ARIPUC VOL. 6.

ANNUAL REPORT

of the
Institute of Phonetics
University of Copenhagen

COPENHAGEN 1972

F1

ARIPUC VOL. 6.

ANNUAL REPORT

of the
Institute of Phonetics
University of Copenhagen

COPENHAGEN 1972

KBHVNS. UNIV. OFFSET AFD. 6133/72
ISBM NR. 87-505-0247-6.

CONTENTS

(continued)

(continued)

PERSONNEL OF THE INSTITUTE OF PHONETICS
1971

Permanent Staff:

Eli Fischer-Jørgensen (professor, director of the Institute)
Jørgen Rischel (lecturer and amanuensis of general phonetics)
Hans Peter Jørgensen (amanuensis of general phonetics)
Oluf M. Thorsen (amanuensis of general and French phonetics)
Børge Frøkjær-Jensen (amanuensis of experimental phonetics)
Hans Basbøll (amanuensis of general phonetics)
Ole Kongsdal Jensen (amanuensis of general and French phonetics)
Karen Landschultz (amanuensis of general and French phonetics)
Carl Ludvigsen (M.Sc.)
Poul Thorvaldsen (B.Sc.)
Svend-Erik Lystlund (technician)
Inger Østergaard (secretary)
Aase Thiim (secretary)

Part Time Teachers of General Phonetics:

Kirsten Gregersen (teaching assistant)
Steffen Heger (teaching assistant)
Peter Holtse (teaching assistant)
Birgit Hutters (teaching assistant)
Nina G. Thorsen (teaching assistant)
Paul Glenting, dr. med. (teaching assistant)
Jørgen Paulsen, dr. med. (teaching assistant)

Guest Research Workers:

Niels Bak (Ranum Teacher Training College)
Marie-Hélène Galvagny (Paris)
Philip Mansell (Essex)
Hideo Mase (Japan)

PUBLICATIONS BY STAFF MEMBERS 1971

Hans Basbøll, "A Commentary on Hjelmslev's Outline of the Danish Expression System" (I), Acta Linguistica Hafniensia, XIII,1971, p. 173-211.

Eli Fischer-Jørgensen,* "Untersuchungen zum sogenannten festen und losen Anschluss", Kopenhagener germanistische Studien,1,1969 = Festschrift Peter Jørgensen, p. 138-164.

Eli Fischer-Jørgensen, "Resumé af forelæsninger over psykoakustik og auditiv fonetik" ("Summary of lectures in psychoacoustics and auditory phonetics"; mimeographed), 1971.

Børge Frøkjær-Jensen,
Kaj Lauritzen, "Fonetisk-akustisk analyse af recurrensstemmer før og efter pædagogisk og operativ behandling", Nordisk Tidsskrift for Tale og Stemme, 1970,2, p. 57-69 (published 1971).

Børge Frøkjær-Jensen,
Carl Ludvigsen,
Jørgen Rischel, "A Glottographic Study of Some Danish Consonants", Form & Substance, Copenhagen 1971, p. 123-140.

Kirsten Gregersen, "Phonetic Investigation of Internal Open Jucture in British English", Language and Literature, vol. 1, 1971, p. 66-72.

*) By mistake these publications were not mentioned in the appropriate volume of Aripuc.

Hans Peter Jørgensen,* "Über den Intensitätsverlauf beim sogenannten losen und festen Anschluss im Deutschen", *Kopenhagener germanistische Studien*,1,1969 = *Festschrift Peter Jørgensen*, p. 165-186.

Ole Kongsdal Jensen,
Karen Landschultz,
Oluf M. Thorsen, *"Fransk fonetik"* ("French phonetics", to be continued; mimeographed), 1971.

Karen Landschultz, "Quantité vocalique en francais - relations quantitatives des voyelles accentuées suivies d'une consonne fricative", *Revue Romane*, tome VI, fasc. 1, 1971, p. 25-51.

Jørgen Rischel, "Fonologiske grundbegreber" ("Basic concepts in phonology", to be continued; mimeographed), 1971.

Jørgen Rischel, "Some Characteristics of Noun Phrases in West Greenlandic", *Acta Linguistica Hafniensia* XIII, 1971, p. 213-245.

Nina G. Thorsen, "Voice Assimilation in /t/ and /d/ in British English", *Language and Literature*, vol. 1, 1971, p. 35-41.

Supplementary note

The Festschrift to Eli Fischer-Jørgensen: *Form & Substance*, Copenhagen 1971, can be purchased from Akademisk Forlag, Store Kannikestræde 8, DK-1169 Copenhagen K.

LECTURES AND COURSES IN 1971

1. Elementary phonetics courses

One-semester courses (two hours a week) in elementary phonetics (intended for all students of foreign languages except French) were given by Hans Basbøll, Kirsten Gregersen, Steffen Heger, Peter Holtse, Birgit Hutters, Hans Peter Jørgensen, Karen Landschultz, Jørgen Rischel, and Nina G. Thorsen. There were 2 parallel classes in the spring semester and 17 in the autumn semester.

Two-semester courses (two hours a week) in general and French phonetics (intended for all students of French) were given through 1971 by Ole Kongsdal Jensen, Karen Landschultz, and Oluf M. Thorsen. There were 7 parallel classes in the spring semester and 8 in the autumn semester.

2. Practical training in sound perception and transcription

Courses for beginners as well as courses for more advanced students were given through 1971 by Jørgen Rischel and Oluf M. Thorsen. (The courses which are based in part on tape recordings and in part on work with informants, form a cycle of three semesters with two hours a week.)

3. Instrumental phonetics

Courses for beginners as well as courses for more advanced students were given through 1971 by Børge Frøkjær-Jensen. (The courses form a cycle of three semesters with two hours a week.)

4. Phonology

Courses for beginners as well as courses for more advanced students were given through 1971 by Hans Basbøll. (The courses form a cycle of three semesters with two hours a week, the subject of the three semesters being elementary phonological method, trends in phonological theory, and generative phonology.)

5. Other courses

Eli Fischer-Jørgensen lectured on auditory phonetics and on the preparation of auditory tests in the spring and autumn semesters.

Eli Fischer-Jørgensen gave a course in the analysis of sonagrams and mingograms of speech in the autumn semester.

Jørgen Rischel lectured on Greenlandic phonetics in the spring semester.

Jørgen Rischel lectured on the production of synthetic speech (using the speech synthesizer of the Institute) in the autumn semester.

Carl Ludvigsen gave a course in statistics in the spring and autumn semesters.

Carl Ludvigsen gave a course in mathematics and electronics in the spring and autumn semesters.

Hans Basbøll gave a course in Danish phonology in the spring semester.

Paul Glenting lectured on neurology in the autumn semester.

Jørgen Paulsen lectured on the anatomy and physiology of the speech organs in the autumn semester.

6. Seminars

The following seminars were held in 1971:

Kirsten Gregersen presented her study on juncture in English.

Speech therapist Kai Lauritzen: "Fonetik og talepædagogik" ("phonetics and logopedics").

Philip Mansell, M.A. (Essex): "A phonetic component and a phonological component of a speech production model".

Philip Mansell, M.A. (Essex) gave a course in electromyography at the Institute, June 14-18.

Professor Fujimura (Tokyo) lectured on investigations of the glottis and on investigations of stop consonants.

Professor Sawashima (Tokyo): "Fiberoptics in Speech Research". In connection with this seminar professor Sawashima presented a film showing velum movements (photographed through the nasal cavity) and a film showing the vocal chords.

Professor Fujimura (Tokyo) lectured on Korean stop consonants and presented a film on dynamic palatography.

Steffen Heger: "Udtalen af dansk R og dets påvirkning på foregående og følgende vokal" ("the pronunciation of Danish r and its influence on the preceding and the following vowel").

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

Eli Fischer-Jørgensen participated in the 7th International Congress of Phonetic Sciences in Montreal, August 23-28, 1971.

Eli Fischer-Jørgensen gave some lectures at the universities of Trondheim and Bergen in September 1971.

Jørgen Rischel gave a course in phonology for postgraduates at the Institute of Phonetics, Lund, in the spring semester.

Jørgen Rischel visited the "Instituut voor perceptie onderzoek", Eindhoven.

Jørgen Rischel participated in the annual meeting of the Societas Linguistica Europaea, September 1971, Leiden.

Jørgen Rischel and Hans Basbøll participated in the Third Scandinavian Summer School of Linguistics in Copenhagen, August 1971.

Børge Frøkjær-Jensen and Carl Ludvigsen participated in the "Speech Symposium" in Szeged (Hungaria), August 26-29, 1971.

Børge Frøkjær-Jensen and Carl Ludvigsen participated in the "9th Acoustic Conference" (Physiological Acoustics and Psychoacoustics) in High Tatras (Czechoslovakia), August 31 - September 4, 1971. Børge Frøkjær-Jensen participated in the round table discussion at which he gave a paper on "Acoustic Parameters in Speech Research".

Carl Ludvigsen participated in the "7th International Congress on Acoustics", Budapest, August 1971.

Børge Frøkjær-Jensen read two papers before the "Danish Association of Logopedics and Phoniatrics" in 1971:
1. "Experimental Phonetics in Logopedics and Phoniatrics",
2. "A Status Report on Glottography".

Børge Frøkjær-Jensen gave some lectures in experimental phonetics at the Institute of Phonetics, Lund, in June 1971.

Oluf M. Thorsen lectured on problems in French phonology at a meeting attended by students and teachers in French Philology, Copenhagen, March 1971.

Hans Peter Jørgensen spent the autumn semester participating in courses at the Pedagogical Institute of the University of Copenhagen.

The teachers and senior students of the Institute participated in a course (about 50 hours in all) held by the Pedagogical Institute of the University of Copenhagen in the spring semester.

INSTRUMENTAL EQUIPMENT OF THE LABORATORY

The following is a list of the instruments that have been purchased or built since January 1st, 1971.

1. Instrumentation for speech synthesis

(See description p.IXff).

2. General-purpose electronic instrumentation

1 impulse precision sound level meter, Brüel & Kjær, type 2204
1 attenuator set, Hewlett Packard, type 350 D
1 band-pass filter set, Brüel & Kjær, type 1615
1 digital multimeter, Philips, type PM 2422

3. Equipment for EDP

1 tape punch, Facit, type 4060
1 control unit, Facit, type 5106

A FORMANT-CODED SPEECH SYNTHESIZER

Jørgen Rischel and Svend-Erik Lystlund

1. Introduction

The speech synthesizer of the Institute of Phonetics has been completed in all essentials according to the design plan outlined some years ago (see the brief mention in Lystlund and Rischel 1968). It may be expedient, therefore, to give a survey of the apparatus, and the reasoning behind its design.

The synthesizer is a formant-coded device controlled by means of a set of time-varying DC-voltages. At present, these parameters are supplied by a function generator.

2. The synthesizer proper

2.1. Phonetic strategy

The phonetically interesting aspects of the synthesizer layout are shown in Fig. 1. It is seen that there is a generator of quasi-periodic pulses which functions as a voice source, and a generator of random noise. The transfer function of the vocal tract is approximated by a system of resonators all coupled in parallel. Each of these resonators (henceforth referred to as formant filters) takes care of a local section of the spectrum, comprising one formant peak.

The repetition rate of the voice pulses (F_o) can be varied over a wider or narrower range according to the compromise between requirements on intonation range and accuracy of resolution which seems preferable in each case.

The contributions from the voice and noise sources are led

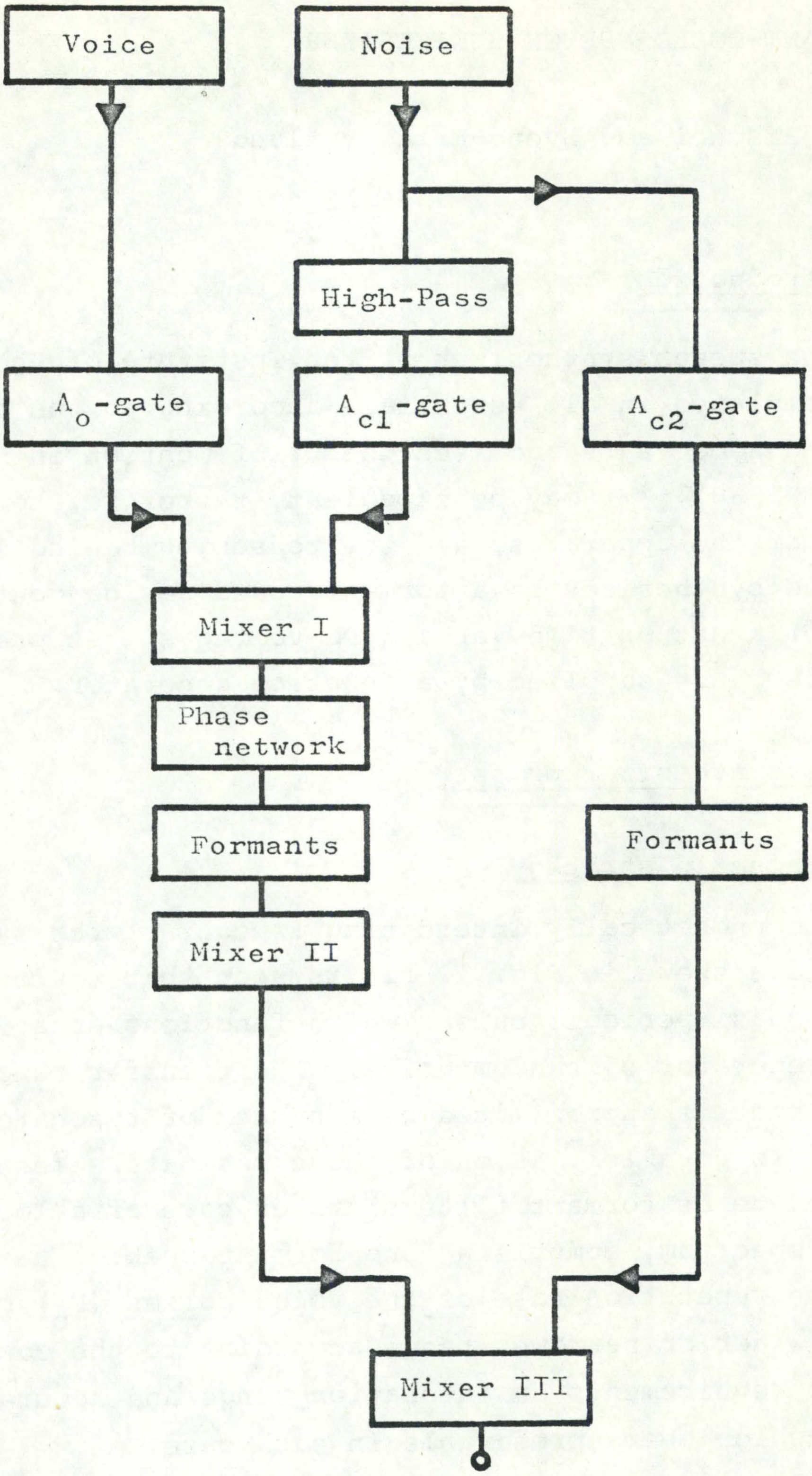

Voice
Noise
High-Pass
Λ_o-gate
Λ_{c1}-gate
Λ_{c2}-gate
Mixer I
Phase network
Formants
Formants
Mixer II
Mixer III

Fig.1

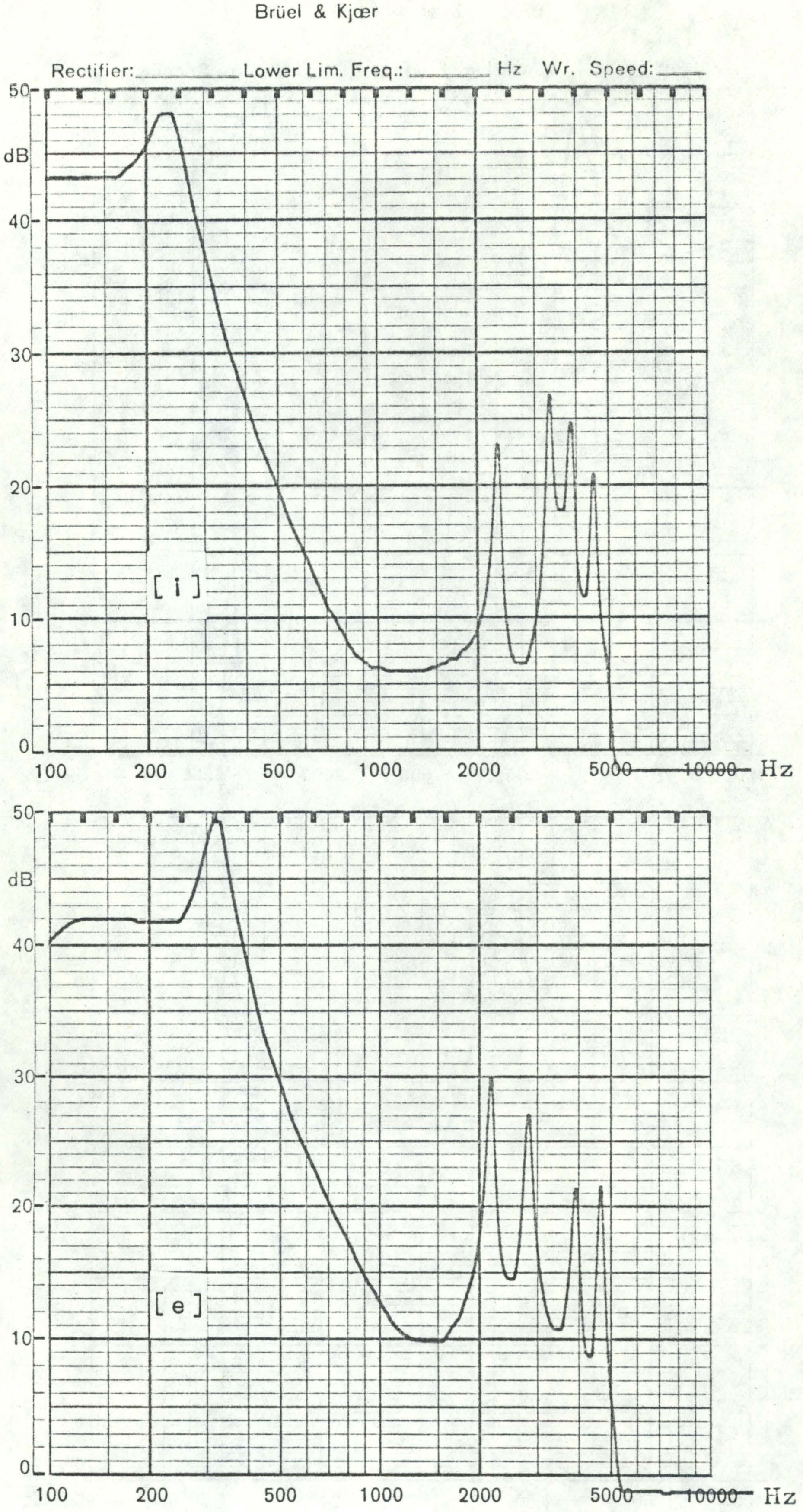
via gates (providing a continuous amplitude control) to each formant filter, and the contributions from these filters are combined by a summation. The individual contribution of each formant filter can be varied continuously (and independently of the other filters) in terms of resonant frequency, bandwidth, and level. The ranges of variation are chosen empirically to be suitable for synthesis of various sound types, but since they can be modified to satisfy special needs there is no point in listing them here.

There is a total of nine formant filters in the system. Five of these are intended to imitate the five lowest resonant modes of the vocal tract, viz. F_1 to F_5. In addition there is a low frequency pole, "F_{sub}" serving to shape the spectral region below F_1, and another pole, "F_{nas}". Both of these can be used to improve the match of natural sounds which have an excess of formant peaks compared to ideal vowels, e.g. vowels with "split" formants or nasal consonants. However, F_{sub} is constantly employed to approximate the low frequency boost of the human voice, which is thus contained conceptually in the transfer function as a low frequency pole.

In addition to the formant filters mentioned so far, there is a separate configuration of two high-frequency formants, nominally F_6 and F_7. In contradistinction to the lower formants, which can be supplied with voice pulses (via the A_o-gate) or noise (via the A_{c1}-gate), or both, the two highest formants can be supplied with noise only, which is introduced via a separate gate common to them (A_{c2}). F_6 and F_7 are coupled in such a way that they form a block of energy in the upper part of the spectrum, the upper and lower limits of this region being determined by the frequency locations of the resonant peaks. The spectral contribution of this subsystem falls off very rapidly below the lower resonance, so that it can be used to give a rough approximation of the high-frequency characteristics of fricatives, whereas the fine structure of the lower frequency region, i.e. roughly below 4500 cps, is represented more faithfully by the lower formants.

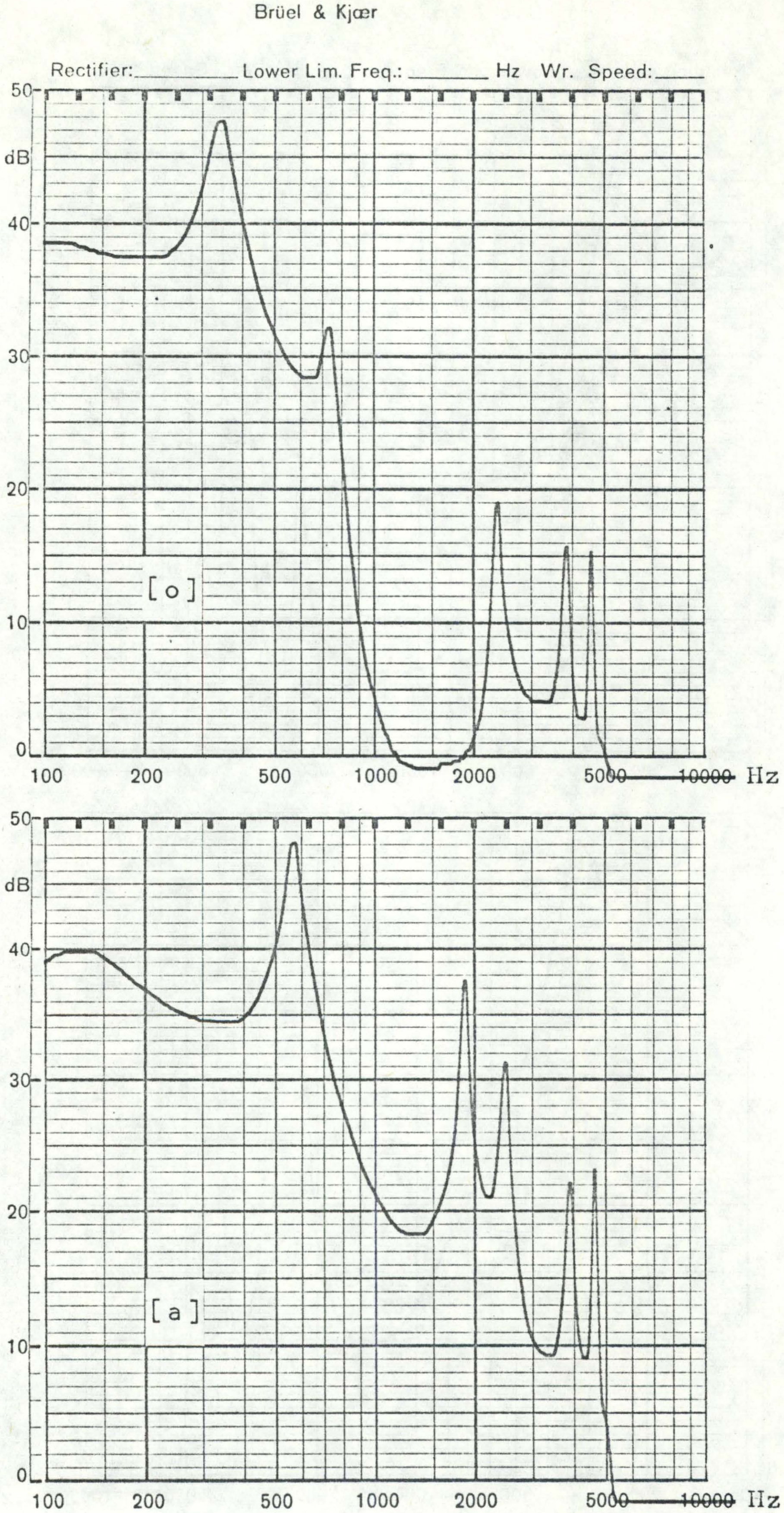
The parallel arrangement of formant filters has an important consequence in vowel synthesis: the residue of a formant at the frequencies of other formants is (almost) negligible on a summation basis, i.e. changes in formant frequencies will not automatically provide the changes of formant levels inherent in human speech. Without such level variations the quality of synthetic speech is quite poor and unsuitable for many research purposes. Recent synthesizers intended for high-quality speech are mostly of the series coupled type (i.e. with a cascade arrangement of the formant filters), which automatically provides the level variations desired. If, however, the amplitude of each formant is controlled, and if the circuits are given appropriate frequency and phase characteristics, vowels of high quality can be produced on a parallel coupled synthesizer, as we have experienced over the last years (also see Fig. 2). For various reasons we have found it worth while to try out the potentialities of such a device for phonetic research purposes.

The predictability of formant levels from formant frequencies should clearly be made use of in a synthesizer designed to produce high quality speech in an <u>economic</u> way. From this point of view the series coupled synthesizer is preferable since the parameters required to operate it contain less redundancy (cf. Fant 1959 p. 47). It is worth noting, however, that this argument applies only with essential reservations. Nasal coupling, leakage through the glottis, and local constrictions in the oral cavity may introduce anti-resonances and additional resonances which considerably influence the relative prominence of different parts of the sound spectrum. This effect can, at least to some extent, be approximated in a parallel connected synthesizer simply by adjusting the formant levels, whereas the



Spectral envelopes of synthetic vowels

Fig. 2a



Spectral envelopes of synthetic vowels

Fig. 2b

series coupled synthesizer is less versatile in this respect (the approaches available with the latter, e.g. a change of formant bandwidths or a modification of the voice source spectrum, are of course equally available with the parallel connected device). Hence, for research on the fine structure of human speech sounds, and its relevance to perception, the parallel coupled synthesizer may be a useful tool.

If we turn now to consonants, a transfer function involving poles only is disputable for liquids and entirely inadequate for fricatives and stops. A series coupled synthesizer must, therefore, be composed of several parallel branches, e.g. one for vowellike sounds and [h], one for nasals, and one for friction noise, as in the Swedish Ove II and Ove III. In the parallel coupled synthesizer one single system can, in principle, be used for the different sound types since formants can emerge or vanish according to the settings of their individual level controls. By strongly attenuating some of the lower formants, for example, one can approximate the transmission characteristics of voiceless fricatives. This means that there is no discontinuity in the formant specification of the spectral properties of adjacent sounds. Thus F_2 of a vowel may (with appropriate bending, and with addition of noise or replacement of voice by noise) continue as F_2 of an adjacent fricative without any switch of parameter, F_3 of a vowel may (with appropriate weakening) continue as F_3 or F_4 of an adjacent nasal, etc. Formants may be suppressed or emerge from nothing, but the formants which are continuous in the sonagram will be controlled consistently by the same set of parameters.

This means that the synthesizer can be operated rather directly from sonagrams by simple spectrum imitation (note that formants which are invisible on sonagrams will be largely negligible in parallel synthesis, since their frequency location does not influence the general shape of the spectrum).

This may be a useful feature in connection with experiments where the perceptual importance of spectral details are at issue.

As a systematic approach to speech synthesis we do, however, employ pre-calculation of sound spectra on the basis of the theory of speech production.

Synthesis by means of a parallel arrangement of formant filters poses a major problem with regard to voiced fricatives. A sound like [z] requires a vowellike formant filtering of the contribution from the voice source (with a strong predominance of F_1, the higher formants being weakened because of the low frequency of F_1), whereas the contribution from the noise source should be characterized by an attenuation of lower formants. This effect of having the voice and noise sources located at different places in the human speech organs cannot be imitated in a straightforward manner. There are two ways to remedy the situation. One is to shape the spectrum of the noise source before it is processed by the formant filters. Another possibility is to apply noise to higher formants only. In our present setup we take both of these measures. The noise applied to the lower formants is filtered in such a way that it is strongly attenuated at very low frequencies, i.e. a low F_1 has practically no noise component in it. Moreover, the highest formants are supplied with noise via a separate gate (the strategy may have to be improved further on this point). Finally, the low frequency boost of the voice source is enhanced by the low-frequency pole (F_{sub}), which is located in a region where the noise is maximally attenuated.

2.2. Electronic design

The synthesizer is based on the heterodyne principle presented in Rischel (1967). The pulses from the voice source are processed by a system of filters and modulators which shifts the useful frequency band from 0-5 kc to 8-13 kc (carrier frequency 8 kc), and back again. The formant filters are inserted into this system at a point where the signal occurs with frequency transposition (as the upper sideband from modulator I). The noise signal is applied directly to the system of formant filters, i.e. frequency transposed only once, whereas the voice signal must be frequency transposed twice in order to ensure a preservation of the harmonic relationship among its components.

Since the formant filters are tuned to much higher frequencies by this technique than they would otherwise be, the relative variation in formant frequencies is drastically reduced. This means that the formant levels vary very little with frequency variation, because the Q, and hence the bandwidth, remains practically constant. This is a necessary prerequisite for a calibration of the scales of the control system (see 3. below). Moverover, we have found it possible and expedient to accomplish the frequency variation by varying the bias voltage applied to capacitance diodes in resonant circuits. This is a very simple method (though it requires a resistance network shaping the control voltage in order to get a reasonable frequency scale). With the high-quality components (particularly inductances) now available, it has been no problem to obtain sufficiently high Q's, i.e. formant bandwidths of the order of 50 cps or less.

The design appears from Fig. 3, which is an elaborated version of the block diagram of Fig. 1 (with inclusion of the heterodyning system). A single formant circuit is shown in Fig. 4.

The formant circuits are preceded by phase inverters (in actual practice the phase correction is more complicated than suggested in the block diagram), and moreover, the circuitry includes extra lowpass filters. The reason for this increased complexity is that a parallel arrangement of formant filters

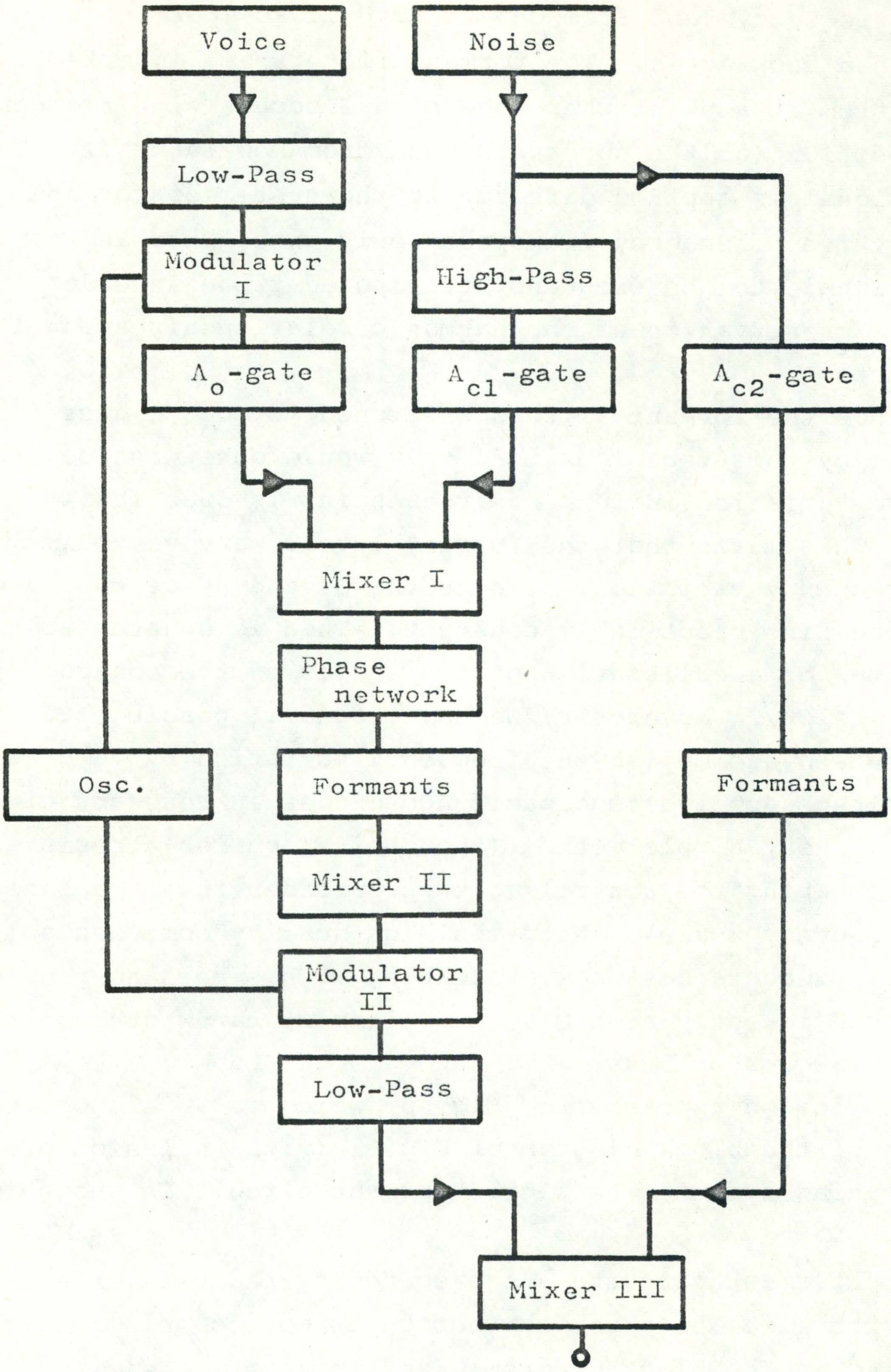
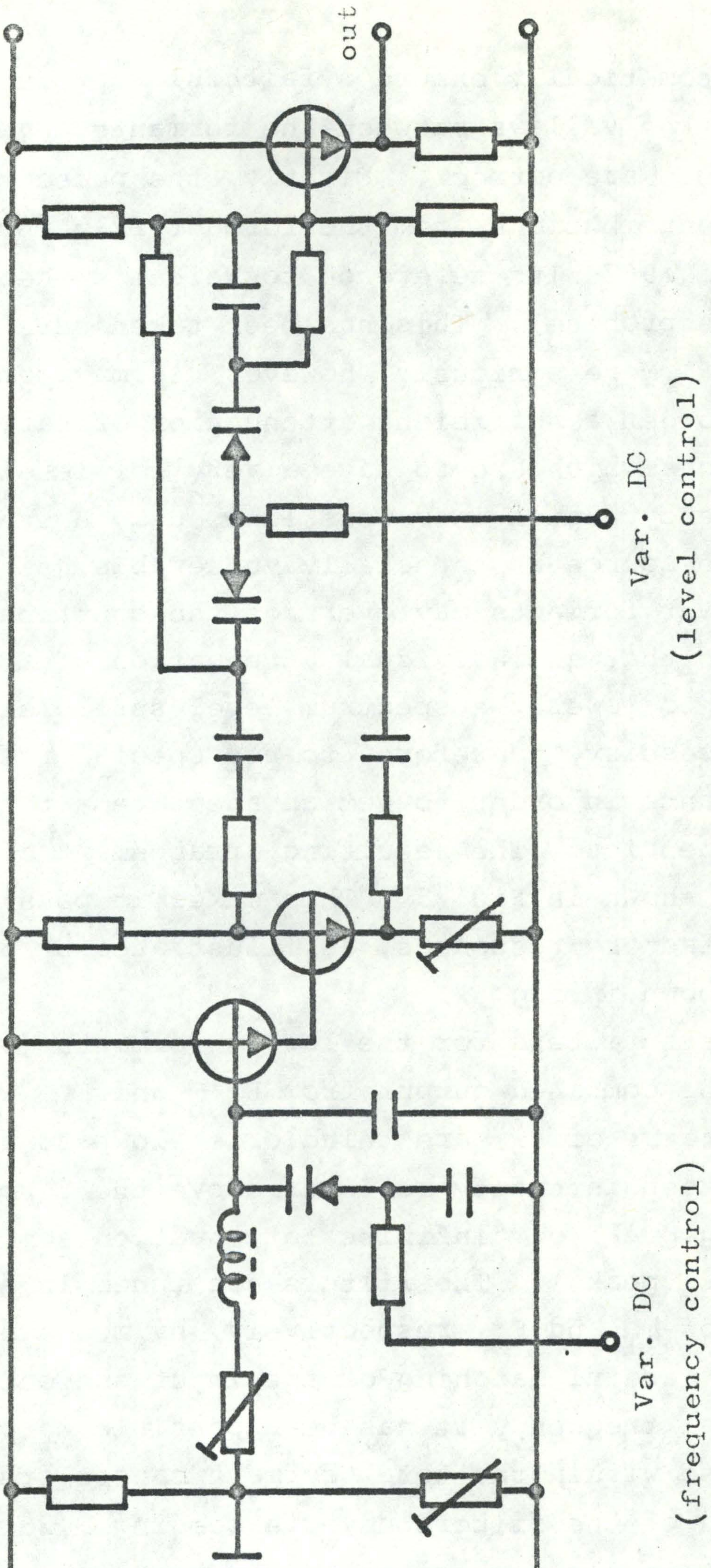

Voice
Noise
Low-Pass
Modulator I
High-Pass
Λ_o-gate
Λ_{c1}-gate
Λ_{c2}-gate
Mixer I
Phase network
Osc.
Formants
Formants
Mixer II
Modulator II
Low-Pass
Mixer III

Fig.3



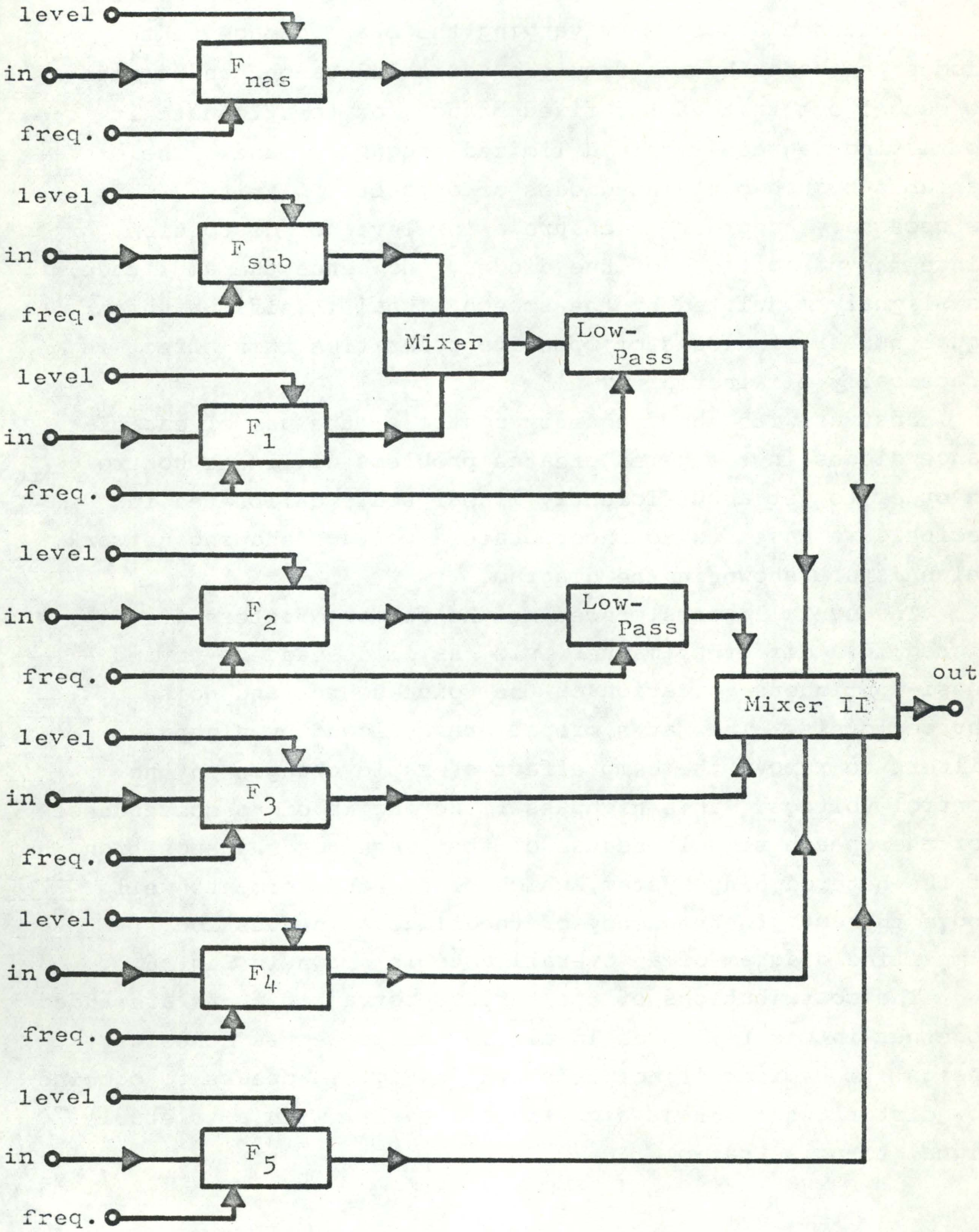
Formant circuit

Fig.4

does not automatically ensure a faithful reproduction of the spectrum in the valleys between the formants, even if the formant levels are correct. Firstly, the phase relationships among the contributions from the formant filters must be correct (see Rischel 1967 with reference to Weibel's theoretical treatment of this problem). This has been taken care of in our synthesizer. More seriously, however, it may sometimes be difficult to get a sufficient attenuation of harmonics in frequency regions which are to have a very low level, e.g. the upper frequency region in vowels like [u]. A synthesizer using a flat voice source is especially vulnerable, since the residues from the lower formants may override the contributions of higher formants, so that an individual attenuation of these latter does not suffice to lower the spectrum level sufficiently. We have found it necessary, therefore, to use special filters for the lowest formants in order to get an adequate attenuation at higher frequencies. The resulting arrangement of the formant circuitry is shown in Fig. 5. This makes it possible to get quite satisfactory responses, as illustrated by the spectrum envelopes shown in Fig. 2.

The extra filters for the lowest formants (one of which processes the combined output from F_{sub} and F_1, whereas the other takes care of F_2) are third order lowpass filters which provide a steep intensity roll-off above the formant peaks (there being peaks of "infinite" attenuation some 1 kc above the F_1 and F_2 peaks). The filters are tuned in accordance with the tuning of F_1 and F_2, respectively, by means of capacitance diodes. By careful matching of the input and output impedances at an optimal frequency it has been possible to obtain a very good response within the total formant ranges (the ripple in the passband of the filters never exceeding 0,25 dB).

Formant levels are controlled by means of continuously variable gates. Each such formant gate consists of a bridge



Arrangement of formants in heterodyne system

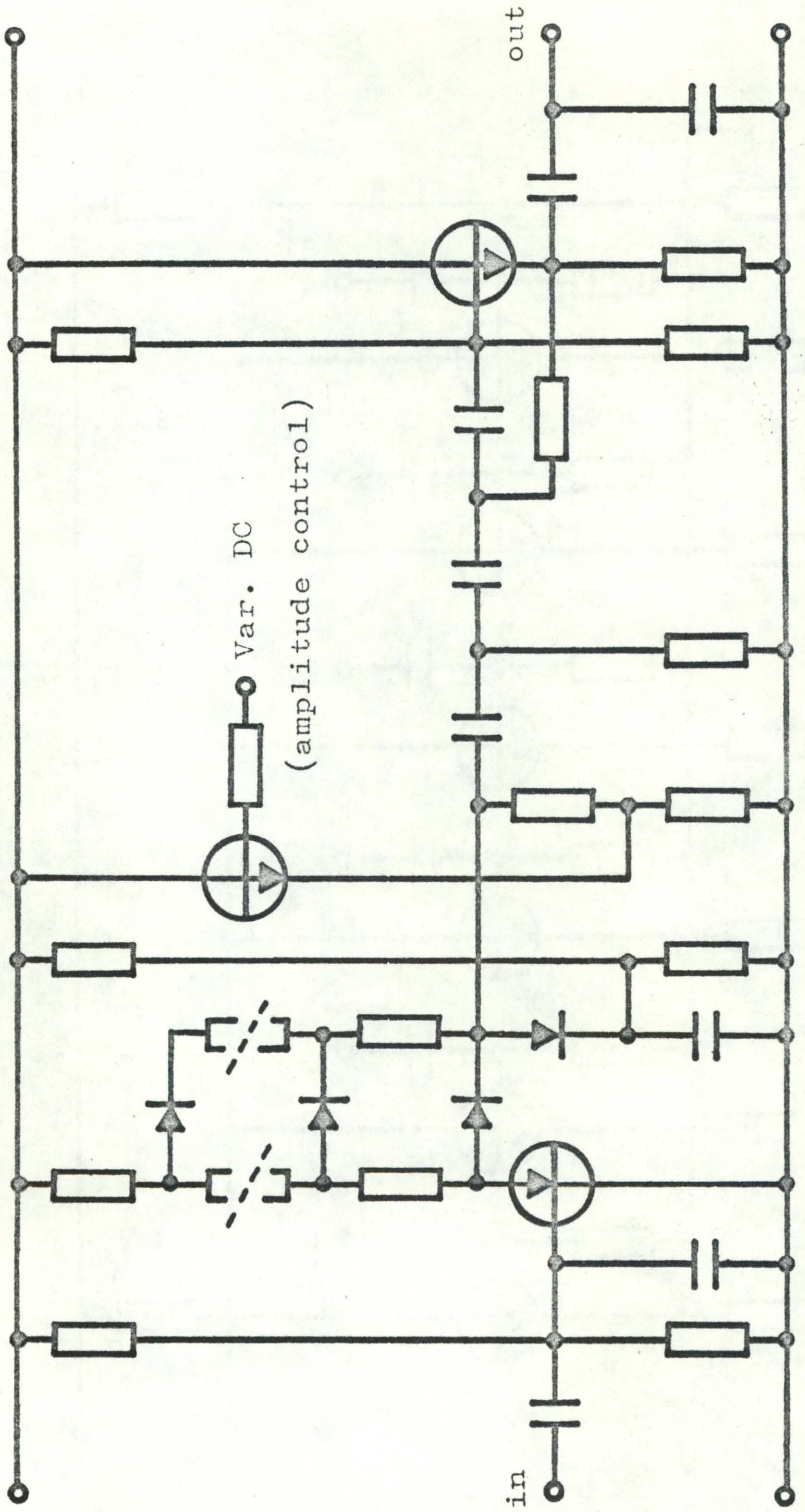
Fig.5

circuit (see Fig. 4) one branch of which is represented by a fixed capacitance whereas the other contains a configuration of capacitance diodes. By varying the bias voltage to the diodes the contribution from the variable branch can be made to cancel out that of the fixed branch, or to attenuate it to varying degrees within a limited frequency band. The variable branch contains diodes of opposite polarity. This is necessary in order to ensure a low level of distortion, since the capacitance of the diodes, and hence the amplitude, is slightly modulated by the speech signal itself. With an equal number of diodes of opposite polarities this effect is practically eliminated.

Just as with the frequency controls, the use of capacitance diodes in the gates creates problems of linear control. In order to get a sufficiently linear scale calibrated in decibels we have had to incorporate a voltage shaping network, which is not shown in the diagram.

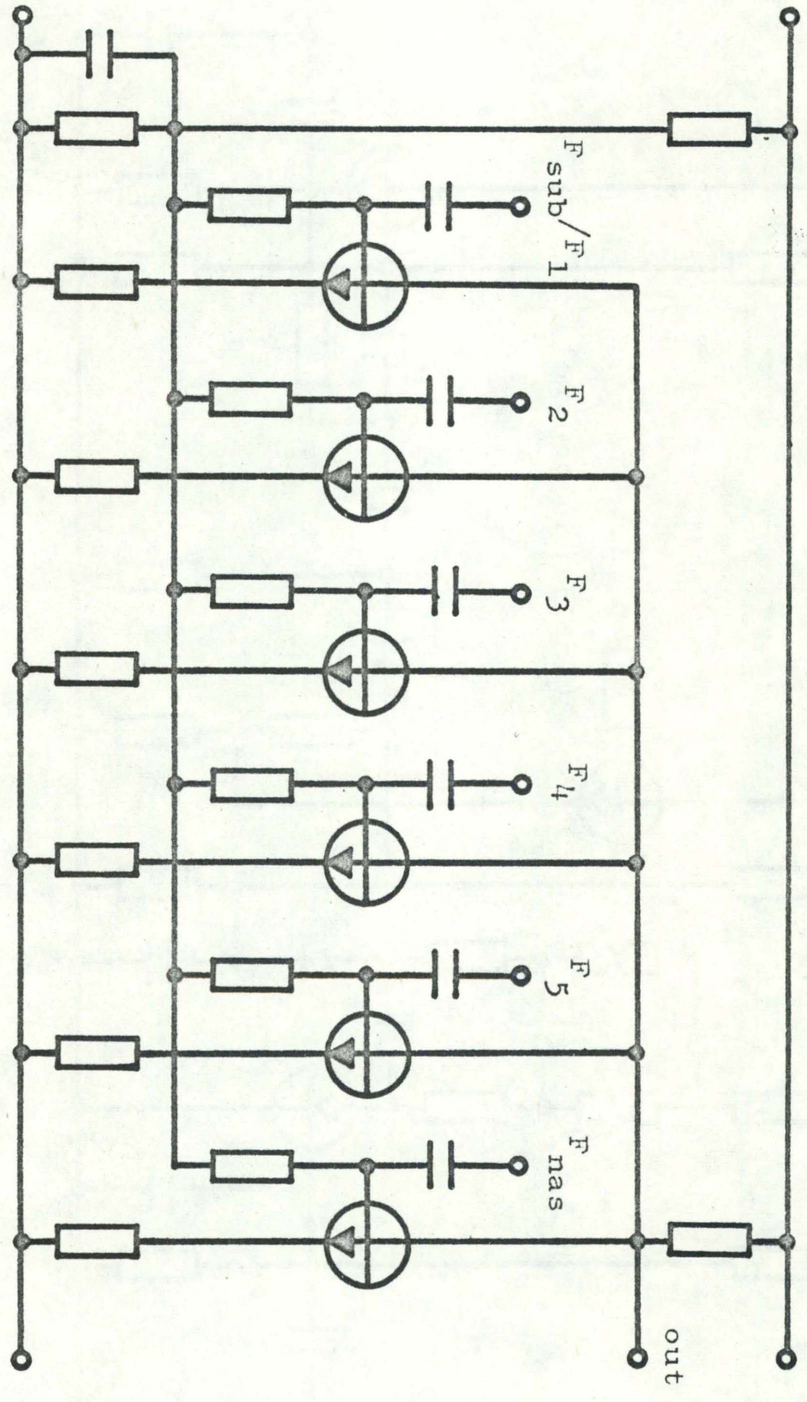
The overall gates preceding the formant filters are designed in a different manner, viz. as diode ladders giving a quasi-continuous variation of the voice source and noise source levels. The gates proper are followed by highpass filters to remove the bump effect of rapid changes in the control voltage. This highpass filtering is of no consequence for the speech signal because of the frequency transposition of the heterodyning system, which places the formants well above the cutoff frequency of the filters in question. - A simplified diagram of an overall gate is shown in Fig. 6.

The contributions of the various formant filters are added together in mixers, three in all in the present synthesizer design. The mixer principle shown in Fig. 7 appears to combine low distortion (even at high signal levels) with a tolerable signal-to-noise ratio.



