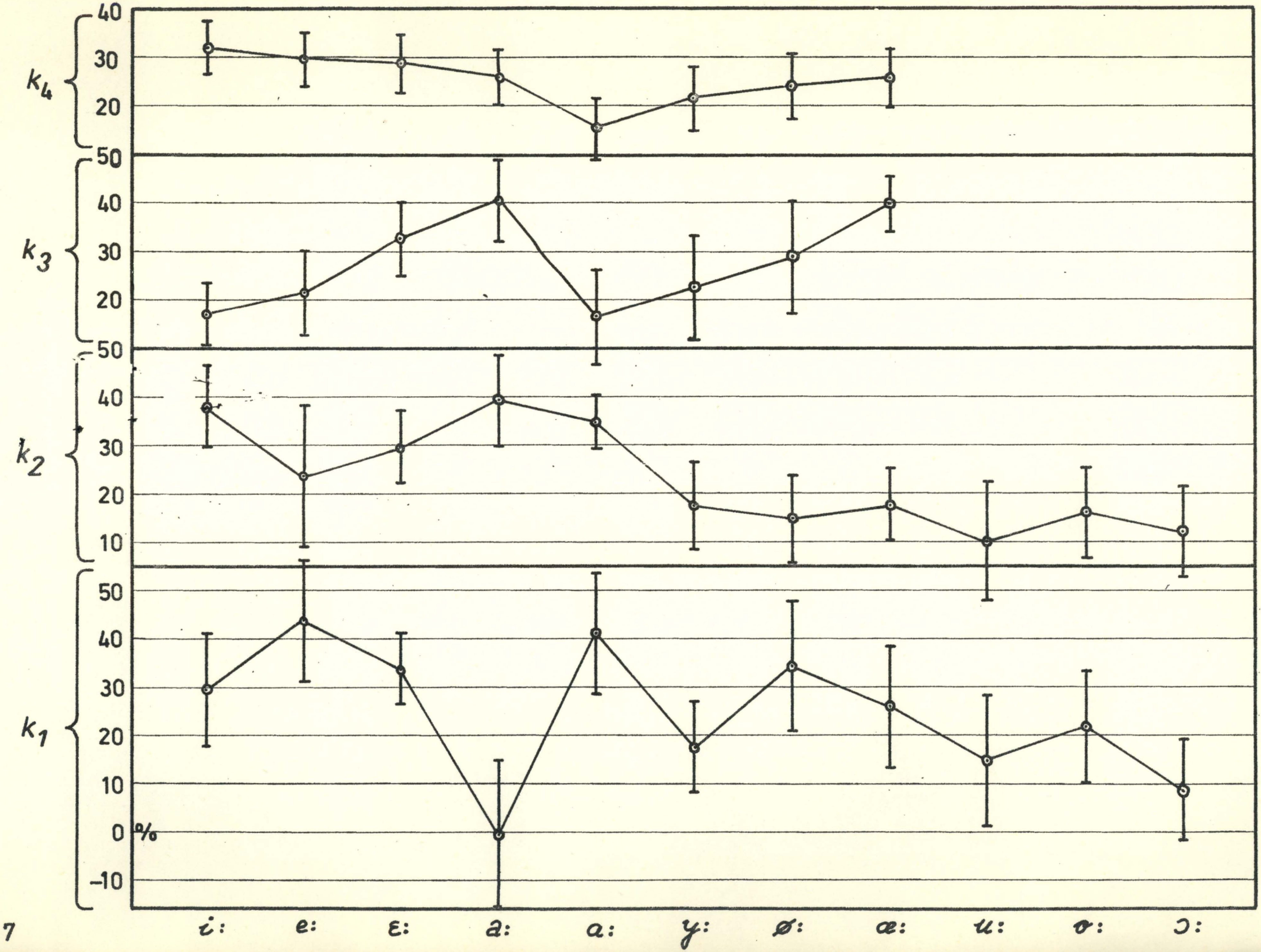


Fig. 6 The average female/male formant frequency ratio. $k_n = \left(\frac{F_n(\text{female})}{F_n(\text{male})} - 1 \right) \cdot 100\ \%$

The statistic error on each k_n-value is shown by means of vertical lines indicating $\pm$ 2 times the standard deviation of k_n.


k4
40
30
20
k3
50
40
30
20
k2
50
40
30
20
10
k1
50
40
30
20
10
0 %
-10
i:
e:
ε:
ə:
a:
y:
ø:
æ:
u:
o:
ɔ:
FIG. 7

1940 by Chiba and Kajiyama (3). In 1959 - 1966 some data were published by G. Fant ((4), (5), and (6)) based on 7 male and 7 female subjects. These data included measurements of the formant levels, and so did the published data by G.E. Peterson and Barney in 1952 (7). The material of these authors was based upon 33 male subjects, 28 female subjects, and 15 children, and it is thus a rather extensive collection of data. As pointed out by the authors, too, this material shows statistically significant differences between the speakers with a 99 % confidence limit for the averages of formant frequencies and levels.