Amplitude gate

Fig.6



Mixer circuit

Fig.7

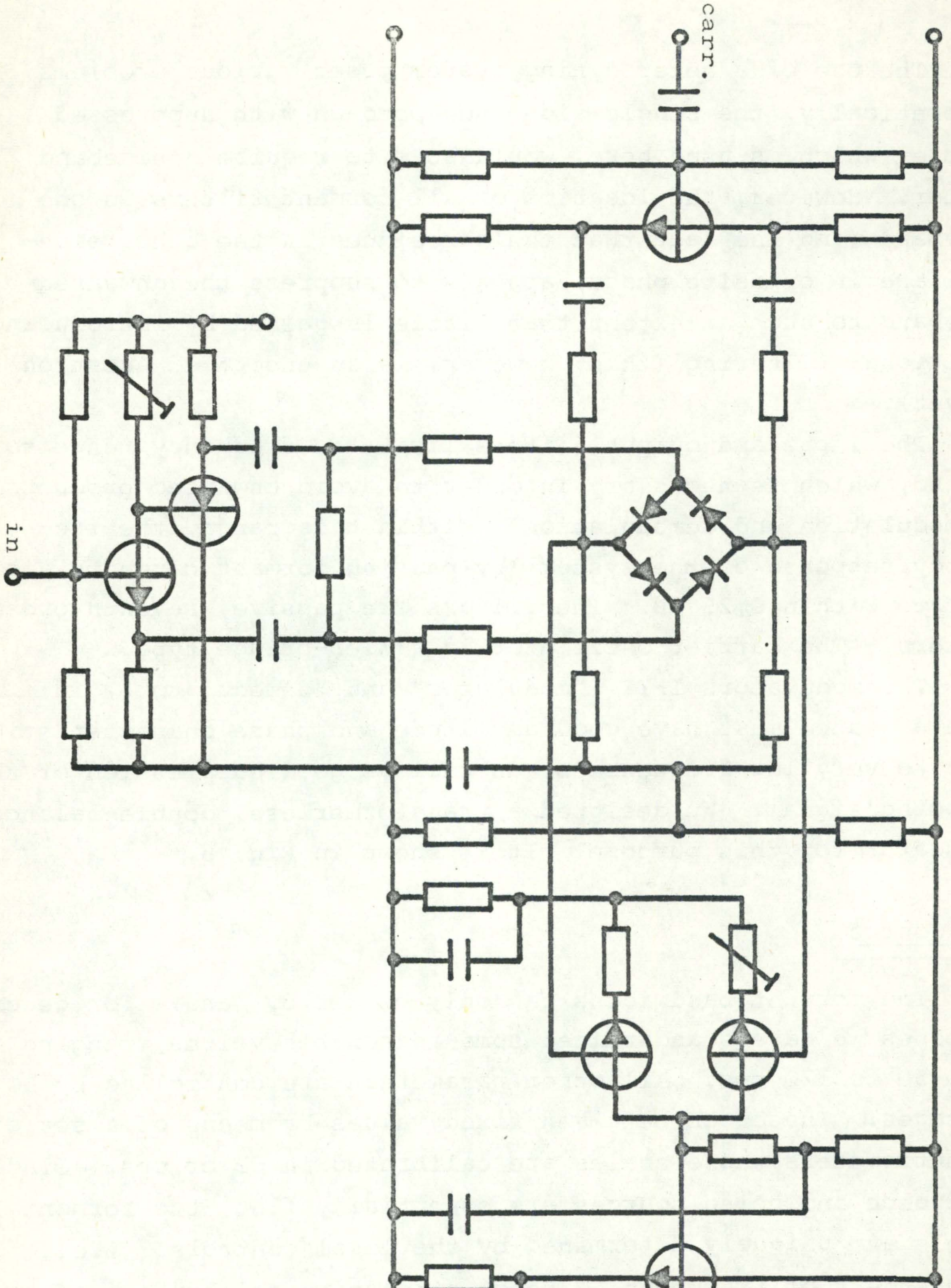
The use of a heterodyning system poses various problems. Theoretically, the single sideband approach with suppressed carrier which is used here, would seem to require a sideband filter. However, the location of all formant filters in one sideband (and the fact that their residues in the other sideband are in opposite phase) appears to suppress the unwanted sideband to such an extent that little is gained by introducing additional filtering (this, however, is an unsettled question as yet).

The input and output filters limit the frequency range to 0-5 kc, which is necessary in order to avoid unwanted products of modulation and demodulation. Within this range, the frequency response of the system (by-passing formant circuits) is flat to within 0,25 dB. The filters are passive, seventh order filters. The carrier oscillator is a Wien-bridge type.

The modulators I-II ("modulator" and "demodulator") in such a system must have good amplitude and phase characteristics down to very low frequencies, as well as good suppression of the incoming signal. We designed a transformerless, double-balanced modulator for this purpose. It is shown in Fig. 8.

3. Control

The formant bandwidths can only be set by hand. For general purposes we have standardized some reasonable values, ranging from 50 to 120 cps. All other parameters are controlled by DC voltages. These can be given fixed values by means of a set of potentiometers whose scales are calibrated in dB or cps. Since the voice and noise sources are essentially flat, the formant levels are uniquely determined by the level controls (i.e. independent of frequency). Hence a relative calibration of each formant level control in dB is possible.



Modulator circuit

Fig.8

There are at present 20 parameters which can be controlled in this way, viz. voice source frequency (F_o) and gate (A_o), noise source gates (A_{c1}, A_{c2}), resonant frequencies of nine formants, and levels of seven formants (F_6 and F_7 sharing the A_{c2} control at present). Although the possibility of controlling all of these parameters independently was introduced in order to make the synthesizer versatile, it is clear that they need not all be varied in order to synthesize some particular stretch of speech. Parameters that are not varied within a stimulus can be given a fixed value, while the others are controlled externally. Moreover, each formant can be switched off separately, and for measuring purposes the voice and noise sources can likewise be disconnected from the system.

A dynamic control for the synthesis of connected speech is obtained by means of a function generator constructed according to the principles outlined in Rischel (1969). Since the said paper gives a rather detailed account of the strategy, a brief presentation may suffice here. In its present shape the function generator supplies 16 time-varying voltages (parameters). Each of these varies in a piecewise linear fashion (though with a slight smoothing). The parameter values are specified in successive steps, 20 in total. For each such step there is a column of potentiometers, one for each parameter. Some parameters are used constantly for specific purposes (e.g. A_o, F_1, etc.); their scales are calibrated in cps or dB, so that the parameter values can be set directly. Others are left open for varying use. For each of the 20 successive steps the duration can be varied in the range 5-100 ms. There is another temporal parameter, viz. transition time, which specifies the duration of a linear transition from the parameter value of the preceding step to the value of the step under consideration. If the transition time coincides with the duration of the step we get a constant rate of change (or invariance if the two steps have similar settings). If, however, the transition time is made shorter, we get a ramp followed by a steady-state portion, the latter increasing in duration as the former is shortened. In this way the duration of transitions can be varied within a synthesized item without affecting its overall duration. (Similarly, the durations of the steps can be increased without affecting the transitions). If two formants have different transition times (e.g. if F_2 moves after F_1 has fulfilled its transition), a faithful reproduction of the transitional stage requires two steps, the one transition being completed in one step, whereas the other takes two.

The stability of the voltage levels produced by the function generator is good. For precision work the settings are controlled by means of external measuring apparatus, the function generator being "locked" to keep the values of each step as long as desired.

The main limitation of the function generator is that there is a limit to the amount of speech that can be synthesized in one sweep. The theoretical maximum is 2 seconds, the practical limit being determined by the precision with which the user wants to reproduce short-time temporal variations. The synthesizer is intended primarily for synthesis of short stimuli which are to be varied systematically with respect to one or several parameters, i.e. vowels, syllables, single words or very short phrases.

In future there will be another option, viz. to control the synthesizer via a computer, which will make it possible to synthesize longer stretches of connected speech.

4. Operation and applications

As mentioned above, the synthesizer can be set according to spectrographic evidence. A useful shortcut can be obtained, however, by having a set of standard sound types with precalculated data. Such a set of numerical data on vowels is being generated by Mr. Peter Holtse.

The synthesizer has been in use for some time for the synthesis of speech-like stimuli of various kinds, as required by researchers inside or outside the Institute of Phonetics. It is presently being used particularly for research on the discrimination of vowels. There has been a set of pilot experiments concerned with voice source characteristics, including dynamic control of the waveshape of the voice source, and shaping of the low frequency region by means of filters. Such experiments will be taken up again in the near future.

Acknowledgements

The speech synthesizer was financed by grants from the State Research Foundation for the Technical Sciences (Statens teknisk-videnskabelige Fond).

References

Fant, Gunnar 1959: Acoustic Analysis and Synthesis of Speech with Applications to Swedish, Ericsson Technics No. 1 (Stockholm).

Lystlund, Svend-Erik and Jørgen Rischel 1968: "Speech synthesizer", ARIPUC 2/1967, p. 34.

Rischel, Jørgen 1967: "Instrumentation for vowel synthesis", ARIPUC 1/1966, p. 15-21.

Rischel, Jørgen 1969: "Constructional work on a function generator for speech synthesis", ARIPUC 3/1968, p. 17-32.

SPECTROGRAPHIC ANALYSIS OF ENGLISH VOWELS[1]

Peter Holtse

1. Introduction

The British English system of stressed vowels is generally agreed to consist of eleven phonologically distinctive elements of relatively pure quality. These eleven monophthongs are traditionally divided into two classes: one comprising five relatively long vowels /i:, a:, o:, u:, ə:/, and one consisting of six short vowels /I, e, æ, ʌ, ɔ, U/.

It was, however, recognised quite early that the phonological opposition between any two vowel phonemes is never dependent on a difference in duration alone. Thus the difference in duration is always accompanied by a certain difference in quality. And it was in fact shown by Gimson (1945-49) that of the two factors: quality and duration, the former is probably the more important for the distinction between pairs of vowels such as /i:/-/I/ and /u:/-/U/ which had otherwise been considered to differ mainly in duration.

The quantitative differences between the English vowels were measured as early as 1903 by E.A. Meyer. Meyer described in details how the exact duration of a given vowel is not only determined by the class to which it belongs (long or short) but also by its quality (open vowels are longer than close vowels), and by the nature of the following consonant (vowels are longer before voiced than before voiceless consonants and longer before e.g. fricatives than before stop consonants).

[1] The contents of the article is part of a thesis work for the degree of cand.phil.

Similar principles were later shown to be operating in American English, first by R-M. S. Heffner in a number of articles and later with improved experimental technique by House and Fairbanks (1953), Peterson and Lehiste (1960), and House (1961). It is true that Heffner found only slight indications of a difference between the so-called "long" and "short" vowels. But Peterson and Lehiste found the vowels to be clearly divided in two groups corresponding to the traditional division. The only exception was the vowel /æ/ which according to Peterson and Lehiste, but contrary to traditional descriptions, should be classified as "long".

The number of investigations dealing with the same problems in British English are surprisingly small. One finds references to various kymographic experiments but no major study before Wiik (1965). On the whole Wiik's results from five informants are in good agreement with the findings of E. A. Meyer except that Wiik classifies /æ/ as neutral with respect to the "long" "short" opposition.

As to the qualitative differences between the vowels American English is again quite well provided with instrumental investigations (e.g. Peterson and Barney (1952), and Fairbanks and Grubb (1961)). The British English vowels, however, have only been described in a short article by Wells (1963) and in the more detailed study by Wiik (1965).

The aim of the present paper is to provide additional analytical data on both the formant frequencies and the duration of British English vowels. This material should further provide a useful basis for a discussion of how quality and quantity may interact as factors distinguishing different vowel phonemes.

2. Procedure

2.1. Material

The material of the investigation consisted of the following list of CVC-words:

[i:]	heat	heed	[ɔ]	pot	pod	cod
[I]	hit	hid	[o:]	port	pawed	cord
[e]	set	head	[U]	put	could	hook
[æ]	hat	had	[u:]	coot	cooed	hoop
[ʌ]	cut	cud	[ə:]	hurt	heard	
[a;]	heart	hard				

This list contains examples of each of the eleven monophthongs before a voiced and a voiceless consonant: the environmental factor which has the greatest influence on the duration of the vowel.

The effect of the initial consonant on the duration of the vowel was considered to be negligible. Therefore only the possible influence from the initial consonants on the vowel formant frequencies was taken into consideration, and the test words were chosen so as to minimize this influence as far as possible.

2.2. Informants

Six male native speakers of Standard English[2] acted as informants. All six of them were judged by competent observers to speak a very close approximation to RP.[2] Three of them were university students, two were university lecturers, and one was a civil servant. One informant was 71 years of age, the remaining five were between the age of 20 and 30.

[2] Cf. Abercrombie (1965) p. 10-16.

2.3. Recordings

The test words were arranged in six randomized lists of isolated words which were read aloud by each of the informants. Two of these read two of the lists twice. After mispronounced words had been discarded the number of recorded words totalled 906.

The recordings of two informants were made in the recording studio of the Department of Linguistics, University of Edinburgh, using a Revox tape recorder and EMI Recording Tape. The recordings of the other four informants were made in the studio of the Institute of Phonetics, University of Copenhagen, using a Lyrec Professional Recorder and Scotch Magnetic Tape.

The recordings were analyzed on a Kay-Electric Sona-Graph. One wide band, one narrow band, and at least one cross section spectrogram were made of each of the recorded words. At least once every time the sound spectrograph had been used a calibration was made using a 100 Hz and a 500 Hz tone.

3. Formant Frequencies

3.1. Formant frequency measurements

The vowel formants were identified as the regions of relatively high intensity and the centre frequencies of these regions were measured. Wherever possible four formants were measured. But as could be expected, the back vowels proved rather difficult in this respect, and in a number of cases only formants one and two were in fact visible on the spectrograms.

The formant frequencies were measured with an accuracy of approximately 25 Hz. In view of the importance of very small differences in the low frequency regions a not altogether successful attempt was made to raise the accuracy of the F1-measurements to 10-15 Hz. This was done by using the individual harmonics of the narrow band spectrogram as a scale after the fundamental frequency at the point had been determined.

In all cases the formants were measured in the middle of a period of relatively steady state.

The results were fed into a computer and run through a standard program (XFON) which for each of the eleven vowels calculated the arithmetic mean, standard deviation, standard error of mean, and 95 and 99 pct. confidence limits.[3]

3.2. Vowel diagrams

A general idea of the auditory quality of a vowel is provided by showing the two most important formants, F1 and F2, in a two dimensional diagram. This is actually a very crude representation, and we might wish to improve it by somehow taking at least F3 into account as well. In the present study this has been done in a few cases in the form of a double-diagram with F1 as a common vertical axis to the two horizontal axes F3 and F2 (see Figs. 2 and 3).

The other possible solution: computing a weighted average of F2 and F3, seemed less attractive since we have at the time being no satisfactory way of taking both formant frequency and formant intensity into account.

In all the diagrams the frequencies of the formants have been given in Mels since this scale provides the best approximation so far to the way frequency is perceived by the ear. It should, however, be borne in mind that the Mel scale has been

[3] The computer programs were written by cand.scient. J. E. Knudsen whose help is gratefully acknowledged.

found from experiments with simple sounds or narrow band noise and we only assume that it is applicable to complex sounds as well.

The vowel diagrams show the mean formant frequencies marked with a cross. The dispersion of the individual measurements of a given vowel is indicated by an ellipse drawn through four points each lying at a distance of two standard deviations from the mean, and measured along the F1- and F2-dimensions. Fig. 1 shows an example of the dispersion of a set of individual measurements within their 2s-ellipses.

Since the scale of the diagrams was drawn in Mels the standard deviations used to construct the 2s-ellipses had to be expressed in Mels as well. Otherwise the ellipses would have been skewed. Therefore all the individual measurements were converted from Hz into the corresponding Mel-values and run through the standard XFON-program once more. These calculations were used in drawing the vowel diagrams.

Traditionally the so-called "phoneme area" of a given vowel has been represented in the vowel diagram simply by a line drawn round all the individual measurements. It has, however, always been a problem how far one or two extreme values could be excluded from the envelope. Using the 2s-ellipses appears to be a satisfactory way of solving this problem. On the other hand the effect on the average of extreme values is still quite marked as far as this is calculated as the arithmetic mean. And it has been suggested to me by Eli Fischer-Jørgensen that a possible solution to this problem might be to use either the geometric mean or the median and quartiles as a measure of central tendency instead of the arithmetic mean.

One serious drawback in the use of 2s-ellipses is that any tendency to correlation between the two formants is completely obscured since it will only appear as an increase of dispersion. Obviously this difficulty ought to be taken

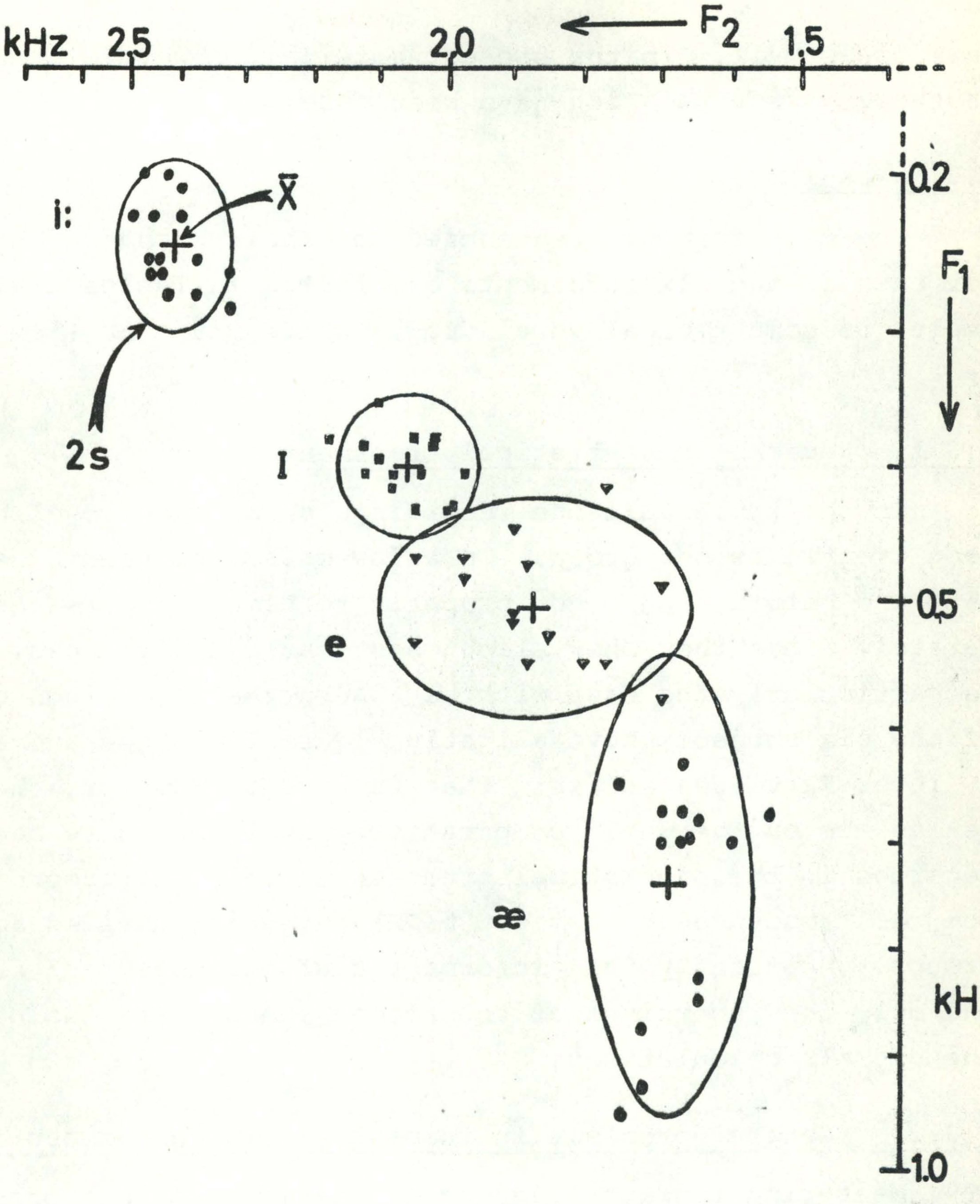


Fig. 1. Example of the variation of individual measurements. The axes of the ellipses correspond to a distance of two standard deviations from the mean.

care of as well. In the present material, however, no such instances of correlation have been observed.

3.3. Results

The mean formant frequencies and their standard deviations for each of the six informants are listed in Tables I-VI. Examples of some typical vowel diagrams are given in Figs. 2 and 3.

3.3.1. Remarks on statistical treatment

In the tables all the recordings of a given vowel have been treated as one group. In a few cases the vowels before [-t] and before [-d] ought properly to have been treated separately since they show slight systematic differences. This is particularly the case with [e] where the recordings of four of the six subjects have slightly higher F1 values and slightly lower F2 values in [set] than in [hed]. However, the differences are on the whole comparatively small and have been disregarded in the statistical treatment. Only the recordings of [o:] as pronounced by subject 5(DH) have been divided into two groups. Apparently this informant distinguishes /ɔe/ from /o :/, and only the recordings of the second phoneme are included in the overall calculations.

3.3.2. General problems in formant finding and measuring

Measuring formant frequencies presents various practical problems. Ladefoged (1962) describes the general ones like e.g. the difficulty in measuring very low first formants or the problems caused by a high fundamental frequency etc. Another kind of difficulty, also mentioned by Ladefoged, is the appearance of so-called spurious formants at frequencies where no formants should be or the possible absence of normal formants at other frequencies.

The present material shows many examples of these phenomena. Thus subject 1(AW) consistently has what looks like two first formants in the vowels [æ] and [ʌ], one at about 900 Hz and another at 700 Hz. And in the same vowels subject 2(GG) has an extra F2: Besides the normal F2 at 1700 Hz in [æ] there is an extra formant at 13-1400 Hz. Similarly in [ʌ], where F2 is normally found at 1250 Hz, an extra formant appears at 1700 Hz. The same tendency although less pronounced is found in the vowels of the other four informants.

In view of the consistency with which these spurious formants appear, it seems doubtful if they should be completely disregarded. If, as appears to be the case, vowel quality is perceived by some sort of weighting of frequency areas rather than by identification of specific formant frequencies, these very strong extra formants must certainly have som influence on the auditory quality of the vowels. Therefore the traditional vowel diagrams, which take into account only the "official" formants, may give rather a distorted impression of the auditory distances between the vowels. In the present case [æ] and [ʌ] may in fact be perceived as much closer in quality than is indicated by the vowel diagrams.

Another example of the same problem is found in [ʌ] as pronounced by informant 6(RD). In all his recordings of [ʌ] before [-d] an extra formant is found at about 500 Hz. If this formant has any effect it must give the vowel a much more centralised quality than we would expect from the normal diagrams.

Furthermore, the apparently "split formants" provide some interesting problems. Many recordings of subject 4(DC) have two distinct formants in the F3 area. Both of them take their origin from the same frequency position and are clearly distinct from F4 and F2. In other cases, especially among the back vowels, F3 is readily identified but a rather strong extra formant may then appear at about 1500 Hz.

TABLE I

Results of formant frequency measurements. Informant 1(AW). The table shows mean values ($\overline{X}$), standard deviation (s), and number of tokens (N) for each of the first four formants.

		F1	F2	F3	F4	N
[i:]	$\overline{X}$:	243	2419	2953	3666	16
	s:	(26.8)	(49.6)	(61.8)	(52.3)	
[I]	$\overline{X}$:	394	2063	2745	3783	15
	s:	(21.9)	(55.0)	(48.3)	(105.5)	
[e]	$\overline{X}$:	503	1873	2712	3975	15
	s:	(47.5)	(110.4)	(50.8)	(126.8)	
[æ]	$\overline{X}$:	740	1688	2542	3767	16
	s:	(107.1)	(59.8)	(69.9)	(66.9)	
[ʌ]	$\overline{X}$:	646	1234	2409	3545	16
	s:	(33.4)	(32.7)	(54.7)	(70.8)	
[a:]	$\overline{X}$:	654	1070	2480	3506	16
	s:	(31.6)	(64.7)	(95.4)	(84.9)	
[ɔ]	$\overline{X}$:	574	927	2520	3457	16[4]
	s:	(45.7)	(34.7)	(50.2)	(96.1)	
[o:]	$\overline{X}$:	450	788	2368	3183	15[5]
	s:	(23.3)	(62.6)	(87.9)	(78.6)	
[U]	$\overline{X}$:	432	1042	2281	3316	16
	s:	(34.4)	(84.5)	(60.2)	(80.0)	
[u:]	$\overline{X}$:	285	1131	2312	3383	16[6]
	s:	(26.7)	(68.0)	(56.6)	(50.6)	
[ə:]	$\overline{X}$:	503	1464	2539	3667	16
	s:	(39.9)	(38.7)	(37.6)	(48.9)	

[4] F4 only 15 ex.

[5] F3 only 14 and F4 only 12 ex.

[6] F3 and F4 only 15 ex.

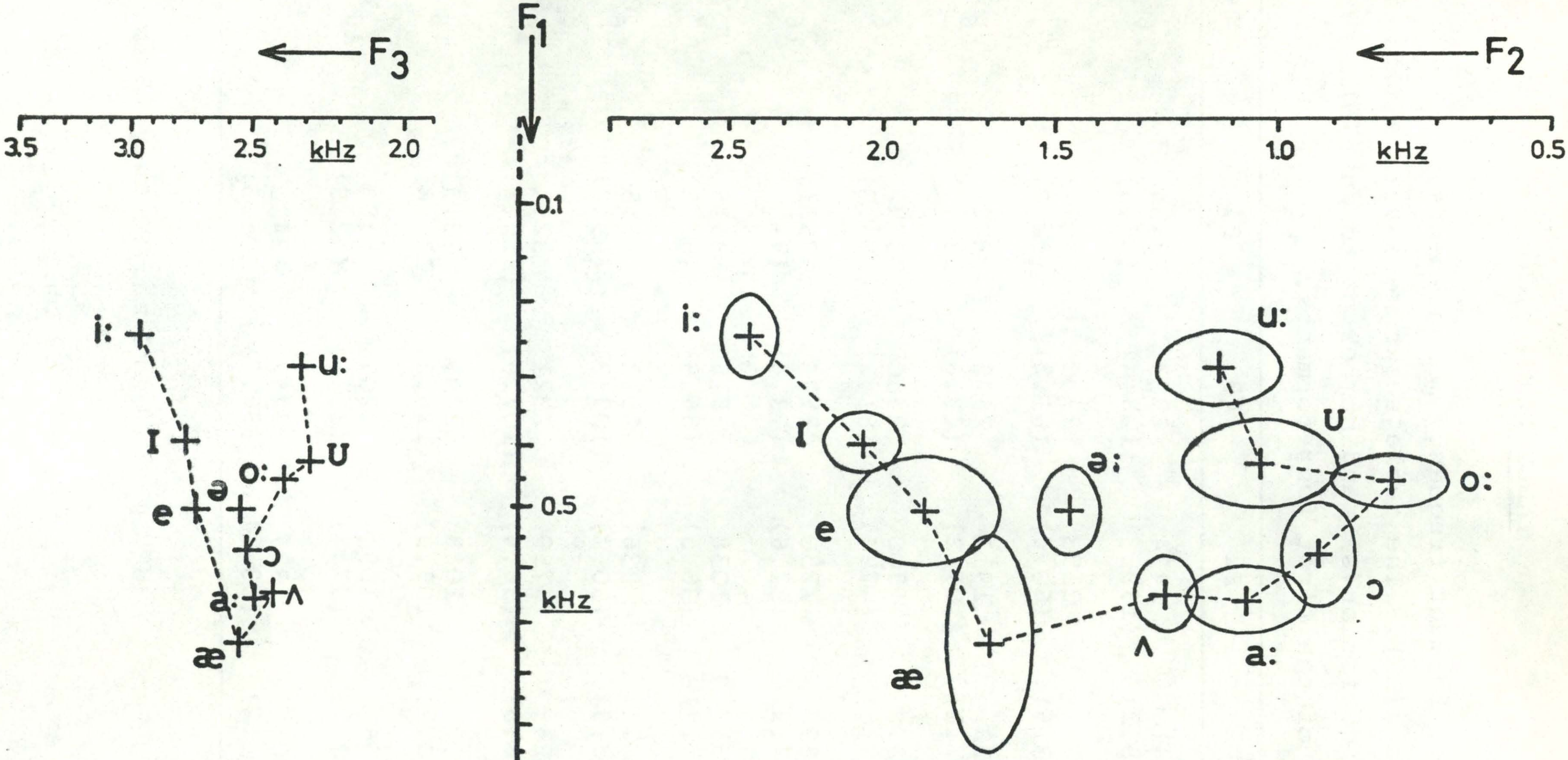


Fig. 2. Vowel diagram Informant 1(AW)

TABLE II

Results of formant frequency measurements. Informant 2(GG). The table shows mean values (X̄), standard deviation(s), and number of tokens (N) for each of the first four formants.

			F1	F2	F3	F4	N
[i:]	-	X̄:	241	2434	2933	3534	16
		s:	(18.2)	(139.3)	(128.4)	(39.6)	
[I]	-	X̄:	377	2153	2700	3697	16
		s:	(23.5)	(60.4)	(61.2)	(82.0)	
[e]	-	X̄:	532	1911	2622	3764	16
		s:	(47.1)	(107.2)	(43.6)	(134.5)	
[æ]	-	X̄:	887	1700	2680	3792	16
		s:	(35.0)	(47.4)	(81.2)	(106.3)	
[ʌ]	-	X̄:	789	1250	2533	3531	16
		s:	(67.9)	(31.6)	(73.4)	(91.5)	
[a:]	-	X̄:	761	1038	2616	3515	16[7]
		s:	(73.7)	(36.5)	(56.9)	(87.5)	
[ɔ]	-	X̄:	617	936	2513	3515	16[8]
		s:	(55.6)	(30.2)	(109.3)	(69.3)	
[o:]	-	X̄:	424	759	2321	3254	16[9]
			(28.8)	(36.4)	(107.6)	(119.2)	
[U]	-	X̄:	402	1088	2299	3313	16
		s:	(20.7)	(109.9)	(116.2)	(124.8)	
[u:]	-	X̄:	256	1091	2207	3483	16[10]
		s:	(17.7)	(77.4)	(85.8)	(42.5)	
[ə:]	-	X̄:	508	1377	2452	3420	16
		s:	(31.2)	(23.2)	(86.3)	(101.3)	

[7] F4 only 15 ex.

[8] F3 and F4 only 15 ex.

[9] F3 only 12 ex.,
F4 only 14 ex.

[10] F3 only 15 ex.

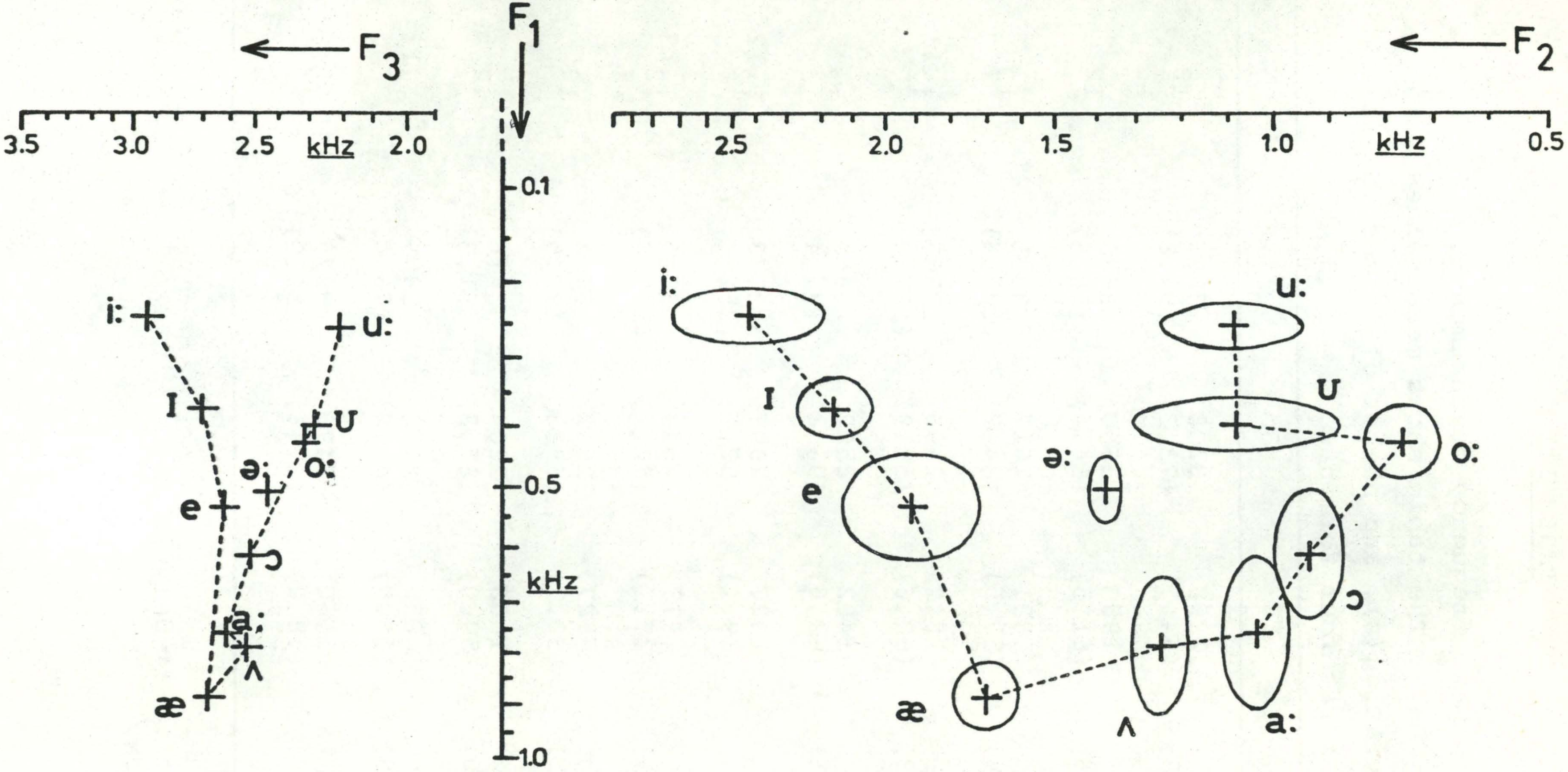


Fig. 3. Vowel diagram Informant 2(GG)

TABLE III

Results of formant frequency measurements. Informant 3(CLB). The table shows mean values ($\bar{X}$), standard deviation(s), and number of tokens (N) for each of the first four formants.

		F1	F2	F3	F4	N
[i:]	- $\bar{X}$:	238	2258	3208	3652	12
	s:	(26.8)	(54.7)	(73.3)	(94.8)	
[ɪ]	- $\bar{X}$:	368	1983	2650	3690	12
	s:	(27.6)	(61.5)	(54.4)	(58.8)	
[e]	- $\bar{X}$:	493	1881	2617	3835	12
	s:	(29.4)	(57.5)	(34.2)	(185.4)	
[æ]	- $\bar{X}$:	648	1842	2490	3807	12[11]
	s:	(46.3)	(62.6)	(55.9)	(137.4)	
[ʌ]	- $\bar{X}$:	727	1402	2502	3865	12
	s:	(62.1)	(62.6)	(96.8)	(97.4)	
[a:]	- $\bar{X}$:	672	1027	2607	3573	12[12]
	s:	(62.7)	(29.1)	(152.9)	(252.1)	
[ɔ]	- $\bar{X}$:	624	977	2573	3388	12[13]
	s:	(33.9)	(27.1)	(82.2)	(195.5)	
[o:]	- $\bar{X}$:	489	727	2642	3458	12
	s:	(36.5)	(60.7)	(69.3)	(92.5)	
[U]	- $\bar{X}$:	411	1000	2204	3416	12
	s:	(35.0)	(50.0)	(85.8)	(61.3)	
[u:]	- $\bar{X}$:	275	835	2158	3409	12[11]
	s:	(30.5)	(75.0)	(49.2)	(76.0)	
[ə:]	- $\bar{X}$	548	1360	2473	3579	12
	s:	(21.8)	(48.2)	(37.6)	(152.9)	

[11] F4 only 11 ex.

[12] F3 only 11 ex., F4 only 10 ex.

[13] F4 only 10 ex.

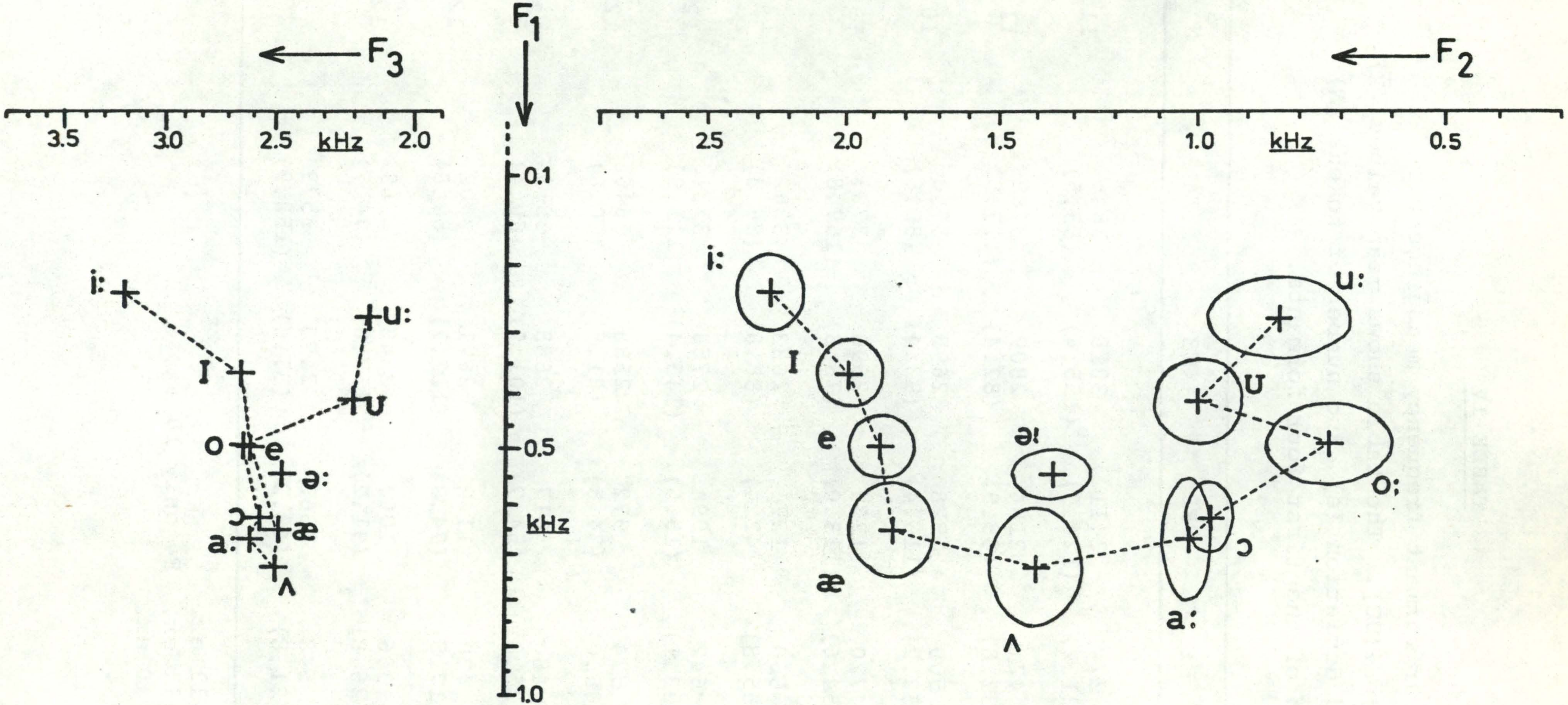


Fig. 4. Vowel diagram Informant 3(CLB)

TABLE IV

Results of formant frequency measurements. Informant 4(DC). The table shows mean values ($\overline{X}$), standard deviation (s), and number of tokens (N) for each of the first four formants.

		F1	F2	F3	F4	N
[i:] -	$\overline{X}$	274	2411	3218	3846	11
	s:	(21.2)	(124.7)	(155.4)	(83.5)	
[ɪ] -	$\overline{X}$	421	2184	2809	3789	11
	s:	(33.8)	(88.9)	(82.4)	(117.5)	
[e] -	$\overline{X}$:	500	1928	2695	3758	10
	s:	(32.2)	(116.4)	(93.4)	(80.8)	
[æ] -	$\overline{X}$:	720	1739	2590	3729	13[14]
	s:	(54.6)	(113.0)	(78.8)	(169.8)	
[ʌ] -	$\overline{X}$:	660	1271	2633	3802	12
	s:	(55.8)	(42.4)	(88.8)	(88.8)	
[a:] -	$\overline{X}$:	642	1094	2756	3731	12[15]
	s:	(33.1)	(15.5)	(145.4)	(111.9)	
[ɔ] -	$\overline{X}$:	574	972	2588	3646	12[16]
	s:	(45.0)	(34.5)	(81.5)	(70.6)	
[o:] -	$\overline{X}$:	480	817	2635	3519	12
	s:	(25.8)	(54.7)	(101.9)	(60.4)	
[U] -	$\overline{X}$:	470	1142	2671	3602	12
	s:	(17.5)	(74.9)	(113.2)	(80.8)	
[u:] -	$\overline{X}$:	306	1094	2248	3269	12[14]
	s:	(26.8)	(91.2)	(57.8)	(74.7)	
[ə:] -	$\overline{X}$:	548	1360	2473	3579	12
	s:	(21.8)	(48.2)	(37.6)	(152.9)	

14) F4 only 11 ex.
15) F3 only 11 ex., F4 only 10 ex.
16) F4 only 10 ex.

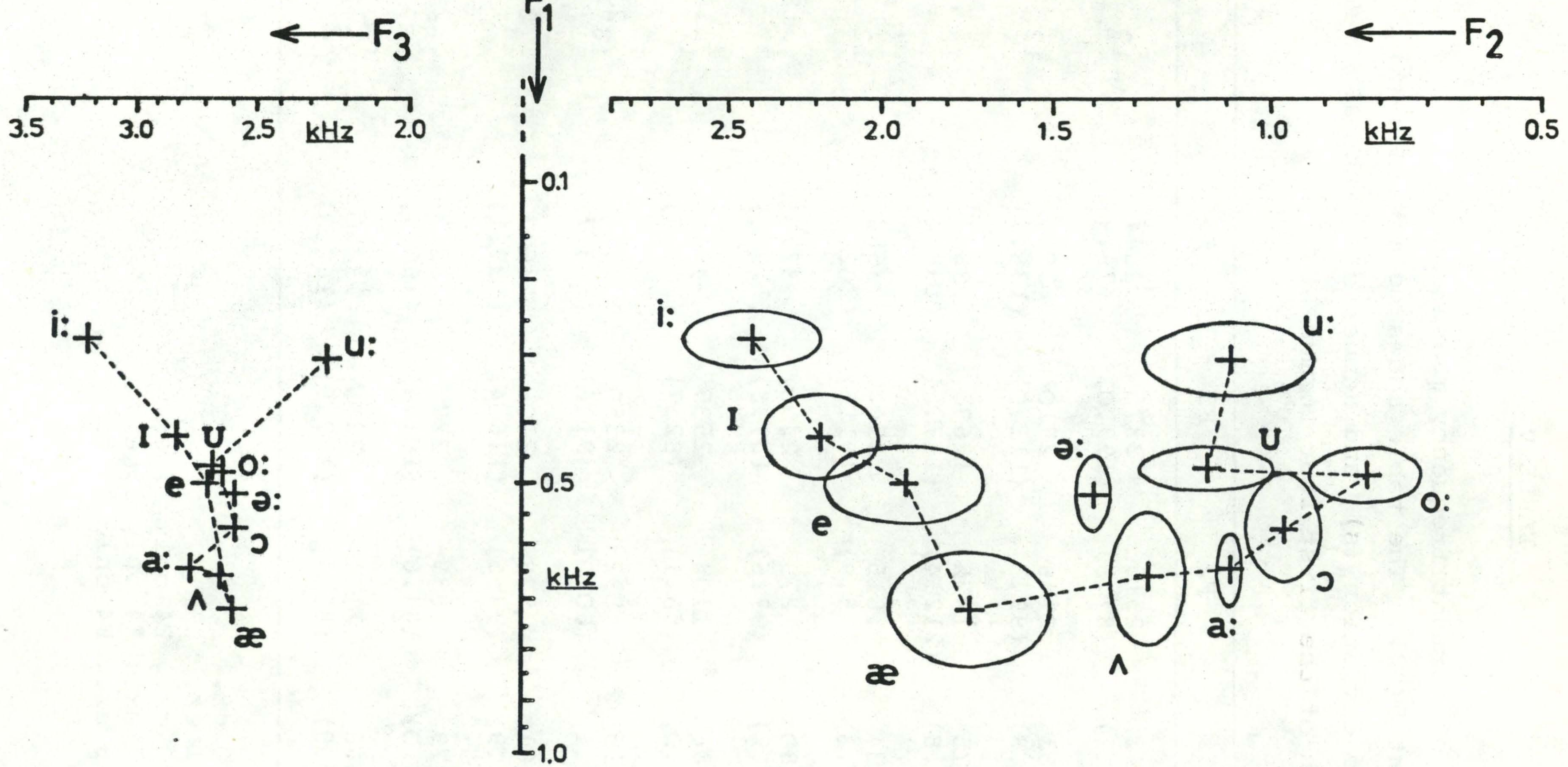


Fig. 5. Vowel diagram Informant 4(DC)

TABLE V

Results of formant frequency measurements. Informant 5(DH). The table shows mean values ($\bar{X}$), standard deviation (s), and number of tokens (N) for each of the first four formants.

		F1	F2	F3	F4	N
[i:] -	$\bar{X}$:	243	2515	3373	3844	12
	s:	(17.1)	(87.5)	(199.0)	(79.9)	
[ɪ] -	$\bar{X}$:	351	2146	2710	3815	12
	s:	(52.8)	(99.3)	(111.0)	(138.4)	
[e] -	$\bar{X}$:	475	1975	2635	3804	12
	s:	(39.8)	(212.4)	(147.5)	(138.4)	
[æ] -	$\bar{X}$:	634	1815	2542	3688	12[17]
	s:	(38.3)	(55.9)	(91.3)	(202.5)	
[ʌ] -	$\bar{X}$:	680	1279	2450	3723	12[18]
	s:	(74.2)	(97.6)	(157.8)	(179.0)	
[a:] -	$\bar{X}$:	686	1104	2535	3741	12[19]
	s:	(88.0)	(53.1)	(82.7)	(200.7)	
[ɔ] -	$\bar{X}$:	615	892	2350	3815	18[20]
	s:	(48.4)	(60.0)	(91.4)	(113.7)	
[o:] -	$\bar{X}$:	361	538	2306	3607	18[21]
	s:	(75.9)	(94.5)	(116.4)	(137.1)	
[ʊ] -	$\bar{X}$:	379	889	2227	3713	18[21]
	s:	(35.0)	(134.8)	(101.4)	(166.9)	
[u:] -	$\bar{X}$:	263	1101	2235	3750	18[19]
	s:	(16.9)	(83.4)	(102.9)	(162.8)	

[17] F4 only 10 ex.
[18] F4 only 11 ex.
[19] F3 only 10 ex., F4 only 11 ex.
[20] F3 only 17 ex., F4 only 15 ex.
[21] F3 only 16 ex., F4 only 14 ex.

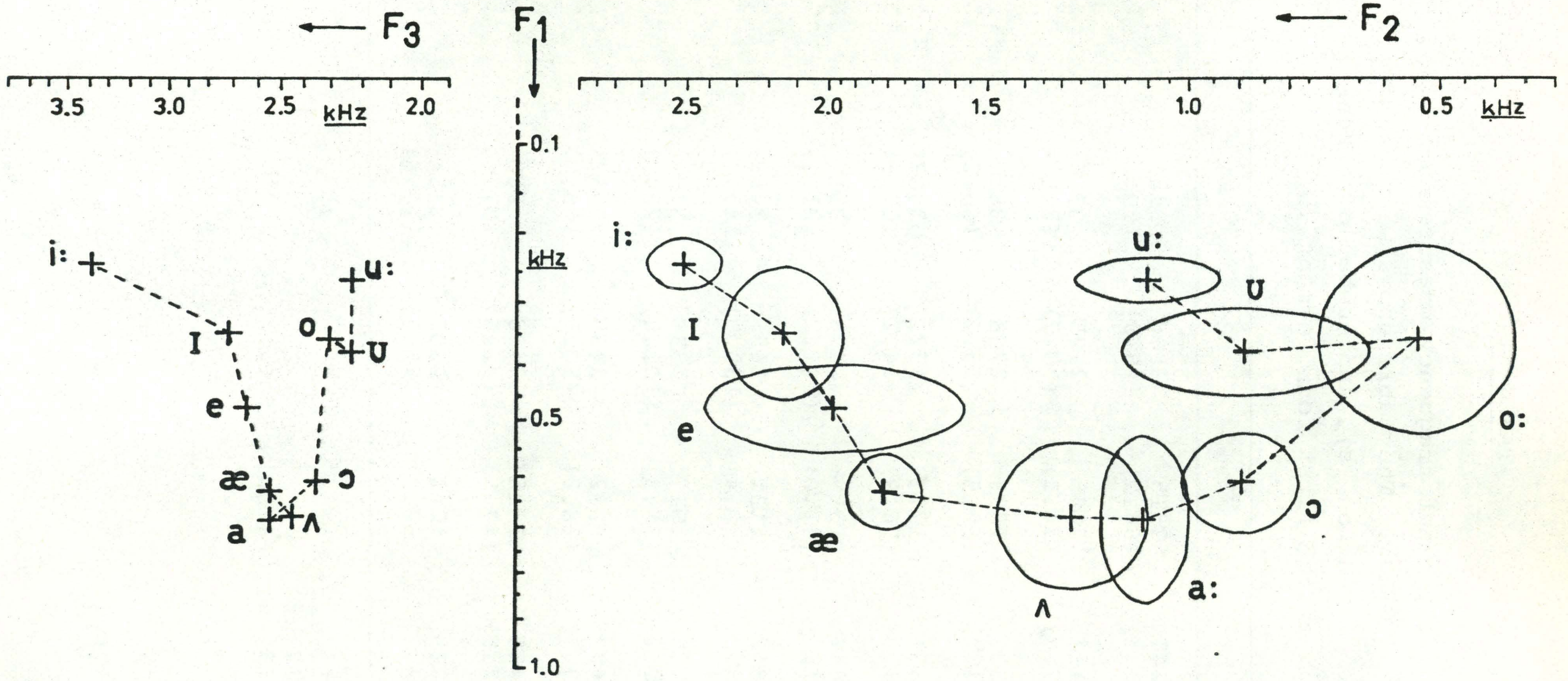


Fig. 6. Vowel diagram Informant 5 (DH)

TABLE VI

Results of formant frequency measurements. Informant 6(RD). The table shows mean values ($\bar{X}$), standard deviation (s), and number of tokens (N) for each of the first four formants.

		F1	F2	F3	F4	N
[i:] -	$\bar{X}$:	245	2515	2808	3988	12
	s:	(11.1)	(66.9)	(46.9)	(90.1)	
[ɪ] -	$\bar{X}$:	414	2085	2725	3706	12
	s:	(37.4)	(74.2)	(70.7)	(70.0)	
[e] -	$\bar{X}$:	526	1898	2708	3810	12
	s:	(24.7)	(69.5)	(56.7)	(65.2)	
[æ] -	$\bar{X}$:	831	1773	2721	3802	12[22]
	s:	(66.8)	(56.9)	(54.2)	(67.5)	
[ʌ] -	$\bar{X}$:	842	1329	2498	3514	12[22]
	s:	(51.0)	(41.0)	(141.2)	(121.1)	
[a:]	$\bar{X}$:	840	1131	2456	3468	12[22]
	s:	(65.3)	(30.4)	(139.0)	(145.4)	
[ɔ] -	$\bar{X}$:	565	1017	2333	3407	18[23]
	s:	(20.4)	(38.3)	(101.8)	(128.0)	
[o:] -	$\bar{X}$:	489	892	2435	3552	12
	s:	(19.4)	(28.9)	(50.5)	(95.0)	
[ʊ] -	$\bar{X}$:	438	1018	2438	3280	18[24]
	s:	(32.5)	(47.6)	(92.6)	(107.4)	
[u:]	$\bar{X}$:	269	1244	2303	3354	18[23]
	s:	(37.6)	(82.9)	(75.7)	(112.6)	

[22] F4 only 11 ex.

[23] F4 only 17 ex.

[24] F3 only 16 ex., F4 only 15 ex.

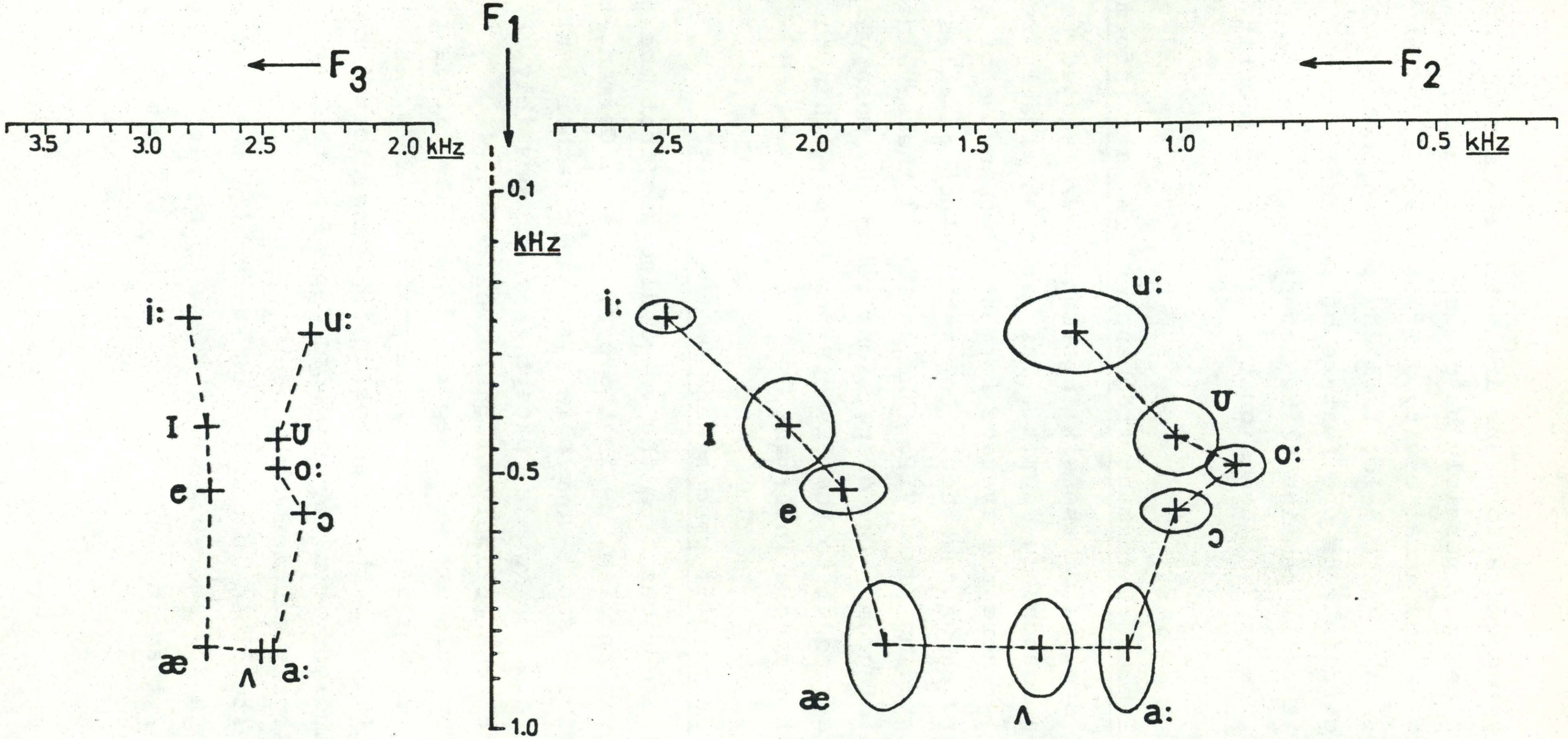


Fig. 7. Vowel diagram Informant 6 (RD)

Apparently the extra formants are primarily associated with the open vowels and may possibly be caused genetically by a slight coupling to the nasal cavity. If they are found at all their positions vary quite a lot from one person to the other, and they must certainly be one feature of personal voice quality. But it still remains to be investigated how far their presence has any influence on the linguistic identity of the vowels.

3.3.3. Overall mean of formant frequencies from six informants

The vowel systems of the six informants were judged to be comparable to a reasonable degree. This is most clearly revealed by Fig. 8 where the average formant frequencies of all six subjects are shown together. The mean values of all six systems are listed in Table VII. And in Fig. 9 these values are shown in a diagram. The dispersions round the mean values are approximately equal to the dispersions found within the vowel system of a single informant.

3.3.4. Qualitative differences between the vowels

It is quite obvious from the vowel diagrams that there is no tendency in English, as is the case in e.g. German, to keep the relatively short vowels in a separate rather centralised system. All the vowels, except /ə:/ and possibly /U/, are placed along the perifery of the vowel chart.

The front vowels are well separated and show little or no overlapping in quality. The exact position of /ʌ/ is somewhat doubtful. In the diagrams it is found closest to /a:/ but the auditory quality may also depend on the extra formants previously mentioned.

/ɔ/ is quite close in quality to /a:/ while there is some distance between /ɔ/ and /o:/. /U/ seems to be closer in quality to /o:/ than to /u:/.

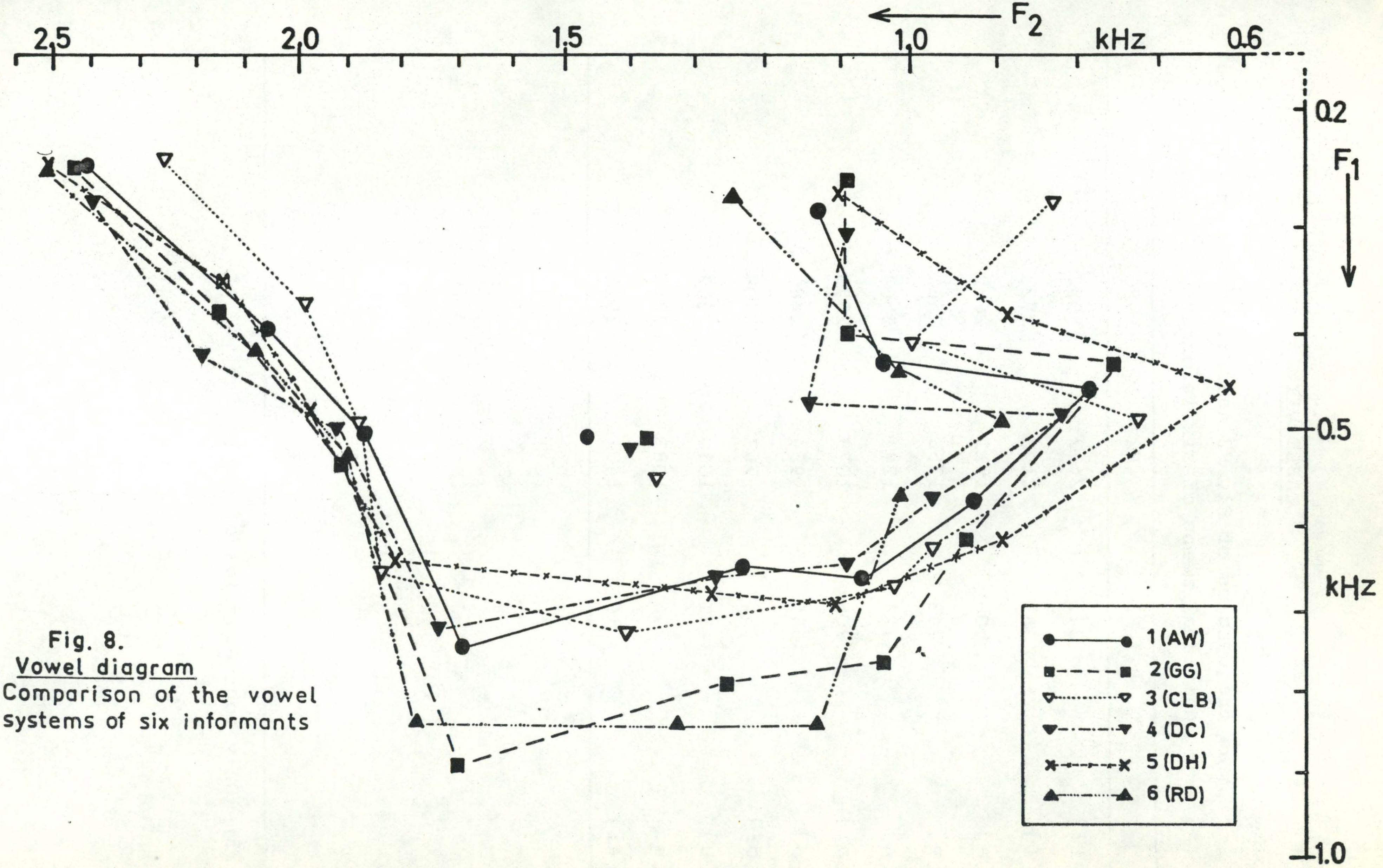


Fig. 8.
Vowel diagram
Comparison of the vowel systems of six informants

TABLE VII

Mean frequencies of F1, F2 and F3. Based on the individual means of six informants.[25]

	F1	F2	F3
[i:]	247	2422	3082
[I]	387	2102	2723
[e]	505	1911	2682
[æ]	743	1756	2561
[ʌ]	724	1294	2504
[a:]	709	1079	2573
[ɔ]	595	954	2480
[o:]	465	767	2409
[U]	421	1030	2352
[u:]	276	1083	2243
[ə:]	514	1401	2511

[25] Only four examples of [ə:].

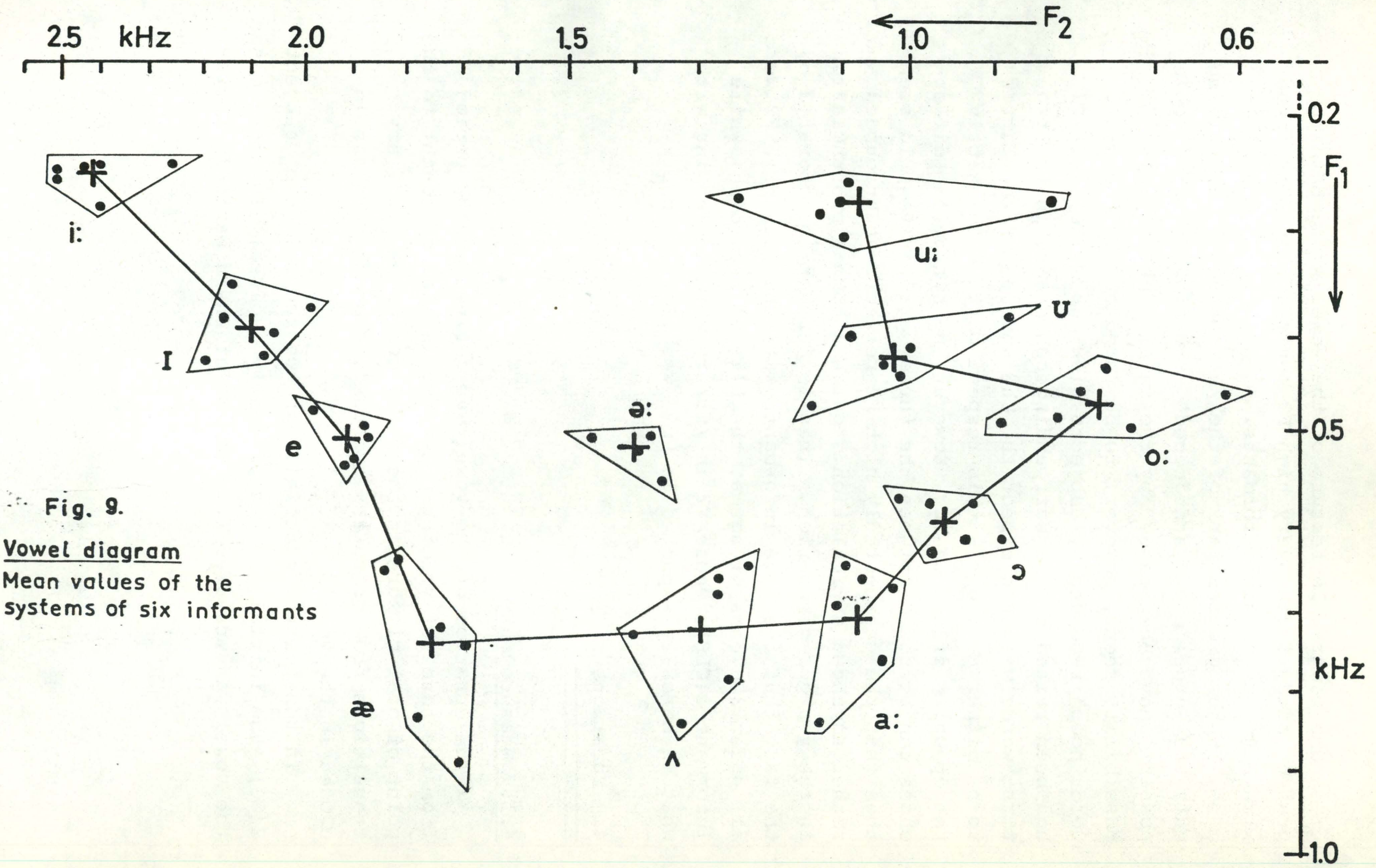


Fig. 9.

Vowel diagram

Mean values of the systems of six informants

One remarkable feature of the vowel systems in the present study is the very high F2 positions of /u:/ found with five out of six informants.

In the greater part of the /u:/-material, containing only the words "coot" and "cooed", the high position could possibly have been caused by the initial [k-]. This possibility seems, however, to be ruled out by the extra word "hoop" recorded by subjects 5(DH) and 6(RD). If F2 had been raised by the surrounding consonants in the two first words we would expect the [h] and [p] of the latter to be either neutral in this respect or to have exactly the opposite effect on F2. However, no systematic differences can be found between the three words. It would seem then that the unexpectedly high F2 reflects an unrounded and rather advanced articulation: quite a common pronunciation in the younger generation. This is further corroborated by the fact that the five informants with the very high F2 were all under 30 years of age at the time of recording, while informant 3(CLB), whose F2 in /u:/ is more in keeping with what is normally expected in [u], was 71 years old.

4. Durations

4.1. Measurements

The duration of the vowel was defined as the period from the onset of voicing after the initial consonant to the point of oral closure in the final consonant, i.e. any transitions from or to the surrounding consonants are included in the vowel.

In some cases, especially in the recordings of subject 2(GG), the initial [h] was partially voiced. In these cases the vowel was measured from the point where periodic vibrations appeared in the higher formants since the energy of the [h] appeared to be concentrated in the low frequency regions.

Before final voiceless consonants the oral closure was frequently obscured in the spectrograms by a rather strong glottal stop. Segmenting these words proved quite difficult since the glottal closure was often preceded by a period of creaky voice. Generally the vowel was then measured to the point when energy disappeared from the upper formants. But in a few cases the result was not wholly satisfactory.

The durations were measured with an accuracy of about 5 msec. And the results were run through the standard XFON-program, each vowel before [t] and before [d] taken separately.

4.2. Results

The results of the measurements are listed in Tables VIII-XIII for each of the six informants separately. The overall averages are listed in Table XIV.

4.2.1. Division into long and short vowels

It is generally observed that there is no clearcut difference between English so-called long and short vowels. This is also true of the present material. Thus if all the vowels before [-t] are pooled (Fig.10a),the curve comes very close to a normal distribution. Dividing the vowels into traditionally long and short vowels produces two distinct distributions but still leaves a considerable overlapping of the two groups (Fig. 10b).

4.2.2. Normalising duration with respect to quality

Part of this overlapping is caused by the tendency of close vowels to be shorter than open vowels. Therefore

TABLE VIII

Vowel durations in cs. Informant 1(AW).
The table shows mean value, standard deviation and number of tokens.

	Vowel before [t]			Vowel before [d]		
	$\overline{X}$	s	N	$\overline{X}$	s	N
[i:]	12.2	1.41	8	27.4	1.02	8
[ɪ]	7.6	1.28	7	11.2	1.53	8
[e]	9.8	0.39	7	13.1	1.35	8
[æ]	10.8	0.88	8	15.3	1.79	8
[ʌ]	8.9	1.36	8	11.5	2.00	8
[a:]	17.5	1.04	8	29.0	1.91	8
[ɔ]	9.5	1.49	8	14.9	2.23	8
[o:]	15.7	1.19	8	28.6	2.54	8
[U]	8.4	1.66	8	11.3	1.46	8
[u:]	12.1	1.16	8	28.3	2.36	8
[ə:]	15.8	1.00	8	27.0	0.96	8

TABLE IX

Vowel durations in cs. Informant 2(GG)
The table shows mean value, standard deviation and number of tokens.

	Vowel before [t]			Vowel before [d]		
	$\overline{X}$	s	N	$\overline{X}$	s	N
[i:]	12.7	1.19	8	21.3	1.39	8
[I]	10.2	0.84	8	12.8	1.69	8
[e]	11.7	0.53	8	14.4	1.33	8
[æ]	13.1	1.46	8	17.1	1.36	8
[ʌ]	10.9	0.98	8	14.2	1.25	8
[a:]	17.4	0.35	8	23.5	1.25	8
[ɔ]	12.3	1.60	8	17.6	1.62	7
[o:]	16.5	1.32	7	23.1	1.35	8
[U]	11.1	0.78	8	13.2	0.92	8
[u:]	12.7	1.44	8	21.4	1.57	7
[ə:]	17.3	1.39	8	22.0	1.41	8

TABLE X

Vowel durations in cs. Informant 3(CLB). The table shows mean value, standard deviation and number of tokens

	Vowel before [t]			Vowel before [d]		
	$\overline{X}$	s	N	$\overline{X}$	s	N
[i:]	13.3	1.73	6	33.8	2.04	6
[ɪ]	7.8	0.82	6	12.9	1.59	6
[e]	10.2	1.03	6	15.6	0.59	6
[æ]	12.6	0.59	6	20.9	1.83	6
[ʌ]	9.2	1.63	6	14.9	1.20	6
[a:]	19.3	1.41	6	34.3	2.66	6
[ɔ]	11.8	0.94	6	18.3	0.93	6
[o:]	18.0	1.52	6	35.3	3.31	6
[ʊ]	8.3	1.25	6	14.9	1.53	6
[u:]	15.2	1.47	6	35.6	1.39	6
[ə]	17.3	1.70	6	32.7	1.78	6

TABLE XI

Vowel durations in cs. Informant 4(DC). The table shows mean value, standard deviation and number of tokens

	Vowel before [t]			Vowel before [d]		
	$\overline{X}$	s	N	$\overline{X}$	s	N
[i:]	16.0	2.21	6	32.8	1.68	5
[I]	12.4	0.74	5	20.8	1.99	6
[e]	17.3	1.86	5	24.5	5.08	4
[æ]	19.3	2.28	5	25.0	2.88	6
[ʌ]	13.0	2.55	6	21.1	1.28	6
[a:]	25.4	3.26	6	37.9	6.28	6
[ɔ]	16.3	1.21	6	21.6	1.80	6
[o:]	23.0	1.87	6	38.1	2.20	6
[U]	13.4	1.77	6	22.5	3.62	6
[u:]	15.5	1.61	6	35.7	2.81	6
[ə:]	22.3	2.21	6	39.1	2.08	6

TABLE XII

Vowel durations in cs. Informant 5(DH) The table shows mean value, standard deviation and number of tokens

	Vowel before [t]			Vowel before [d]		
	$\overline{X}$	s	N	$\overline{X}$	s	N
[i:]	9.8	1.72	6	32.8	4.80	6
[ɪ]	6.1	1.16	6	10.8	0.82	6
[e]	7.6	0.81	6	17.0	1.70	6
[æ]	9.6	0.86	6	15.7	2.07	6
[ʌ]	8.0	0.89	6	14.3	1.61	6
[a:]	17.3	1.78	6	34.8	2.75	6
[ɔ]	7.9	0.67	6	15.0	0.84	6
[o:]	15.9	1.53	6	38.0	2.17	6
[ʊ]	5.9	0.67	6	11.8	2.12	6
[u:]	10.6	1.36	6	38.4	8.21	6

TABLE XIII

Vowel durations in cs. Informant 6(RD).
The table shows mean value, standard deviation and number of tokens

	Vowel before [t]			Vowel before [d]		
	$\overline{X}$	s	N	$\overline{X}$	s	N
[i:]	12.9	1.53	6	24.7	3.68	6
[I]	9.3	0.98	6	16.1	1.43	6
[e]	10.5	0.89	6	15.4	1.93	6
[æ]	14.2	1.21	6	20.1	1.07	6
[ʌ]	11.0	1.48	6	18.5	2.32	6
[a:]	22.2	1.51	6	29.9	3.34	6
[ɔ]	13.9	1.72	6	20.1	1.72	6
[o:]	21.8	1.51	6	-	-	0
[U]	11.2	1.33	6	17.0	1.79	6
[u:]	13.3	1.54	6	25.0	2.03	6

TABLE XIV

Mean vowel durations before [t] and [d]. Based on the individual mean values from six informants

	Duration before [t]	Duration before [d]
[i:]	12.8	28.8
[I]	8.9	14.1
[e]	11.2	16.7
[æ]	13.3	19.0
[ʌ]	10.2	15.7
[a:]	19.8	31.6
[ɔ]	11.9	17.9
[o:]	18.5	32.6
[U]	9.7	15.1
[u:]	13.2	30.7
[ə:]	18.2	30.2[26]

[26] Only four examples of [ə:].

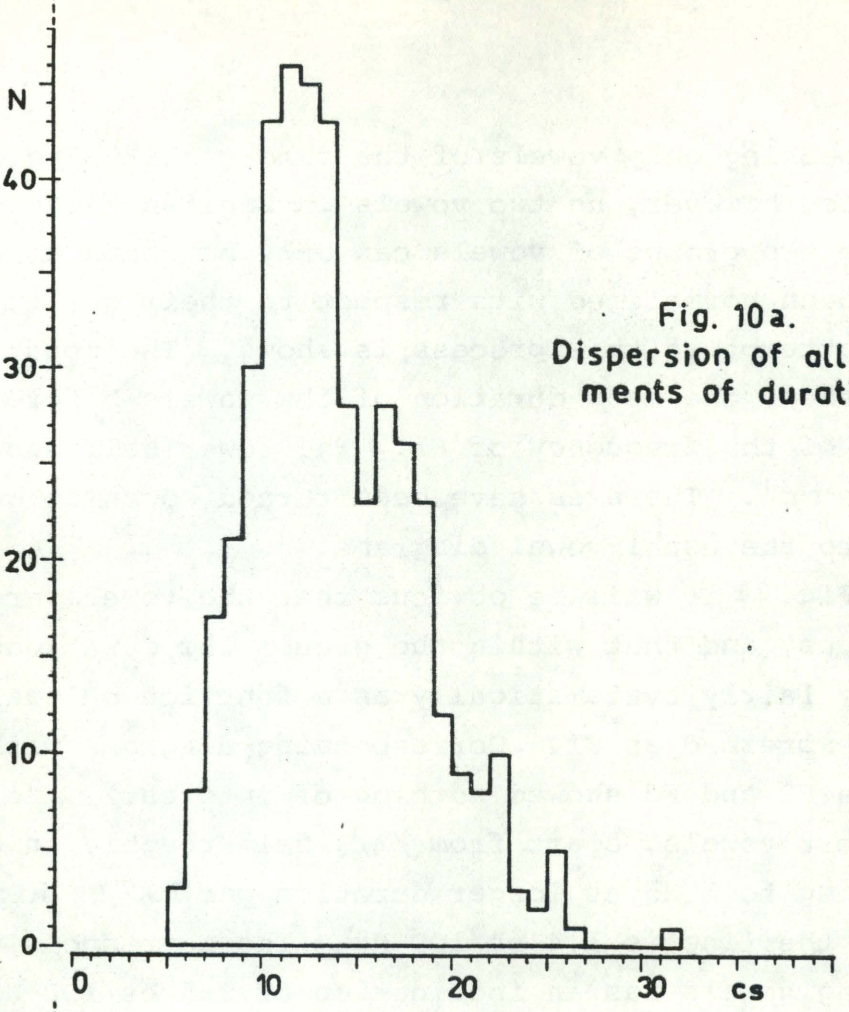


Fig. 10a.
Dispersion of all measurements of duration

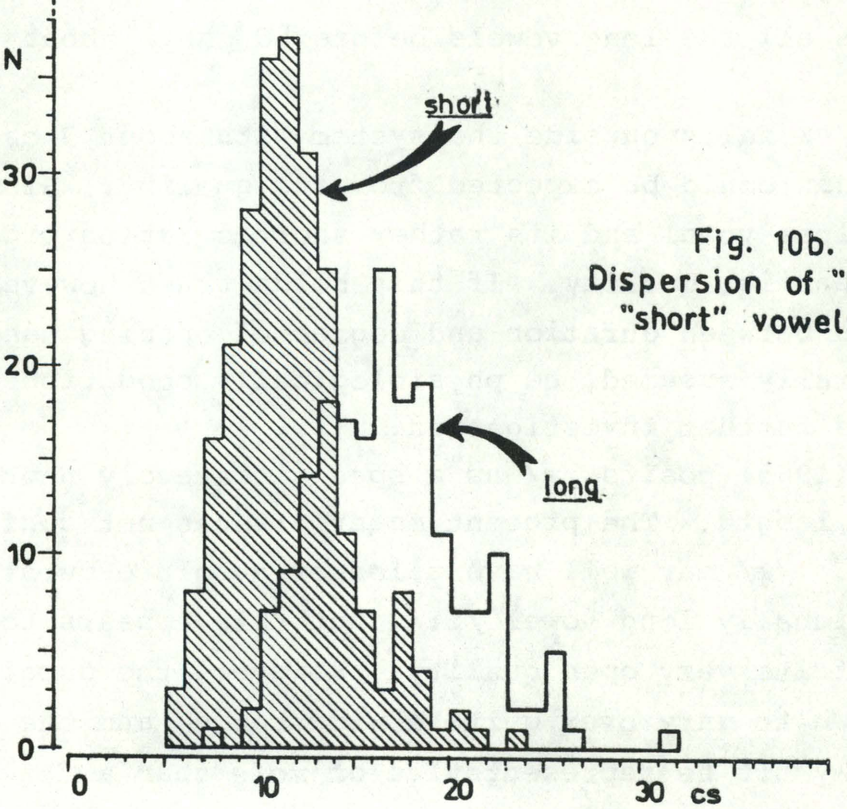


Fig. 10b.
Dispersion of "long" and "short" vowels

strictly speaking only vowels of the same quality are comparable. Since, however, no two vowels in English have the same quality the two groups of vowels can only be compared after they have been normalised with respect to their quality. In Fig.11 an attempt at this process is shown. The upper half of Fig.11 shows the mean duration of the vowels before [t] as a function of the frequency of F1. The lower half shows the same before [d]. The axes have been turned so as to give easy reference to the usual vowel diagrams.

From Fig. 11 it will be obvious that the vowels are divided in two groups, and that within the groups the durations of the vowels vary fairly systematically as a function of quality (in this case expressed as F1. Corresponding diagrams with duration versus F2 and F3 showed nothing of interest). Before [t] all the short vowels, apart from /ʌ/, fall roughly on a line corresponding to 1.25 cs longer duration per 100 Hz higher F1. Before [d] the line is 1.4 cs/100 Hz. The corresponding line for the long vowels has an inclination of 1.5 cs/100 Hz before [t] whereas all the long vowels before [d] have about the same duration.

Only /ʌ/ falls outside the system with about 3 cs shorter duration than could be expected from its quality. Historically /ʌ/ is a close vowel and its rather short duration might reflect its earlier quality. If this holds true, however, the correlation between duration and degree of opening cannot, as it is generally assumed, be physiologically conditioned. This point needs further investigation.

Wiik (1965) posits /æ/ as a special category neutral with respect to length. The present material does not confirm this hypothesis. /æ/ may well have a longer absolute duration than the traditionally long vowel /i:/, but this appears to be a function of its very open quality. However, the duration of /æ/ is known to vary over quite a wide range, and the present material may not be representative of more than a fraction of

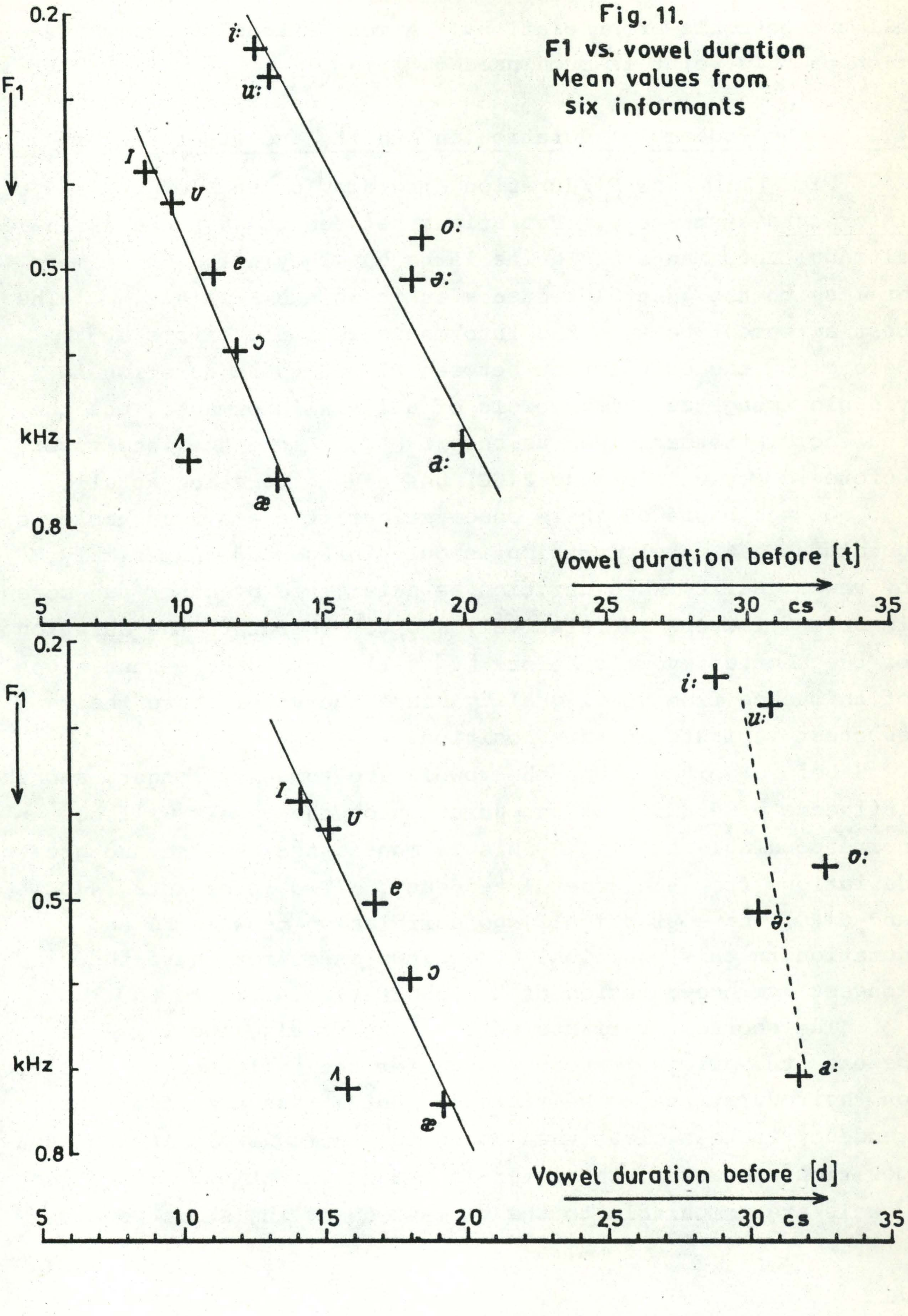


Fig. 11.
F1 vs. vowel duration
Mean values from
six informants

all the possible pronunciations. A more extensive investigation of this point is much needed.

4.2.3. Dependency of duration on quality and other factors

Figs.12-15 show Fl/duration diagrams for each of the six informants separately. Variations between the systems of individual informants are quite large but they all conform more or less to the general tendency as it is shown in Fig. 11. The best agreement between the informants is found before [t]. Before [d] the correlation between Fl and vowel duration is visible among the short vowels of all six informants, but the tendency is not as clear as before [t]. Among the long vowels before [d] only informant 2(GG) has any correlation at all.

On the basis of these observations it seems reasonable to conclude that vowel duration is only influenced appreciably by vowel quality when duration as determined by other and more important factors is relatively short. Therefore the duration of the shortest vowels before [t] will show the clearest signs of influence from vowel quality since the vowels have their shortest variants in this position.

Before voiced stops the vowels are somewhat longer, and the influence of quality on the duration of the vowels will be correspondingly smaller. This is manifested as more and greater deviations from the general tendency. (Two informants, 4(DC) and 6(RD), have practically no correlation between Fl and duration in this position. These two informants have the longest average duration of the short vowels before [d].)

The shortest variants of the long vowels, the variants before [t], are just short enough for the influence of quality on their durations to be visible. But at the same time the tendency is less clear than among the short vowels in the same position. However, the inter-informant variations of the long vowels are comparable to the variations of the short vowels before [d]: In this position the average duration of the short vowels is about equal to the average duration of the long vowels before [t].

Finally the long vowels before [d] are so long that any influence from quality on vowel duration has disappeared. Only the vowels of informant 2(GG), whose long vowels before [d] are very short indeed, show traces of the tendency. .

4.2.4. Differences between long and short vowels

The mean duration of all the short vowels[27] amounts to 10.9 cs before [t] and 16.4 cs before [d]. The mean duration of the long vowels is 16.5 cs before [t] and 30.8 cs before [d].

These figures show that the two groups are not affected by the surrounding consonants in the same way. Thus in the present material the long vowels are 86.6 pct longer before [d] than before [t], while the short vowels are only 50.5 pct longer.

Another interesting point noticed by Wiik (1965) is that the lengthening effect of the voicing in the final consonant is of the same order of magnitude as the lengthening caused by the opposition between long and short vowels. This is confirmed in the present investigation. But it should be remembered that this rule applies only to the mean values of short and long vowels. Within these groups the durations of the individual vowels vary as a function of quality, and therefore no general percentage "lengthening" can be given which applies to all the individual vowels.

Thus it will be seen from Fig.11 that as long as the vowels are short enough for the influence of quality to be noticable the "lengthening effect" consists roughly of an added absolute value, the magnitude of which is independent of the "normal" duration of the vowel in question. For instance, the short vowels are all about 5 cs longer before [d] than before [t].

[27] Including /æ/.

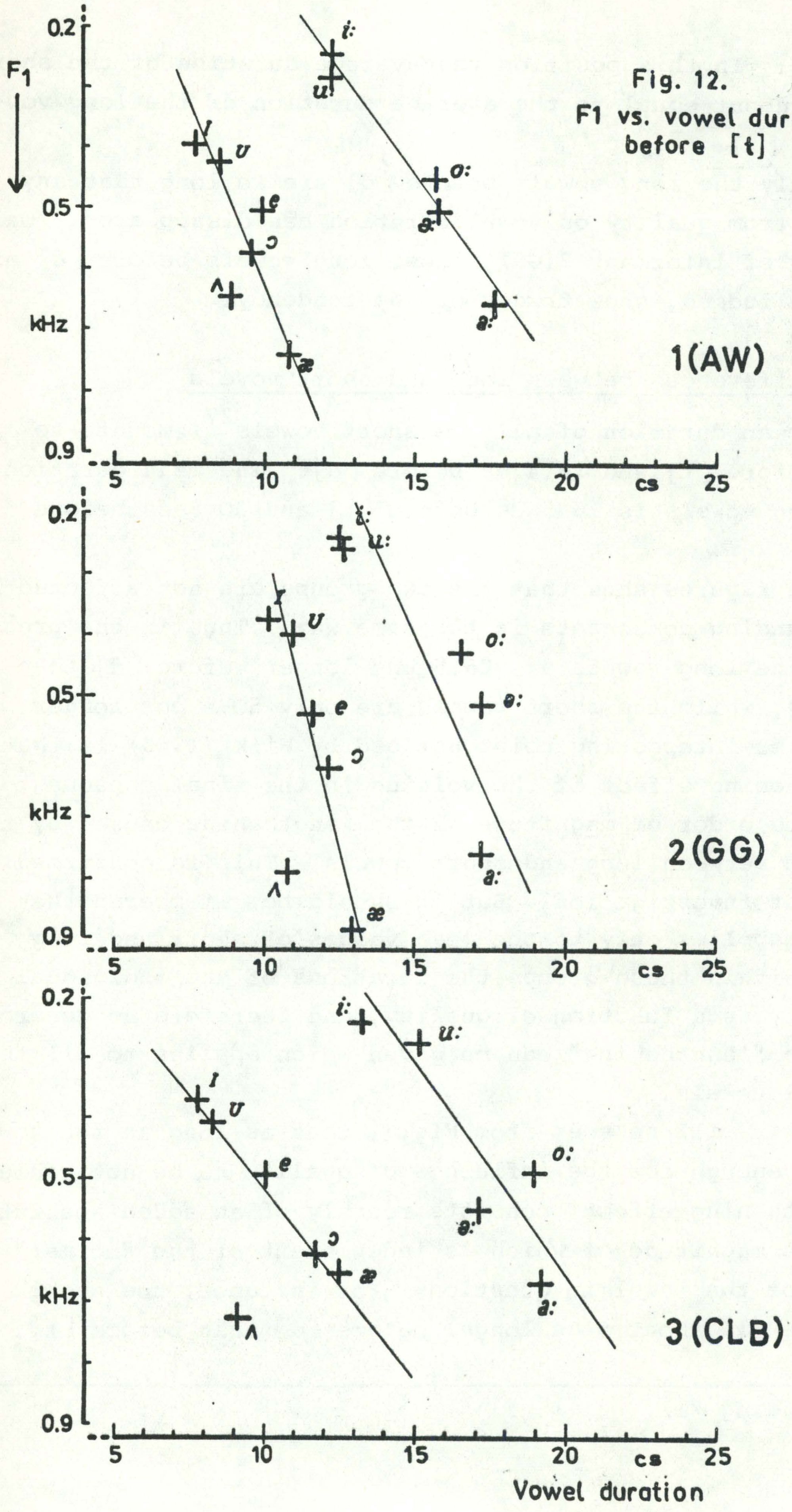


Fig. 12.
F1 vs. vowel duration
before [t]

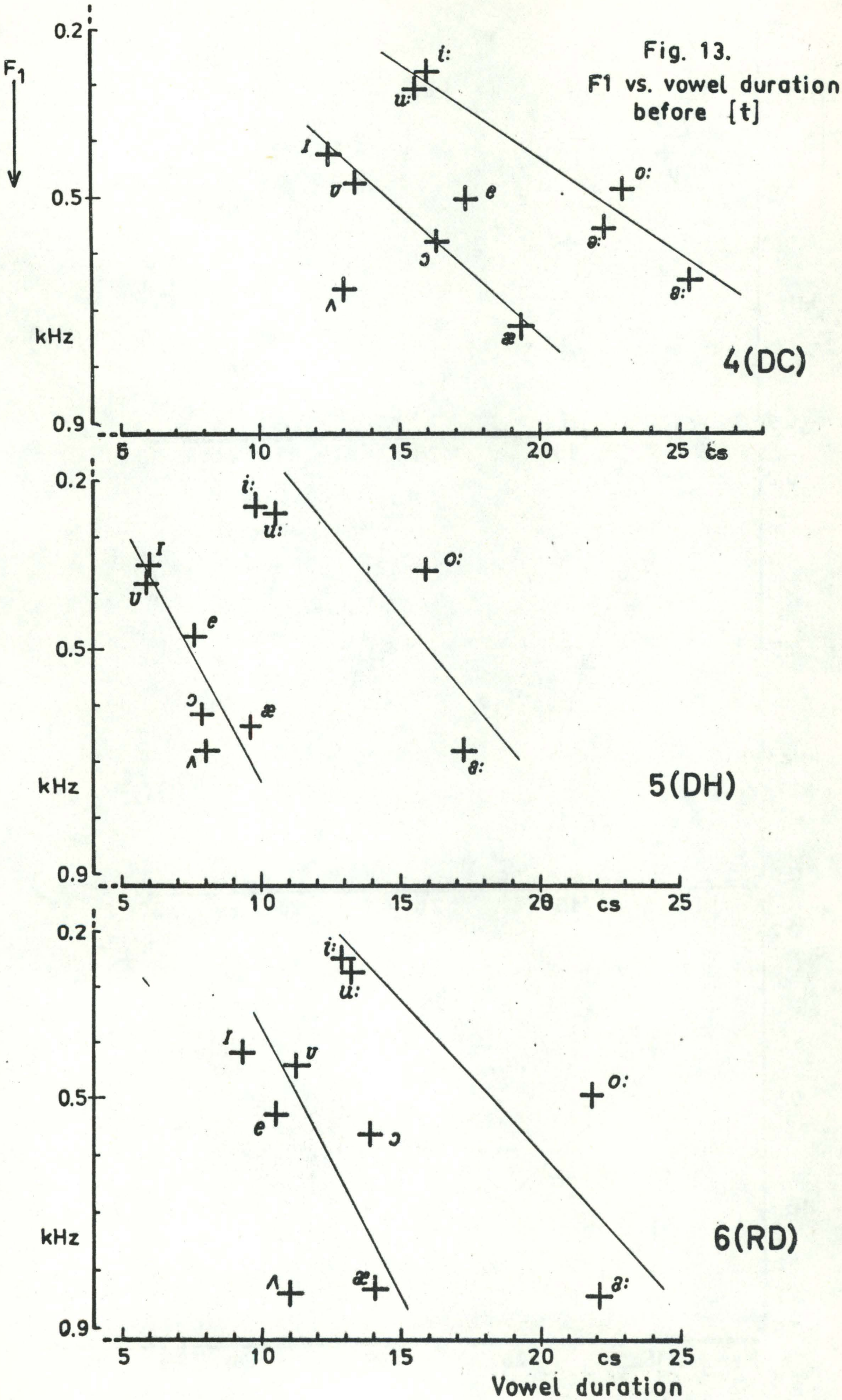


Fig. 13.
F1 vs. vowel duration
before [t]

Fig. 14. F1 vs. vowel duration before [d]

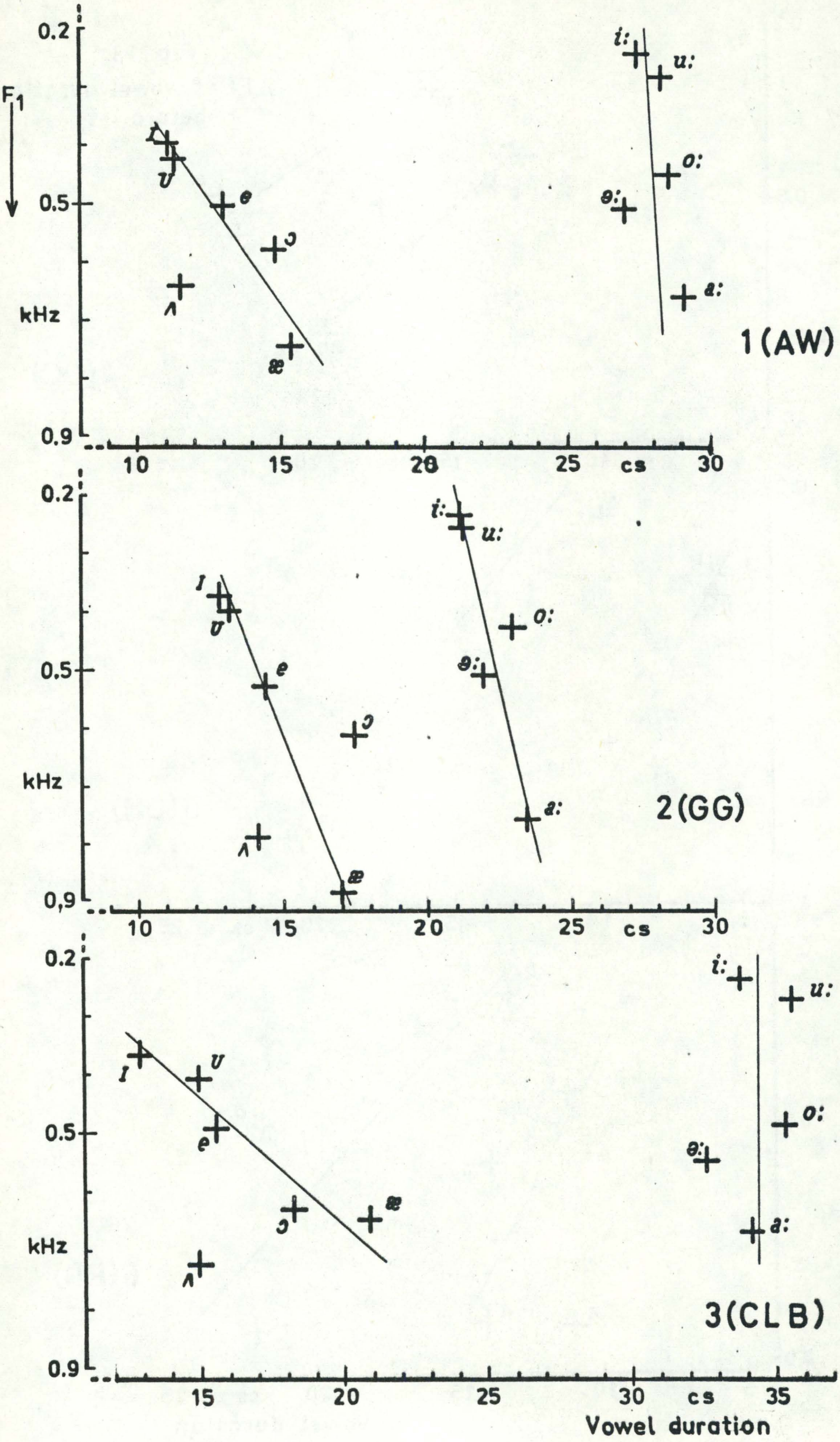
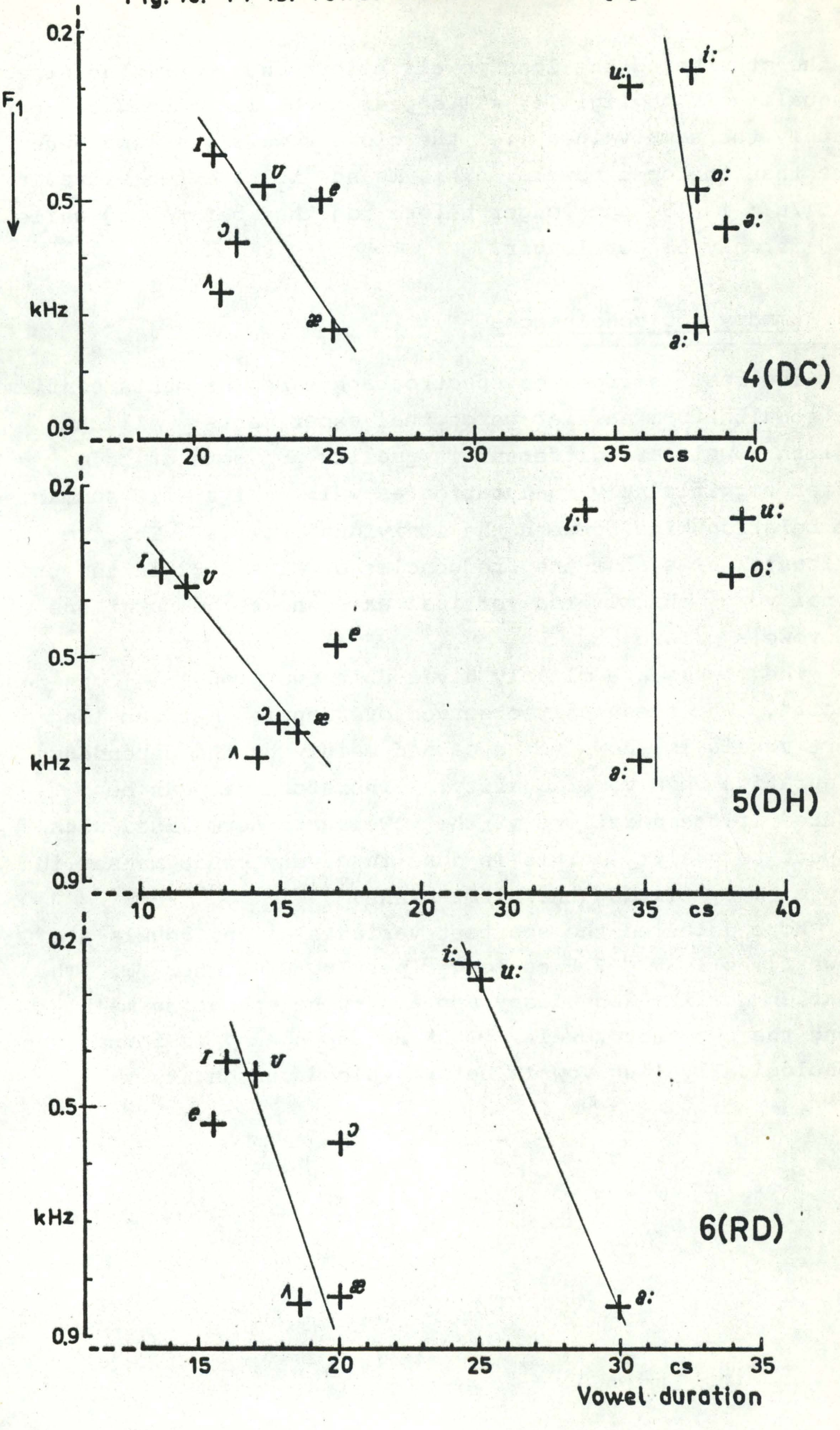


Fig. 15. F1 vs. vowel duration before [d]



On the other hand the long vowels before [d], where the effect of quality is negligible, all appear to be lengthened to approximately the same value, i.e. the close vowels are lengthened more than the open vowels. This means, to take two extremes, that /u:/ is 133 pct longer before [d] than before [t] while [a:] is only 59 pct longer!

5. Summary and conclusions

The results from the spectrographic measurements confirm the results from earlier perceptual experiments: All the English vowels are different in quality and some of them differ significantly in duration as well. Figure 16 summarises the relationships between the individual vowels. The two horizontal axes show the frequencies of F1 and F2 in the traditional vowel chart. The vertical axis shows the durations of the vowels before [t].

The vowels are clearly divided in two groups: "long" and "short". The frequently observed overlapping between long and short vowels is shown to be caused mainly by the dependency of duration upon vowel quality. Therefore it can be reduced if the durations of the vowels are normalised with respect to quality. This is done in a very crude manner in figure 16 by tilting the F1/F2-plane.