The Swedish, English, and Danish data correlate fairly well for most of the vowels. In all three sets of data we find that k_1 is high in the open back vowel [ɑ:] and low in the rounded back vowels [o:] and [ɔ:]. Also the close rounded front vowel [y:] has a low k_1. In contradistinction to the data of Fant and Peterson/Barney my material shows a high k_1 in the unrounded front vowels [i:], [e:], and [ɛ:]. All authors have found that k_2 is lowest in back vowels and highest in the unrounded front vowels. Notice the k scale factor in [a:] and [ɑ:] especially in Fig. 7 (comparison between children's and male vowel formants): [a:] and [ɑ:] have significantly different k_1 and k_3 scale factors, where k_1 for [a:] is slightly below 0 % and k_1 for [ɑ:] is higher than 40 %, whereas k_3 for [a:] is higher than 40 % and k_3 for [ɑ:] is only 16.5 %. Finally, my data suggest that k_4 is of a rather constant nature.

3. Discussion concerning Requirements on Confidence Limits:

It is clearly shown in Fig. 6 and Fig. 7 that there is in general a significant difference between male formant frequencies and female or children's formant frequencies. All the average formants in female or children's speech are higher than the corresponding male average formants, which may be seen from the formant ratio tables below the main tables of statistic calculations, Fig. 4 and Fig. 5. We do not need any statistic calculations in order to observe this. Quite another problem appears when we oberve the differences between the vowels, i.e. the k scale factors for different vowels. Is the difference between the k scale factors significant ?

Because of the overlapping in Fig. 6 we must conclude that a request for 95 % confidence limits (which is satisfied by k_n ± twice the standard deviation of k_n) is too hard a request for a comparison of the female/male formant ratio if it is based on 10 male and 9 female recordings only as is the case in my test material for this paper. Especially in the k_1 range we have a strong overlapping (it is virtually possible to draw a straight line through all the k_1 dispersion ranges indicating a common 7 % k_1 scale factor). A 68 % significance level (which is satisfied for k_n ± the standard deviation of k_n) would be more realistic because of the reduced overlapping ranges, but it has on the other hand very little meaning to speak of a significance level which excludes about one third of the cases.

If we compare the 6 children's vowels with the male vowels, Fig. 7, we can observe greater differences between the various children/male formant ratios than between the female/male formant ratios. This has been pointed out earlier by several investigators.

In spite of the greater standard deviations in Fig. 7 caused by the smaller number of samples in the children's group, we may probably expect a better significance level because of the greater differences between the k scale factors. This is seen in the diagram where the k scale factor variations are greater and the overlappings between the vowels therefore fewer than in Fig. 6.

These rough statistic calculations may give rise to the essential question whether or not the phoneticians normally operate with a too restricted material for phonetic investigations. If we want a reliable significance in a certain material we ought to operate with a probability level which equals or is better than 0.05 as a criterion of significance. This presupposes a sample number which must be so great that the range constituted of at least ± twice the standard deviation is small in relation to the differences between the phenomena under consideration.

Acknowledgement:

We want to thank the Northern Europe University Computing Center for ready help with the program tests and run of programs.

References:

(1) W.J.Dixon, Biomedical Computer Programs, Health Science Computing Facility, University of California.

(2) B. Frøkjær-Jensen, "The Danish Long Vowels", ARIPUC No. 1 (1966).

(3) Chiba and Kajiyama, The Vowel (Tokio, 1941).

(4) Gunnar Fant, "Acoustic Analysis and Synthesis of Speech with Application to Swedish", Ericsson Technics 15 (1959), pp. 60-86.

(5) Gunnar Fant, Acoustic Theory of Speech Production ('s-Gravenhage, 1960).

(6) Gunnar Fant, "A Note on Vocal Tract Size Factors and Non-Uniform F-Pattern Scalings", Quarterly Progress and Status Report 4/1966, Speech Transmission Laboratory, RIT, Stockholm.

(7) G.E.Peterson and H.L.Barney, "Control Methods Used in a Study of the Vowels", Journal of the Acoustic Society of America 24 (1952), pp. 175-184.

----

(8) On request the changed BMD-01D program can be obtained from our Institute as Fortran II source deck.

(9) On request the TALCOM program can be obtained from our Institute as Fortran IV source deck.