More detailed the shortest variants of the vowels are shown to be 1.25 - 1.5 cs longer per 100 Hz higher F1. The duration-quality dependency appears to be operating mainly among the shortest vowels. Thus no dependency is found among phonologically long vowels before voiced consonants.

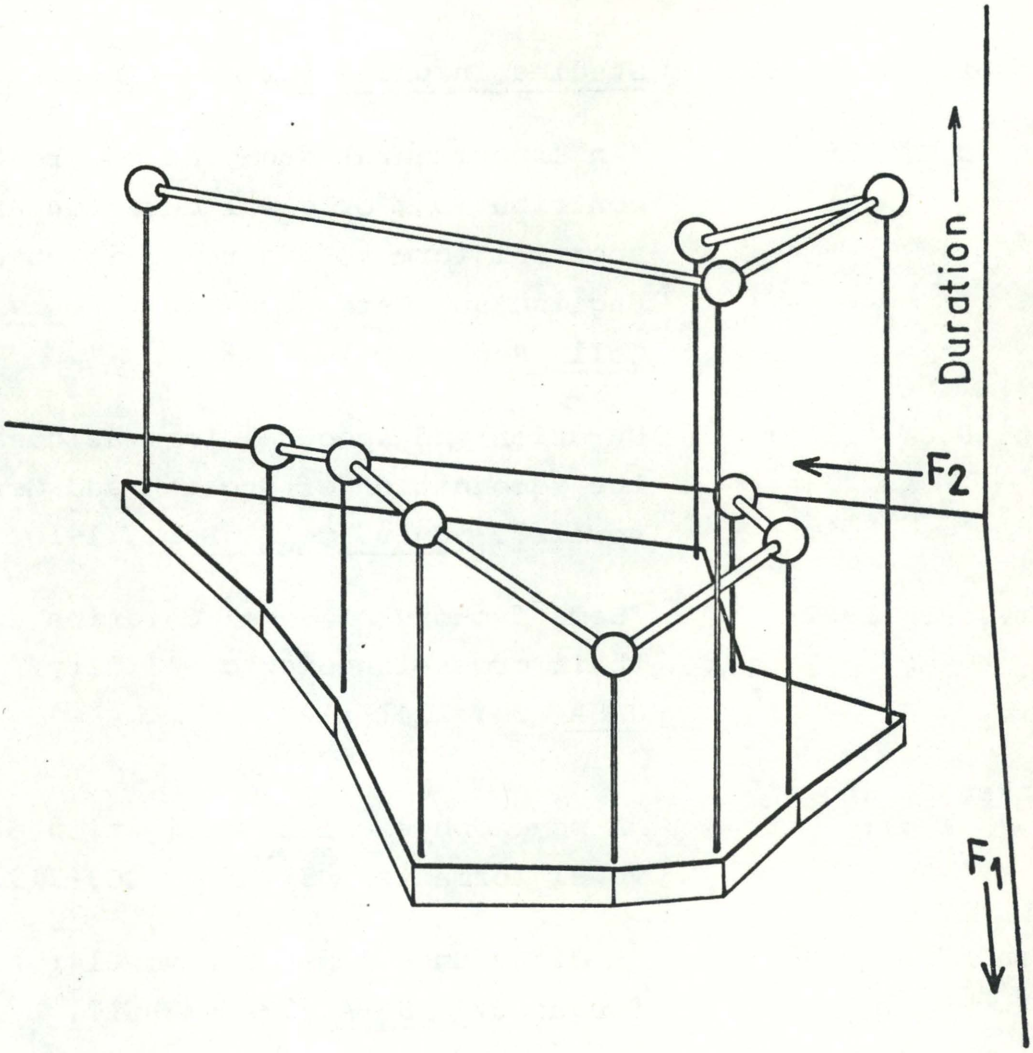


Fig. 16.

Three dimensional sketch of the distances between the individual vowels. The axes show the frequencies of F1 and F2 and the durations of the vowels before [t]. The scales used are about the same as those found in Figs. 12-15.

References

Abercrombie, D. 1965: *Studies in phonetics and linguistics*

Bennett, D.C. 1963: "An experimental study of the relative contributions of vowel duration and spectral form to the recognition of English and German vowels". *Univ. Coll. Rep.*, 10-14.

Bennett, D.C. 1965: Duration and spectral form as cues in the recognition of English and German vowels". *Univ. Coll. Rep.*, 1-10.

Delattre, P. 1962: "Some factors of vowel duration and their cross-linguistic validity" *JASA* 34, 1141-1143.

Fairbanks, G. and Grubb, P. 1961: "A psychophysical investigation of vowel formants" *JSHR* 4, 203-219.

Flanagan, J.L. 1955: "A difference limen for vowel formant frequency" *JASA* 27, 613-617.

Frøkjær-Jensen, B. 1963: "De danske langvokaler", *Tale og Stemme*, 2, 59-75.

Gimson, A.C. 1945-49: "Implications of the phonemic/chronemic grouping of English vowels" *Acta Linguistica*, 94-100.

Gimson, A.C. 1970: *An introduction to the pronunciation of English* (2nd ed.)

Heffner, R-M.S. 1937ff: "Notes on the length of vowels I". *American Speech* 128-135, II *A.S.* (1940 a) 74-80, III *A.S.* (1940 b) 377-381, IV *A.S.* (1941) 204-208, V *A.S.* (1942) 42-49, VI *A.S.* (1943) 208-216.

Heike, G. and Hall, R.D. 1969: "Vowel patterns of three English speakers: A comparative acoustic and auditive description" *Linguistics*, 54, 13-38.

House, A.S. 1961: "On vowel duration in English" *JASA* 33, 1174-1178.

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

Jones, D. 1956: The pronunciation of English (4th ed.).

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

Ladefoged, P. 1962 *The nature of vowel quality* (Coimbra).

Meyer, E.A. 1903: "Englische Lautdauer" *Skrifter utg. af Kgl. Hum. Vetenskaps-Samf. i Uppsala*, VIII, 3.

Peterson, G.E. and Barney, H.L. 1952: "Control methods used in a study of the vowels" *JASA* 24, 175-184.

Peterson, G.E. and Lehiste, I. 1960: "Duration of syllable nuclei in English" *JASA* 31, 428-435.

Stevens, K.N. 1959: "Effect of duration upon vowel identification" *JASA* 31, 109 (abstract)

Stevens, S.S. and Volkman, J. 1940: "The relation of pitch to frequency" *Am. Journ. Psych*. 53, 329-353.

Wells, J.C. 1963: "A study of the formants of the pure vowels of British English" *Univ. Coll. Rep*., 1-6.

Wiik, K. 1965: *Finnish and English vowels*.

FORMANT FREQUENCIES OF LONG AND SHORT DANISH VOWELS[1]

SR: 8b
Lyd og lydsystem

Eli Fischer-Jørgensen

1. The material

The spectrograms on which the present formant measurements are based were taken in MIT in 1952. They have since then been amply used for teaching purposes, but the measurements have never been published. - The material is very restricted, but since there were nine speakers, who all show the same tendencies, it may nevertheless be of some value. It consists of a series of isolated long vowels i:, e:, ɛ:, a:, y:, ø:, œ:, u:, o:, ɔ: and a word list containing these vowels and the corresponding short vowels in similar surroundings, i.e. in the stressed syllable of a dissyllabic word and, as far as possible, preceded by /h/ or zero and followed by the consonant l (or at least a dental), namely: ile, hele, hæle, hale, hyle, øde, høne, hule, Ole, åle, ilde, inde, hælde, halve, hylde, øllet (participle), hønse, hulde, hullet (participle), holde; in traditional phonemic transcription: / i:lə, he:lə, hɛ:lə, ha:lə, hy:lə, ø:ðə, hœ:nə, hu:lə, o:lə, ɔ:lə, ilə, enə, hɛlə, halə, hylə, øləð, hœnsə, hulə, holəð, hɔlə /. (In the recordings of one of the subjects the words with back vowels had a labial consonant after the vowel). Moreover the word hare, containing the variant [ɑ:] of /a:/ before /r/ was included.

The list was spoken only once by each informant, with the exception of EFJ, who has spoken the list twice, and of JK and VL, who have spoken the isolated vowels twice. In these cases an average of the two recordings was taken. The informant PD did not speak the words in isolation.

[1] This paper is a shortened version of a paper which will appear in a Festschrift in the autumn 1972.

2. The speakers

The list was spoken by one female and eight male speakers of Standard Danish, born between 1905 and 1930. The speech of five of the informants can be characterized as Copenhagen Standard, mainly on the basis of intonation features, one speaker has certain Funish characteristics, and the speech of the remaining three subjects has hardly any local features.

3. Recordings

The recordings were made on professional tape recorders. In almost all cases both narrow and wide band spectrograms were taken, and in half of the cases they were supplemented by sections.

4. Results

The average frequencies of the first three formants of the male speakers are given in the following table.

Formant frequencies of eight male speakers

1. isolated long vowels
2. long vowels in words, and
3. short vowels in words

		/i/			/e/			/ɛ/		
		F_1	F_2	F_3	F_1	F_2	F_3	F_1	F_2	F_3
1.	lg. is.	225	2197	3189	286	2186	2865	371	2049	2635
2.	lg. w.	219	2169	3238	286	2231	2922	373	2099	2722
3.	sh. w.	239	2208	3070	295	2200	2878	386	2007	2607

		/y/			/ø/			/oe/		
1.	lg. is.	235	1919	2235	289	1671	2107	402	1462	2202
2.	lg. w.	240	1903	2234	283	1694	2131	384	1563	2137
3.	sh. w.	256	1756	2105	303	1594	2113	403	1556	2278

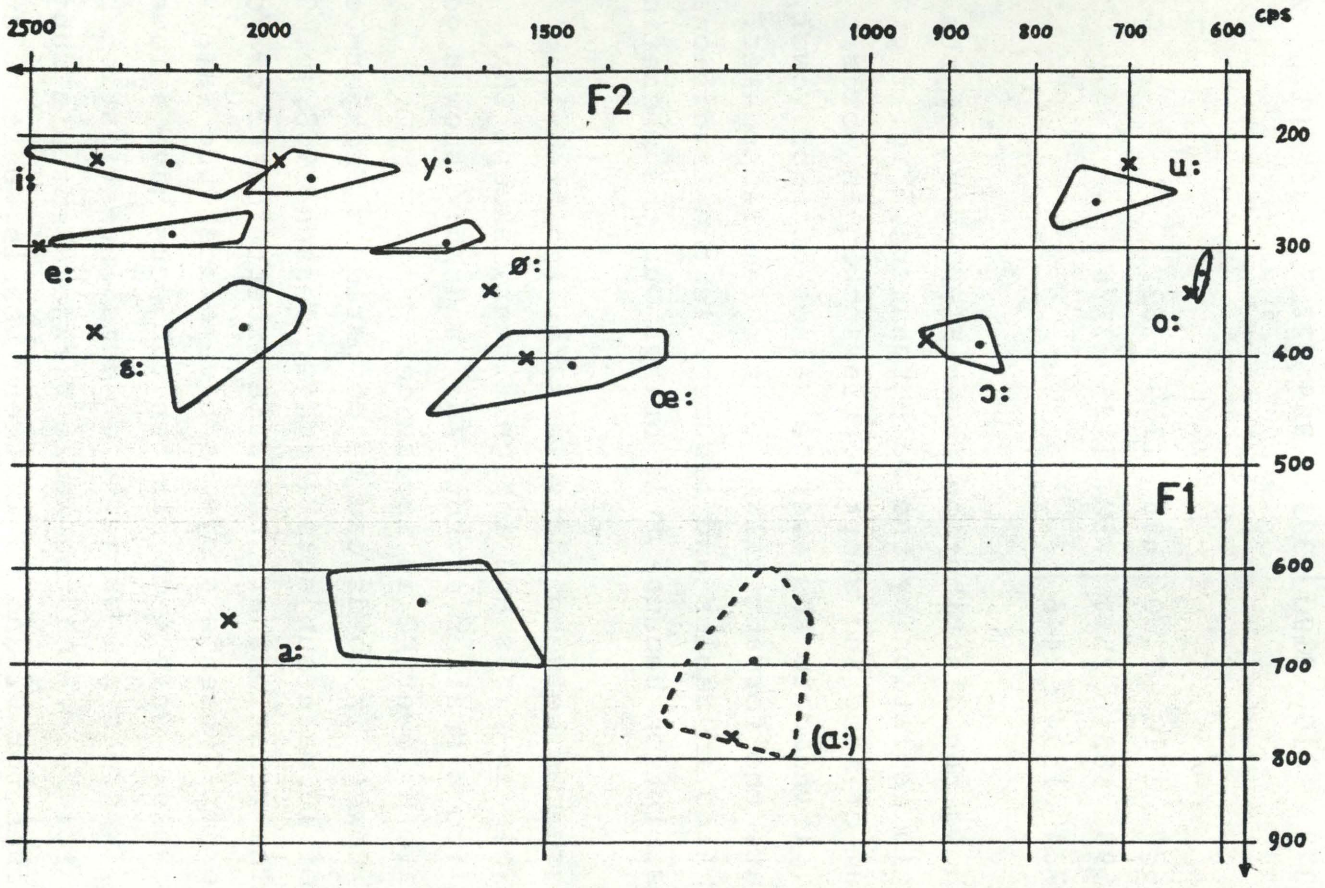


Fig.1. Danish long vowels spoken in isolation by 7 male speakers.
x = formants of the female speaker.

	/u/			/o/			/ɔ/		
1. lg. is.	257	732	2117	323	623	2375	387	865	2392
2. lg. w.	254	781	2013	308	659	2413	376	912	2333
3. sh. w.	273	819	2190	399	969	2207	588	1126	2298

	/a/			/ɑ/		
1. lg. is.	630	1710	2490	696	1179	2632
2. lg. w.	598	1831	2518	750	1185	2568
3. sh. w.	734	1594	2446			

Fig. 1 shows the distribution in an F1-F2-plot of the long vowels spoken in isolation by the male speakers. Fig. 2 gives a similar diagram of long and short vowels spoken in words. The points in Fig. 1 indicate the average values of each vowel, the crosses indicate the formant values of the female speaker. The transcription, also in tables and figures, is the traditional phonemic transcription because this brings out the phonetic deviations.

The differences between the values for isolated long vowels and vowels in words are insignificant, except that /oe:/ /a:/ and (to a smaller extent) /u:/ and /ɔ:/ show a general tendency to have a lower F2 in isolation.

The long vowels in words can be compared to Frøkjær-Jensen's measurements of 10 male subjects (Frøkjær-Jensen 1968, p. 35 and diagram I). The relations are the same except for F2 of /o:/ and /u:/. Frøkjær-Jensen's subjects have almost the same value for the two vowels: Only 4 of his ten subjects have a lower value in /o:/, whereas all subjects of this investigation have a lower value for /o:/. On the whole Frøkjær-Jensen's subjects have lower values for F2 and F3 in front vowels, particularly front unrounded vowels, whereas the deviation for F1 are very small, except for /a:/, which has a somewhat lower F1.

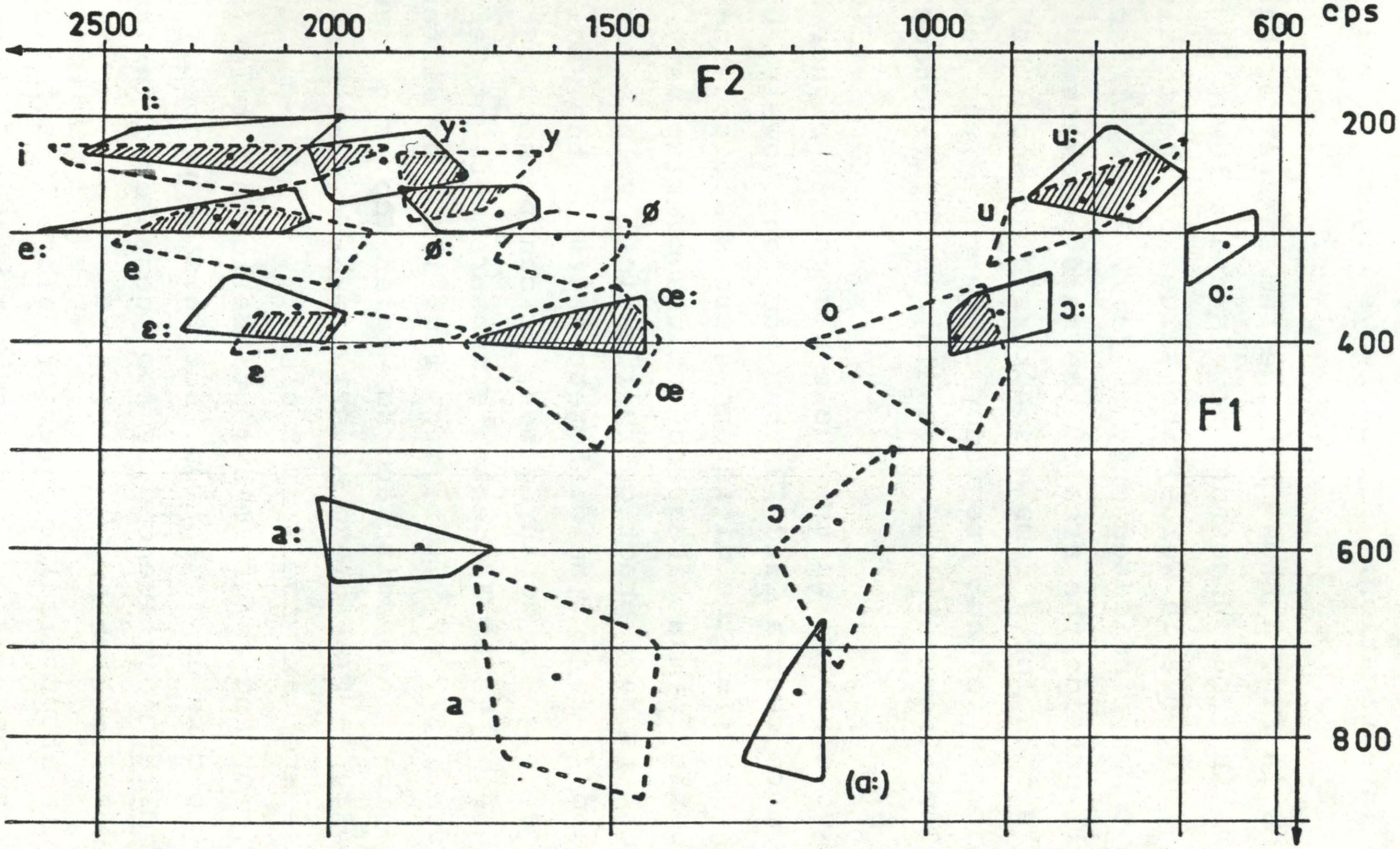


Fig. 2. Danish long and short vowels spoken in words by 8 male speakers.

In both groups some subjects have a lower F2 in /i:/ than in /e:/, but the tendency is somewhat stronger for the subjects of the present investigation. F3 is always higher in /i:/ than in /e:/.

The crosses in Fig. 1 show that the female speaker has higher F2 values in front unrounded vowels and a lower F1 in /ø:/. In Fig. 2 (vowels in words) the values of the female speaker have not been indicated; only in /a:/, /a/, /ɑ:/ and /ɔ/ do they fall outside the area of the male speakers. Frøkjær-Jensen's data for 9 female speakers also show the greatest increase in formant frequency compared to the male speakers for F2 of front unrounded vowels, (Frøkjær-Jensen 1967, 4 and p. 6).

The short vowels have on the whole a somewhat higher F1 and a slightly more centralized F2 than the long vowels, but apart from /o/ and /a/ the differences are small and inconsistent, and the formant frequency values overlap very much. Some speakers have practically identical values for /i-i:/ /e-e:/ and /ɛ-ɛ:/. The differences are on the whole much smaller than e.g. in North German, as can be seen by a comparison with Hans Peter Jørgensen's measurements of German vowels (Jørgensen 1967, p. 77 ff, e.g. Fig. 1). The difference is also smaller than in Swedish and Norwegian (cp. the diagram for Swedish vowels in Elert 1970 p. 67). There is no basis for talking of tense and lax vowels in Danish.

Short /a/, /o/ and /ɔ/ form exceptions to this rule (short /ø/ also seems to make an exception, but this is probably due to the example <u>øllet</u> ('influenced by too much beer-drinking'), which invites to affective pronunciation).

Short /o/ is of particular interest, since it coincides more or less with long /ɔ:/. It has even a higher F1 than /ɔ:/ in the speech of 6 of the nine speakers and higher F2 in the speech of 7 speakers. Three have a slightly lower F1 in /o/, but their speech is rather conservative.

The cross-over of short /o/ and long /ɔ:/ raises a phonological problem. Short [o] might be identified with /ɔ:/ rather than with /o:/. This problem, which is complicated by the fact that short [o] is found in the cases of shortened [o:] and in some foreign words, is treated by Jørgen Rischel (Rischel 1969, p. 180). He also treats the phonological problem of describing the four degrees of aperture in Danish vowels.

The lowering and fronting of the short back vowels might be considered as the result of a tendency to greater phonetic distance between phonemes. But in this case one should also expect a lowering of the long back vowels and of the front vowels, and such a tendency is not found. On the contrary, in modern Copenhagen pronunciation, long /ɛ:/, /œ:/ and /a:/ are even more close than for most speakers in the present study. (The youngest of the speakers shows a certain tendency in this direction). But even these rather conservative speakers show a conspicuous crowding of phonemes in the upper part of the diagram.

This can also be shown by placing the average values of Danish long isolated vowels in an acoustic diagram of Jones' cardinal vowels, obtained by calculating the average frequencies of formant 1 and 2 in Ladefoged's measurements of the cardinal vowels spoken by a number of British male phoneticians (Fig. 3). It can be seen that almost all Danish vowels are placed in the upper third of Jones' diagram, and according to IPA conventions Danish /ɛ:, œ:, ɔ:/ should thus be written [e:, ø:, o:], and the long /a:/ should be written [ɛ:]. This crowding explains why almost all foreigners find it difficult to distinguish the various degrees of aperture in Danish vowels.

In Figs. 1, 2 and 3 the frequency values are indicated in hz (cps), but the scale is the mel scale. This is in accordance with normal conventions, and it facilitates a

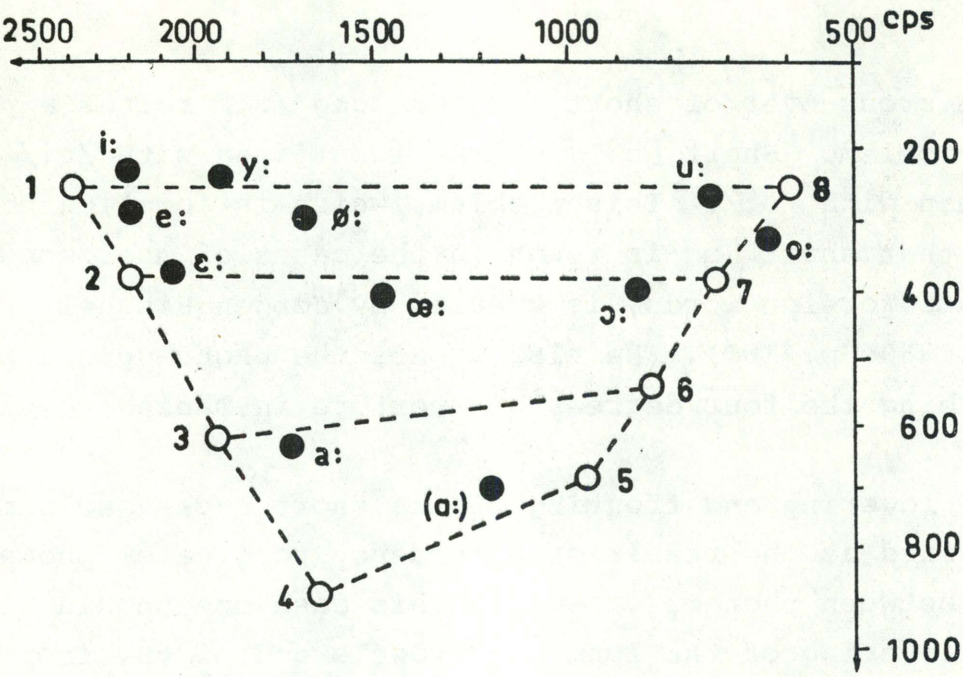


Fig. 3. Isolated Danish long vowels placed in an acoustic cardinal vowel diagram (mel scale).

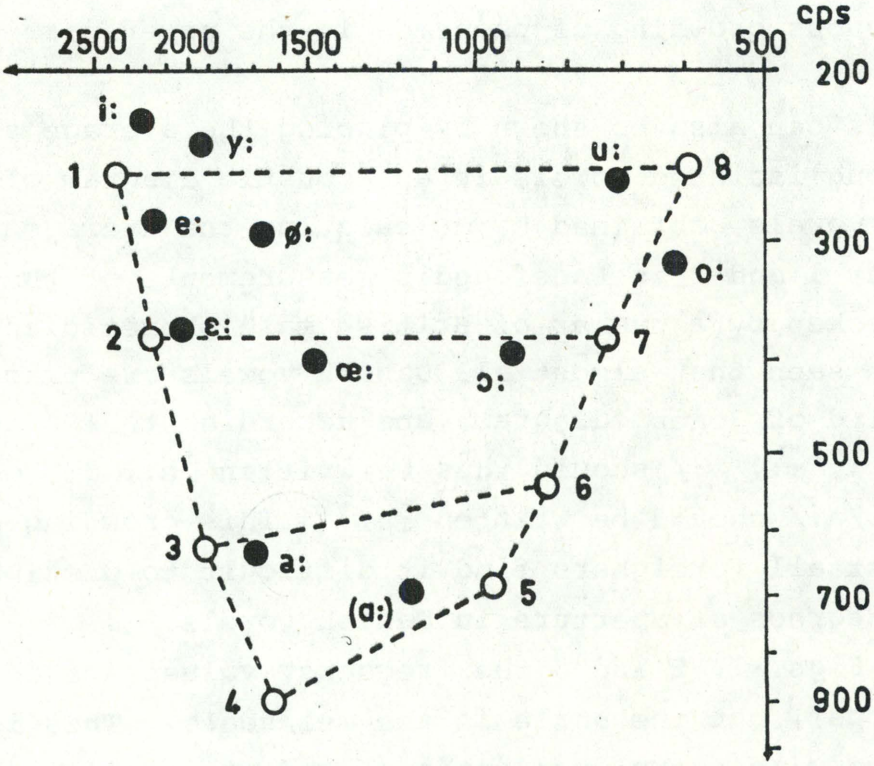


Fig. 4. Isolated Danish long vowels placed in an acoustic cardinal vowel diagram (log. scale).

comparison with the diagrams in the papers by Frøkjær-Jensen and Hans Peter Jørgensen. But there is some evidence that a logarithmic scale would be in better accordance with the auditory impression. The logarithmic scale (see Fig. 4) makes the figure look more like the normal cardinal vowel quadrangle, which is based more or less on auditory impressions of equal distances, and more like the figures found in more recent articles on auditory dimensions of vowels. Moreover, an auditory similarity test applied to the six Danish vowels [i: y: u: ɛ: œ: ɔ:] presented in triads gave the result that the answer to twelve out of 59 questions about relative similarity were in better agreement with a logarithmic scale than with the mel scale. There were no counter examples to this.

References

Elert, C.C. 1970: Ljud och ord i svenskan. (Uppsala).

Frøkjær-Jensen, Børge 1967: "The Danish Long Vowels", ARIPUC 1/1966, p. 34-46.

Frøkjær-Jensen, Børge 1968: "Statistical Calculations of Formant Data", ARIPUC 2/1967, p. 158-59.

Jørgensen, Hans Peter 1967: "Acoustic Analysis of Tense and Lax Vowels in German", ARIPUC 1/1966, p. 77-86.

Ladefoged, Peter, 1962: The Nature of Vowel Quality. Revista do Laboratório de Fonética Experimental. (Coimbra).

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

KINESTHETIC JUDGEMENT OF EFFORT IN THE PRODUCTION OF STOP CONSONANTS

SR:
8b. Lyd og lydsystem

Eli Fischer-Jørgensen

1. Introduction

Traditionally ptk are considered to be "stronger" than bdg irrespectively of the dominant phonetic difference: (i) voicelessness vs. voicing (in the narrow sense of vibrations of the vocal chords), (ii) aspiration vs. lack of aspiration, (iii) fortis vs. lenis (in the narrower sense of articulatory force) or (iv) a combination of two or all three of these differences.

In agreement with this tradition Jakobson-Fant-Halle (1952) and Jakobson-Halle (1956, 1962) combine the three differences under one feature "tense-lax", whereas Chomsky-Halle (1968) keep the three differences apart as three features, but with somewhat dubious phonetic descriptions.

In an earlier volume of this report (EFJ 1968a) I have given arguments for considering voicing, aspiration and tenseness (in the sense of fortis-lenis) as three independent phonetic features. According to this conception tense stops should be characterized by a longer closure period and a stronger organic pressure in the supraglottal cavity than lax stop consonants, whereas intra-oral air pressure is considered to belong mainly to the voicing feature.

Tenseness alone seems to be relevant in Swiss German stops, and aspiration alone in Danish stops, but I am not sure that the voicing opposition can be found without a concomitant difference of fortis versus lenis. In French the latter seems to be the primary feature (see e.g. Malmberg 1943). But even if three independent features must be recognized, the relations aspirated-unaspirated and fortis-lenis might still have something in common, which might be called "strength" in a vague sense. It is evident that acoustically aspirated stops are strong. Physiologically they are, however, not tense in the sense used here. But how is it from the kinesthetic point of view?

2. English stops

Malécot considers American English ptk to be stronger than bdg, not only in the vaguer sense of "strong", but in the more precise sense of "fortis", i.e. having a more tense articulation in the supraglottal cavity, (Malécot 1955). This assertion is based partly on a test in which 125 students were asked to pronounce English consonants in pairs in the environment a-a and decide which of the two (e.g. p or b) required more effort (1955), partly on physiological measurements.

In the psychological test ptk were on the whole indicated to require more effort than bdg. Malécot thinks that the answers were based entirely on the action of the supraglottal organs. He bases this hypothesis (1955) on measurements showing a higher intra-oral air pressure in ptk than in bdg, and on the finding of Rousselot that t has a higher tongue-pressure compared to d. (Like Jespersen and others Malécot assumes that a higher organic pressure is needed to maintain the hold against a stronger intra-oral pressure.) His hypothesis is supported by later measurements of duration showing that ptk have a longer duration of the closure than bdg, (Malécot 1966 a and 1966 b), and of organic pressure showing that pt have a tendency to higher organic pressure than bd, although this latter difference is not significant (1966 b). Similarly Harris-Lysaught-Schwey (1965) found a tendency to stronger EMG-activity of the lips in p than in b, but no stable difference. Lubker-Parris (1971) also found a certain tendency to higher organic pressure and higher EMG-activity in p than in b, and, like Malécot, they found a constant difference in intra-oral air pressure.

On the other hand Tatham-Morton (1968 a and b) did not find any difference in the activity of the orbicularis oris between British English p and b and in Lisker's investigation of American stop consonants a clear difference in intra-oral air pressure and duration was found only in the position after a stressed vowel, whereas there was hardly any difference initially in the syllable before a stressed vowel (Lisker 1965 and 1966).

Probably Lisker's subjects had voiceless bdg in this position, whereas Malécot's and Lubker-Parris' subjects had voiced bdg. This is not stated clearly by any of the authors, but it is true of the few curves they give as illustrations (Malécot 1966 a p. 68, Lisker 1966 5.4). It is well known that there is great variability of voicing initially in English, and it is important to know whether the stops in question were voiced, since, as I have shown earlier (EFJ 1963 and 1968 a), there is a close connection between intra-oral air pressure and voicing (but not between intra-oral pressure and organic pressure).

The instability of the length difference also appears from a small number of measurements of English stops which I made some years ago. Four speakers, two British and two Americans, spoke a series of words containing stop consonants medially before stressed i a u (of the type "the part, the peal, the pool, the bark, the bean, the boom"). Voicing and duration were measured on mingograms. All had partly or fully (75-100 %) but rather weakly voiced bdg in this position.

One of the British speakers (CB) spoke the list four times in different order, which gives 12 examples of each consonant. He had a significant difference of length between ptk and bdg with a mean difference of 23 msec for p/b, 26 for t/d, and 32 for k/g. The other three speakers spoke only 6 examples of each consonant. One of the American speakers had a mean difference of 14 msec between the durations of the closures of ptk and bdg, but complete overlapping. The two others had practically no difference. (There was, by the way, a clear tendency to have longer closures in labials than in dentals, and longer in dentals than in velars).

Now, to return to Malécot's assumption, we do not know whether the subjects used in his psychological test had voiced or voiceless bdg, and consequently we do not know whether they had higher air pressure in ptk than in bdg. Moreover, as the differences in closure duration and organic pressure are unstable, the conclusion that the subjects must have based their impression on the action of the supraglottal organs has very little foundation. On the other hand, we are pretty sure that they had aspirated ptk, and we might just as well draw the conclusion that their impression was based on this difference. A similar experiment with Danish stops might throw some light on this question.

3. Danish stops

Danish ptk and bdg are distinguished in syllable initial position only. Both are voiceless, but ptk are aspirated (and t affricated) whereas bdg are unaspirated. As for the difference fortis-lenis it is small and hardly of any perceptual importance, but phonetically bdg are slightly more fortes than ptk, in the sense that they have a tendency to higher organic pressure than ptk (this difference is significant for some subjects, but not for all)[1] and that there is a small, but stable and significant difference in the duration of the closure. As for intra-oral air pressure the pressure of ptk is only about 5 % higher than that of bdg, and only at the end

[1] An electromyographic investigation of the lip muscles is in progress.

of the closure,[2] which is less than the DL for kinesthetic judgement of air pressure found by Malécot (1966 a). Therefore, if Danish subjects find that ptk require more effort than bdg, the impression cannot be based on the supraglottal cavities.

In the years 1959-62 192 Danish students of philology (in their first term, before they had learnt anything about stop consonants) were asked to answer a questionnaire containing the following questions:

"I. Which of the two syllables in each pair requires greater effort in pronunciation? (underline the one that requires most effort, or put an equation mark if you cannot feel any difference).

a) ba or pa, b) da or ta, c) ga or ka.

II. In which of the two syllables do you apply more force to the airstream? (underline or put an equation mark)

a) ba or pa, b) da or ta, c) ga or ka.

III a) Are the lips pressed together with more strength and tension in ba or pa?

b) Is the tongue tip pressed against the upper part of the mouth with more strength and tension in da or ta?

c) Is the dorsum of the tongue pressed against the palate with more strength and tension in ga or ka?

(underline or put an equation mark)."

[2] See e.g. EFJ (1968 a). The measurements have not been published in detail, except for a bilingual subject (EFJ 1968 b). The difference of duration has been corroborated by the measurements of Danish p and b in Frøkjær-Jensen-Ludvigsen-Rischel (1971). Weaker organic pressure for aspirated stops has already been found by Rousselot for Armenian (1897 I p. 596), and the same was found for Gujarati (EFJ 1968 a, p. 96).

In an accompanying instruction it was emphasized that I was interested in their personal impression only, not in any prejudices or theories, which might probably be wrong, and they were asked to pronounce the pairs, eg. *ba*-*pa*, *ba*-*pa* a couple of times, and the same in inverse order: *pa*-*ba*, *pa*-*ba*, then *papapapa* and *babababa* at different speeds, and finally some word pairs like *pande*-*bande*, before making any decision. It was also emphasized that each question should be answered without regard to the others, and that it was by no means certain that the pairs would behave in a similar way. - Only 26 % gave the same answers to the labial, dental and velar pairs, which shows that they have really tried to make a personal judgement.

Two groups (60 subjects in all) were asked to answer some more specific questions about the articulation of *t* and *d*. As the answers to these questions required some phonetic training, it may be more or less due to chance that 55 % found that the tongue tip was more advanced in *d* than in *t* (whereas 33 % found it more advanced in *t*), and that 64 % found that the tip of the tongue was raised during the closure of *d* (and 31 % that it was lowered), whereas the percentages were equal for *t* (47-47). It can, however, not be due to chance that 85 % found that the teeth were closer together in *t* than in *d*.

The answers to the questions about effort gave the following results:

I Greater general effort

p 64 %	t 45 %	k 58 %
b 29 %	d 38 %	g 27 %
= 7 %	= 17 %	= 15 %

II More forceful airstream			II Stronger organic pressure		
p 83 %	t 69 %	k 76 %	p 28 %	t 17 %	k 20 %
b 11 %	d 15 %	g 11 %	b 60 %	d 78 %	g 64 %
= 6 %	= 16 %	= 13 %	= 12 %	= 5 %	= 16 %

In Fig. 1 the same information is given in graphical form.

It is evident that a great majority of the subjects feel that the airstream is more forceful in *ptk*, and the organic pressure stronger in *bdg*. In both cases there is a significant difference between *ptk* and *bdg*. This is in good agreement with physiological measurements of Danish stops.

Since *ptk* have a stronger airstream and *bdg* a stronger organic pressure, it is understandable that there is a less pronounced majority in the answers to question I about general effort. Nevertheless the differences between the reactions to *p* versus *b* and *k* versus *g* are significant. The uncertainty about *t-d* can be explained by the strong affrication of *t* which implies that it has a relatively weaker organic pressure and less strong airstream, and this is also reflected in the answers to questions II and III (*t* also has a shorter closure than *p* and *k*). The answers show that the feeling of general effort is influenced both by organic pressure and by airstream, but more by airstream. If the influence was about equal, we should expect *t* to have less general effort than *d*, and *p* and *k* to have a smaller majority for stronger effort.

In order to see whether there might be geographical differences in the answers, the subjects were divided into four groups according to geographical origin: (i) Copenhagen and suburbs (131), Zealand apart from Copenhagen (25), Funen (8), Jutland (28). The configurations were, however, almost the same in all groups with small variations only.

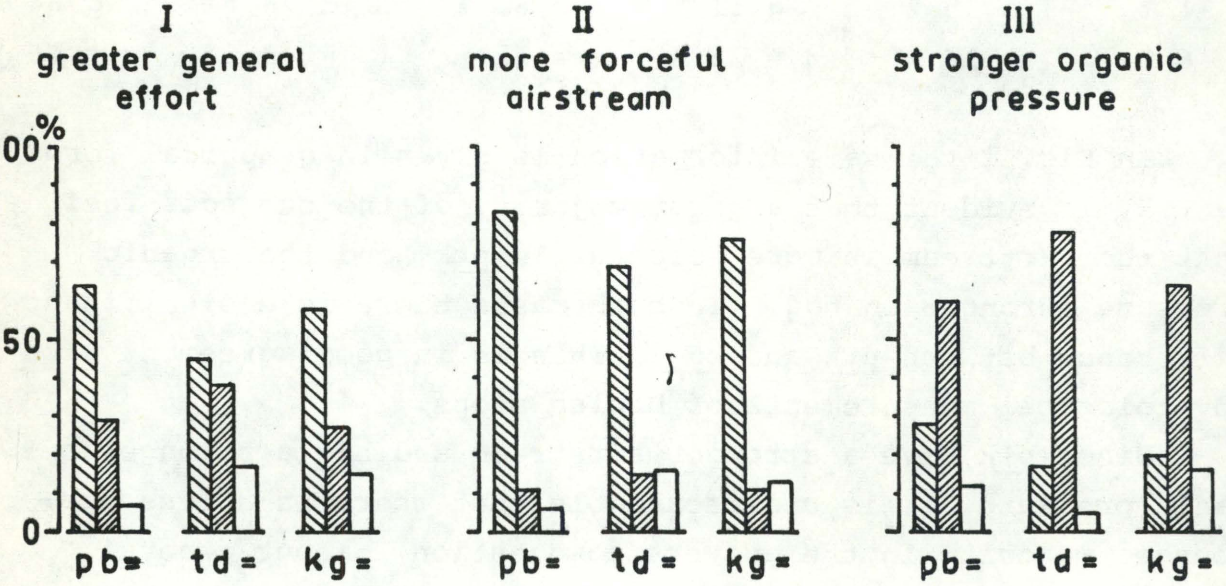
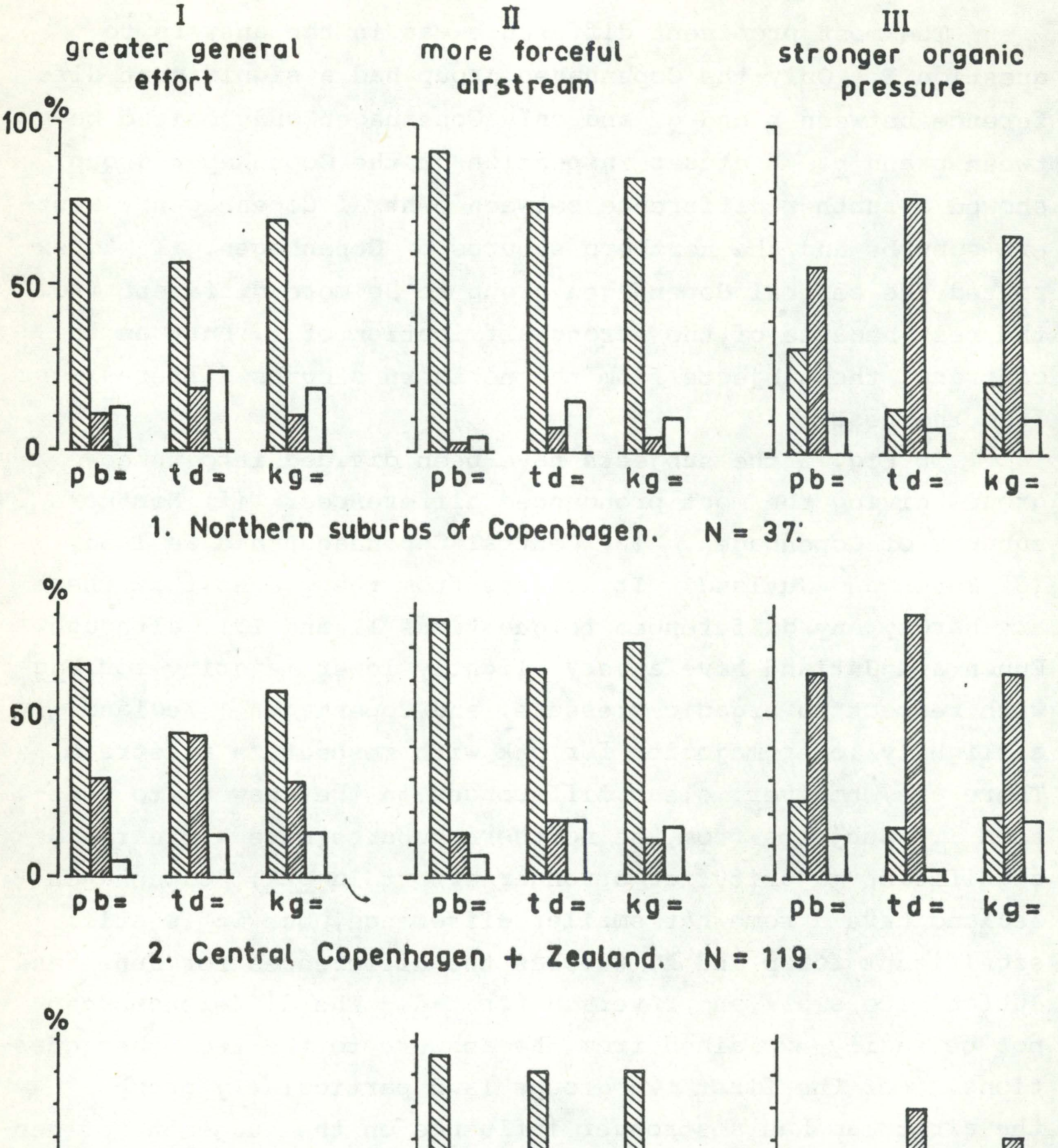


Fig. 1.
Answers to the kinesthetic test (in percentage). N=192.

The most prominent difference was in the answers to question I. Only the Copenhagen group had a significant difference between *k* and *g*, and only Copenhagen and Zealand between *p* and *b*. A closer inspection of the Copenhagen group showed a further difference between central Copenhagen + western suburbs and the northern suburbs of Copenhagen. I had expected the central Copenhagen group to be more different from the rest because of the strong affrication of *t*, but, on the contrary, the subjects from the northern suburbs differed most from the rest.

In Fig. 2 the subjects have been divided into three groups giving the most pronounced differences: (1) Northern suburbs of Copenhagen, (2) Central Copenhagen and Ze land, (3) Funen and Jutland. - It appears from the graphs that there are hardly any differences to questions II and III, although Funen and Jutland have a very slightly lower majority for *bdg* with respect to organic pressure, and Copenhagen + Zealand a slightly lower majority for *ptk* with respect to airstream. There are, however, clear differences in the answers to question I. Subjects from the northern suburbs have a clear and significant majority for stronger effort in *ptk*; Copenhagen + Zealand have a somewhat smaller difference, but it is still significant for *p* and *k*, whereas the differences for Funen and Jutland are small and reversed for *t-d*. - The differences cannot be easily explained from the answers to the two other questions. For the first two groups (and particularly group 1) the airstream has a stronger influence on the judgement of general effort than the organic pressure, for group 3 (Funen and Jutland) the airstream does not seem to have a decisive influence on the judgement of general effort, although it is felt very clearly as a separate phenomenon (question II). Group 2 (central Copenhagen) has a strong affrication of *t* and partly of *k*, which impedes the airstream, whereas members of group 1 have often less affrication and therefore a less impeded air-



1. Northern suburbs of Copenhagen. N = 37.

2. Central Copenhagen + Zealand. N = 119.

3. Jutland + Funen. N = 36.

Fig. 2. Answers to the kinesthetic test (in percentage) divided into three geographical groups.

stream, which may have dominated the impression of general effort. The answers of group 1 to the question of overall effort may be taken as a support of the view that this group is characterized by a reaction against the specific features of Copenhagen speech (affrication, and also some features of vowel quality) a view which is contended by some phoneticians. As the consonants of the subjects have not been recorded, this must remain a hypothesis.

4. General discussion

The main result of the experiment with Danish stops, i.e. that most of the subjects consider ptk to require more effort than bdg, and that this answer cannot be explained by any fortis features of the supraglottal cavity, makes it very improbable that Malécot's subjects should have reacted to the activity of the supraglottal cavity only.

The result of the Danish test also gives some support to the traditional designation of ptk as stronger than bdg, not only acoustically, but also kinesthetically, even in cases where they have less energy in the supraglottal cavities.

It is, however, a problem why the stronger airstream is felt as an effort. In question II the subjects were asked to indicate whether they put more force to the airstream in pa than in ba etc. The answer "ptk" can, however, hardly be interpreted to indicate more than the feeling of a stronger airstream. But question 1 was a direct question concerning articulatory effort, and, as demonstrated above, the answers to this question were clearly influenced by the presence of a stronger airstream. Now, if aspiration, as generally assumed, is only due to the fact that the glottis is open during the closure and has not yet been closed at the moment of explosion, then the feeling of effort should be due to laryngeal adjustments. Frøkjær-Jensen-Ludvigsen-Rischel (1971) have brought evidence for the assumption that the slight opening of the glottis in b can be explained by aerodynamic conditions, whereas the large opening in p must require a new neural command. The former is called a "passive opening-closing-gesture", and the latter an "active opening-closing-gesture". According to Rischel (personal communication) the terms "active" and "passive" were meant to imply only that p clearly involves a new neural command, whereas no such assumption seemed necessary for b on the basis of the glottographic material used in the said investigation. A very preliminary processing of some limited electromyographic recordings, made in collaboration with H. Hirose at the Haskins laboratories, does not seem to corroborate the assumption of the passive b-gesture. Both p and b show a clear relaxation of the interarytenoid muscle and an increased activity of the posterior cricoarytenoid, although both phenomena are generally more pronounced and of longer duration in p than in b (this is particularly true of the relaxation of the interarytenoid). The difference between p and b is smaller and the difference between b and the surrounding vowels greater than in the English examples described by Hirose (1970). It is not very probable that this relatively modest difference of activity should be felt as a clear difference in effort. Chomsky-Halle (1968) make the assumption that aspirated stops have heightened subglottal pressure, but I do not know of any physiological evidence for this assumption. As there is hardly any resistance in the glottis, the subglottal pressure must equal the supraglottal pressure in ptk, a pressure which is almost the same as in bdg. It is difficult to see any reason why the pressure below the glottis should be lower in bdg. But it is possible that the simple fact that the speaker loses more breath in syllables with ptk, particularly if they are repeated, is felt as a strain and contributes to the impression of effort. An investigation of respiratory muscles might perhaps throw some light on these problems.

References

Abramson, A.S. and Lisker, Leigh 1967: "Laryngeal Behavior, the Speech Signal and Phonological Simplicity", Haskins SR 11, p. 23-33.

Chomsky, N. and Halle, H.1968: The Sound Pattern of English (New York, Evanstone, London).

Fischer-Jørgensen, Eli 1963: "Beobachtungen über den Zusammenhang zwischen Stimmhaftigkeit und intraoralem Luftdruck", Zs. f. Phon., Sprachw. und Komm.f. 16, p. 19-36.

Fischer-Jørgensen, Eli 1968a: "Voicing, Tenseness and Aspiration in Stop Consonants, with Special Reference to French and Danish", ARIPUC 3, p. 63-114.

Fischer-Jørgensen, Eli 1968b: "Les occlusives françaises et danoises d'un sujet bilingue", Word 24, p. 112-153.

Frøkjær-Jensen, B., Ludvigsen, C. and Rischel, J.1971: "A Glottographic Study of some Danish Consonants", Form and Substance, p. 123-140.

Harris, K.S., Lysaught, G. and M.C.Schwey 1965: "Some aspects of the Production of Oral and Nasal Labial Stops", Haskins SR 2.1.

Hirose, H. 1971: "An Electromyographic Study of Laryngeal Adjustment during Speech Articulation. A Preliminary Report", SR Haskins 25/26, p. 107-116.

Jakobson, Roman, Fant, G. and M. Halle 1952: Preliminaries to Speech Analysis (MIT Technical Report No. 13).

Jakobson, Roman and Halle, M.1956: Fundamentals of Language (Mouton, s'Gravenhage).

Jakobson, Roman and Halle,M. 1962: "Tenseness and Laxness", Selected Writings I, p. 550-555.

Lisker, L. 1965a: "Supraglottal Air Pressure in the Production of English Stops", Haskins SR 4.3. and Language and Speech 13, p. 215-230.

Lisker, L. 1966: "Measuring Stop Closure Duration from Intra-Oral Pressure Records", Haskins SR 7/8.5.

Lubker, J. and Parris, O.J.1971: "Simultaneous Measurements of Intra-oral Pressure, Force of Labial Contact and Labial Electromyographic Activity during Production of the Stop Consonant Cognates p and b", JASA 27, p. 625-633.

Malécot, A. 1955: "An Experimental Study of the Force of Articulation", Studia Linguistica 9, p. 35-44.

Malécot, A. 1966a: "The Effectiveness of Intra-Oral Air-Pressure-Pulse Parameters in Distinguishing between Stop Cognates", Phonetica 14, p. 65-81.

Malécot, A. 1966b: Mechanical Pressure as an Index of 'Force of Articulation'", Phonetica 14, p. 169-180.

Malmberg, M. 1943: Le système consonantique du français moderne. Etudes de phonétique et de phonologie. Etudes romanes de Lund VII.

Rousselot, J.P. 1897: Principes de phonétique expérimentale I (Paris), 2.e éd. 1924, p. 596 ff.

Tatham, M.A.A. and Morton, K.1968a: "Some Electromyographic Data towards a Model of Speech Production", University of Essex, Language Centre, Occasional Papers 1, p. 1-24.

Tatham, M.A.A. and Morton, K.1968b: "Further Electromyographic Data towards a Model of Speech Production", ibid., p. 25-59.

Tatham, M.A.A. and Morton, K. 1968a. "Some Electromyographic Data towards a Model of Speech Production", University of Essex Language Centre Occasional Papers 1, p. 1-24.

Tatham, M.A.A. and Morton, K. 1968b. "Further Electromyographic Data towards a Model of Speech Production", ibid., p. 25-39.

SRi
lyd og lydsystem

PERCEPTUAL STUDIES OF DANISH STOP CONSONANTS

Eli Fischer-Jørgensen

The experiments upon which these studies are based were carried out in the years 1954-64, (mainly 1954-56), and most of the material was processed at that time. The publication was postponed because the studies were planned to form part of a monograph on Danish stop consonants including acoustic, physiological and perceptual descriptions, a plan which I later had to give up. A summary of the acoustic results was given in the paper "Acoustic Analysis of Stop Consonants" Miscellanea Phonetica II, 1954, p.42-59. The main part of the perceptual results are published on the following pages, and the results of the physiological measurements will be published at a later date.

I am grateful to the students who helped me with the processing of the data, particularly to Henny Pontoppidan Lauritzen, Brita Waage-Bech, Lise Preisz and Rune Jacobsen.

I also want to express my gratitude to Statens Videnskabsfond, whose support (and whose patience) has made this work possible.

PERCEPTION OF FOREIGN STOP CONSONANTS BY DANISH LISTENERS

Eli Fischer-Jørgensen

1. Introduction

Danish ptk and bdg are distinguished only syllable initially. Both sets are voiceless. ptk are articulated with somewhat less energy in the supraglottal cavities than bdg, but the main difference is one of aspiration. Earlier measurements (EFJ 1954) of the duration of the open interval (in the sense of the distance from the explosion to the beginning of the vowel, including affrication and aspiration) have given the result that the average in stressed position is 66 ms for p, 79 for t and 74 for k (with individual averages ranging from 53 for p to 98 for t) and 14, 17 and 23 ms for b, d, g respectively. The open interval of t is slightly longer than that of k in Danish because of the strong affrication of t. The order of the unaspirated stops (labial-dental-velar) is more normal. As in other languages, the open interval is normally somewhat longer before close vowels than before open vowels. It is rarely longer than 35 ms in bdg and shorter than 40 ms in ptk.

It might be expected that these durations would influence the perception of stops in foreign languages. This was tested by experiments carried out in 1958 and 1961. The results have not been published until now.

2. The material

The material used in the test consisted of 137 different CV syllables (where C is a stop consonant, and the vowel mostly i, a or u) cut out of words spoken by five Danish, one German, two British English, one French, one Dutch, one Chinese, and four speakers of Indian languages, viz. Urdu, Hindi, Marathi, and Malayalam. The recordings were made earlier for other purposes. The words were chosen so as to be representative of the languages (but the English speakers had relatively short aspirations) and, as far as the Danish stops and the Indian voiced aspirates are concerned, to cover a certain range of variation. The voiced aspirated stops from Indian languages represented various types. Most of them were totally voiced (though often with a very short cessation of vibrations at the explosion, a few had voiceless closure or voiceless aspiration. Some of the Danish examples were chosen to represent unusually short aspirations (35 and 40 ms for p).

The English, German and Danish words had been spoken in a context consisting of a preceding word ending in an unstressed vowel. The English bdg-sounds were voiced, the Danish and German ones voiceless. The different types of stops represented were: (a) voiced bdg 23 (English, French, Dutch, Indian), (b) voiceless bdg 29 (Danish, German), (c) unaspirated ptk 28 (French, Dutch, Indian, Chinese), (d) aspirated ptk 55 (Danish, English, German, Indian, Chinese), (e) voiced aspirated 18 (Indian), and, (f) aspirated affricates 2 (Chinese).

These syllables were combined into a test tape (each syllable was repeated three times). The signal to noise ratio was not quite good in the Danish, German and some Indian examples, but this does not seem to have influenced the results. The order was quasi-random. The French examples of ptk (nine in all) were repeated in different environments, once after a syllable with bdg, once after a syllable with aspirated ptk, in order to check the influence of the context. There was a slight influence in four of the nine examples, but only in one case was the difference significant.

Spectrograms and mingograms (including intensity and pitch curves) were taken of all examples. Moreover four phoneticians listened to the test and characterized the consonants in terms of voicing, aspiration and impression of fortis-lenis.

3. The listeners

The test tape was played over a loudspeaker in a sound-treated room to groups of listeners. The listeners, 82 in total, were Danish students of philology in the beginning of their first term. All of them were familiar with English, and most of them with German and French, but they had not yet learnt any phonetics. More naive subjects might have been preferable, but as it can be seen from the results, they were not much influenced by their knowledge of foreign languages. They were asked to identify the initial consonants of the syllables with one of the six Danish stops, and to write down the result on answer sheets.

4. The results

The main results (apart from the answers to voiced aspirates) can be seen in Fig. 1. The examples with a majority of bdg-answers are placed above the base line, the examples with a majority of ptk-answers below the line. The horizontal scale indicates the duration of the open interval and the vertical scale the number of examples. The various stop categories are distinguished on the graph.

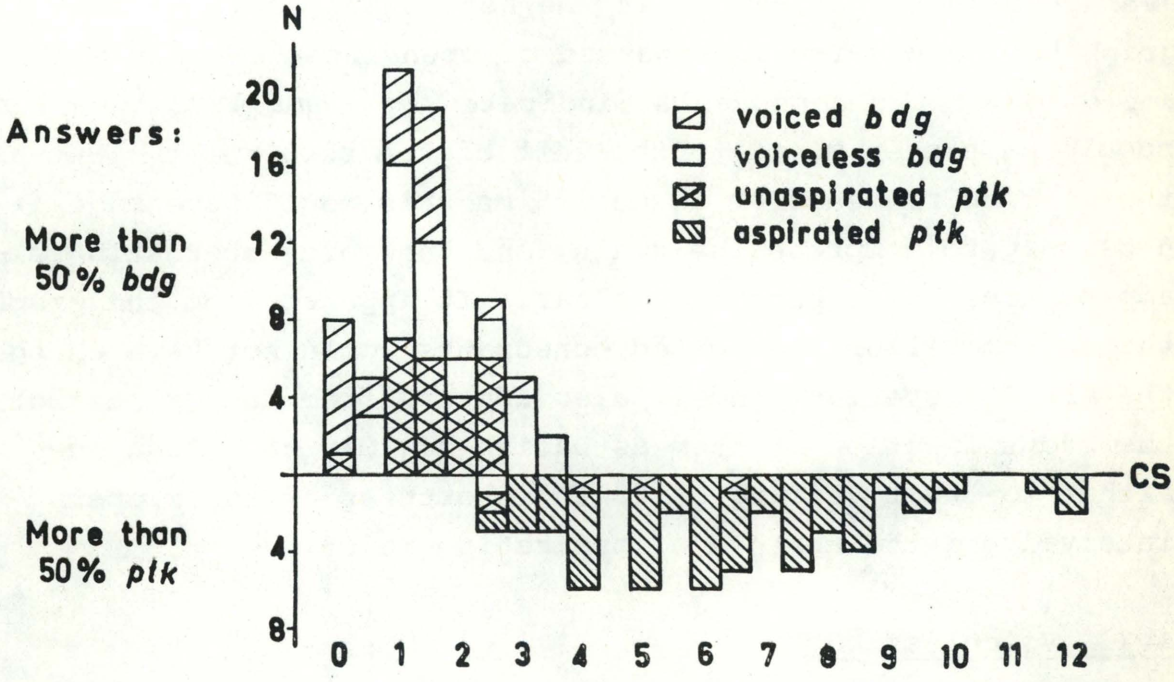


Fig. 1. Identification of foreign stop consonants by Danish listeners compared to the duration of the open interval.

4.1. Voiced stops

All voiced stops were identified as Danish bdg with an average majority of 99 %. If the horizontal scale of the graph had been VOT-value instead of open interval, the examples of voiced consonants (indicated by a small oblique line) should have been moved to the left of the zero point. But as the voicing ranges from 30 to 195 ms this would have made it necessary to compress the scale, and the more interesting examples would have been less clear. It appears from the graph that a removal of the voiced consonants would not have changed the area of overlap, and it also appears from the graph that the identification of them as Danish bdg may have been due either to their voicing or to the shortness of their open interval and the absence of aspiration noise.

4.2. Voiceless bdg

All voiceless bdg-stops were identified as Danish bdg. The average for all examples was 93 % bdg (with the exception ga left out it was 95 %). This means that voicing is not a necessary cue; the small difference in open interval is sufficient for explaining the slight difference in the majority (99 and 95 %) for the two categories. If the voiceless bdg-sounds are divided into two groups (5-20 ms and 25-35 ms) the averages are 99 and 87 %, respectively.

The only exception to the bdg-answers is a Danish example of ga (subject OT) with an open interval of 25 ms. We shall return to this problem below.

4.3. Unaspirated ptk

The unaspirated ptk-sounds were heard by the majority as Danish bdg in 24 out of 28 examples. It appears from the graph that three of the four exceptions can be explained by a much longer open interval. These were Dutch ki (65 ms, 99 % k), French ku (50 ms, 98 % k) and Indian ka (40 ms, 59 % k). Only one example (French pu 25 ms, 77 % p) has overlapping with the examples identified as bdg. The average majority for those identified as bdg is 90 %. The average bdg-majority for all unaspirated stops is 80 %.

4.4. Aspirated ptk

All instances of aspirated ptk were identified with Danish ptk, and as it can be seen in Fig. 1, their open interval in 48 out of 55 examples was longer than that of the other stops. 7 examples overlap with the stops heard as bdg. The two consonants with 12 ms open interval were Chinese affricated aspirates.

4.5. Overlapping cases

The great majority of answers can be explained by the length of the open interval. But there is a number of overlapping cases between 25 and 35 ms. The individual examples are given together with the answers in tables I and II.

Table I

Overlapping of Labials

answers	25 ms	30 ms	35 ms
b		(Da.) by 98 % (G.) bu 65 %	
p	(Fr.) pu 77 % (E.) pa 70 %	(E) pi 100% (Da.) pa 100% (Da.) pa 100%	(Da.) pi 82 % (Da.) pi 90 %

Table II

Overlapping of velars

answers	25 ms	30 ms	35 ms
g	(E.) gi 94 %	(E.) gu 82 %	(G.) gu 95 %
	(Da.) gy 76 %	(Da.) ga 90 %	(Da.) gu 88 %
	(G.) ga 71 %	(G.) gi 96 %	
	(I.) ka 100%		
	(Ch.) ka 98 %		
	(Du.) ku 99 %		
	(Fr.) ka 81 %		
	(Fr.) ke 52 %		
k	(Da.) ga 74 %		(E.) ka 93 %

The 15 consonants heard as b or g in this area are all sounds generally characterized as bdg or unaspirated ptk. The 9 consonants heard as p and k are, with two exceptions, sounds normally characterized as aspirated ptk. This seems to indicate that a supplementary criterium besides length of open interval must be aspiration noise. It cannot be tenseness in the sense of fortis articulation, since Danish ptk are not fortis in this sence, whereas French ptk are, and since French ptk are normally heard as bdg, whereas Danish ptk are heard as ptk.

As for the labials the Danish examples were chosen as extreme values both of b and p, and I wanted to see whether they were identified correctly. As a matter of fact they were. A check on the stimuli by inspection of curves and by listening reveals that the p-explosions are followed by a certain aspiration, whereas the Danish b-sounds are not (moreover, by had some voicing at the beginning of the closure).

The German bu has a very slight aspiration noise and has only 65 % majority. There is also aspiration noise after the two English p-examples, and French pu. Labials outside the area of overlap have normally more than 90 % majority in the answers. There is however an example of Dutch pu (15 ms) with only 57 % majority for b. Two Danish examples of pa (35 and 25 ms) were left out because they were so weak that they were mostly heard as f.

There is no overlapping of t and d, which may partly be due to the fact that there are no examples of 30 and 35 ms. All answers to dental stops below and above these values show more than 90 % agreement on d and t respectively, except for French tu (25 ms, 52 % d), and French te (20 ms,52 % d). There is one other example of unaspirated t at 20 ms (Chinese ta heard as d with 91 % majority);it has a weaker explosion.

As for the velars, the boundary seems to be slightly higher than for the labials (about 35 for velars and 25-30 for labials), which is in agreement with the difference observed in the actual length of the open interval. English ka (35 ms) has some aspiration. The Danish example ga (25 ms, 74 % k) is astonishing. It is not aspirated, but it has a strong explosion and is characterized by the phoneticians as fortis in contradistinction to ga (30 ms), spoken by a different Danish speaker and heard as ga. It has a pitch fall in the start, but so has Dutch and Chinese ka, which are heard as ga. Similarly, the Danish ga of 30 ms, heard as ga, has a slight fall in pitch at the start. On the whole, it does not seem possible to explain the overlapping cases on the basis of pitch differences.

Only the French speaker has a clear difference. Nor can they be explained by differences in formant transitions. In the exceptional Danish ga-example the intensity of the explosion may play a role. (But the explosion of German ga (25 ms) heard as ga seems to be just as intense).

4.6. The answers of the phoneticians

Four phoneticians listened to the test tape and characterized all examples as *ptkbdg* and moreover as aspirated or unaspirated, voiced or unvoiced and fortis or lenis. In view of the small number of listeners it has no sense to give a detailed account of the answers.

The most interesting question is whether the phoneticians were able to distinguish between voiceless *bdg* and unaspirated *ptk*. It turned out that they made a distinction, but it coincided only partly with the expected distinction between German and Danish *bdg* on one hand and French, Dutch, Indian, and Chinese *ptk* on the other (the Chinese sounds are, by the way, often described as voiceless *bdg*). Of the 112 answers to Danish and German *bdg* there were 23 % *ptk*-answers and 77 % *bdg*-answers, and of the 116 answers to unaspirated *ptk* there were 53 % *ptk*-answers and 47 % *bdg*-answers (for the French stops 69 % *ptk*-answers). The answers to fortis-lenis were not quite the same, since particularly one of the listeners often combined *bdg*-answers with fortis-answers. There were 37 % fortis- and 62 % lenis-answers to German and Danish *bdg*, and 60 % fortis- and 40 % lenis-answers to unaspirated *ptk* (for the French stops there were 80 % fortis-answers). This means that there was a difference in the answers to the two groups, both in the designations as *bdg* or *ptk* and in the reaction to fortis-lenis, and this difference is statistically significant, but there is great overlapping.

As regards the designations as *bdg* or *ptk* the length of the open interval has some influence. There are hardly any *ptk*-answers to sounds with 5-10 ms open interval, and no *bdg*-answers to the three sounds with 40, 50 and 65 ms open interval, but from 15 to 35 ms there is complete overlapping for the unaspirated *ptk*, and from 25 to 35 for voiceless *bdg*. For the values 15 and 20 ms there are very few ptk-answers to voiceless bdg, but many to unaspirated ptk. There must therefore be some further criteria. On the whole, k-answers are more frequent than p- and t-answers. This means that the longer open interval of velars influences the judgement. (One might have expected a compensation in the perception because the longer duration of the open interval in velars is a universal, mechanical phenomenon).

The answers to the fortis-lenis question are much less dependent on VOT-values for the consonants in question.[1] There are evidently other cues. The number of examples and listeners, and the irregularities in a material consisting of natural speech do not permit any clear decision concerning these cues. There do not seem to be differences in pitch within the vowel or in formant transitions, corresponding to the answers, but there seems to be a certain correlation with the intensity of the explosion. This needs further investigation.

4.7. Aspirated b d g

The answers to aspirated Indian bdg have not been included in Fig. 1 because they would spoil the neat distribution.

There are 18 examples. Five of these were heard as bdg, but not with a very great majority (from 51 to 79 % with an average of 62 %). The remaining 13 were identified as Danish ptk, with a majority varying from 56 to 100 and an average of 83 %).

The duration of the open interval (i.e. the aspiration) is not decisive here. The five examples with a bdg-majority have

[1] The aspirated stops are characterized as fortes. This may depend on the VOT-value, but it is more probable that it is due to the intensity of the aspiration noise.

an aspiration ranging from 40 to 100 ms with an average of 77 ms, and the thirteen examples heard as ptk have an aspiration ranging from 30 to 110 with an average of 64 ms. The two examples with a clear bdg-majority have 100 and 40 ms aspiration, the four with unanimous ptk-identification have 45, 60, 70 and 80 ms.

Voicing may be of some importance. The five examples with a bdg-majority are fully voiced (except for an interval of 10 ms at the explosion in one of the examples).

As for the 13 examples which gave ptk-answers, three are fully voiced, six have a short voiceless pause at the explosion of about 10-20 ms, three have voiceless closure, and one has voiceless aspiration. But there is no evident correlation between the degree of voicing and the percentage of bdg-answers. There are seven examples with more than 88 per cent ptk-answers; two of these are fully voiced.

The intensity of the aspiration noise seems more important. Of the seven examples with the highest number of ptk-answers (above 88 %) six are characterized by strong noise, of the other cases only one has strong noise. (The intensity has not been measured but only judged from spectrograms and intensity curves.)

4.8. Conclusion

When Danish listeners are asked to identify foreign stops as Danish ptk or bdg, their main criterion is aspiration, which plays a role both in terms of the duration of the open interval and in terms of the degree of noisiness. Voicing plays a very subordinate role, if any.

Apart from the Indian voiced stops, for which the aspiration noise seems to be the most important cue, the duration of the open interval shows a rather clear boundary, which is located somewhere in the range 25-35 ms according to the place of articulation. This is in agreement with measurements of natural

Danish consonants. For voiceless sounds this measure is the same as VOT-value, and the Danish values are in good agreement with the VOT-values found by Lisker and Abramson for American English in experiments with synthetic speech (see e.g. Abramson and Lisker 1965 and Lisker-Abramson 1970). The phoneticians were, partly, able to distinguish voiceless _bdg_ and unaspirated _ptk_.

References

Abramson, A.S. and Lisker, Leigh 1965: "Voice Onset Time in Stop Consonants: Acoustic Analysis and Synthesis", SR. Haskins 1, 1-7.

Lisker, Leigh and Abramson, A.S. 1970: "The Voicing Dimension. Some Experiments in Comparative Phonetics", Proceedings 6, Int. Congr. Phon. Sc. Prague 1967, p. 563-67.

Fischer-Jørgensen, Eli 1954: "Acoustic Analysis of Stop Consonants", Miscellanea Phonetica II, p. 43-59.

IDENTIFICATION OF UNRELEASED DANISH STOP CONSONANTS

Eli Fischer-Jørgensen

1. Experiments 1 A and B (1954-56)

1.1. Material and listeners

A series of Danish nonsense syllables of the structure VC and CVC, where V was short *i*, *ɛ*, *a*, *u*, and C was *b*, *d* or *g*, were spoken by me and recorded on tape, in 1954. In experiment 1A there was one example of each of the four vowels before each of the three consonants, thus 12 VC syllables in all. In experiment 1B the same four vowels were found after and before the same three consonants, which gives 36 different CVC syllables. Initial *bdg* were voiceless, final *bdg* almost voiceless, with a few periods of weak voicing in the start. The vowel length was mostly 90-100 ms, in a few cases of *a* 110-120, and in some cases of *i* or *u* only 70-80 ms.

The syllables were combined into two tests (A (VC) and B (CVC)), where each syllable was presented once. The order of the syllables was random, and the distance between the syllables was 4-5 sec. The test was given to three groups of students of philology in their first term in the years 1954, 1955 and 1956. There were 94 listeners in all. This means that there were 94 answers to each vowel-consonant combination in test A, and 282 in test B, i.e. 4512 answers in all.

The listeners were told that they were going to hear a list of meaningless syllables containing the vowels *i*, *a*, *u* + unreleased *b*, *d* or *g* in test A, and *bdg* + the same vowels + unreleased *bdg* in test B. They were asked to write down what they heard and allowed to put a questionmark if they could not identify the consonant. The syllables had been recorded on a professional tape recorder but were played back on a semi-professional tape recorder via a loudspeaker. The quality was not very good, but as a control the same listeners were asked to identify three additional lists, viz. two CV lists containing ptk or bdg + the vowels i, ε, a, u, and one VC list containing the same vowels + released ptk. There were hardly any mistakes in these lists, and it was also extremely rare that there were mistakes in the initial consonants in test B. The mistakes in the final consonants in test A and B must, therefore, be due to the lack of explosions.

1.2. Results

The results of the two tests are given in Table I and Table II and in graphical form in Fig. 1.

TABLE I

Answers (in percentage) to test 1 A (VC)

Answers:	b	d	g	?		b	d	g	?		b	d	g	?
ib	90	3	0	7	id	24	40	11	25	ig	43	10	35	11
εb	65	23	5	8	εd	11	74	9	7	εg	9	5	78	8
ab	77	2	16	4	ad	25	58	12	5	ag	1	1	91	7
ub	70	4	13	13	ud	4	88	1	7	ug	37	2	46	15

TABLE II

Answers (in percentage) to test 1B (CVC)

Answers:	b	d	g	?		b	d	g	?		b	d	g	?
ib	92	2	1	5	id	55	30	6	9	ig	46	16	26	13
ɛb	75	12	4	9	ɛd	31	33	21	12	ɛg	6	8	76	11
ab	60	7	24	9	ad	19	40	32	9	ag	7	1	86	6
ub	42	7	37	14	ud	2	91	3	5	ug	20	4	58	18

"?" in the tables covers the answers: questionmark, minus sign, or no indication of the consonant in question or of the whole syllable, e.g. bu?, bu÷, bu or nothing for bud.

Some subjects have written ptk finally instead of bdg. As there was, of course, no aspiration at the end, this is simply an influence from orthography. Phonologically there is no distinction between [p] and [b] finally.

On the whole the answers to test A (the VC list) are better than the answers to test B (the CVC list). The percentage of correct answers for all environments taken together are in test 1 A b 76 %, d 65 % and g 63 %, in test 1B b 70 %, d 49 % and g 62 %. The difference is small for b and g, but considerable for d. Particularly the syllables with d-d are heard incorrectly in test 1 B (only 36 % correct answers). One of the purposes of list B was to find out whether there would be perceptual dissimilations in the answers. This seems to be the case for d-d, but there is no clear tendency for the syllables with b-b and g-g. This difference may, of course, be due to chance, but the question deserves further investigation.

b after u is heard better in test A than in test B. In test B it is often heard as g, whereas in test A g is often heard as b in this environment.

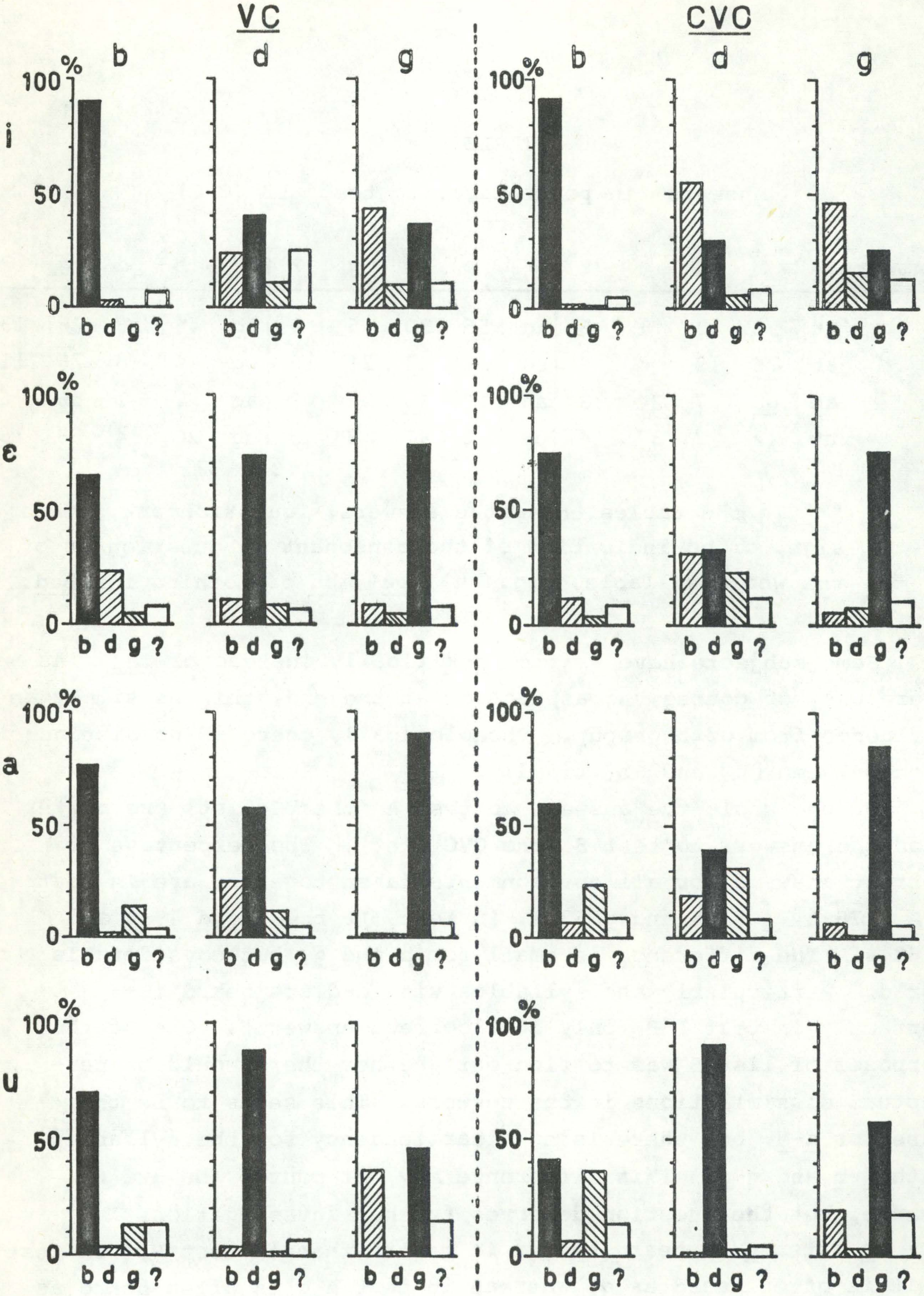


Fig. 1. Identification of final unreleased *bdg*. Test IA (VC) and IB (CVC).

Apart from these differences the answers to test A and B have many features in common. The main tendencies common to both are that b is identified most correctly after i, d after u, and g after ɛ and a.

Repeated listening has revealed a very slight explosion noise in id and ud in test A, but not in test B. id is heard somewhat better in test A, but there is no difference between the answers to ud in test A and B. This slight explosion noise can therefore not have improved the tendencies found in the answers.

The differences just mentioned between b, d and g can be easily explained when the extent of the transitions is taken into account. The transitions of F2 and F3 are particularly clear in ib and almost non-existent in ub (which has the lowest percentage in test B). Before d, on the other hand, the vowel u has an extensive transition, but i very little, and before g, ɛ and a have a characteristic converging transition of F2 and F3, whereas the transitions are slight in u and i.

The high number of correct g-answers after a is not astonishing, since the quality of the whole vowel is influenced in this case (a is more retracted). For this reason ɛ has been included in the test as a better representative of open vowels, but the answers to ɛg are not much poorer than the answers to ag. a is also somewhat more retracted before b than before d, not so much in my speech as in Copenhagen speech, but clearly visible in the spectrograms.

If the examples with a are kept aside, the percentage of correct answers for g are somewhat lower: 54 and 53 % instead of 63 and 62 in tests A and B.

The confusion between b and g after u is explicable from the fact that they have a similar and very small transition of F2. After i and to a lesser extent after ɛ, d is often heard as b. This can be explained by the common direction of their F_2 transition. As the extent of this transition is larger in b than in d, it is possible to hear b for d rather than the opposite. Such an explanation can, however, not be given for the cases of g heard as b after i. On the whole, it seems that the labial is often chosen in cases of doubt. It is, in some sense, the most neutral consonant.

2. Experiments 2 A and B (1964)

In 1964 I undertook a new experiment with VC syllables, this time including all the Danish short vowels i y u e ø o ɛ œ ɔ a. Each of the syllables with these vowels followed by unreleased b, d and g were spoken twice by me and by the speaker JR. The tests were run in the same way as in experiment I. My recording (A) was played to two groups, one consisting of 20 students in their first term, and the other of 10 phoneticians. The recording of JR (B) was only played to the 10 phoneticians. The number of correct answers to test A was for the student group b 67%, d 64% and g 48%, which was not quite as good as test 1. For the phoneticians the number of correct answers was higher: b 79%, d 78% and g 58%. In test 2 B (spoken by JR and played to the phoneticians only) the number of correct answers was still higher: b 76%, d 82% and g 88%. The difference is very large for g. One of the reasons why JR's recording gave more correct answers is that he spoke somewhat more slowly than I did. His vowels were 10-20 ms longer, and he had a less steep rise and decay of the intensity. Spectrograms of the syllables with g show more pronounced transitions in some cases.

The answers to the syllables with i ɛ a u can be compared with test 1. In Fig. 2 the percentage of correct answers to the syllables with these four vowels are given for all the tests. The stimuli for tests 1 A and B and test 2 A were

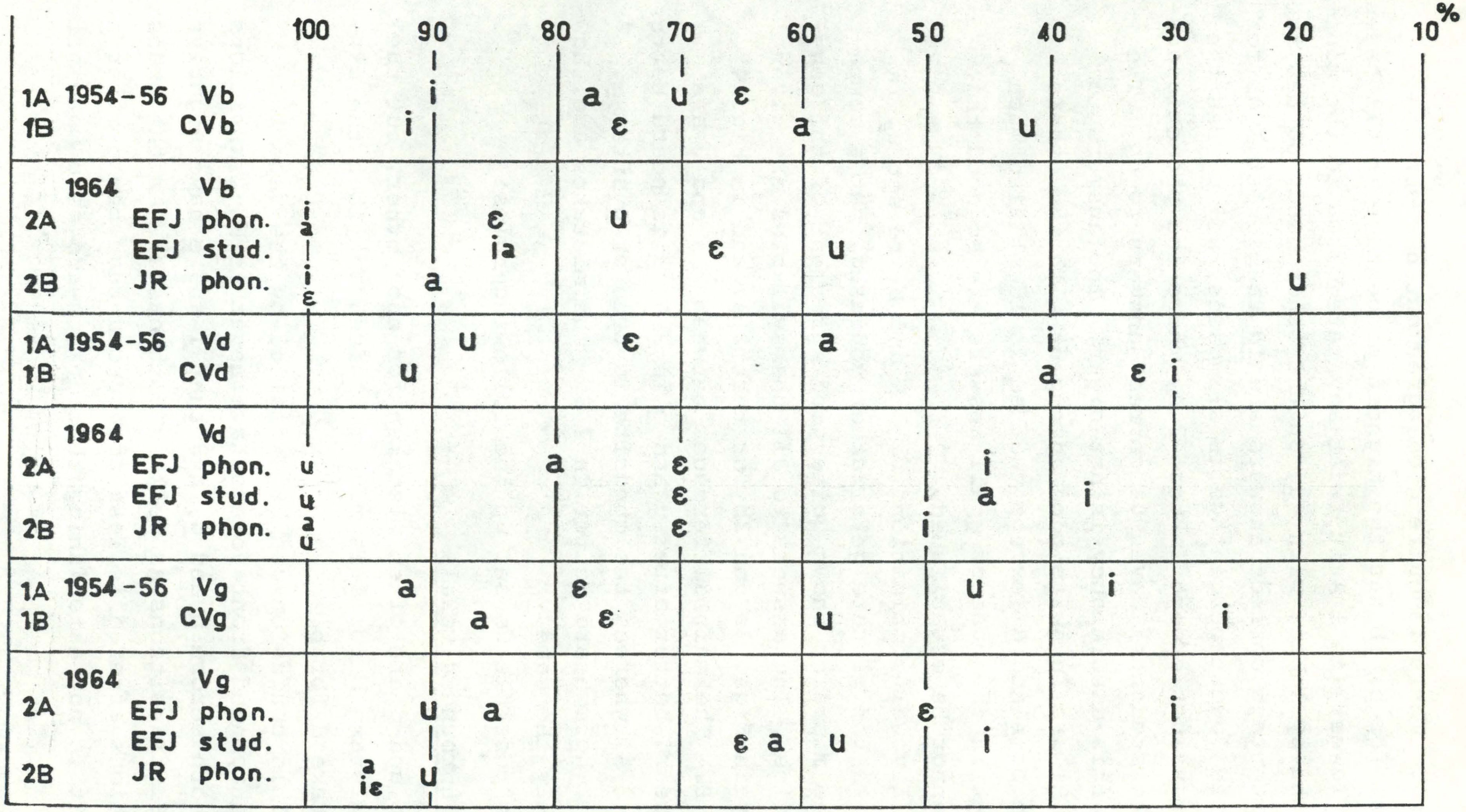


Fig. 2. Percentage of correctly identified unreleased *bdg* in Test 1AB and Test 2AB.

spoken by me. There is quite good agreement between the answers. As for b, ib has the highest percentage in all cases and ub thé lowest (in 1 A it is at any rate close to the minimum). In the syllables with -d, on the other hand, ud has the highest percentage correct answers and id the lowest. As for g, ig has the lowest percentage in all cases, and ug has the next lowest values except in test 2 A, where the phoneticians (but not the students) have 90% correct answers to ug. The reason for this discrepancy is not clear. The answers to test 2 B (JR) are in quite good agreement with the others in the case of b and d (except that ad is identified much better) but in the case of g all answers are so good that an order cannot be established.

If all vowel combinations in 2 A and B are set up in this manner, the picture gets rather confused, but a comparison of the mistakes shows quite good agreement of the general tendencies. The answers to all the syllables are given graphically in Figs. 3a and 3b for test 2 A, and in Fig. 4 for test 2 B. The main differences between 2 A and B are that b is heard better after u and o in 2 A, d is heard better after œ in 2 A, whereas the opposite is true of b after œ. g is on the whole heard better in 2 B. The mistakes are, however, roughly the same.

It is possible to set up the following rules:

(1) after unrounded vowels

(1a) b and d are often mistaken for each other, but the mistake b for d is much more frequent than the mistake d for b.

This confusion can be explained by the fact that unrounded front vowels have negative F_2-transitions both before b and d, and that i and e have negative F_3-transitions as well in both cases. As the transitions after d is less pronounced than those after b, it is not astonishing that d is heard as b more often

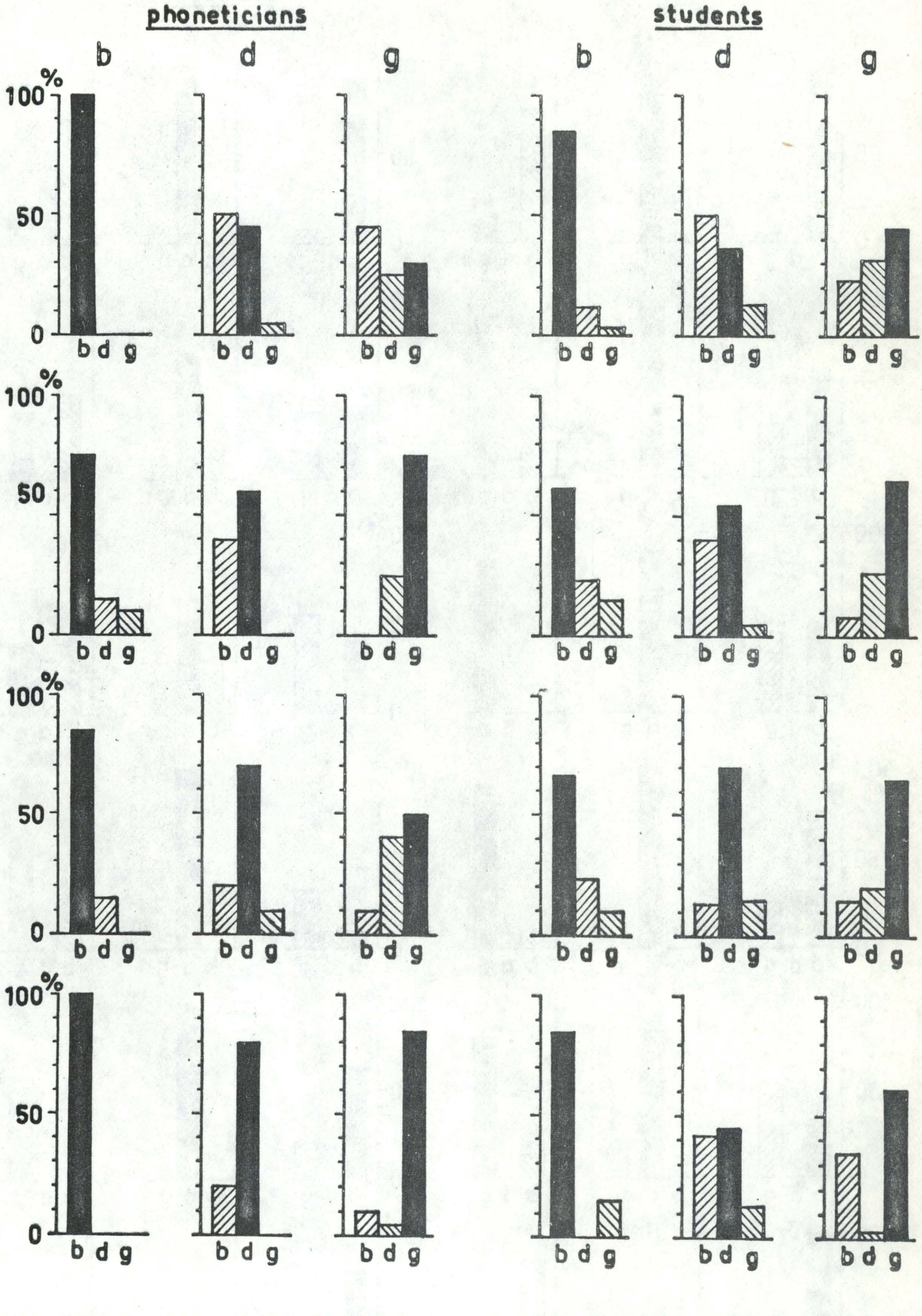


Fig. 3a. Identification of final unreleased *bdg* 1964, spoken by EFJ.
Listeners: 10 phoneticians & 20 students.

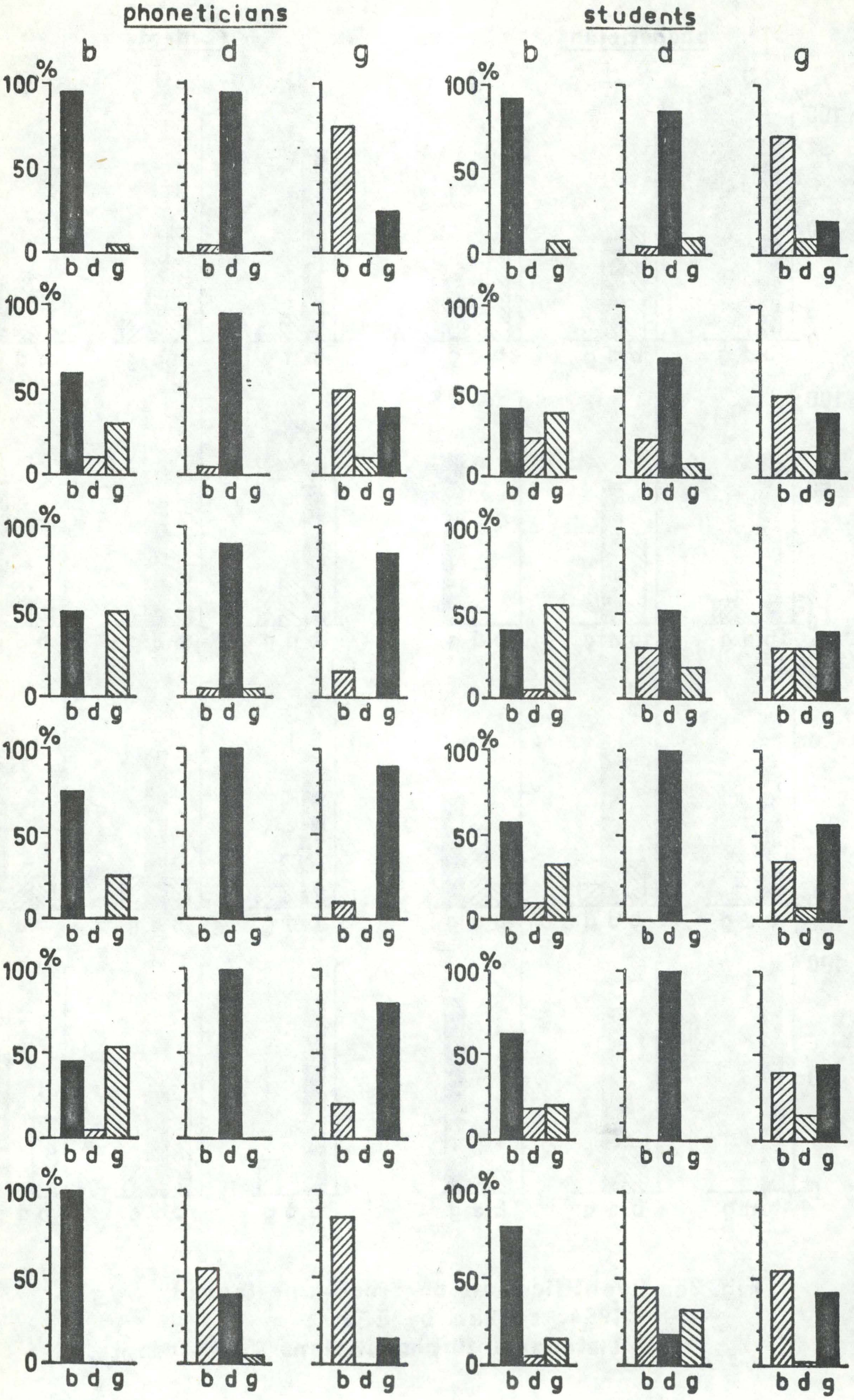


Fig. 3b. Continuation of 3a.

than the opposite. A further reason is general tendency,mentioned above, to choose labial in case of doubt.

(1b) g is heard either as b or d (it seems to be heard more often as d after i and a, and more often as b after e and ε, but these differences may be due to chance. They are difficult to explain).

(2) after rounded vowels

(2a) b and g are often confounded. The mistakes go both ways, except after y and ɔ where g heard as b is much more common than the opposite.
It is easy to understand that b and g are confounded after rounded back vowels, since they have almost identical transitions (straight or slightly positive F_2-transitions, straight or negative F_3-transitions).
It is not quite as evident why they are confounded after rounded front vowels; but it may be explained by the fact that they have similar transitions of F_3, and that both may have negative F_2-transitions, although this is more stable in combination with labials than in combination with velars.

(2b) d is often heard correctly after rounded vowels, if not, it is heard as b.
It is easy to understand that d is heard correctly after rounded back vowels (u and o), but more difficult to understand why it is often heard as b after rounded front vowels and after ɔ. Here again, the main reason may be that the labial is the neutral consonant, most often heard when there is no reason to hear others. The transitions may be fairly level in rounded front vowels in combination with d.

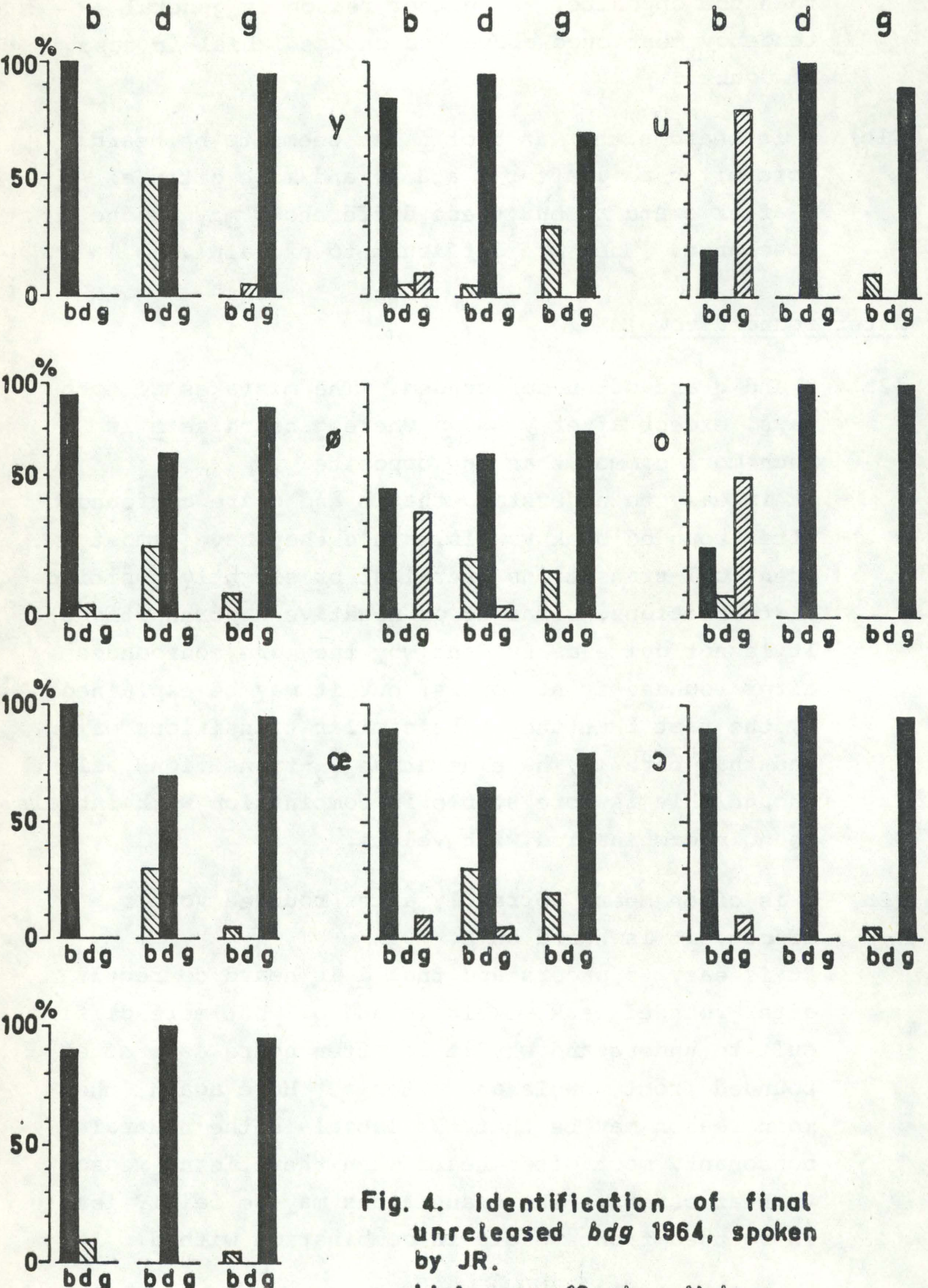


Fig. 4. Identification of final unreleased *bdg* 1964, spoken by JR. Listeners: 10 phoneticians.

The tendencies found in test 2 are in agreement with the results of test 1, and they are for the main part confirmed by experiments with removal of initial explosions (see the article on "Tape cutting experiments with Danish stop consonants" later in this volume).

3. Comparison with other investigations

Since the experiments described here were started, a good deal of perceptual tests with unreleased stop consonants have been carried out, e.g. Halle-Hughes-Radley (1957), Householder (1956), Malécot (1958), Wang (1959), (all based on American English), Andresen (1960) (British English), Malécot-Lindheimer (1966) (French) and Kurt Johansen (1969) (Swedish).

The main result of the present investigation, i.e. that unreleased stop consonants are more easily identified when the transitions are extensive (e.g. ib, ud, ag) than when they are small (e.g. id, ig, ug), appears as a tendency in many of the above mentioned studies. In Halle-Hughes-Radley's test labial stops are identified better after i than after u, alveolar stops better after u than after i, and velars have a particularly high degree of correct identification after I and ʌ (which have pronounced formant transitions), and a low degree of correct identification after i: and u:. Andresen finds that p and k are identified better after I than after ɔ, t better after ɔ than after I, and Malécot finds the same for ɛ and ɔ, although the answers to ɔd and ɔt are only clear in the test with French listeners. Householder's results are not very clear. Wang-Fillmore examined consonant perception in noise and found that initial bilabials are identified with much higher correctness before i than before u, alveolars better before u than before i, and velars very poorly both before i and u, whereas all have relatively high percentages of correct identification before a and ɛ.

Both Wang-Fillmore and Halle et al. mention that an extensive transition may be more important for perception than a small transition. This has been contended by Delattre (1958), who maintains that the decisive thing is that the transition points to the locus of the consonant, and this is true of most transitions, with the exception of the transitions of rounded back vowels in combination with velars; in this combination the explosion is necessary for the identification. Delattre tries to interpret Householder's and Halle's results in accordance with this view, but he can only do so by leaving out *i* as "a special case". Malécot's result can be quoted in support of Delattre (*ɔ* + velar was difficult to recognize without release), but his material is too restricted. The present investigation demonstrates (i) that velars can be recognized without release after *u o ɔ* (cp. test 2 B) and (ii) that other combinations may be more difficult (e.g. *id*, *ig*). It is thus not sufficient that the transition points to the locus.

As for the mistakes made in identification, rule 1a (confusion of labial and alveolar after unrounded vowels) is on the whole supported by the results of Householder, Halle et al. and Malécot. Rule 1b (*g* is heard either as *b* or *d*) is not really a rule but rather lack of a rule, but the same irregularity can be seen in other investigations. Rule 2a (*b* and *g* are often confounded after rounded vowels) is only partly confirmed; velars are often heard as labials in Halle's and Householder's material, but not often vice versa.

Rule 2b (*d* is often heard correctly after rounded vowels; if not, it is heard as *b*) is true of long vowels in Householder's material, after short vowels *t* is heard either as *p* or *k*.

It can be concluded that whereas the absolute number of mistakes is evidently dependent on the precision with which the speaker articulates, the relative number in different combinations is subjected to some more general tendencies which are not in agreement with Delattre's hypothesis, and that there are also certain common traits in the type of mistakes made.

References

Andresen, N.S. 1960: "On the Perception of Unreleased Voiceless Plosives in English", Language and Speech 3, p. 109-121.

Delattre, P. 1958: "Unreleased Velar Plosives after Back-rounded Vowels", JASA 30, p. 581-582.

Halle, M., Hughes, G.W. and Radley, J.P.A. 1957: "Acoustic Properties of Stop Consonants", JASA 29, p. 107-116.

Householder, F.W. 1956: "Unreleased ptk in American English", For Roman Jakobson, p. 235-244.

Johansen, K. 1969: "Transitions as Place Cues for Voiced Stop Consonants: Direction or Extent?", Studia Linguistica XXIII 11, p. 69-81.

Malécot, A. 1958: "The Role of Releases in the Identification of Released Final Stops", Language 34, p. 370-380.

Malécot, A. and Lindheimer, E. 1966: "The Contribution of Releases to the Identification of Final Stops in French", Studia Linguistica XX, 2, p. 99-109.

Wang, W. S-Y. 1959: "Transition and Release as Perceptual Cues for Final Plosives", Journal of Speech and Hearing Research 2, p. 66-73.

Wang, W. S-Y. and Fillmore, Ch.J. 1959: "Intrinsic Cues and Consonant Perception", Journal of Speech and Hearing Research 4, p. 130-136.

TAPE CUTTING EXPERIMENTS WITH DANISH STOP CONSONANTS IN INITIAL POSITION

Eli Fischer-Jørgensen

1. Introduction

The experiments described in the present paper were inspired by the perceptual tests with stop consonants carried out by the Haskins group in the early fifties (e.g. Liberman et al. 1952, Cooper et al. 1952, Liberman et al. 1954) and by the tape cutting experiments of Carol Schatz (1954). I decided to work with natural speech, not only because we had at that time no synthesizer in Denmark, but also because it might be of interest as a necessary supplement to the work with synthetic speech. Most of the experiments were carried out in 1954-55, but they were supplemented at some points in 1972 (particularly concerning removal of explosions and transitions of bdg). The purpose was to find out what part of the acoustic sequence, bursts or aspirations and transitions, was the most important for the perception, and I wanted to use at least three different vowels, because I had a suspicion that the results might differ for different vowels. Some of the results were mentioned very briefly in a paper on the commutation test (1956), but a more detailed report has not been given until now. In the meantime others - and particularly the Haskins group - have worked with similar questions, but the results of the present experiments may still be of interest, not only because of the specific information they give about a language of the Danish type, but also because they throw some light on more general problems, particularly the importance of vowel transitions and some aspects of the locus theory.

2. Material and listeners

2.1. The material

The material consisted of Danish words with initial stop or fricative consonant followed by either a, i or u, viz. the words [panə, tanə, kanə, kalə, banə, danə, galə, bas, das, gas, pi:lə, ti:lə, ki:lə, kilə, bilə, dilə, gilə, pu:ə, tu:ə, ku:ə, bu:ə, du:ə, gu:lə, - fanən, sanə, hanə, halə, fi:lə, si:lə, hi:lə, hilə, fu:ə, su:ə, hu:ə, hu:lə]. They are all existing Danish words, with the exception of [hi:lə]; two cases ([hanə] and [ti:lə]) are names.

The tests with fricatives will only be mentioned here in so far as they form part of the investigation of stop consonants. Only initial consonants were investigated, since Danish ptk and bdg are only distinguished in syllable initial position.

The words were spoken by me on tape several times, so that typical examples could be chosen for the tests on the basis of spectrograms. Two or three examples of each word were used in the test.

Fig. 1 shows schematic spectrograms of the stop consonants used in the test including the transition part of the following vowel. It appears from the figure that Danish ptk are strongly aspirated, and that t is affricated (more so before close than before open vowels, and particularly before i). It is possible to distinguish burst (in the following called explosion) from aspiration in p and k, and often explosion, fricative phase (with high frequency noise) and aspiration phase in t, e.g. in ta. But the two latter segments were not separated in the experiments, and they will be combined under the designation "aspiration" in this report. It may be difficult to separate the explosion of t from the following fricative phase, but in my pronunciation it is possible. The explosion is somewhat lower in frequency than the fricative phase and very weak. It should be borne in mind that

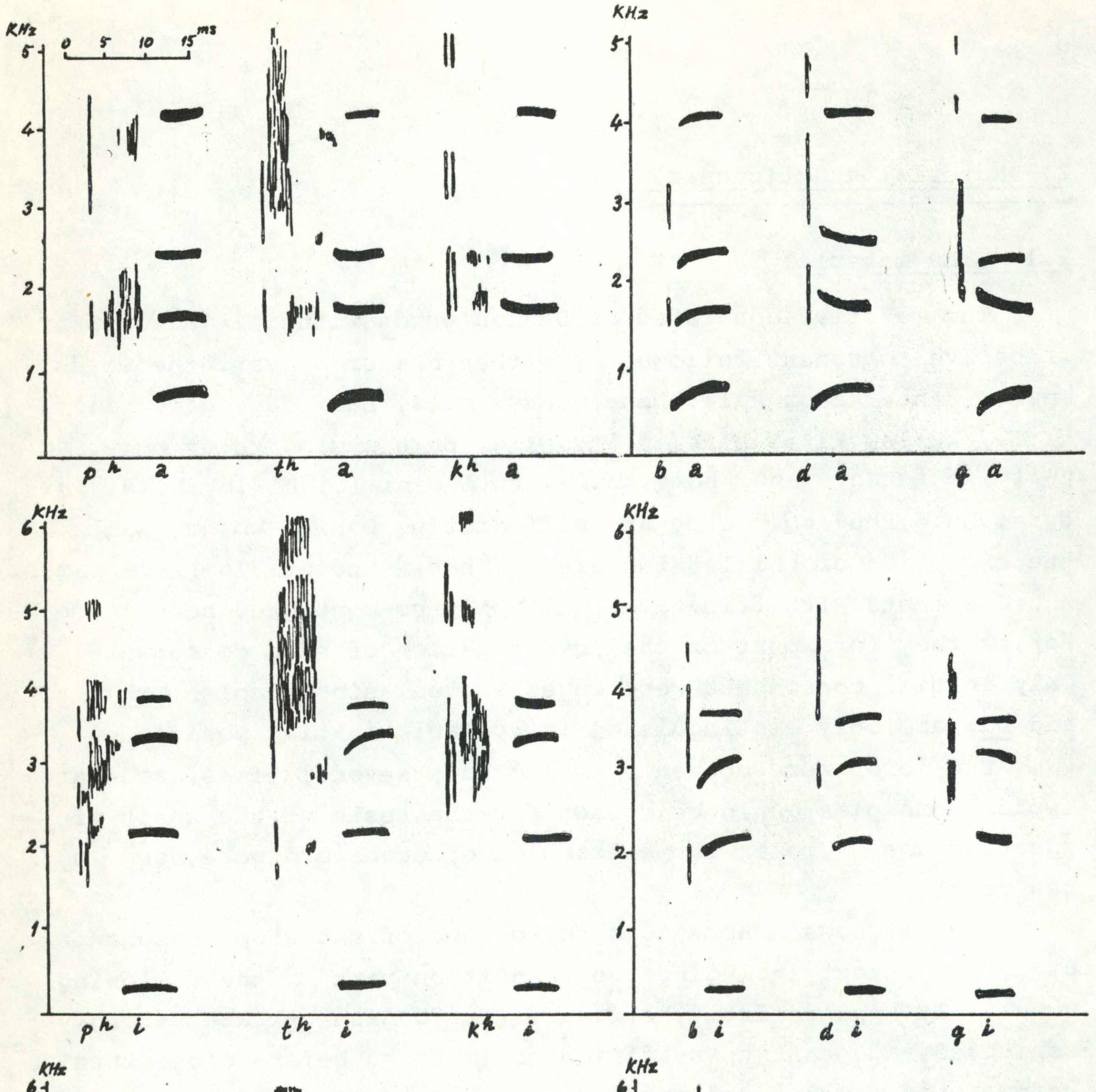


Fig. 1. Schematic spectrograms of Danish sound sequences used in cutting and splicing experiments.

in languages where *t* is unaffricated, a segmentation into a relatively short explosion noise (corresponding in frequency to the fricative phase of Danish *t*) and an aspiration phase (or nothing, if the consonant is unaspirated) will be more natural, and the removal of the explosion will not give the same result in these languages as in Danish. The duration of the aspiration in the words with *ptk* utilized in the experiment was (in ms) *pa* 60-80, *ta* 70-85, *ka* 60-70, *pi* 50-75, *ti* 75-90, *ki* 60-70, *pu* 55-70, *tu* 70-75, *ku* 55-65. The duration of the open interval in the words with *bdg* was (in ms) *ba* 5-10, *da* 10, *ga* 10-20, *bi* 10, *di* 10-15, *gi* 15-20, *bu* 10-15, *du* 10-15, *gu* 10-20. Exchanges of explosions may have involved small changes in the open intervals of *bdg*, of 5-10 ms, but this has hardly influenced the identification of the place of articulation, and the differences were too small to influence the distinction between *ptk* and *bdg*.

There is very little transition to be seen after *ptk* because most of the movements of the speech organs from the consonant to the vowel take place during the aspiration phase (this is clearly seen, e.g. in *tu*). After the strongly affricated *t* in *ti* there is, however, a pronounced transition of F_3, because the constriction of the vocal tract continues during the fricative phase, so that the movement to the vowel starts later than after *p* or *k*. Danish *i* is very fronted, so that the front cavity may become responsible for F_3, and this formant shows often more extensive transitions than F_2. The velars are more strongly influenced by the place of articulation of the following vowel than in English; for this reason there are hardly any vowel transitions in *i* and *u* after velars.

The fact that *bdg* are voiceless simplifies the cutting experiments.

The experiments consisted in removal of explosions, aspirations and transitions, and exchange of these segments, but always before the same vowel (in this respect the tests differ from those of Carol Schatz). All cutting and splicing was done by hand in 1954. For the supplementary test in 1972, some cutting was done by means of an electronic segmentator, but the splicing had to be done by hand anyway. Some cuttings done by hand were repeated by means of the segmentator. The two methods gave the same results. In all cases the cutting was sharp. Spectrograms were taken before and after each cutting and splicing.

The test played in 1954-55 consisted of 500 different items, the supplementary test 1972 contained 92 items. There were, however, only 1-3 examples of each single phenomenon (e.g. b-explosion removed before i), and the results must therefore be taken with some reservation. It would also have been preferable to have more than one speaker. But 592 words and 21 (20) listeners gave 12,340 answers, and without access to a computer the processing was rather time consuming.

2.2. Listeners and testing procedure

Each word was repeated three times on the test tape with half a second between the repetitions, and there was a pause of five seconds before the next test word. The tape was played over a loudspeaker in 1955 to a group af 21 listeners and over headphones in 1972 to a group of 20 listeners. They were all phoneticians or dialectologists. The listeners were asked to identify the words as existing (or at least possible) Danish words, and to make a note if they found the pronunciation abnormal. It has not been possible to find any differences in the answers indicating that the frequency of the word or even its existence played a role. On the contrary, although the subjects were asked to use normal Danish spelling, they sometimes made spelling mistakes showing that they had heard a Danish phoneme sequence and written it down in a way which was possible according to Danish spelling rules, but not the one used for the word in question.

In a certain number of cases they indicated that ptk were unaspirated. Less sophisticated listeners would probably have written bdg instead (cp. the paper on the identification of foreign stops by Danish listeners in this volume). But the answers given in this test have the advantage of giving some indication of cues that may be used in distinguishing voiceless bdg from unaspirated ptk.

Unfortunately there was some low-frequency noise on the tape on which the test words were recorded in 1955, but none on the tape used for intervals between the words. This had the result that the listeners sometimes heard an h or an f (particularly before u) where an initial consonant had been cut off, whereas the listeners to the supplementary test in 1972 more often heard ? or nothing.

3. Results concerning the distinction between ptk and bdg

In this section we are only concerned with the distinction between the two categories ptk and bdg, and not with possible mistakes in place of articulation within these two categories.

3.1. Identification of bdg after removal of explosions and/or transitions

In Table I the percentual identifications of bdg before i, u and a are indicated for four conditions: (1) unchanged, (2) both explosion and transition removed, (3) explosion removed, and (4) transition removed. The same is given in graphical form in Fig. 2. There were 6-9 different examples in each case, and as there were 20 or 21 listeners, the possible number of answers reach from 120 to 189. The actual numbers are indicated in the table.

TABLE I [1]

Identification (in %) of bdg, unchanged and after removal of transition and/or explosion

		i				a				u		
		ptk	bdg	x		ptk	bdg	x		ptk	bdg	x
	N				N				N			
1.+ex.+tr. B$\underline{V}$V	144	-	99	1	186	1	98	1	186	-	97	3
2.-ex.-tr. (B$\underline{V}$)V	180	1	14	85	140	-	1	99	183	4	25	71
3.-ex.+tr. (B)$\underline{V}$V	144	1	48	51	186	3	86	11	186	2	56	42
4.+ex.-tr. B($\underline{V}$)V	120	-	97	3	183	2	42	56	120	-	91	9

(1) bdg are identified almost 100 % correctly in the unchanged words. Most exceptions are due to one example of [bu:ə]with very weak explosion, which has sometimes been heard as [fu:ə] because of the noise on the tape, and to one example of [dilə], sometimes heard as [gilə].

(2) If both explosions and transitions are removed, no stop consonant is heard before a, and only a small number before the other vowels. 25 % before u is however somewhat surprising. The consonant heard here is always b, and most examples belong to three individual words from the supplementary test, where the cutting has produced a slight click.

[1] Here and in the following tables and figures the following abbreviations are used: B = bdg, P = ptk, V = vowel, $\underline{V}$ = vowel transition, H = aspiration, x = 0,?, h or f, () = removed, / = splicing between sounds from different words.

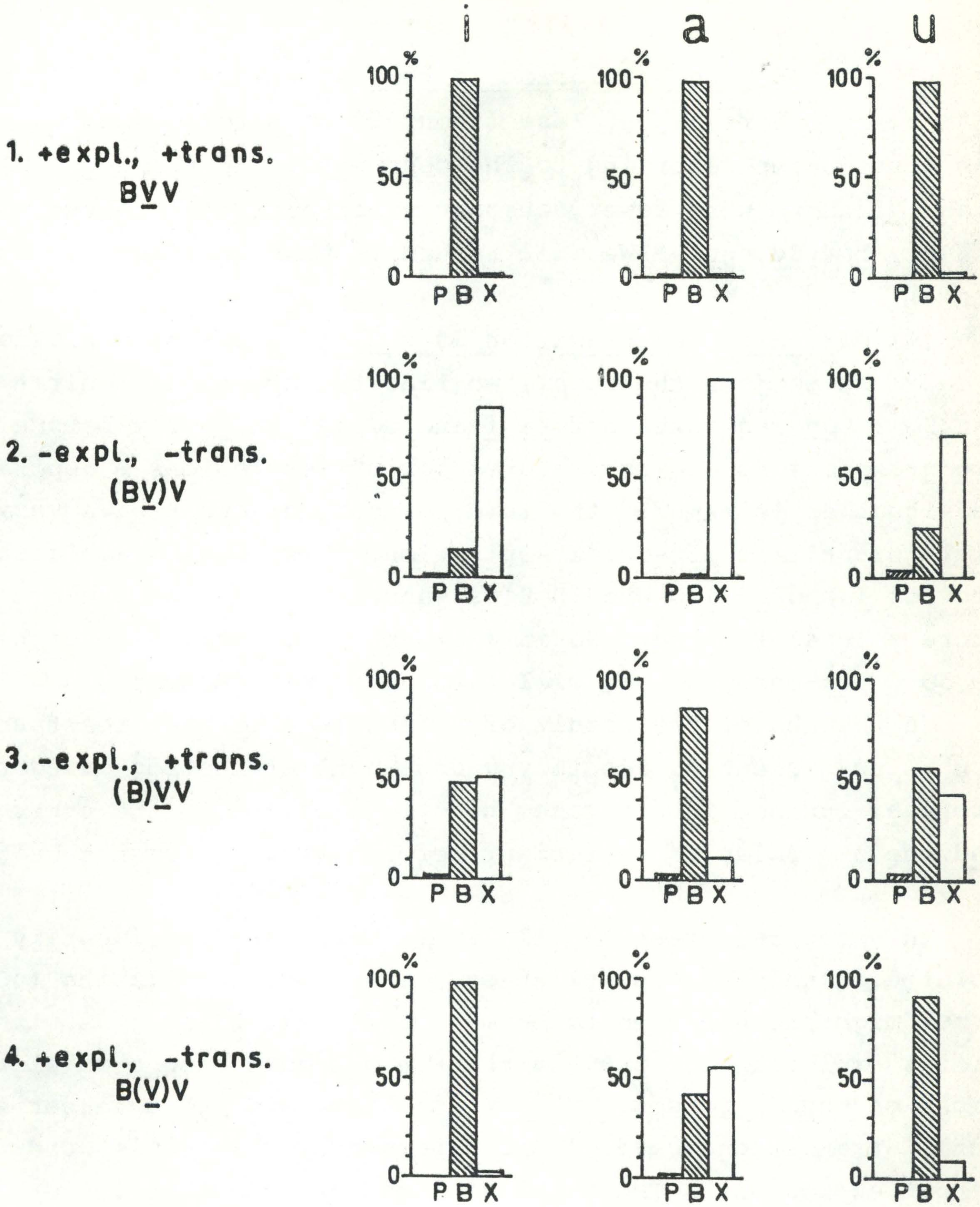


Fig. 2. Identification of *bdg* after removal of explosion and/or transition.

P = *ptk*-explosion
B = *bdg*-explosion
H = aspiration
V = vowel
V̲ = vowel transition
X = other answers (0, *h, f*)
( ) = removed

(3) If only the explosion is removed, there are still 86 % bdg-answers before a, but less (about 50 %) before i and u, with some differences according to the consonant. In the words with di and gi there were fewer consonant-responses than in the words with bi, bu, du, gu. (We will return to this question in section 4).

(4) If only the transition is removed, and the explosion moved so as to join the vowel, we find the opposite result: more than 90 % bdg-responses before i and u, and only 42 % before a. If bdg-explosions are substituted for ptk-explosions + aspiration the situation is almost the same, since the vowels have very little transition after the aspiration. The results are also the same for i and u (100 and 90 % bdg-answers). The percentage bdg before a is 43 % (almost as in (3)), but instead of zero there are 50 % ptk-answers. We will return to this question in 3.7.

On the whole, the result of these tests is that the transition is sufficient to identify a consonant of the bdg-category before a, but not in all cases before i and u. On the other hand, the explosion is sufficient before i and u, but not before a.

In a separate test the listeners were asked to identify isolated transitions as syllables. These short syllables (40 - 80 ms) might be expected to be more difficult than words with the initial explosion cut out. a-transitions gave also a smaller number of bdg-responses, but u-transitions and (to a lesser extent) i-transitions gave more bdg-responses than whole words without explosions. The u-transitions were somewhat longer than the others.

The results are given in Table I a.

TABLE I a

Identification of stop consonant on the basis of isolated bdg-transitions

i				a				u			
N	ptk	bdg	x	N	ptk	bdg	x	N	ptk	bdg	x
61	-	63	37	81	-	79	21	61	-	61	39

3.2. Identification of ptk after removal of explosions and/or aspirations

In table II and in Fig.3 the perceptual identification of ptk before i, a and u is indicated for four conditions: (1) unchanged, (2) both explosion and aspiration removed, (3) explosion removed, and (4) aspiration removed. The same is given in graphical form in Fig.3. There were from 3 to 9 different examples of each case, and 21 listeners, which gives from 63 to 189 answers, as indicated in the table.

TABLE II

Identification (in %) of ptk, unchanged and after removal of explosion and/or aspiration

	i				a				u			
	N	ptk	bdg	x	N	ptk	bdg	x	N	ptk	bdg	x
1.+ex.+asp. PHV	126	99	-	1	189	98	-	2	126	100	-	-
2.-ex.-asp. (PH)V	63	-	21	79	63	61	17	22	63	2	65	33
3.-ex.+asp. (P)HV	63	100	-	-	126	98	-	2	189	74	7	19
4.+ex.-asp. P(H)V	63	-	97	3	63	27	70	3	63	-	100	-

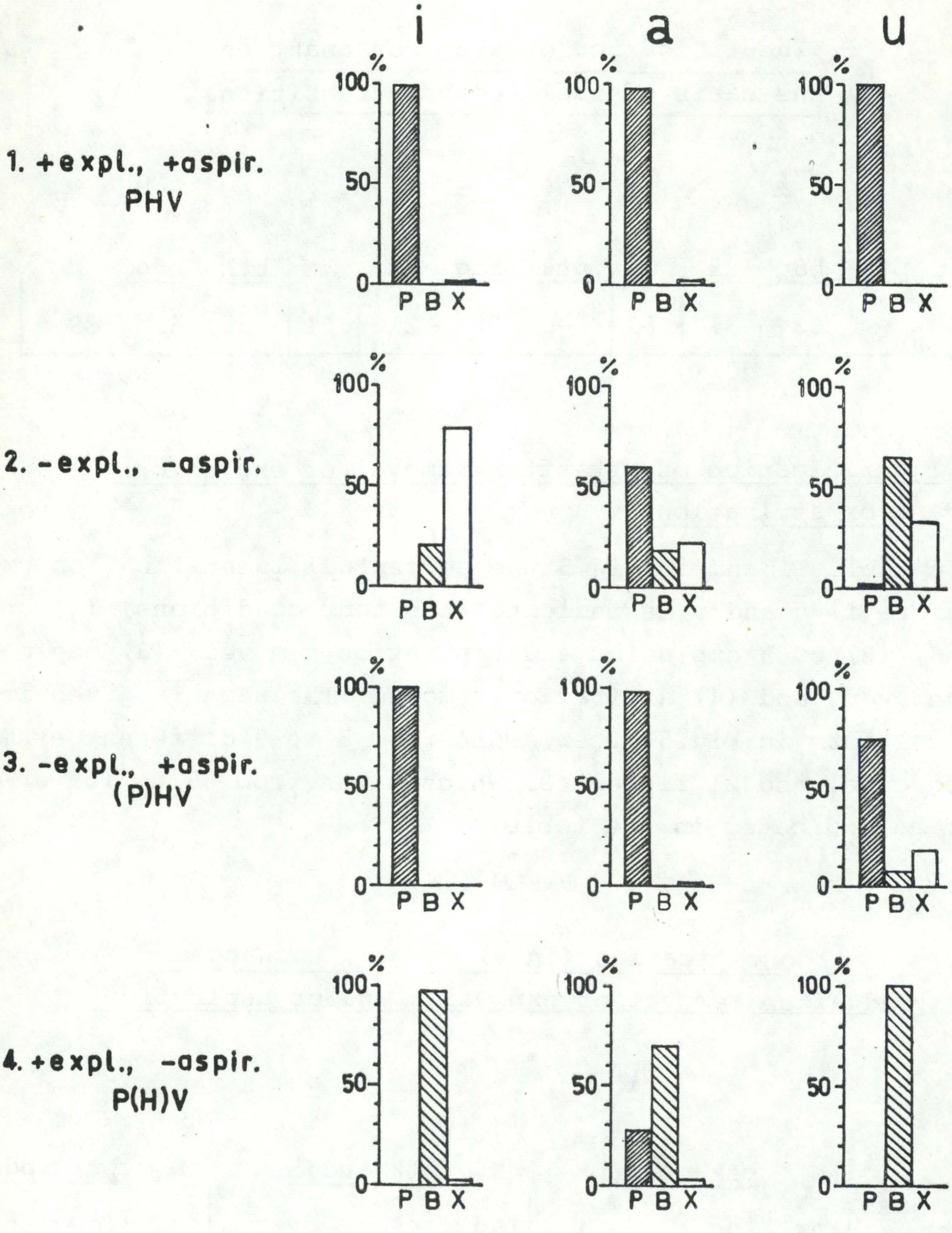


Fig. 3. Identification of *ptk*, unchanged and after removal of explosion and/or aspiration.

(1) *Unchanged words* are heard correctly in almost 100 % of the cases. The very few exceptions concern the word [kanə], sometimes heard as [panə].

When both explosion and aspiration are present, the start of the following vowel is not very important. Substitution of *ph*, *th* and *kh* for *bdg* + transition has been tried before *a* only. The result was 93 % *ptk*. Substitution of *ph*, *th*, *kh* for *bdg*-explosions (i.e. with the transitions retained)before *i* and *u* gave almost the same result (97 %), when one example is kept aside, namely *kh* for *g* in g*a*lde, which gives only 52 % *k*. (This *k* had a rather weak aspiration, and *ga* had a strong transition). *ph*, *th*, *kh* can also be identified in isolation.

(2) When *both explosion and aspiration are removed*, the result is in most cases zero before *i*, as might be expected, but there are 65 % *bdg*-answers before *u* (almost exclusively *b*), and before *a* there are 61 % *p*-answers and 17 % *b*-answers. An abrupt vowel start without pronounced transition is often interpreted as labial + vowel.

The labial seems to be the most neutral of the stops. (See also the paper on unreleased stop consonants in this volume). But a labial before an *i* involves an extensive transition of F_2 or F_3, and it is understandable that no consonant is heard when such transitions are absent. There are also more *b*-answers before *u* than before *i* when both explosions and transitions are removed from words with *bdg* (see 3.1), though the difference is less pronounced in this case. The problem of *p*-answers before *a* will be treated in 3.7.

(3) If *only the explosions are removed*, the words are still heard as starting with *ptk* in 90 % of the cases. And if one example of *pu* with a very weak aspiration is kept apart, the percentage *ptk*-answers is 97. This means that the aspiration is sufficient for the identification of the *ptk*-category. (*ka* is heard as *pa*, however).

(4) If only the aspiration is removed, and the explosion moved so as to join the vowel, the normal response is bdg, with a few exceptions before a.

The conclusion is that the aspiration is very important, whereas the explosion without aspiration is not sufficient to evoke ptk-responses.

3.3. Pause or aspiration noise?

The question may be raised whether a pause between explosion and vowel start is sufficient for the identification of ptk, or whether aspiration noise is necessary.

In order to investigate this question, the aspiration was removed in some cases and the explosion placed at a distance of 60-70 ms from the start of the vowel. The results are given in table III.

TABLE III

Identification (in %) of
ptk-explosion + pause + vowel

	N	ptk	bdg	x
i	63	38	29	33
a	63	65	28	7
u	63	33	41	26
aver.		45	33	22

A comparison with table II, 4 shows that the introduction of a pause of 60-70 ms between explosion and vowel start increases the ptk-responses considerably, but there is also an increase in zero-responses, and a majority of ptk-answers is only found before a. Before i and u the answer depends on the consonant-vowel combination. pi, ki, ku have 62 % ptk-answers, pu, tu, ti only 10 %. This means that, on the whole, the poorest ptk-identifications are found with consonants having a weak explosion and requiring a strong aspiration in normal speech (tu, ti), whereas consonants with stronger explosions and a normally weak aspiration (ku, ka) are identified more easily as ptk under these conditions. In ki the strong explosion may be the only cause for the good identification.

The general conclusion is that a simple pause is not sufficient for the identification of Danish ptk.

Before u a distance of 30 ms was also tried out. Although 60-70 ms should be the optimal distance, there were, on the average, almost as many ptk-answers when the distance was 30 ms (36 %). This apparent similarity is, however, due to an increase of pt-answers and a decrease in k-answers accompanied by an increase of g-answers. The explanation is probably that if weak explosions (as in pu, tu) are removed far from the vowel, they do not fuse with it to form a syllabic unit, and sometimes they are not heard at all.

3.4. Aspiration noise replaced by fricative noise from f,s,h

In order to see how critical the type of noise is, an experiment was carried out consisting in replacing the aspiration noise by approximately 50 ms fricative noise cut out of s, f, h taken from the same vowel environment; f from fi was thus substituted for the aspiration of pi, ti, ki etc.

The results are given in table IV.

TABLE IV

f, s, h substituted for aspiration in ptk (answers in %)

	N	ptk	bdg	affr.	x
f	63	49	9	13	29
s	63	99	-	1	-
h	63	80	13	-	7

It can be seen from the table that s-noise evokes the highest number of ptk-answers, f-noise the lowest number. It should be remembered that in this section we are not concerned with the place of articulation, only with the difference between ptk and bdg. As a matter of fact s-noise gives t-responses in 92 % of all cases. - Without a preceding explosion an abruptly starting s-noise of 40-70 ms is heard 100 % as t. The explosion is thus of no importance here.

As for the effect of f-noise it depends chiefly on the following vowel. There are 81 % ptk-answers before u and only 34 % ptk-answers before a and i. The lower part of the f-noise is apparently related to the aspiration noise before u. Differences according to preceding consonant are less pronounced. The highest percentage is reached after k (68), the lowest after t (30), probably because a stronger explosion is needed, when the noise of the aspiration is not quite appropriate. The extremes are k/f/u (95 % k) and t/f/i (0 % ptk). p/f/u gives only 71 % pu (the other answers are fu and pfu). Abruptly starting f-noise without explosion gave f in almost all cases. The noise is apparently not strong enough to give the impression of a stop consonant until just before the vowel (25 ms f gives b), h substituted for aspiration noise gives ptk-responses in most cases, the exceptions are k/h/i and t/h/i, which are heard as gi, because they require a stronger noise.

An abruptly starting h without preceding explosion may be heard as p before a, but before i and u it is rarely heard as a stop.

3.5. The importance of the duration of the aspiration

A certain number of test words, both words with initial ptk and words with initial fricatives, had been cut off at different distances from the vowel start; but the number of examples is not sufficient and the cutting points not sufficiently parallel to allow precise conclusions. Shortening of the aspiration with retained explosion has not been tried. If my tests had been conducted after the appearance of the papers of Lisker and Abramson on VOT-value, I would have given more consideration to this problem. Moreover, cutting of natural sounds will give less clear results than synthetic sounds, because the aspiration noise is not homogenous in the time domain. The noise has, for instance, very often increasing intensity in pa and (rather suddenly) decreasing intensity in ta. And if the cutting point lies after the fricative phase of ta, the result is pa.

As mentioned above, only short durations of f-noise give stop-answers. Durations of 75,70 and 50 ms give f-answers, 15, 20 and 25 ms give b-answers for fi and fa, but f for fu. (Here the noise of the tape may have played a role). - Cuttings of h give very irregular results, from which no conclusions can be drawn.

The fricative s and the stop consonants give relatively clear results, but the number of examples are too small. In table V the answers of the majority for each consonant and duration is given. Underlining indicates a majority of more than

90%. - The answer f for pu is due to an example with very weak aspiration.

TABLE V

Answers to p, t, k and s cut off at different distances from the vowel start

	10	15	20	25	30	35	40	45	50	55	60	65	70 ms
pi				g									p
pu	f					f			f	p			
pa							p		p		p		
ti				d					t				t
tu	d				f				t		t		t
ta							t		t		t	t	
ki					g/o				k		k		
ku	b/f					p		k		k	k		
ka						k	p/k	k					
si			d						t				
su		d					t				t	t	
sa			d				t						

In spite of the restricted number of examples a pretty clear dividing line appears between 30 and 35 ms for g/k and between 25 and 35 ms for b/p and d/t. This is in complete agreement with the results from identification of foreign stop consonants by Danish listeners (see the report in this volume). But although a dividing line can be found for the majority of answers, there are still ptk-answers at much shorter VOT-values (see 3.7.).

3.6. Influence of explosion type (ptk or bdg)

The explosions of ptk and bdg are not very different. The only relatively stable difference seems to be that the explosion of k is somewhat stronger than that of g and that the explosion of t is somewhat weaker and lower in frequency than that of d.

Thus it is not astonishing that it makes no great difference to the identifications whether the explosions belong to one or the other type. Other cues are much stronger and override the small differences which may exist between the explosions.

This appears from a great number of identifications, which are combined in table VI. The examples are ordered in comparable pairs, where the only difference lies in the explosions. Only ptk- and bdg-answers are included, not zero, h etc.

TABLE VI

Comparison between answers (in %) to stimuli with ptk- and bdg-explosions

		N	ptk	bdg
1a	BVV	516	-	98
1b	P/(B)VV	189	4	96
2a	PHV	441	99	-
2b	B/(P)HV	189	98	-
3a	P(H)V	189	7	90
3b	B/(PH)V	189	17	78
4a	P/f/V	189	49	9
4b	B/f/V	231	55	11
5a	P/s/V	189	99	-
5b	B/s/V	189	92	-
6a	P/h/V	189	80	16
6b	B/h/V	231	8o	11

It is evident from the responses listed in table VI that the type of explosion has been without any influence. The same appears from three other cases (all with 126 answers), viz.(1) substitution of aspiration for transition in bdg-words, which gives 99% ptk-answers, (2) substitution of transition for aspiration in ptk-words, which gives 100% bdg-answers, and (3) aspiration placed before the transition in bdg-words, which gives 98% ptk-answers.

In all these cases the difference between the explosions has been overridden by the strong cue: aspiration/lack of aspiration, or there has been fricative noise.

If there is only a pause between the explosion and the vowel start, the type of explosion has some effect. As mentioned in 3.3., some of the stimuli consisted in explosion noise + pause + vowel. If ptk-explosions are replaced by bdg-explosions in this context, there is a small decrease of ptk-answers. It has, however, only been tried before the vowel u. The results are given in table VII.

TABLE VII

Substitution of bdg for ptk before
pause with aspiration removed. (Answers in %)

		N	ptk	bdg	x
1. 60-70ms	P-(H)u	63	33	41	26
pause	B-(PH)u	63	19	37	44
2. 30-40ms	P-(Hu)	63	37	49	14
pause	B-(PH)u	63	33	33	33

The difference is appreciable only with the longer pause, and here it is only due to the difference between *g* and *k*. In the case of the shorter pause, both *p*- and *k*-explosions give some more *ptk*-answers than *b*- and *g*-explosions, but for *d/t* the opposite is the case.

If the explosions in words with *bdg* are moved the same distances away from the vowel,there are more *bdg*-answers, but this depends primarily on the character of the vowel start (see 3.7.).

3.7. Importance of the vowel start

Although stimuli with a short open interval are normally heard as *bdg*, there is a good number of cases with *ptk*-answers. They are somewhat more common with velars than with alveolars and labials. This may be due to the stronger explosion of velars, since cases without explosion show the same number of *ptk*-answers for alveolars and velars.Much more evident is a difference according to the nature of the following vowel: *ptk*-answers are almost exclusively found in words with *a* and hardly ever in words with *i* and *u*. Finally *ptk*-answers are only found in cases where an aspiration has been cut off, not in cases where the vowel originated from a *bdg*-word. In table VIII ,and in Fig.4 the different types of stimuli without aspiration are indicated with the percentages of *ptk*- and *bdg*-answers (only the stimuli with *a* are considered, since the others show none or at most 3% *ptk*-answers).

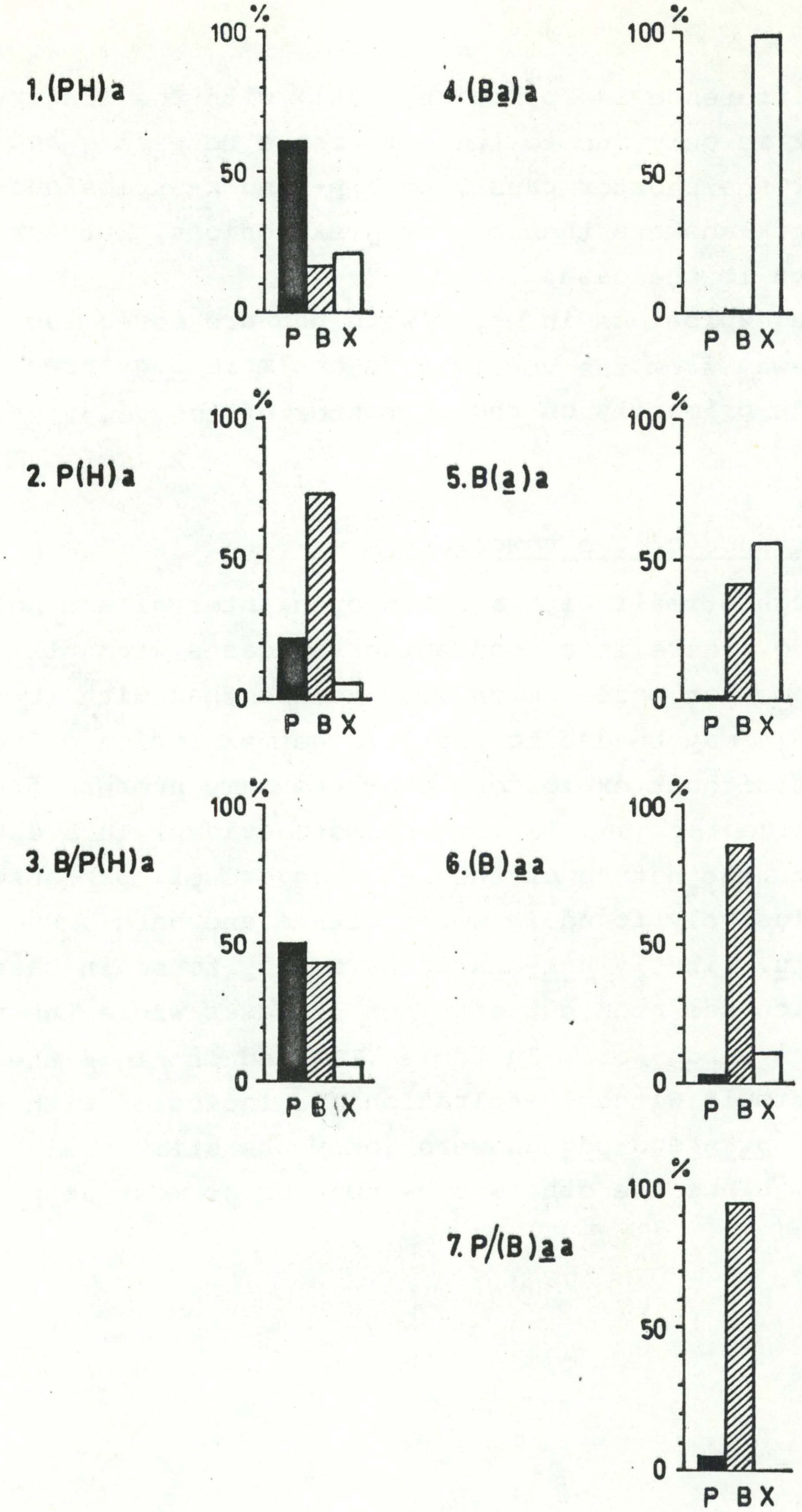


Fig. 4. Answers to stimuli with the vowel a without aspiration, 1-3 in original *ptk*-words, 4-7 in original *bdg*-words.

TABLE VIII

Answers (in %) to stimuli with the vowel a without aspiration. 1-3 original ptk-words, 4-7 original bdg-words.

	N	ptk	bdg	x
1. (PH)a	63	61	17	21
2. P(H)a	63	22	73	6
3. B/(PH)a	63	50	43	7
4. (Ba)a	140	-	1	99
5. B(a)a	183	2	42	56
6. (B)aa	186	3	86	11
7. P/(B)aa	186	5	95	-

It appears from table VIII that only the cases where an aspiration has been removed show a considerable number of ptk-answers, whereas the number of ptk-answers is insignificant if the vowel has followed a bdg-stop. A comparison betweeh e.g. 3 and 7 shows that the explosion is irrelevant (there are 50% ptk-answers after a bdg-explosion, and only 5 after a ptk-explosion). The duration of the open interval is at most 35 ms.A closer inspection of the words with velars shows the intervals 20 ,25 and 35 ms in the words with a majority of k-answers, and 10, 25, 30 and 35 in words with g-answers. Thus, the difference in the responses cannot be due to differences in open interval. Moreover, in (1) there is no explosion, and consequently no interval.

Two other explanations are possible: (a) The presence of transitions favours bdg-answers, whereas lack of transitions or rudimentary transitions favour ptk-answers. This would also explain why only the vowel a is involved. F_1 of a has a more pronounced transition than F_1 in i and u,and these may be necessary for the identification of *bdg*. However, this explanation is hardly sufficient.

It is true that (1), (2) and (3), which have many *ptk*-answers, are characterized by very rudimentary transitions because of the original aspiration, and that, on the other hand, (6) and (7), which show a majority of *bdg*-answers have full transitions. But (2) has a majority of *bdg*-answers in spite of its very slight transitions, and in (4) and (5), where the transitions have been removed altogether, there are no *ptk*-answers, but (in 5) some *bdg*-answers.

(b) The *ptk*-answers in (1-3) must therefore be due to a different cue in the vowel start. An inspection of spectrograms revealed that the vowel start after aspirated stops, particularly in the vowel *a*, is more irregular than after unaspirated stops. There are often one or two vibrations at a very low frequency before the start of the first formant, and at the same time there may be noise at the frequency of F_2-F_4, and these formants may continue to be somewhat noisy after the start of the first formant. This can be explained physiologically by the fact that the vocal chords are wide open at the explosion and may start to vibrate before they have reached complete closure. The early start may be more frequent before open vowels than before close vowels because the free passage in open vowels will cause the pressure above the vocal chords to drop more quickly. In Danish *bdg*, on the other hand, the vocal chords are close together already at the explosion, and a breathy start should not be expected (see Frøkjær et al. 1971).

The hypothesis that this vowel start is of importance for the identification is corroborated by a closer inspection of (1) and (2) and of individual examples of (2). At first sight it seems paradoxical that there are 61% *ptk*-responses to stimuli where both explosion and aspiration have been removed, and only 27% *ptk*-responses to stimuli where the explosion has been retained. But the cuts were slightly different in the two cases. In the latter case the cuts were closer to the vowels, in the former the cuts were placed a little earlier, and the irregular start with low frequency vibrations was retained. Moreover, besides the examples with the cut close to the vowel, experiment (2) contained two other examples of the same words where the cuts were placed 15 ms earlier in ta and 10 ms earlier in ka. This addition of 15 and 10 ms changed the responses drastically.The ptk-answers for ta were changed from 19 to 43% and for ka from 52 to 95%. The same words were cut 20 ms later with the result that the ptk-answers decreased to 14% both for ta and ka. (Only the words with the middle position of the cuts were included in the averages). The spectrograms (Fig.5) of the three different cuts show that it is not a question of vowel transitions. There are no transitions in the examples (c) heard as bdg by the majority of the listeners. - The very start of the vowel thus seems to be crucial.

In about half of the cases of ptk-answers to stimuli with removed aspiration, the subjects designated the consonants as unaspirated ptk, not normal Danish ptk-sounds. An inspection of some spectrograms, however, did not reveal a similar start of the unaspirated French and Dutch stops. Probably the glottis is less open in unaspirated stops than in aspirated stops, and a breathy start of the vowel is less probable, cp. that Slis and Damsté (1967) have found a wider glottis opening in voiceless fricatives than in voiceless stops in Dutch, whereas no such difference is seen in Danish.

The stimuli with pause between the explosion and the vowel are of interest in this connection. Unfortunately only words with u and i were used both with ptk- and bdg-stops and substitutions of bdg-explosions for ptk-explosions was only tried with u. But a comparison of individual words confirms the influence of the vowel start for these vowels, although less clearly than in the case of a. The answers are given in table IX. The alveolars are left out because the very weak t-explosion is only able to evoke a few ptk-responses under these conditions.

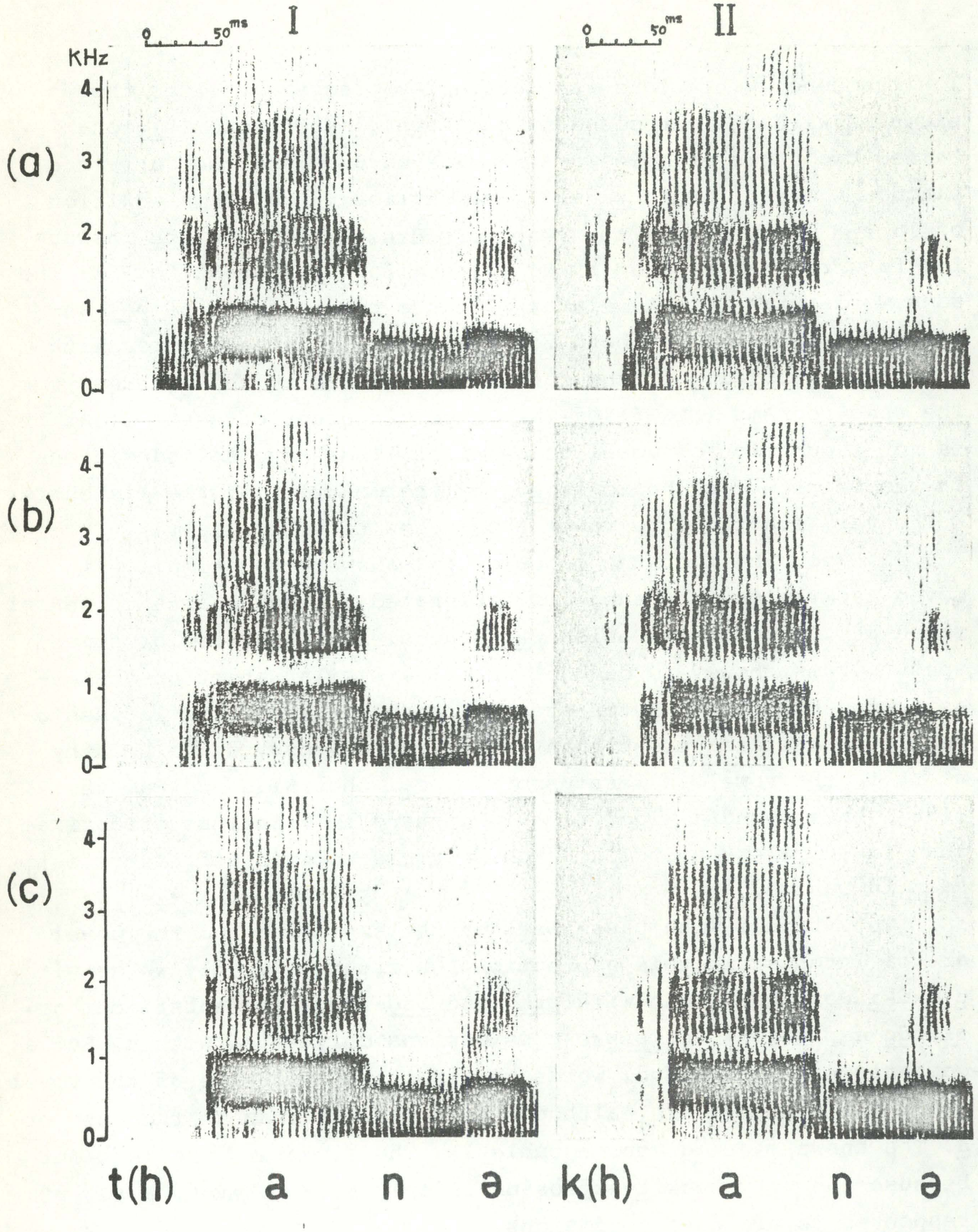


Fig. 5. Spectrograms of [t(h)anə] and [k(h)anə] with aspiration cut out (female voice, played at half speed, with expanded scale):
(a) cut at start of vibrations: I, 43% *ptk*, 48% *bdg*: II, 95% *ptk*, 5% *bdg*.
(b) 15 and 10 ms more cut off: I, 19% ", 81% " ; II, 52% ", 48% " .
(c) further 20 ms cut off: I, 14% ", 81% " ; II, 14% ", 57% " .

TABLE IX

Answers (in %) to different types of words with pause (60-70 ms) between explosion and vowel start (N=21)

	ptk	bdg	x		ptk	bdg	x
p-(h)i	71	5	24	k-(h)i	43	38	19
p-(b)i	52	48	-	k-(g)i	-	95	5
b-i	43	57	-	g-i	19	76	5
p-(h)u	9	48	43	k-(h)u	71	29	-
b-(ph)u	14	29	57	g-(kh)u	38	38	24
b-u	-	81	19	g-u	29	62	9

The answers to the words with i show that stimuli with original aspiration have a relatively high percentage of p- and k-answers, and words without original aspiration have a relatively high percentage of bdg-answers, whereas the difference in explosion type is almost irrelevant.

The words with u show that a high percentage of bdg-answers is found only in words without original aspiration, and that the explosion type has some effect for the velars (k is stronger than g).

4. Identification of place of articulation

4.1. Identification of place of articulation in bdg

4.1.1. Identification of unchanged bdg

In 3.1. it was mentioned that the identification of unchanged bdg-words was almost 100% correct, as far as the category of stops was concerned. As for the place of articulation, it is 100% correctly identified for b and g (there are a few f-answers for b), for d it is correctly identified in 92% of the cases. The mistakes for d are: 2% p before a, 6% b before u, and 17% g before i. The number of answers were 165 for b and d, and 186 for g.

4.1.2. Identification of bdg-words after removal or exchange of both explosion and transition

A. It was mentioned in 3.1. that when both explosion and transition are removed the most common answer was zero or ?. There were, however, a certain number of bdg-answers. The place of articulation heard in these cases is indicated in table X.[1]

[1] Here, and in the following tables, a few ptk-answers are included in the bdg-answers, since only the place of articulation is relevant. x covers 0, ?, and sometimes h and f. Since the f-answers were caused by noise on the tape they are not counted as correct answers for labial place of articulation.

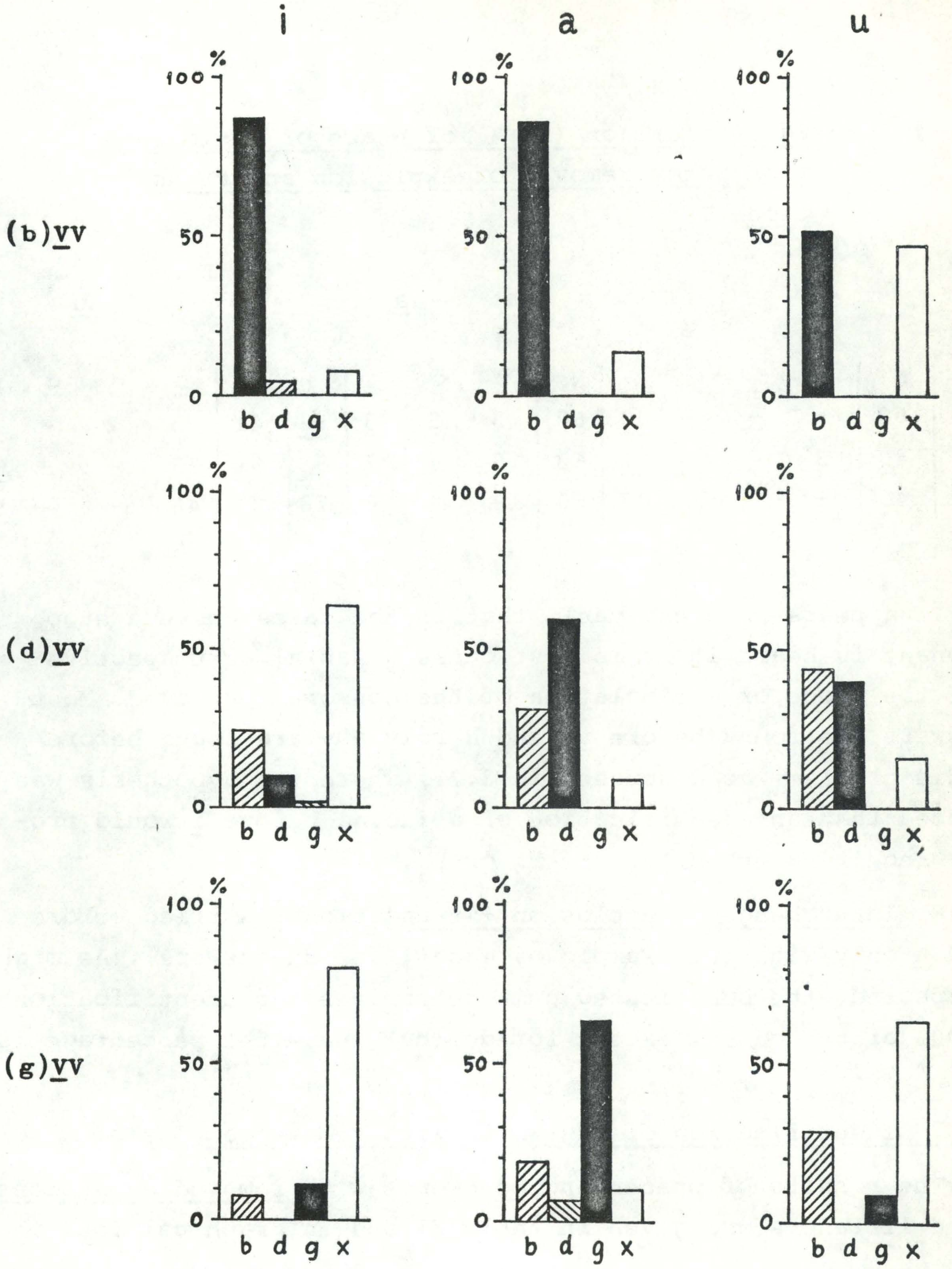


Fig. 6. Identification of <u>bdg</u> + <u>iau</u> after removal of the explosion.

<u>V</u> = transition, X = 0 (h, f)

TABLE X

Identification (in %) of place of articulation after removal of explosion and transition

	i					a					u				
	N	b	d	g	x	N	b	d	g	x	N	b	d	g	x
(bV)V	60	-	-	-	100	60	3	2	3	92	41	3	2	-	95
(dV)V	40	3	-	-	97	60	8	4	5	83	61	33	-	-	67
(gV)V	4o	-	-	-	100	60	13	4	5	78	81	48	-	-	52

It appears from the table that in the cases where a stop consonant is heard it is in most cases a labial, irrespectively of the place of articulation of the consonant removed. Most b-answers are found before u, and hardly any are found before i. This problem was discussed in 3.2., where the hypothesis was advanced that an identification of a labial before i would probably require a transition of F_2 or F_3.

B. Interchange of explosion + transition was tried before i and u only, with one example of each (i.e. 21 answers). As might be expected, the substituted unit determines the identification in 100% of the cases, except for du (bu) where the percentage is 95.

4.1.3. Identification of bdg-words with removed explosion

The results of presenting bdg-words with removed explosions to the listeners are given in table XI and in graphical form in Fig.6

TABLE XI

Identification (in %) of bdg-words after removal of the explosion

	i					a					u				
	N	b	d	g	x	N	b	d	g	x	N	b	d	g	x
(b)VV	62	87	5	-	8	42	86	-	-	14	62	52	-	1	47
(d)VV	62	24	10	-	64	82	31	60	-	9	62	44	40	-	16
(g)VV	61	8	-	12	80	62	19	7	64	10	62	29	-	7	64

The table and Fig.6 show that when the vowel is a, it is in most cases possible to remove the explosion and still get correct answers. Before i only b is heard correctly, before u b and d are recognized in about 50% of the cases, whereas g cannot be identified.

If the transition is cut out of words and presented alone, it is still possible to hear stops in many cases. The results are given in table XII.

TABLE XII

Identification (in %) of transitions cut out of bdg-words (N=40)

	i				a				u			
	b	d	g	x	b	d	g	x	b	d	g	x
(b)V(V)	90	-	-	10	70	-	-	30	82	-	5	13
(d)V(V)	42	-	3	55	15	35	3	47	25	60	-	15
(g)V(V)	7	3	43	47	12	3	45	40	45	-	20	35

A comparison with table XI shows certain deviations. da and ga are perceived less correctly, bu, du, and gi better than in words. This I cannot explain. Perhaps short transitions are heard better in isolation.

On the other hand the agreement is quite good between the test with removal of explosions in words and the test with unreleased bdg (see the report in this volume). In table XIII the two tests are compared. The percentages given for unreleased bdg are averages of the VC and the CVC test.

TABLE XIII

Comparison between correct identifications of words with removed initial bdg and words with unreleased final bdg

(b)i	87	i(b)	91	(d)i	10	i(d)	35	(g)i	12	i(g)	45
(b)a	86	a(b)	69	(d)a	60	a(d)	57	(g)a	64	a(g)	89
(b)u	52	u(b)	56	(d)u	40	u(d)	90	(g)u	7	u(g)	51

On the whole, the identification is better in final than in initial position. This is not surprising since the open interval of 10-30 ms in initial bdg cuts off part of the transition. Both finally and initially b is identified more correctly in combination with i and a than with u, d more correctly in combination with u and a than with i, and g more correctly in combination with a than with i and u.

As for mistakes, the number of zero-answers is greatest in the test with initial bdg. In most of the remaining mistakes a labial was heard. This was also the most common mistake in the test with final unreleased stops, but in some cases b was heard as g after a rounded vowel.

4.1.4. Identification of bdg-words after removal of the transition

The result of removing the transition in bdg-words is seen in table XIV and in Fig. 7.

TABLE XIV

Identification of bdg-words after removal of the transition

	i					a					u				
	N	b	d	g	x	N	b	d	g	x	N	b	d	g	x
b(V)V	40	70	8	10	12	61	18	-	-	82	40	82	-	-	18
d(V)V	40	-	80	20	-	61	23	21	-	56	40	60	35	-	5
g(V)V	40	13	5	82	-	61	25	-	44	31	40	5	-	90	5

A comparison between Figs. 6 and 7 shows striking differences between the results of removing the explosion and of removing the transition. In the first case all consonants were identified pretty well before a, in the second case they are identified most poorly before a. The opposite is true of i and u, except that bi was also identified well without explosion, and du is not heard better with explosion than with transition. The percentages of correct identifications for the different vowels with removed explosions are: a 70%, i 36%, and u 34%, with removed transitions a 24%, i 77%, u 69%.

A glance at the schematic spectrograms in Fig.8 will give an explanation to these differences. In ba, da, and ga the transitions are clearly different, and the explosions very similar. Therefore the transitions are more important for the perception,

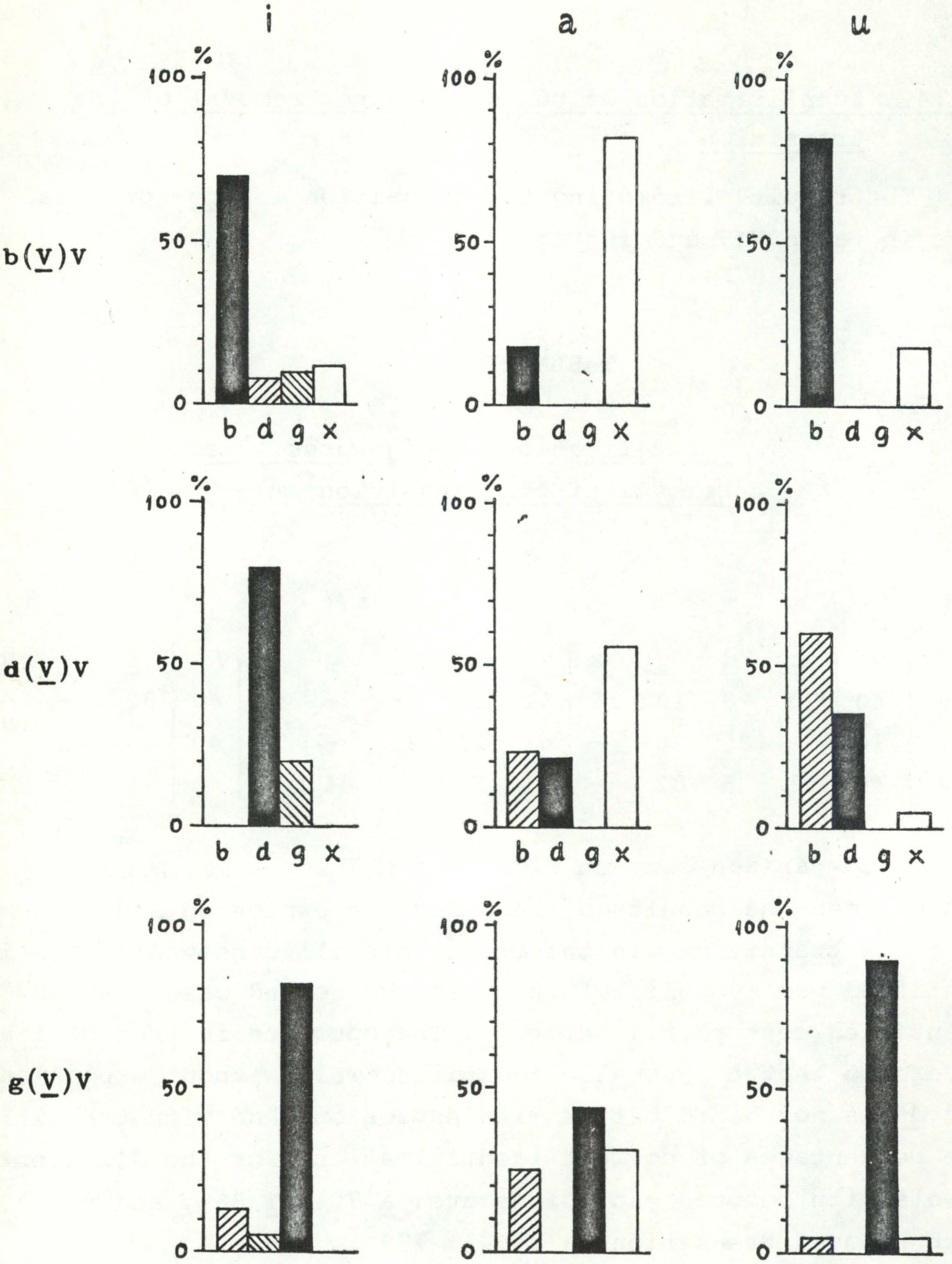
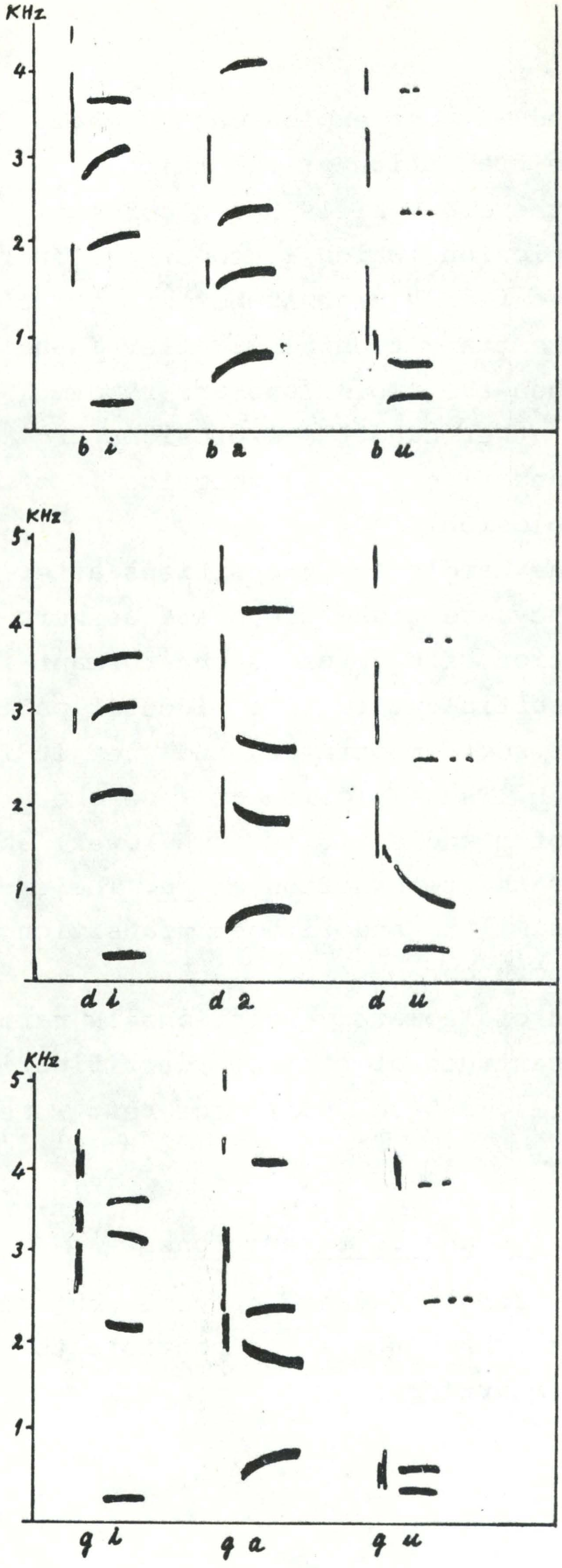


Fig. 7. Identification of bdg + iau after removal of the transition.
V = transition, X = 0 (h, f)



(1) -explosion

	i	a	u
(b)	b	b	$\underset{\dots}{b}$/$\boxed{0}$
(d)	$\boxed{\underset{\dots}{0}}$	$\underset{\dots}{d}$	$\boxed{\underset{\dots}{b}}$/d
(g)	$\boxed{0}$	$\underset{\dots}{g}$	$\boxed{\underset{\dots}{0}}$

(2) -transition

	i	a	u
(b)	b	$\boxed{0}$	b
(d)	d	$\boxed{\underset{\dots}{0}}$	$\boxed{\underset{\dots}{b}}$
(g)	g	$\underset{\dots}{g}$/$\boxed{0}$	g

(3) exchange of explosion

decisive for perception:
explosion
○ transition
□ neither

	i	a	u
b/(d)	b	$\textcircled{d}$	$\textcircled{\underset{\dots}{d}}$
b/(g)	b	$\underset{\dots}{b}$ $\textcircled{g}$	$\underline{b}$
d/(b)	d	$\textcircled{b}$	$\textcircled{\underset{\dots}{b}}$
d/(g)	d	$\textcircled{g}$	$\boxed{b}$
g/(b)	$\underset{\dots}{g}$	$\textcircled{b}$	g
g/(d)	g	$\textcircled{d}$	g

... = only 50-70% majority

Fig. 8. Schematic spectrograms of *bdg + iau* together with a survey of the most frequent answers to stimuli with (1) removed explosions, (2) removed transitions, and (3) interchanged explosions.

and the words can be identified without explosions but not without transitions. In bi the transition of the third formant is very pronounced, and therefore bi is heard correctly without explosion. But the explosion (which is somewhat lower than in di and gi) also gives a fairly high number of correct answers without transition. The transitions in i after d and g are very small, therefore when the explosions are removed no consonant is heard. On the other hand the explosions are sufficiently different to be decisive (the g-explosion is of longer duration than the d-explosion).

As for the vowel u, it has hardly any transitions after b and g. Without explosions they are therefore heard as zero or as b; d has a clear transition of F_2, but as the formant is rather weak, it is not quite sufficient to allow identification. The transition is, however, of some importance, and when it is cut out, d is mostly heard as b. The explosion of g before u is quite different from that of b and d (it is much lower) and it is therefore sufficient for the recognition of gu. The explosions of b and d are very similar, and without transition b is the most common answer in both cases.

The place of articulation of isolated explosions is rarely identified correctly. The percentages of correct identifications were: 29% for g(u), 33% for b(i), 30% for d(i), the rest were below 12%.

4.1.5. Interchange of explosions and of transitions

When both explosions and transitions are present, but one of them taken from a different consonant, it is possible to throw more light on their relative importance.

A. Interchange of explosions

The result of interchanging explosions in bdg-words is given in table XV and in graphical form in Fig.9.

TABLE XV

Identification (in %) of bdg-words after interchange of explosions (N=21 for i and a, 42 for u)

	i				a				u			
	b	d	g	x	b	d	g	x	b	d	g	x
b(d)	71	24	5	-	5	95	-	-	33	64	-	2
b(g)	80	10	10	-	57	5	38	-	86	-	-	14
d(b)	-	100	-	-	100	-	-	-	68	10	-	21
d(g)	-	76	24	-	-	19	81	-	95	5	-	-
g(b)	29	10	57	5	76	10	14	-	-	-	100	-
g(d)	-	10	90	-	-	81	19	-	-	29	71	-

It is possible to summarize the results in a few simple rules:

(1) Before i the explosion is decisive

(2) Before a the transition is decisive, except for b(g)

(3) Before u the results are different according to the individual combinations.

These rules are given in schematic form in table XVI

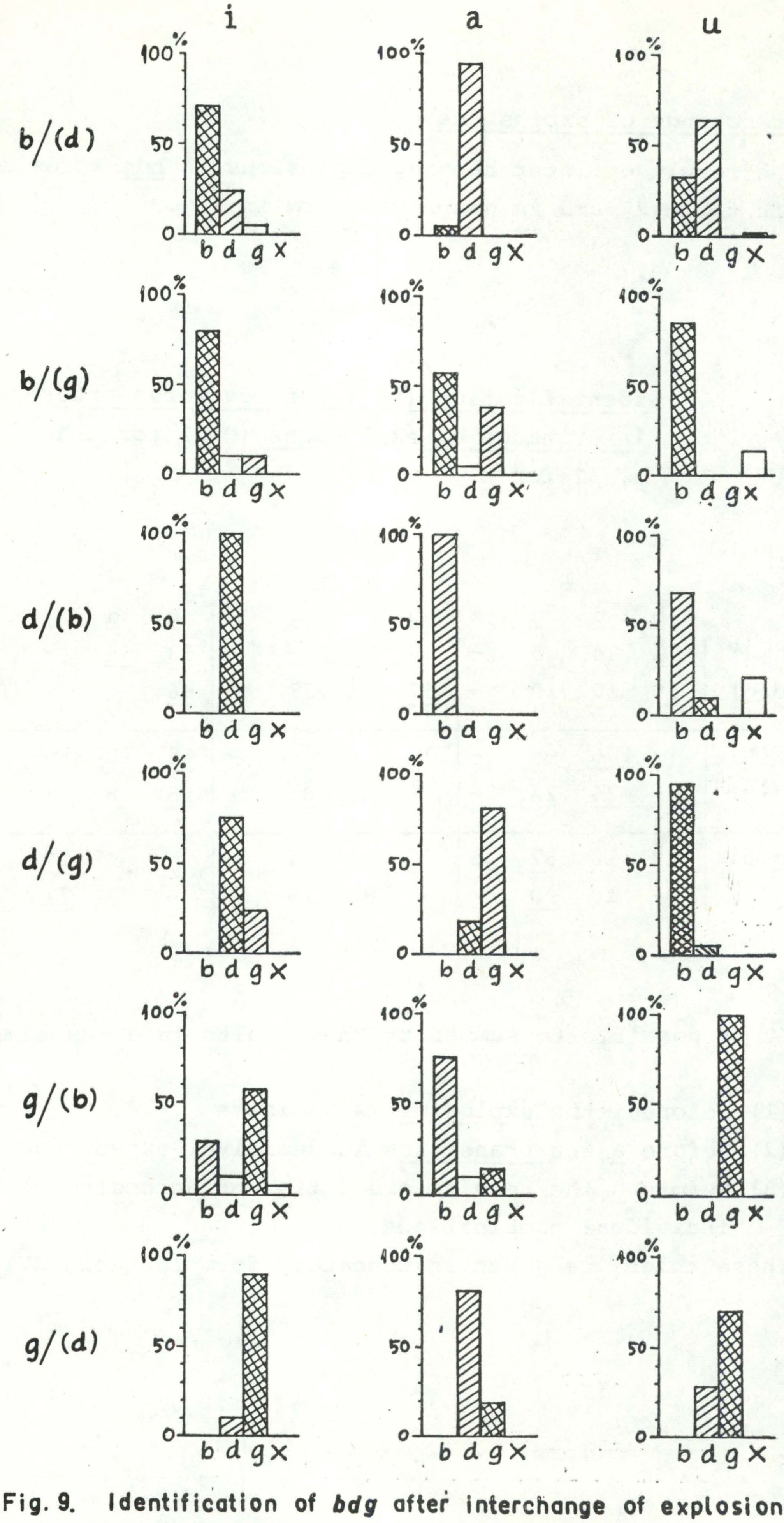


Fig. 9. Identification of *bdg* after interchange of explosions.
▩ = explosion, ▨ = transition.

TABLE XVI

The relative importance of explosions and transitions when explosions are interchanged in bdg-words (+=decisive, - not decisive)

	i		a		u	
	expl.	trans.	expl.	trans.	expl.	trans.
b/(d)	+	-	-	+	-	+
b/(g)	+	-	+	-	+	-
d/(b)	+	-	-	+	-	+
d/(g)	+	-	-	+	-	-
g/(b)	+	-	-	+	+	-
g/(d)	+	-	-	+	+	-

The schematic spectrograms in Fig.8 can again be used to explain the results.

As for the vowel i, di and gi have very small transitions and therefore the explosion must be decisive. This is also in agreement with the results from removing explosion or transition (Figs.6 and 7). One might have expected the strong transition in bi to have influenced the perception of d/(b) and g/(b) before i, but the explosions of d and g have probably been too high and too strong to allow interpretation as b.

The vowel a has different transitions after b, d, and g, whereas the explosions are not very different. Therefore the transitions determine the responses. This is in agreement with the results of removing explosions or transitions (Figs.6-7), where the transitions alone were found to be sufficient, but not the explosions. The exception b/(g) heard as b is somewhat astonishing.

As far as u is concerned, g has a characteristic explosion which must be decisive for g/(b) and g/(d), although the transition after d diminishes the majority for g somewhat. In b/(g) the explosion must also be decisive, because the explosions of b and g are very different and the transitions very similar. On the other hand, when b and d are interchanged, the transitions must be decisive because they are rather different, whereas the explosions are very similar. Finally, when a d-explosion is placed before a gu-transition,the result is b. This is not as surprising as it may look at first sight. The d-explosion is similar to the b-explosion. Thus it could be either d or b (but not g). The g-transition is similar to that of b, but different from the d-transition. Thus it could be g or b (but not d). The remaining possibility is b.

B. Interchange of transitions

An interchange of transitions should give the same result as an interchange of explosions, but the operation is somewhat more complicated and requires very precise splicing. It was only carried out for the vowel u, and it gave the same result as the interchange of explosions (table XV), but with less pronounced majorities and a small deviation for b/(d) which had the same number of b- and d-answers. The percentages are given in table XVII.

TABLE XVII

Identification (in %) of bdg before u
after interchange of transitions (N=21)

	b	d	g	x
b/(d)	38	38	5	19
b/(g)	67	-	10	24
d/(b)	52	48	-	-
d/(g)	71	19	-	-
g/(b)	-	-	100	-
g/(d)	-	10	90	-

4.2. Identification of place of articulation in ptk

ptk are more complicated than bdg, because the aspiration covers most of the transition time, so that there is very little vowel transition, and because of the affrication of t. The answers therefore cannot be compared directly with those given to bdg-words, and the results apply more particularly to the Danish language than the results for bdg which can probably be generalized to other languages as well.

4.2.1. Identification of unchanged ptk

There were 21 unchanged ptk-words in the test, and thus 441 answers. The words with t were heard 100 % correctly, those with p 99 % (there was one k-answer for pa). For the k-words thé percentage was 91. There were 8 % p-answers (all for ka) and 1 % zero (for ku).

4.2.2. Identification of ptk-words after removal or exchange of both explosion and aspiration

A. When both explosion and aspiration are removed there are still a good number of ptk-answers before a and of bdg-answers before u (see 3.2.). But the place of articulation is identified only in some examples of (th)i in which the opening movement starts rather late because of the strong affrication and which, therefore, have vowel transitions in F_3. Otherwise almost all stops heard are identified as labials. The results for the single consonant-vowel-combinations are seen in table XVIII.

TABLE XVIII

Identification (in %) of ptk-words with both explosion and aspiration removed (N=21)

	i				a				u			
	lab.	alv.	vel.	x	lab.	alv.	vel.	x	lab.	alv.	vel.	x
(ph)V	5	-	-	95	76	-	-	24	62	-	-	38
(th)V	14	19	-	67	86	-	5	9	43	-	9	48
(kh)V	19	-	5	76	67	-	-	33	81	-	5	14

Since labials are heard just as frequently when the removed consonant was alveolar or velar as when it was labial, it would be misleading to say that the labials were correctly identified before a and u. If the vowel starts abruptly and there are some rudimentary transitions indicating a preceding consonant, but no clear indications of the place of articulation, the normal answer is apparently "labial". The exception before i may, as proposed above (3.2. and 4.1.2.), be due to the fact that labials before i require extensive transitions of F_2 or F_3.

On the other hand, the isolated segments consisting of explosion + aspiration can on the whole be identified as far as the labials and alveolars are concerned, whereas the velars are badly identified. This was only tried before the vowels i and a. The percentage correct answers were

ph(i)	67%,	th(i)	90%,	kh(i)	33%
ph(a)	100%,	th(a)	90%,	kh(a)	24%

kh(i) was often heard as t, kh(a) as indefinite noise. This k had a relatively weak aspiration.

B. Interchange of explosion and aspiration as a whole gives the result that the substituted segment determines the identification. This is true in 96-100 % of the cases with the exception of kh/(th)i which gave 81 % k, and 19 % p.

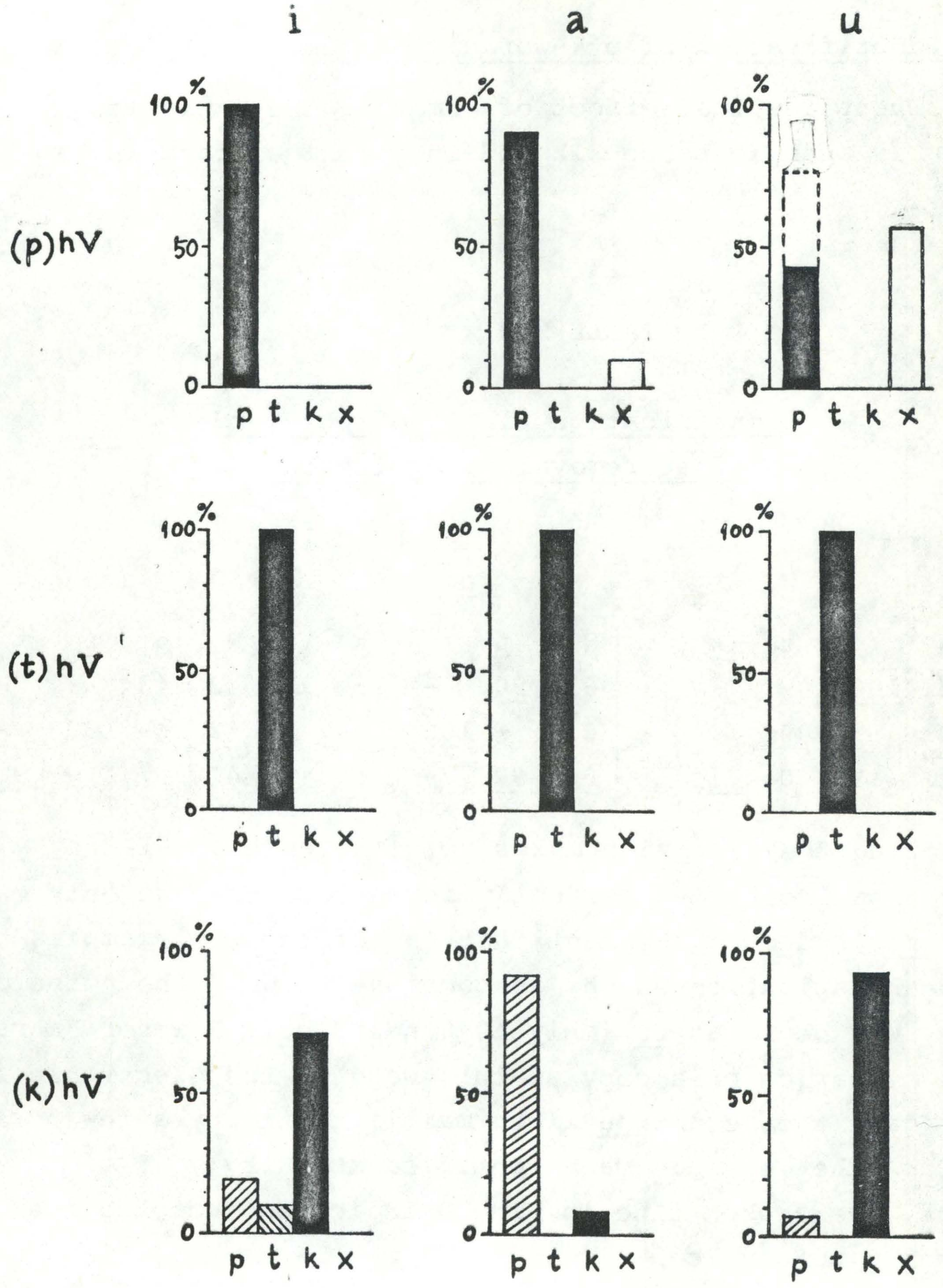


Fig. 10. Identification of *ptk + iau* after removal of explosion.

4.2.3. Identification of ptk-words with removed explosion

A. The perceptual effect of removing the explosion in *ptk*-words is seen in table XIX and in graphical form in Fig. 10.

TABLE XIX

Identification (in %) of ptk-words after removal of explosion

	i					*a*					*u*				
	N	*p*	*t*	*k*	x	N	*p*	*t*	*k*	x	N	*p*	*t*	*k*	x
(p)HV	21	*100*	-	-	-	63	*90*	-	-	10	42	43	-	-	*57*
(t)HV	21	-	*100*	-	-	63	-	*100*	-	-	42	-	100	-	-
(k)HV	21	19	10	*71*	-	63	*92*	-	8	-	42	7	-	*93*	-

Removing the explosion makes no difference at all for *t*, and *p* is also identified correctly in most cases. The only exception: *pu* is due to the fact that one of the two examples had a very weak aspiration which, in combination with the noise on the tape, was heard as *fu* (only 10 % heard *pu*). However, a renewed segmentation of a copy of this word on the electronic segmentator gave the result *pu* (informal listening by a few persons). The other example was identified as *pu* by 76 % of the listeners. The broken line in Fig.10 indicates that the result was probably not typical.

k is identified correctly in most cases before *i* and *u*, but it is heard as *p* before *a*. This answer, which is identical for all three examples, can be explained by the fact that the aspirations of *p* and *k* are very similar, whereas the explosions are different. The explosion of *k* is normally stronger, of longer duration (it often contains two maxima at a short distance) and somewhat more concentrated in frequency than that of p. When this characteristic explosion is lacking and only the aspiration is left, the listeners hear a p (see also Fig.12, where the main results of the tests with ptk can be compared with schematic spectrograms).

B. Interchange of aspirations (with removed explosions) was tried with p-, t- and k-aspirations (one example of each before i a and u). The aspirations were interchanged before the same vowel; thus there were 18 different items. In practically all cases the identification was determined by the substituted segment (100 % for p and t, 95 % for k). The pu-explosion was identical with the one which gave 76 % pu-answers before the interchanges. When substituted for (t)h and (k)h, it was heard 100 % as pu.

The aspiration is thus sufficient for the identification of ptk in all cases except for ka, and it does not matter if the following short vowel transitions are taken from other consonants.

4.2.4. Identification of ptk-words after removal of the aspiration

If the aspiration is removed and the explosion moved so as to join the vowel, most listeners hear bdg (with some ptk-answers before a, cp.3.2.). The same result is obtained when ptk-explosions are replaced by bdg-explosions (see 3.6. and 3.7.). In table XX and Fig.11 the answers are distributed according to place of articulation.

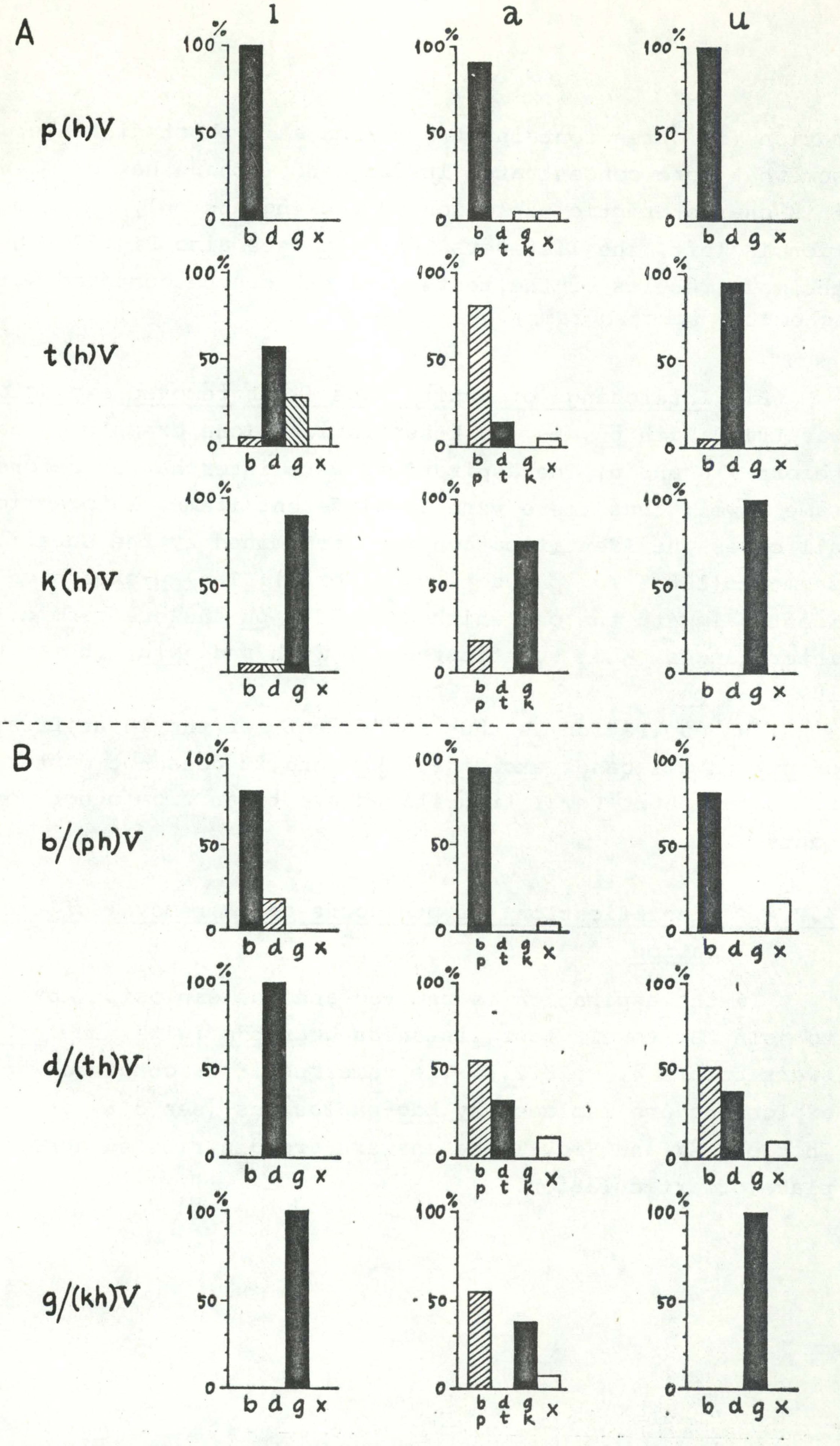
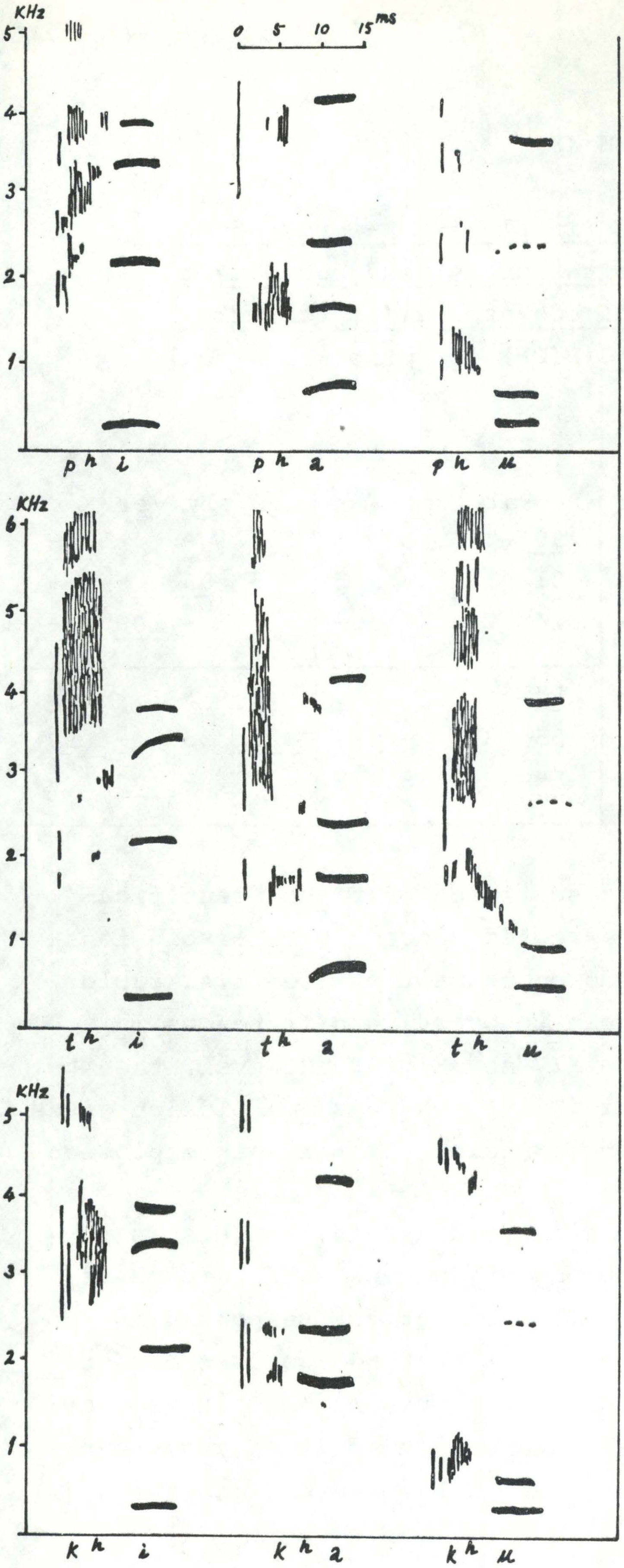


Fig. 11. Identification of place of articulation after removal of aspiration in *ptk* (A) and after subsequent replacement of *ptk*-explosions by *bdg*-explosions (B).



(1) −explosion

	i	a	u
(p)h	p	p	(p)
(t)h	t	t	t
(k)h	k	$\boxed{p}$	k

(2) −aspiration

	i	a	u
p(h)	b	b	b
t(h)	$\underset{\cdots}{d}$ $\boxed{(g)}$	$\boxed{b}$	d
k(h)	g	g	g

(3) exchange of explosion

decisive for perception:
explosion
○ aspiration
□ neither

	i	a	u
p/(t)h	ⓣ	ⓣ	ⓣ
k/(t)h	ⓣ	$\underset{\cdots}{ⓣ}$	ⓣ
p/(k)h	p	p	p
t/(k)h	ⓚ	$\boxed{p}$	$\boxed{p}$
t/(p)h	ⓟ	ⓟ	$\underset{\cdots}{ⓟ}$(t)
k/(p)h	k	k	k

... = only 50 70% majority

Fig. 12. Schematic spectrograms of *ptk+iau* together with a survey of the most frequent answers to stimuli with (1) removed explosions, (2) removed aspirations and (3) interchanged explosions.

TABLE XX

Identification (in %) of ptk-words after removal of aspiration (A) and subsequent substitution of bdg-explosions for ptk-explosions (B).(B(PH) 42, otherwise N=21)

	i				a				u			
	lab.	alv.	vel.	x	lab.	alv.	vel.	x	lab.	alv.	vel.	x
p(h)V	100	-	-	-	90	-	5	5	100	-	-	-
t(h)V	5	57	28	10	81	14	-	5	5	95	-	-
k(h)V	5	5	90	-	24	-	76	-	-	-	100	-
b(ph)V	81	19	-	-	95	-	-	5	81	-	-	19
d(th)V	-	100	-	-	55	33	-	12	52	38	-	10
g(kh)V	-	-	100	-	55	-	38	7	-	-	100	-

Before i the explosion is sufficient for the identification. This is what should be expected since the bdg-explosions were found sufficient with removed transition (see 4.4. table XIII and Fig.7)and were even able to override differences in transitions (see 4.1.5., table XIV and Fig.9). However, the majority is small for t before i (57 %), whereas d-explosion gives 100 % d. This difference can be explained by the weak explosion of t, particularly before i.

Before a most explosions are heard as labial. This is in accordance with the answers to bdg-explosions with removal of transitions in so far as the number of correct responses is approximately the same, but in the case of bdg-words the most common answer was zero, whereas in the present case it is p or b. This can be explained by the presence of rudimentary transitions and the particular vowel start after aspiration (see 3.7.).

Only the *k*-explosion is sufficiently strong to cause a majority of *k*-answers (see also Fig.12). The *d*-explosion is somewhat more efficient than the *t*-explosion because it is stronger.

Before *u* labial and velar explosions are sufficient. This is what could be expected on the basis of the experiments with *bdg*-words (table XIV and Fig.7). But in this case the *t*-explosion is superior to the *d*-explosion (the *t* in question had a relatively strong explosion, and the *d*-explosion contained energy at relatively low frequencies).

Before *u* the *ptk*-explosions were interchanged. But they were still decisive for the identification in 90-100 % of the answers. This also shows that the rudimentary vowel transitions are of very little importance.

4.2.5. Interchange of explosions and of aspirations

In the preceding sections we have seen that the aspiration alone (including affrication) is sufficient to allow identification of *ptk*, with the exception of *ka*, and that the explosion alone is sufficient to allow identification of the place of articulation except for *ta* and partly *ti* (*bdg*-explosions were not sufficient for *da*, *du* and *ga*).

Now the question is what happens when explosion and aspiration are in conflict. This question was investigated by interchanging explosions (this was done before *i*, *a*, and *u*) and aspirations (before *i* and *u* only).

The results of these two experiments are given in table XXI. Fig.13 contains a graphical representation of the former experiment (XXI,A).

TABLE XXI

A. Interchange of explosions in ptk (N=21)

		i				a				u			
expl.	asp.	p	t	k	x	p	t	k	x	p	t	k	x
p	(t)h	-	100	-	-	-	95	-	5	-	100	-	-
k	(t)h	-	95	-	5	-	57	14	29	-	86	5	10
p	(k)h	95	-	5	-	95	-	5	-	90	-	5	-
t	(k)h	10	-	90	-	100	-	-	-	100	-	-	-
t	(p)h	95	-	-	5	100	-	-	-	62	38	-	-
k	(p)h	19	-	81	-	14	-	86	-	-	-	100	-

B. Interchange of aspirations (N=21)

		p	t	k	x		p	t	k	x
p	(t)h	10	90	-	-		-	100	-	-
k	(t)h	-	100	-	-		-	86	10	5
p	(k)h	86	-	14	-		95	-	-	5
t	(k)h	10	-	90	-		71	24	-	5
t	(p)h	95	5	-	-		43	57	-	5
k	(p)h	71	-	24	-		-	-	95	5

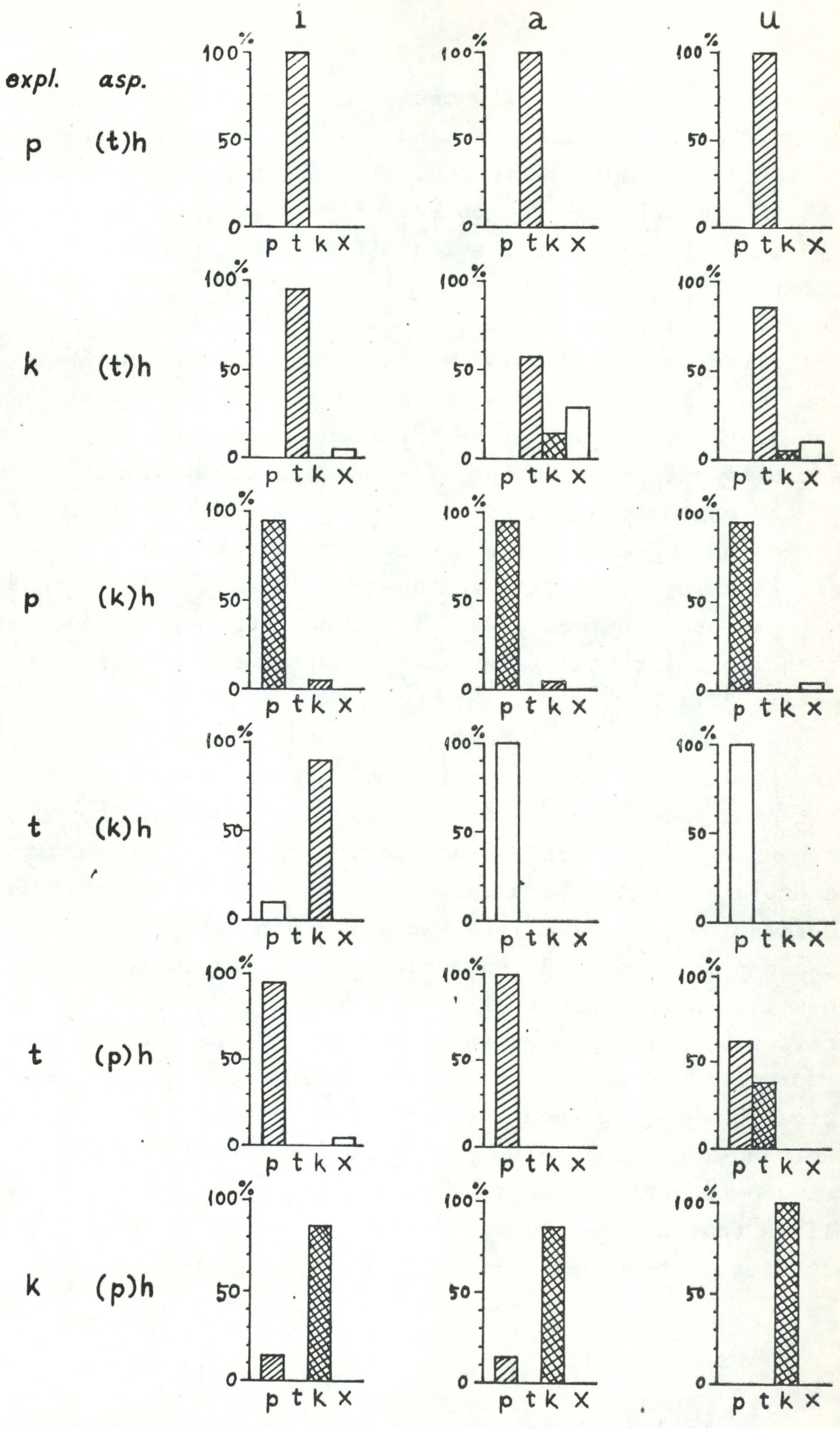


Fig. 13. Identification of *ptk* after interchange of explosions.
▩ = explosion, ▨ = transition

The answers are - with one exception - in good agreement in the two experiments, and the results can be described in relatively simple terms (cp. also the schematic spectrograms in Fig.12): (1) The characteristic affrication is necessary and sufficient for the identification of t.When it is present, the listeners hear a t irrespectively of the preceding explosion. The majority is large except for k-explosion + t-aspiration before a (57 %). Here the k-explosion is so strong that some listeners (19 %) hear a k and others (24 % of the 29 listed under x) hear the consonant cluster kt. When the t-affrication is absent, listeners fail to identify the t, i.e. the explosion is not sufficient. The only exception is t-explosion + p-aspiration + u in experiment B, which has a small majority for t (in experiment A there are 33 % t-answers. The t-explosion before u was relatively strong, and the p-aspiration looks like the later (low) part of the t-aspiration (see also Fig. 12)).

(2) In the other cases, i.e. in the examples with p- or k-aspiration, the explosion is decisive in the sense that p- and t-explosions give p, and k-explosion gives k. The only exception is t-explosion + k-aspiration before i, which gives k. The explanation of these answers is that p- and k-aspirations are relatively similar (see Fig.12), therefore the distinction between p and k must depend mainly on the explosions. k-explosions are characteristically different from both p- and t-explosions by being longer and stronger and (particularly before u) more concentrated. This explosion is necessary for the identification of k. p- and t-explosions are more similar to each other and might both give p or t. Since identification as t is excluded because the characteristic t-affrication is absent, it follows that not only must p-explosion + k-aspiration give p-responses, but so must t-explosion + p- and k-aspirations. The

exception *t*-explosion + *k*-aspiration before *i* heard as *ki* can be explained by the fact that *pi* requires either the explosion or the aspiration to begin at a rather low frequency.

The only evident case of conflict between the two experiments (A and B) is *k*-explosion + *p*-aspiration + *i*. In A there is a clear majority for *k*, which is in accordance with the rules given above. In B, however, there is a clear majority for *p*. The spectrograms look very similar, and when I listened to them again, I found that they also sounded alike;both of them gave the impression of a sound somewhere between *p* and *k*. In such a case the environment in the test may have played a role. The example heard as *ki* came (with one word between) after a clear example of *pi*, the example heard as *pi* came (with one word between) after a clear *gi*. But such a context effect in consonant perception with another word in between, is not very probable. Another explanation may be sought in the intensity level. The example heard as *ki* had a somewhat higher level than the other. Some informal experiments with changes of intensity level showed that this might have an effect on the identification as *k* or *p*.- Reactions to stimuli with two conflicting cues are, on the whole, subject to variations.

The importance of the affrication phase for the identification of *t*, also appears from the fact that a *t*-explosion put before *h* + vowel does not give *t*, but either a labial (before *a* and *u*) or a velar (before *i*). (*p*- or and *k*-explosions before *h* are heard as *p* and *k* respectively in almost all cases). Similarly, if explosion and affrication is cut off in *ta* 4o ms before the vowel start (i.e. after the affrication phase), the listeners hear a labial.

5. Summary of results

5.1. Summary of the results concerning the difference between the ptk- and the bdg-category

The primary cue for distinguishing between Danish _ptk_ and _bdg_ is the presence or absence of aspiration. _ptk_ normally cannot be identified without aspiration, whereas the explosion noise is of very little importance (3.2.). It does not matter whether the explosion is a _ptk_- or a _bdg_-explosion (3.6.) or whether it is there at all (3.1.).The aspiration phase must contain noise; a simple pause is not sufficient (3.3.). The duration of the aspiration noise must exceed 3o ms to give a majority of _ptk_-answers, and a high percentage of _ptk_-answers is not reached until about 45-50 ms (3.4.).

A secondary cue is the vowel start, at any rate for the vowel _a_ (3.7.). After an aspiration there may be some weak low frequency vibrations accompanied by somewhat noisy second and third formants. This breathy start may give a relatively high percentage of _ptk_-answers even at very short distances from the explosion (e.g. 2o ms). These _ptk_-sounds are, however, often described as "unaspirated _ptk_", i.e. not as normal Danish _ptk_-sounds,but different from _bdg_.

Lack of aspiration is insufficient for the identification of _bdg_. This requires the presence of explosion and/or vowel transitions (3.1.). The relative importance of explosion and transition depends on the vowel. Before _i_ and _u_ the explosions are necessary and sufficient, and the transitions constitute a secondary cue. (Transitions alone give only about 50 % _bdg_-answers). Before _a_ the explosion is not sufficient and of minor importance.Here transitions alone give 86 % identification. The rudimentary transitions found after _ptk_ may give some improvement compared to absence of transition (3.2.), but they are unable to give a majority of bdg-answers.

The aspiration-cue of ptk is much stronger than the transition-cue of bdg. If both are combined, the result is ptk (3.2.).

5.2. Summary of the results concerning place of articulation

The contributions of explosions, aspirations and transitions to the identification of place of articulation when each of them is the only present cue is indicated in schematic form in table XXII.

TABLE XXII

Contribution of explosions, transitions and aspirations to the identification of place of articulation (+= more than 70 % identification, ±=50-70 % identification, -=less than 5o % identification)

		i			a			u		
bdg	expl.	+	+	+	−	−	−	+	−	+
	trans.	+	−	−	+	±	±	±	−	−

		i			a			u		
ptk	expl.	+	±	+	+	−	+	+	+	+
	asp.	+	+	+	+	+	−	+	+	+

In the case of *ptk* both explosions and aspirations were found to be sufficient in most cases (but without aspiration the sounds were heard as *bdg*). The only exceptions were *ka* without explosion heard as *pa*, *ta* without aspiration heard as *pa*, and *ti* without aspiration which was often heard as *gi*.

For the identification of *bdg* the explosions were found to be sufficient before *i* and *u* (with the exception of *du*) but not before *a*. The somewhat lower efficiency of explosions in *bdg*-words than in *ptk*-words is probably partly due to the fact that *g* normally has a somewhat weaker explosion than *k*, and that *b* before *a* has a particularly weak explosion; an additional reason is that the explosions in *ptk*-words are probably supported by rudimentary vowel transitions. Why the *d*-explosion before *u* should be inferior to the *t*-explosion, is not quite clear. It may not have been typical (it contained rather low frequency noise).

The vowel transitions in *bdg* constitute a weaker cue than the aspirations in *ptk*, also for the place of articulation. They are, however, important in *bi* and in *ba da ga* (but not quite sufficient in *da* and *ga*). In *bu* the identification without explosion is just above 5o %, in *du* just below 50 %.

Figs. 8 and 12 give a survey of a somewhat different kind. Here the most frequent answers to stimuli with removed explosions or transitions (aspirations) are given together with the answers to exchange of explosions. Schematic spectrograms have been added as a visual support to the explanations.

Detailed explanations have been given in the preceding sections, see particularly 4.1.4. and 4.1.5. for *bdg* and 4.2.3. - 4.2.5. for *ptk*. Most identifications can be explained when it is assumed that extensive transitions and strong explosions (and aspirations) are more efficient than small transitions and weak explosions, and that transitions, aspirations, and explosions of a given consonant which differ characteristically in frequency or intensity from those of other consonants before the same vowel are more efficient than those which differ little from one consonant to the other before the same vowel.

Thus velar explosions differ from labial and alveolar explosions in being stronger, of longer duration, and with a concentration of energy close to F_2 of the vowel. They are therefore relatively strong cues; an exception is the explosion of *ga* which is less concentrated and more similar to that of *d*. Similarly the strong affrication of *t* makes it easily recognizable from *p* and *k*, which have mutually similar aspirations. The relatively pronounced and mutually different transitions of *ba* *da* *ga* constitute a better cue than the relatively similar explosions. As for *i* and *u*, on the other hand, only *bi* and *du* have extensive transitions. Therefore the transition is sufficient in *bi* and of some importance in *du*, though not sufficient in the latter case because the formant in question is weak.

The answer *b* for *d*-explosion + *g*-transition before *u* and the answer *p* for *t*-explosion + *k*-aspiration before *a* and *u* can also be explained in this way. These stimuli lack the characteristic velar explosion and the alveolar transition (or affrication), and the remaining possibility is thus a labial. On the whole, there is a tendency to hear labials when no other strong cues are present.

5.3. The reliability of the results

I have not found it worth while to test all the results mentioned in the preceding sections for significance. As each table contains information both about the percentual identifications and about the number of responses upon which the calculations are based, the reader will be able to judge about the reliability for himself. Moreover, in the cases with several examples the significance is often quite evident. This is particularly true of the responses to the categories ptk and bdg. In other cases,where sometimes only one example has been used (e.g. in some tests, about place of articulation),a significance test might be misleading. If there is a high degree of unaminity among the 21 listeners, the identification may well be statistically significant in the sense that this unaminity cannot be due to chance. It might, however, be due to some atypical characteristics of the example in question, which does not allow of any generalisation. I have therefore found it more important to point to consistencies among different tests and to find explanations of the responses based on the typical features of the sounds in question. If such consistencies are found and such explanations can be given, the probability that the results can be generalized and that they tell something about the perceptual cues, is relatively high. Examples of consistencies are the dominance of the t-affrication before all vowels irrespectively of preceding explosions, or the weakness of the explosion as a cue for bdg before a, both when the transitions are removed, and when explosions are interchanged, also the importance of the transitions in the same cases. Examples of plausible explanations are the specific character of velar explosions, the similar and strong affrication noise after t before all vowels, and the extensive transitions in ba da ga bi.

6. Comparison with the results of other investigations

6.1. The distinction between ptk and bdg

In the literature on stop consonants there has been a certain tendency to talk about the various acoustic characteristics and perceptual cues without reference to any specific language. But ptk and bdg are labels covering rather different sound types, and the different cues and their relative importance must be described for different languages or language types separately. The confusion is considerably increased by the use of the terms "voiced" and "voiceless" in the same sense as bdg and ptk, i.e. as vague labels for a variety of sound types. bdg may not be voiced in the narrower sense of the word; they are voiceless in Danish, and may be voiceless initially for instance in English and German. The results found for Danish stops cannot be directly compared with the results of tests with French and Dutch stops, which are of quite a different type, but they can partly be compared with the results for English, since in both languages the main difference between ptk and bdg initially is one of aspiration. The aspiration is, however, on the whole somewhat longer in Danish, and the strong affrication of Danish t, involving a short and weak explosion followed by a long fricative phase gives some specific results which cannot be generalized to languages without affrication, (with a different segmentation, for instance the first 20 ms counted as explosion, the explosion phase would have been more like that of English t, but the "aspiration" phase would have been different anyway).

The importance of the aspiration cue for the identification of Danish ptk is in good agreement with the results of perceptual tests with English listeners. Aspiration is practically the same as "F_1 cut back" (Liberman et al. 1958) or "voicing lag", which has found to be the most important cue for the distinction between bdg and ptk in English (e.g. Lisker and Abramson 1964). Only it should be emphasized that pure voicing lag is not sufficient; there must be noise at the frequencies of the higher formants in the interval, as is also the case in the experiments by Lisker and Abramson.

There is also good agreement between the crossing point found in the experiment with English listeners and the approximate crossing found in the present experiment and in the experiment with identification of foreign stops. - I do not know of any experiments with breathy start of the vowel.

6.1. Identification of place of articulation

The Haskins group have made extensive investigations by means of synthetic stimuli in order to throw light on the importance of explosions and transitions in stop-consonant perception. They have been able to synthesize stimuli heard as labial, alveolar and velar stops both by means of explosions alone (Liberman et al. 1952, Cooper et al. 1952) and transitions alone (Liberman et al. 1954).

There is, however, particularly in the cases with explosion alone, a good deal of overlapping between the answers. The result of the present experiment, i.e. that the explosions are more important before i and u and the transitions in the case of a, does not appear from the experiments of the Haskins group. But a comparison between synthetic and natural sounds is not quite straightforward. The experiment with synthetic explosions may, for instance, have given relatively many velar responses because the noise employed as a stop cue was of relatively long duration (15 ms) and of concentrated frequency, and therefore more like a velar explosion than like an alveolar or labial explosion. The experiment with synthetic formant transitions shows better discrimination for a than for i and u in the section with F_1 transition and bdg-answers, but not in the section with ptk-answers. di gives a particularly bad result (like in our experiment), but it is improved by the addition of F_3 (Harris et al. 1958 and Hoffman 1958). Also ga gives an unsatisfactory result, but it is improved by adding a third formant (cp. the results Harris et al. for gæ).

On the whole, the transitions in natural Danish words do not seem as efficient as the synthetic transitions, and there is a certain contradiction between our results and the formulation of the locus theory found in the later writings of P. Delattre (e.g. 1958 and 1969). According to Delattre the locus is the frequency toward which the formant transitions must point in order to contribute maximally to the perception of a given place of articulation as found in experiments with synthetic speech. In natural speech the formant transitions do not point to the locus in the case of velars before rounded vowels because the coarticulation, involving rounding already in the consonants, lowers the resonance. Therefore explosions are needed in this case, but not in others. Delattre critisizes Halle and Householder who have assumed that an extensive transition contributes more to perception, and states that the extension of the transition is irrelevant; it may be straight. What matters is that it points to the locus. Delattre's own experiments seem to corroborate this assumption (Delattre 1969), but the results obtained with Danish stops confirm the hypothesis of Halle and Householder (cp. also Wang and Fillmore 1961): that it does matter whether the transition is extensive or not. And this seems quite natural. If the transition is straight, the listener cannot know whether it points to a consonant locus or to nothing. It must have a certain extension in order to be an efficient cue. The results of the experiments with Danish stops show, for instance, that the explosion is not only necessary in the case of velars in combination with rounded vowels, but also in di, gi, bu, and du. In di and gi the transitions alone give only about 10 % d- and g-answers, in bu and du around 50 %. Even in da and ga the percentage reaches only 60 and 64 % without explosion. But the more extensive the transitions are the more they contribute to the identification.

The importance of explosion was also emphasized in Malécot's experiments with English and French final stops (Malécot 1958, Malécot and Lindheimer 1966), but transitions alone gave also a relatively high percentage of identification (after the vowels ɛ and ɔ).

The importance of the initial explosions before *i* and *u* found in the present investigation has also some interest for the general theory of speech perception. On several occasions (e.g. 1969) Liberman has emphasized that speech is characterized by parallel transmission and overlapping cues (since the transitions contain information about both consonants and vowels)and that the characteristic feature of speech perception is therefore the decoding of such overlapping cues. He makes an exception for fricatives (interchange of Danish *f* and *s* before *i*, *a* and *u* also shows that the fricative noise is alone decisive here), but he might have made an exception for some combinations of stops and vowels too. The examples most frequently quoted by Liberman are *bag* and the syllables *di* and *du*. *bag* is in fact a good example, but it is remarkable that the transition in spoken Danish *di* seems to be of no importance for perception. It cannot be denied that overlapping cues are very often found in speech, and it remains the merit of the Haskins group to have demonstrated this fact, but its role in the perception of natural speech may have been somewhat overemphasized.

The finding that the most common answer is labial when no pronounced cues are present has been confirmed by Öhman (1961) and by Wajskop (1971). Wajskop finds that the stimuli without transitions or explosions are most often heard as labials, but if transitions are present the most common mistake is *t*. It is valid for Wajskop's material, where velars are often heard as dentals. It is, however, an observation which can hardly be generalized. *t* is also a common mistake in my experiments with final stops after unrounded vowels, but not in other cases (see the paper on final unreleased stop consonants in this volume).

References

Cooper, F.S., P.C.Delattre, A.M.Liberman, J.M.Borst and L.J.Gerstman 1952: "Some Experiments in the Perception of Synthetic Speech Sounds", *JASA* 24, p. 597-606.

Delattre, P.C. 1958: "Unreleased Velar Plosives after Back-rounded Vowels", *JASA* 30, p.581-582.

Delattre, P.C. 1969: "Coarticulation and the Locus Theory", *Studia Linguistica* XXIII, p.1-25.

Fischer-Jørgensen, Eli 1954: "Acoustic Analysis of Stop Consonants", *Miscellanea Phonetica* II, p. 42-59.

Fischer-Jørgensen, Eli 1956: "The Commutation Test and its Application to Phonemic Analysis", *For Roman Jakobson*, p. 140-151.

Fischer-Jørgensen, Eli 1969: "Voicing, Tenseness and Aspiration in Stop Consonants, with Special Reference to French and Danish", *ARIPUC* 3, p. 63-114.

Frøkjær-Jensen, B., C. Ludvigsen and J.Rischel 1971: "A Glottographic Study of Some Danish Consonants" *Form and Substance*, p. 123-138.

Harris, K.S. and H.S.Hoffman, A.M. Liberman, P.C. Delattre and F.S.Cooper 1958: "Effect of Third-Formant Transitions on the Perception of the Voiced Stop Consonants", JASA 30, p. 122-126.

Hoffman, H.S. 1958: "Study of Some Cues in the Perception of the Voiced Stop Consonants", JASA 30, p. 1035-1041.

Liberman, A.M., P.C.Delattre and F.S.Cooper 1952: "The Role of Selected Stimulus-Variables in the Perception of the Unvoiced Stop Consonants", The Am. Journal of Psych. LXV, p. 497-516.

Liberman, A.M., P.C.Delattre, F.S.Cooper and L.J.Gerstman 1954: "The Role of Consonant-Vowel Transitions in the Perception of the Stop and Nasal Consonants", Psychological Monographs 68, p.1-13.

Liberman, A.M., P.C.Delattre and F.S.Cooper 1958: "Some Cues for the Distinction between Voiced and Voiceless Stops in Initial Position", Language and Speech 1. p. 153-167.

Liberman, A.M. 1969: "The Grammars of Speech and Language", S.R.Haskins 19/20, p. 79-125.

Lisker, L. and A.S.Abramson 1964: "A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements", Word 20, p. 384-422.

Lisker, L. and A.S.Abramson 1970: "The Voicing Dimension: Some Experiments in Comparative Phonetics", 6. Int. Cong. of Phon. Sciences 1967 1, p.563-567.

Malécot, A. 1958: "The Role of Releases in the Identification of Released Final Stops", Language 34, p. 370-380.

Malécot, A. and E.Lindheimer 1966: "The Contribution of Releases to the Identification of Final Stops in French", Studia Linguistica XX, p. 99-109.

Öhman, S. 1961: "Relative Importance of Sound Segments for the Identification of Swedish Stops in VC and CV Syllables", Stockholm QPSR 3, p. 6-14.

Slis, I.H. and P.H.Damsté 1967: "Transilluminations of the Glottis during Voiced and Voiceless Consonants", *Ipo Annual Progress Report* 2, p. 103-109.

Schatz, C.D. 1954: "The Role of Context in the Perception of Stops", *Language* 30, p. 47-56.

Wajskop, M. 1971: "Identification des occlusives intervocaliques en Français", mimeographed.

Wang, W.S-Y. and Ch.J. Fillmore 1961: "Intrinsic Cues and Consonant Perception", *JSHR* 4, p. 130-136.

IDENTIFICATION OF INITIAL STOP CONSONANTS IN SYLLABLES PLAYED AT DOUBLE AND HALF SPEED

Eli Fischer-Jørgensen

1. Stimuli and listeners

The stimuli used in the present investigation consisted in 18 CV-syllables with initial ptkbdg + the vowels i, a or u, spoken by me on tape. They were played in random order, once at double speed and once at half speed, to four groups of students of philology (two in 1954 and two in 1955) over a loudspeaker in a normal classroom. The listeners were asked to write down what they heard. 52 students listened to the syllables presented at double speed and 54 to those presented at half speed. The answers in 1954 and 1955 differed at two points. The listeners in 1954 made more mistakes than the listeners in 1955 for the vowel i at double speed (it was often heard as u),but less mistakes for p at double speed (it was less often heard as t). Probably the loudspeaker used in 1954 had stronger attenuation at high frequencies. But on the whole, the answers were the same in the two cases. There were 312 responses to each vowel and 156 responses to each consonant at double speed, and 324 and 162 respectively at half speed.

2. Identification of syllables played at double speed

2.1. Identification of the vowels

When the syllables were played at double speed the vowels were on the whole correctly identified (except that the listeners in 1954 heard i as u in 32 % of the cases, probably because formants 3 and 4 were raised too much). u has most of the energy concentrated in F1 at 250 Hz, but the doubling to

500 Hz did not change the identification. There were no o-answers and only a few y-answers. It is at first sight surprising that a was heard correctly when its first formant of about 900 Hz and its second formant of 1800 Hz were shifted to 1800 and 3600, respectively. The most probable explanation is that under these conditions the former F_1 functions as F_2, and the former F_2 as F_3, whereas a former subformant at about 300 takes over the function of F_1.

2.2. Identification of the consonants

The consonant identifications at double speed are given in table I and in Fig. 1 A. In the latter case ptk and bdg have been combined under the designation PTK.[1]

TABLE I

Identification (in %) of ptk and bdg played at double speed (N = 52). Answers horizontally.

	i				a				u			
	p	t	k	O	p	t	k	O	p	t	k	O
p	24	63	14	-	36	31	31	3	20	73	2	5
t	2	93	5	-	2	88	7	3	-	100	-	-
k	10	61	27	2	2	49	46	3	2	61	32	5
	b	d	g	O	b	d	g	O	b	d	g	O
b	30	56	2	12	46	46	-	8	22	68	-	10
d	27	64	2	8	29	68	3	-	-	95	5	-
g	2	85	9	5	19	58	22	2	-	58	36	7

1) Thus capital P is here used to cover p and b, not, as in the paper on tape cutting experiments, ptk.

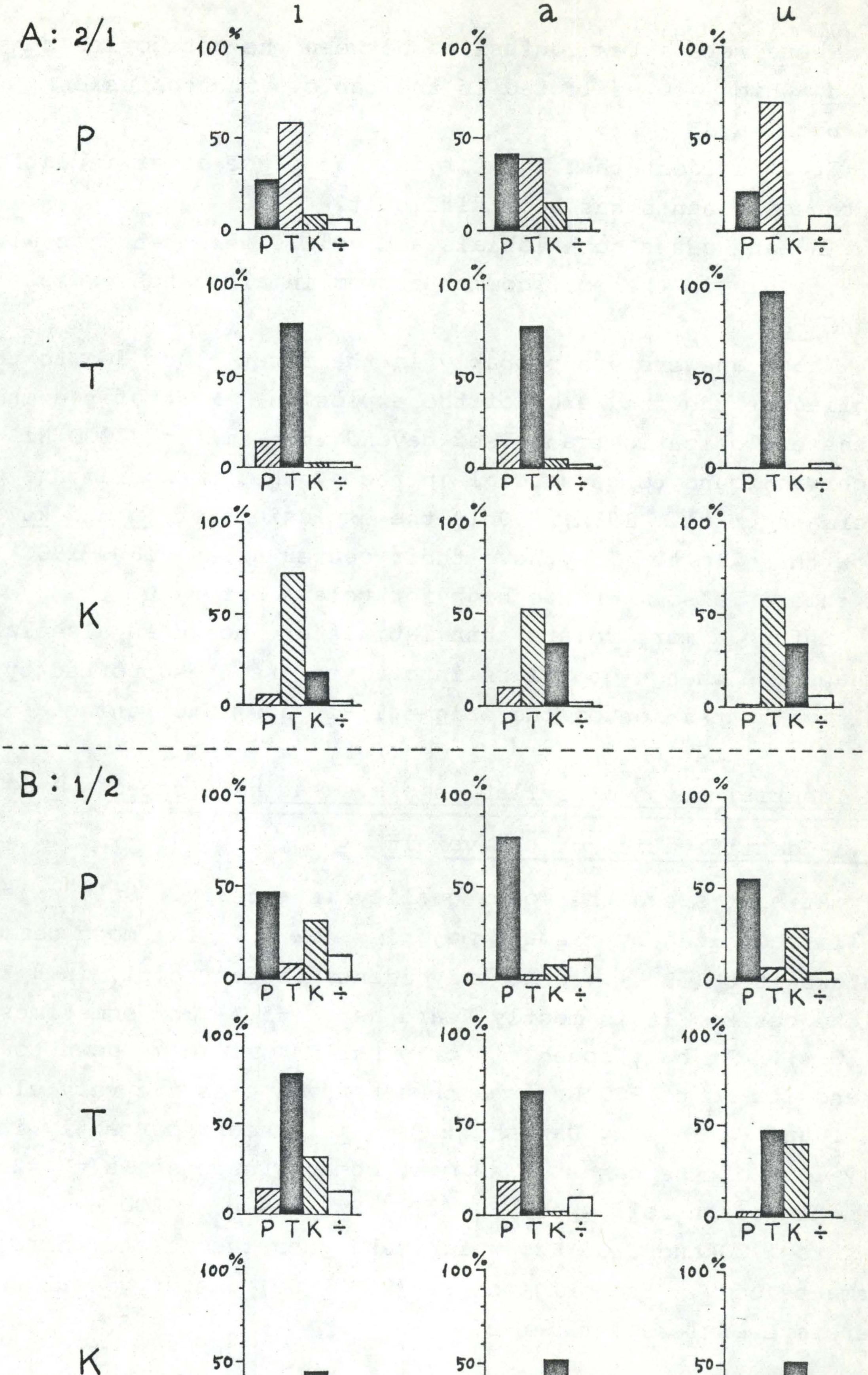


Fig. 1. Identification of labial, alveolar and velar stops (symbolized by P, T and K), played at double speed (A) and half speed (B).

There were a few confusions between the categories ptk and bdg which are neglected in the table. The confusions went both ways.

It is evident that identifying the place of articulation for the consonants has been difficult.

In many cases both labials and velars were heard as alveolars. This is the most common answer in all cases except for pa and ba.

These answers are probably in the first place due to the doubling of the frequency of the explosion. Most of the energy of the explosion is transposed beyond the limit of 3000 Hz which was found to be crucial in the synthetic experiments of Libermann et al. (1952). Only the explosions of gu and ku are below this limit. They have their center around 1800-1900 Hz, but this is also much too high for velars before u.

Before a more velars than labials are heard as alveolars. Perhaps the change of velars into alveolars is supported by the positive transition of the original F_2, now functioning as F_3.

3. Identification of syllables played at half speed

3.1. Identification of the vowels

At half speed the vowel quality is altered. Only u is still recognized by a small majority (52 %). The most common mistake is o (19 %). a is only identified correctly in 8 % of the cases. It is mostly heard as ɔ (65 %) and sometimes as o (10 %). At half speed a's original F_1 will come down to 450 Hz and its F_2 to 900 Hz, and these frequencies are very close to F_1 and F_2 in long Danish ɔ:. - i is heard correctly in only 10 % of the cases. The most common response is y (62 %). When played at half speed F_3 (3400 Hz) and F_4 (4200 Hz) come down to 1700 and 2100 Hz, which make good second and third formants of y. The original F_2 (2200 Hz) was very weak and therefore of less influence.

3.2. Identification of the consonants

The consonant identifications at half speed are given in table II and in Fig. 1 B. In the latter case ptk and bdg have been combined under the designation PTK.

TABLE II

Identification (in %) of ptk and bdg played at half speed (N = 54). Answers horizontally

	i (y)				a (ɔ)				u			
	p	t	k	O	p	t	k	O	p	t	k	O
p	31	13	41	15	69	2	17	13	56	5	26	13
t	2	98	-	-	-	93	-	7	-	59	33	7
k	24	11	56	9	24	11	56	9	20	6	56	18
	b	d	g	O	b	d	g	O	b	d	g	O
b	63	4	22	11	87	4	-	9	54	7	30	9
d	26	57	6	11	39	44	4	13	6	37	46	11
g	31	26	35	7	-	31	48	20	7	20	48	24

At half speed there are some confusions between the categories ptk and bdg,which have been neglected in the table. There are very few bdg-responses to ptk-stimuli, but the opposite case is somewhat more common. The percentages are, however, negligible except for di 11 %, gu 26 %, and gi 44 %. This can be explained by the doubling of the open interval, which for di, gu, gi becomes 40, 50 and 70 ms respectively. The others are shorter.

The place of articulation is identified somewhat better at half speed than at double speed. The highest percentages of answers correspond to the correct consonants except for pi and du, but the highest percentages are often below 50. The general uncertainty also shows up in the higher number of zero-answers as compared to double speed.

In almost 100 % of the cases t in ti and ta are identified correctly because of the high frication noise which in ti reaches above 3000 Hz even at half speed, and in ta at least above 2000 Hz. tu still has some noise concentrated about 3000 Hz, but the lower concentration around 1400 Hz reminds of k before u (1200 Hz).

The characteristic transition in the aspiration of pi (py) gets reduced at half speed, so that it looks very much like ky, in which the aspiration often starts below F_2 of the vowel. Before the other vowels p and k are heard correctly by a small majority. The most common mistake is mutual confusion between p and k.

d is identified correctly less often than t, because it lacks the high fricative noise. It is often heard as b before i (y) and a (ɔ), (b is, on the whole, a common mistake, and the transitions of d look more like those of b than like those of g), but before u the majority has heard g, probably because the frequency of the explosion comes down in the neighbourhood of the g-explosion at half speed (cf. that t is often heard as k before u).

b is relatively correctly identified. Before i (y) and u it is heard as g in a number of cases. This never happens before a (ɔ), nor is g before a (ɔ) heard as b, probably because the transitions are too different.

4. Summary

The identification scores show a greater number of mistakes for the consonants at double speed than at half speed. This can perhaps be explained by the fact that the change in absolute frequency is larger in the former case (a noise of 1000 Hz goes up to 2000 Hz at double speed, but only down to 500 Hz at half speed). For the vowels, however, doubling gives less distortion. But this is because a happens to get formants at the right frequencies, and u and i do not get into the areas of any other vowels. On the contrary, when the speed is halved, i and a get formants corresponding to those of other Danish vowels (y and ɔ).

The perceptual results of doubling and halving the speed for a and i are in good agreement with the results of Chiba and Kajiyama (1941), who, however, consider the preserved relative distances between formants to be a sufficient explanation of the preservation of a at double speed.

The most common mistakes for the consonants are an increase of d- and t-responses at double speed, and confusion between labials and velars at half speed. This indicates that the explosions (and aspirations) play a great role. This is particularly true of ptk, where the transitions are of minor importance. In bdg the transitions are also of some influence, particularly in combinations with a.

The general conclusion of this small experiment, as also of the extensive experiment with tape cutting, is thus that explosions should not be neglected. They play a fairly important role for perception.

References

Chiba, T. and M. Kajiyama 1941: The Vowel.

Liberman, A.M., P.C. Delattre and F.S. Cooper 1952: "The Role of Selected Stimulus-Variables in the Perception of the Unvoiced Stop Consonants", The Am. Journal of Psych. LXV, p. 497-516.

ON UNIVERSALS IN VOWEL PERCEPTION

Birgit Hutters and Peter Holtse

1. Introduction

It has been shown in a number of studies (e.g. Libermann et al. 1957) that consonants are normally perceived in a categorial way. Thus listeners are generally unable to discriminate much better than they can identify the sounds. The picture is less clear among the vowels. Stevens (1968) reports a tendency towards categorial perception of vowels in words. One recent experiment using isolated vowels (Fujisaki 1971) has shown some correspondence between vowel phoneme boundaries and the ability of listeners to discriminate between vowels. But most perceptual studies on isolated vowels have found a tendency towards continuous perception similar to the way non-speech sounds are perceived. For isolated vowels listeners will discriminate much finer differences in quality than they can identify qualities as phonemes.[1]

However, in an interesting study by Stevens, Libermann, Studdert-Kennedy, and Öhman (1969) it is suggested that the perception of vowels is not altogether continuous, but shows peaks and valleys in the discrimination function of a shape similar to the discontinuous discrimination of consonants - although the overall scores are higher than is normally the case with consonants. The peaks and valleys in the discrimination are said to be independent of the linguistic experi-

[1] Nobody seems to have investigated how small differences in vowel quality listeners are in fact able to identify. But they are probably considerably smaller than differences between phonemes.

ence of the listeners. And the theory is advanced that certain areas in the vowel continuum are better suited to contain phonemes since the auditory mechanism is less critical about changes in these areas. We should like to offer some comments on the methods used in this study and the conclusions drawn from them.

2. Comments on the article by Stevens et al. (1969)

2.1. Method of posing the problem to the subjects

The experiment reported by Stevens et al. (1969) contained two series of synthetic vowels, one front unrounded (approximately [i-e-ε]) and one narrow, unrounded to rounded (approximately [i-y-u]). A group of Swedish and American listeners were asked to identify the unrounded series with their own front vowel phonemes. Then the Swedes were asked to identify the rounded series with their phonemes /i/, /y/ or /u/ while the Americans were first played a record of the Swedish vowels and then asked to identify the test vowels with the Swedish phonemes. (The fact that they were using numbers instead of phonetic symbols alters nothing in the basic problem.)

This seems rather an unfortunate way of posing the problem. If the Americans did in fact identify the rounded series with the Swedish phonemes they had heard, they must have acquired a Swedish linguistic background, at least as far as the vowel qualities [i-y-u] were concerned. This would mean that the two groups were no longer representatives of different backgrounds, and the object of the experiment: comparing the influence from different linguistic backgrounds, would have been lost. The proper procedure must have been to ask the American listeners to identify the rounded series with their own narrow vowel phonemes - as far as this could be done with the stimuli at hand.

2.2. The discrimination test

2.2.1. Vowel stimuli

The stimuli of the tests were 25 vowels synthesized with approximately equal logarithmic steps. According to the authors "... there were small deviations from uniform spacing. These deviations arose because the formant frequencies for each stimulus could only be set to within a few cps." (p. 4). The magnitudes of these deviations are best judged when the differences in formant frequencies from one stimulus to the next are expressed in per cent. If the differences are equal logarithmic steps they can be expressed as a constant percentage. How far this is the case with the vowel stimuli under consideration may be judged from table I which gives the percentual differences between the formant frequencies of each vowel. As will be seen the deviations are random, but not inconsiderable.

TABLE I

Percentual distances in formant frequencies between vowel stimuli.[2]

Based on Stevens et al. (1969), Table 1, p. 3.

Vowel number	F1	F2	F3	Vowel number	F2	F3
1- 2	5.6	1.7	2.0	1-R 2	3.1	3.3
2- 3	4.6	1.6	2.0	R 2-R 3	3.1	3.3
3- 4	5.5	2.1	2.3	R 3-R 4	2.8	4.1
4- 5	6.7	1.7	2.1	R 4-R 5	2.7	3.6
5- 6	5.2	2.0	2.1	R 5-R 6	3.0	4.4
6- 7	5.8	1.6	2.0	R 6-R 7	2.8	3.4
7- 8	6.0	1.8	1.4	R 7-R 8	3.4	3.1
8- 9	5.8	1.6	1.8	R 8-R 9	3.1	2.4
9-10	6.0	1.7	1.0	R 9-R10	3.1	2.4
10-11	6.3	2.0	0.5	R10-R11	2.8	1.9
11-12	6.1	1.5	1.0	R11-R12	2.9	1.7
12-13	5.8	2.1	1.0	R12-R13	2.8	1.1

2.2.2. Results of the discrimination tests

For both the unrounded and rounded vowels Stevens et al. report that the valleys of the discrimination curves correspond roughly to the centres of the phoneme areas as previously established from identification tests, but the correspondence is not particularly good. However, there is surprisingly high agreement between Swedish and American lis-

[2] F1-values of the rounded series have not been included in the table since none of the differences exceed 3 Hz.

teners as to where the tops and valleys should be. This is taken as an indication that tops and valleys in the discrimination function are inherent in the perceptual mechanism and not conditioned by linguistic experience. On the contrary, perceptual constraints would favour the placing of vowel phonemes in areas where discrimination is relatively poor.

However, some odd features led us to examine the data more closely. For instance, why did both Swedes and Americans show a pronounced discrimination top between stimuli 5 and 6 in the rounded series? This top is not mentioned very clearly in the article although it is found practically at the top of the Swedish identification function for /y/.

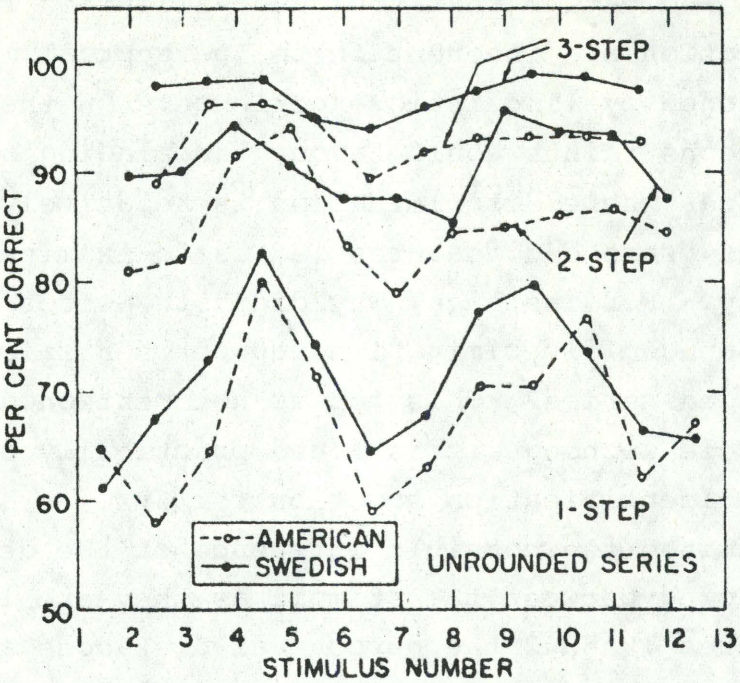
We examined the possible influence of the deviations from uniform spacing between the stimuli as they are listed in table I. In Figures 1 and 2 the percentual distances are compared with the discrimination functions of Stevens et al. (their figs. 6 and 7, p. 10 and 11).

Even to a cursory glance the correspondence between the upper and lower halves of Figures 1 and 2 is quite striking. The top in the discrimination of unrounded vowels between stimuli 4 and 5 corresponds exactly with the large percentual difference between the F1 frequencies of stimuli 4 and 5. And among the rounded vowels the correspondence is even better. The discrimination curves show tops in three places: 3-4, 5-6, and 7-8. The first two tops coincide with large percentual differences in F3, while the third and very small top coincides with a relatively large difference in F2 between stimuli 7 and 8. The relatively poor discrimination between stimuli 8 through 13 may be due to the rather small distances in F3 between these stimuli.

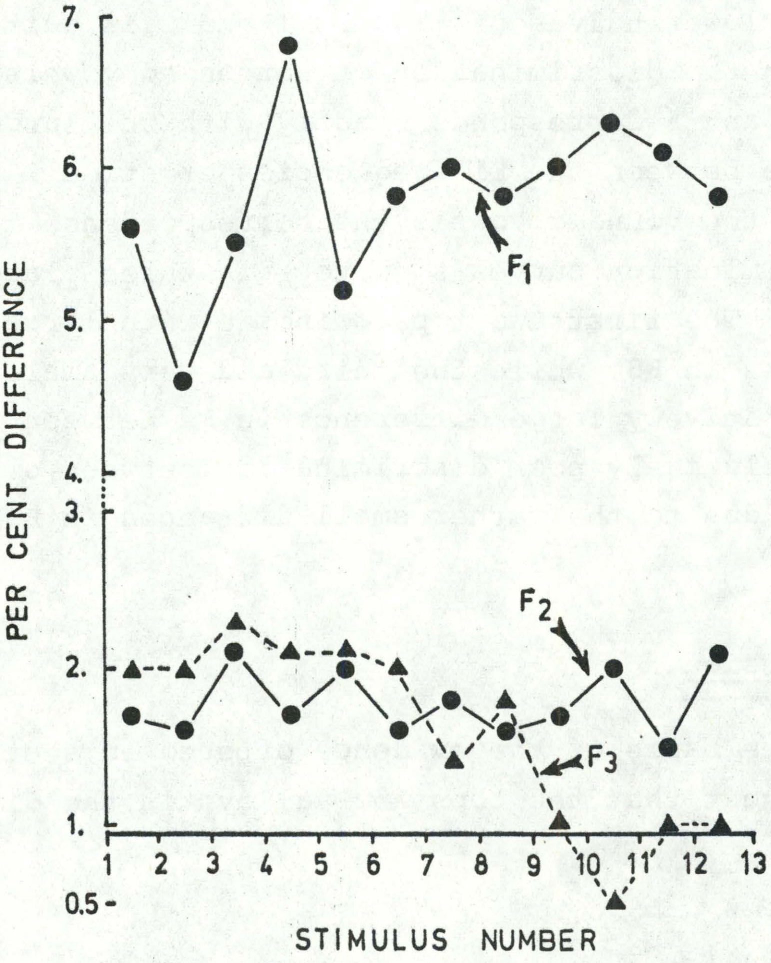
3. Conclusion

On the basis of the evidence offered in section 2.2. we would suggest that the tops and valleys in the discrimination

Fig.1 UNROUNDED VOWELS

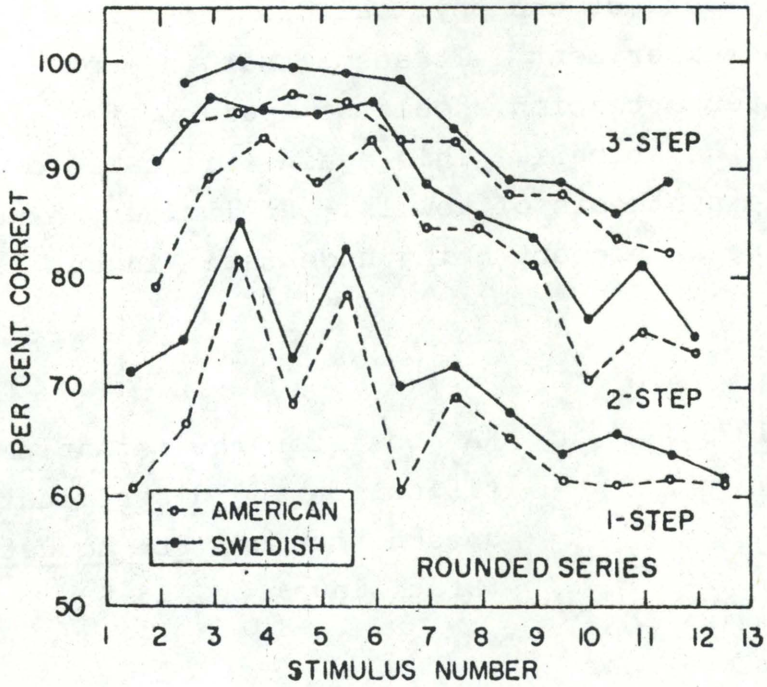


Photocopy of Fig. 6' from Stevens *et al.* (1969)

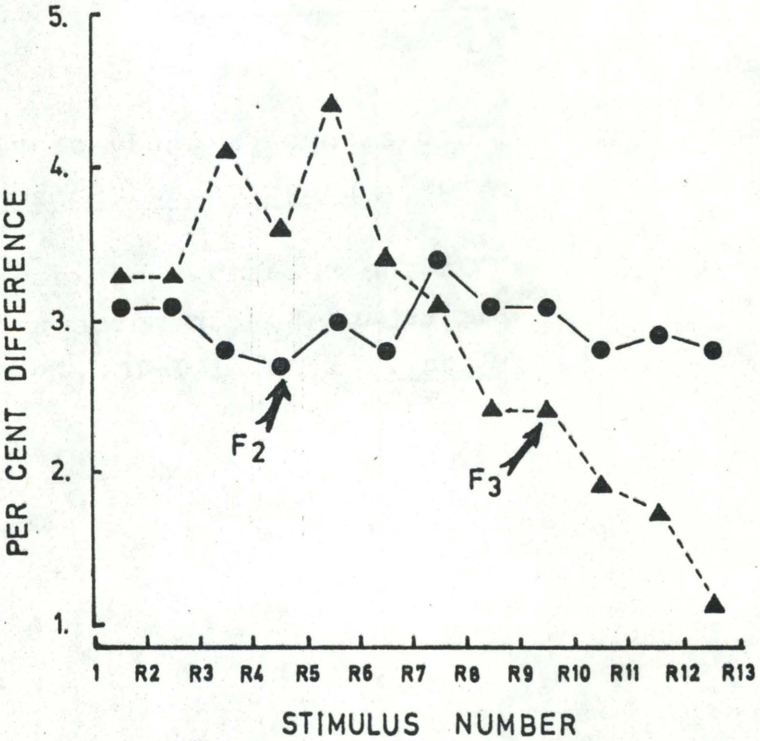


Differences between the vowel stimuli of Stevens *et al.* (1969) expressed in per cent

Fig. 2 ROUNDED VOWELS



Photocopy of Fig. 7 from Stevens *et al.* (1969)



Differences between the vowel stimuli of Stevens *et al.* (1969) expressed in per cent of vowels as reported by Stevens et al. are not due to any inherent universals in the perceptual mechanism. To some extent they may simply reflect the physical distances between the stimuli used in the experiment. It seems that in experiments of this kind greater attention should be paid to the exact distances between the stimuli. This again will call for greater accuracy in the synthesis of vowels. Preliminary experiments of this kind are at present being undertaken in our laboratory.

References

Fujisaki, H. 1971: "A model for the mechanisms for identification and discrimination of speech sounds", 9th Acoustics Conf. (Bratislava), p. 56-59.

Liberman, A.M., K.S. Harris, A.S. Hoffman, and B.C. Griffith 1957: "The discrimination of speech sounds within and across phoneme boundaries", J. Exp. Psychol., 54, p. 358-368.

Stevens, K.N., A.M. Liberman, M. Studdert-Kennedy, and S.E.G. Öhman 1969: "Crosslanguage study of vowel perception", Language and Speech, 12, p. 1-23.

Stevens, K.N. 1968: "On the relations between speech movements and speech perception", Zs. f. Phon., 21, p. 102-106.

SR:
lyd og lydsystem

SOME CONDITIONING PHONOLOGICAL FACTORS FOR THE PRONUNCIATION OF SHORT VOWELS IN DANISH WITH SPECIAL REFERENCE TO SYLLABIFICATION[1]

Hans Basbøll

1. Introduction

The aim of the present (very preliminary) paper is to shed light on what some linguists would earlier have called rules for allophonic variation of short vowel phonemes. Thus the distinctive feature analysis of the Danish vowels is not at issue here (see Rischel 1969 and Austin 1971 for general discussions of that subject), and questions about which vowels are found underlyingly are only briefly touched upon.

The theoretical framework of the present paper is basically that of a generative grammar. However, very little formalism will be used here since current notational conventions are on many points quite inadequate for expressing the generalizations discussed in this article.

The word "phonological" in the title of this paper indicates that differences in pronunciation due to different sociological, geographical or other background, as well as stylistic factors and purely phonetic factors (e.g. physiologically conditioned), are not taken into account.

The language under consideration is a variety of standard Danish. Unless otherwise stated, I think the features of pronunciation that are discussed are common to most non-provincial varieties of standard Danish. As long as no reliable material

[1] I am very indebted to Eli Fischer-Jørgensen and Jørgen Rischel for valuable comments on the manuscript, and to Peter Holtse for many improvements on my English style.

on these problems has been published, it is of course impossible to decide on these matters. But I think that the very existence of the pronunciation patterns discussed here is interesting, even if some other varieties of standard Danish have deviating pronunciations on some points.

Except for section 3. (where the long vowels are used for comparison), the only vowels to be considered here are those which are genuinely short, i.e. phonetically short vowels which are not derived from long vowels. Thus we shall neither discuss the quality of phonetically long vowels derived from underlying short vowels (as in bade [ˈbæːðə], plural of bad [bað]), which have the same quality as the genuinely long vowels, nor the quality of phonetically short vowels derived from underlying long vowels (as in øvrig [ˈøu̯ʁi], cf. øverst [ˈø·ʔvɒsd]). As a consequence of this latter limitation, one interesting problem in Danish phonology concerning vowel quality cannot be adequately dealt with here, viz. that of the two qualities [ɔ] and [ʌ] of shortened /ɔː/ (e.g. på gaden, på den [p^{h}ɔ ˈgæːðən, ˈp^{h}ɔ dən] versus påsmøre [ˈp^{h}ʌˌsmœ·ʔɒ], cf. the isolated word på [ˈp^{h}ɔ·ʔ]).[2]

[2] Although the problem has received some attention in the literature (e.g. in Rischel 1969 p. 190f and 202ff), it seems to me that the processes have not been stated in a very satisfactory way. It may therefore be worth while mentioning that short and long /ɔ/ should have the same opening degree in their underlying form, and that the distinction mentioned above is due to the fact that the vowel shortening in examples like påsmøre - a shortening which evidently belongs to derivational morphology - applies before the rule that opens short /ɔ/ to [ʌ], which again applies before the rule that shortens /ɔː/ in certain syntactically conditioned environments, applying across word boundaries. This is precisely the rule ordering we should expect.

We shall presuppose that the input to the phonological rules contains the following nine short vowels (which will be written in //):

/ i y u
e ø o
ε ɔ
a /

They are normally said to have the following "main allophones":

[i y u
e ø ɔ
ε ʌ
a]

We shall then consider the manifestations [o] and [ɔ] of /o/ (section 2.1.), and [a] and [α] of /a/ (section 2.2.). We shall find that the syllable is a crucial unit in this connection, which motivates a discussion of some principles for syllabification in Danish (section 2.3.). Finally we shall discuss "r-colouring" (section 3.) and some aspects of the phonology of the short rounded front vowels (section 4.).

2. The role of syllabification

The use of the concepts "syllable" and "syllabification" in this connection will be clarified in section 2.3. below.

2.1. Short /o/

Apart from the position before /r/ (see section 3. below), native Danish words contain the following long rounded back vowels [u:, o:,ɔ:] (e.g. in *hule*, *hole*, *åle*) and the following short ones [u, ɔ, ʌ](e.g. in *hulde*, *hulle*, *holde*). These have normally been taken to manifest the phonemes long and short /u/, long and short /o/, and long and short /ɔ/, respectively, for reasons of pattern congruity. It is also well known that the vowel [o] occurs, partly as shortened /o:/, partly in some foreign words which have been treated as deviating from the native pattern (e.g. [ˈfot^{s}o] *foto*, cf. Rischel 1969 p. 180).

It has been pointed out (Basbøll 1969 p. 44) that the short vowels [o] and [ɔ] both occur posttonally in complementary distribution, [o] occurring in open and [ɔ] in closed syllables. At the time I did not, however, fully realize the generality of this principle.

Consider the following words where the underlined letters represent the short vowel phoneme /o/:

[o]: f*ó*t*o*, céll*o*, Vígg*o*, *o*naní, K*o*sángas
[ɔ]: b*ó*nde, céntr*u*m, h*ú*l(le), l*ú*ffe, *u*ndgåelig, m*u*kkerí

We shall give evidence below (section 2.3.) that a single intervocalic consonant belongs to the syllable containing the preceding vowel if the following vowel is shwa, and to the following syllable if its vowel is a "full vowel" (not including weakly stressed posttonal *i* and *e* occurring in endings like *ig*, *isk*, *ing*). Under this supposition the rule seems to be: */o/ is pronounced [o] in open and [ɔ] in closed syllables*, applying to /o/ in both pretonal, tonal, and posttonal position. Notice especially that the [o]-manifestations of short /o/ in the often cited "exception" *foto* are quite regular under this analysis: the syllable boundary occurs after the stressed /o/ (which is thus [o]) and before /t/ (which is thus aspirated and affricated). The fact that there *is* a posttonal /o/ indicates a foreign word structure.

I know of no evidence disconfirming the present hypothesis, but its value of course cannot be determined without regard to the principles of syllabification discussed below (in section 2.3.).

2.2. Short /a/[3]

It is well known that /a/ has several variants: A back vowel in the environment of /r/ (see further section 3. below), a mid vowel [α] occurring before velars, and a front vowel [a] occurring before dentals and in word final position. Before labials some conservative standards have a vowel between [a] and [α], whereas in the variety of standard Danish spoken around the capital the /a/-manifestation before labials is [α]. I shall base the presentation on data taken from this latter norm in the following, but all the arguments to be given below apply *mutatis mutandis* to the other variety mentioned.[4]

The formulation *[a/ is pronounced [α] before non-coronal consonants, otherwise as [a]* accounts for e.g. the following words:

[a]: da, land(e), hat(te), sófa, Aída
[α]: dam(me), lang(e), lak(ke), tap(pe), abstinént

However, in some cases where the above-mentioned rule predicts [α], we actually have [a]:

[a]: Amérika, akadémiker, habilitét, kakofoní, áhòrn

[3] I am very much obliged to Henrik Holmberg for his kind permission to use some material concerning the /a/-variants which he collected in an independent and skillful manner during my course in Danish phonology in the spring of 1969. Henrik Holmberg then advanced the idea that the syllable played a crucial role for the determination of the /a/-variants too (cf. section 2.1. above).

[4] From the generative point of view used in the present paper it is immaterial whether [α] is identical to the "r-coloured a", as is the case in advanced standard Copenhagen.

Compare the following words:

[ɑ]: amfíteater, akkeléje, ábsalon, akcént

It seems reasonable to advance the hypothesis that the above-mentioned rule is correct but applies at the syllable level and not at the word level, or otherwise stated: that the rule has the syllable as its "domain". Note that syllabification follows the principles that are needed anyway for the prediction of the /o/-variants (section 2.1. above), see section 2.3. below.

Notice further that cases of vacillation like:

[a/ɑ]: affrikat, amerikaner,

especially the latter one, support the hypothesis: the form *affrikat* can be syllabified either a.fri... (corresponding to a pronunciation with [a]) or af.ri... (corresponding to a pronunciation with [ɑ]) (see below). The form *amerikaner* is pronounced with [ɑ] if the second vowel is dropped, as it normally is ([ɑmʁiˈkʰæ·ʔnɒ]), otherwise with [a] ([ameʁiˈkʰæ·ʔnɒ] like [aˈme·ʔʁikʰa] *Amerika*; cf. footnote 9).

Compare the words *sandkrabbe* [ˈsanˌkʰʁ̥ɑbə, ˈsaŋˌkʰʁ̥ɑbə] and (the invented) *sangkrabbe* [ˈsɑŋˌkʰʁ̥ɑbə] whose first /a/'s would always be pronounced differently. The pronunciation [ˈsaŋˌkʰʁ̥ɑbə] (of *sandkrabbe*) shows that the /a/-variant used is independent of the operation of the optional nasal assimilation rule applying across a #, or in other words: that the quality of /a/ is determined *before* the mentioned assimilation rule applies. On the other hand, the determination of the /a/-variant presupposes that the nasal before a homosyllabic /g/ has already been specified as velar.

2.3. Principles for syllabification

2.3.1. Introductory remarks

Let me illustrate what I mean by "syllable" and "syllabification" with a German example where the facts are well known.

[s] and [z] only contrast between vowels, [s] being excluded word initially and [z] word finally (details left aside). Hjelmslev[5] proposed to account for this fact by reducing [s] and [z] to one phoneme /s/ and describing the distinction reissen : reisen by means of different syllabification, reissen having the syllable boundary after /s/ and reisen before /s/. This is, however, in itself an empirically rather empty proposal, unless it is seen in connection with other facts of German phonology, as I shall do in the following.

It seems possible to claim that the following generalization holds in German: A stressed syllable must be a possible phonological word, i.e. it is always possible to syllabify a native German word in such a way that a stressed syllable does not violate any phonotactic rule for monosyllabic words. This principle explains why a word having a stressed short vowel followed by a voiced obstruent followed by shwa belongs to a non standard (High) German word type (e.g. Low German words like Ebbe, Kladde, Roggen): The syllable boundary cannot occur before the obstruent (since a stressed monosyllable can never end in a short vowel), nor can it occur after the obstruent (since a German word can never end in a voiced obstruent).

Now this principle (that a stressed syllable must be a possible phonological word), together with Hjelmslev's analysis of [s] and [z] as bound variants, explains why [z] never occurs after short vowels: An impossible phonological word like [hazən] could have the syllable boundary neither before [z] (since a stressed monosyllable can never end in a short vowel),

[5] "Es könnte in derselben Weise [as with [x] and [ç]] nachgewiesen werden, dass das stimmhafte und das stimmlose s (Lenis- und Fortis-s) im Deutschen silbenbedingte Varianten ein und derselben Ausdruckseinheit sind" (1938 p. 156f).

nor after [z] (since the pronunciation of /s/ is [s] and not [z] in syllable final position).

To an imaginary objection that the above-mentioned facts can more simply be stated like this "voiced obstruents do not occur after short stressed vowels", I should make the following points:

(i) The principle "all obstruents are voiceless in the final (i.e. postvocalic) part of the syllable"[6] accounts for the otherwise quite disparate facts that in native words voiced obstruents are found neither word-finally nor after short vowels.

(ii) The use of syllable boundaries to account for the distribution of [s] and [z] correctly predicts the very restricted number of possible contrasts between them. Furthermore, this principle explains why only [z] occurs between a sonorant consonant and shwa (e.g. *Amsel*, *Hülse*), since the syllable boundary goes between the two intervocalic consonants (as can be seen from contrasts like *halben*, *Alpen*, obstruents occurring between a sonorant and shwa are in syllable initial position).

[6] Eli Fischer-Jørgensen has called my attention to forms like *Redner*, *Wagner* (pronounced with a voiced stop), derived from *reden*, *Wagen*, which could be syllabified /rē.də.nər, vā.gə.nər/ with obligatory loss of the first shwa after the rule for devoicing of syllable final obstruents has failed to apply (since its structural description is not met), cf. the fact that words like *Regen* [ʁe:gŋ̍] must be /rē.gən/ where the syllabification presupposes that the phonological form contains shwa. (A Danish parallel to such words is mentioned in section 2.3.2.1. below.) It should be added that words like *Adler*, *leugnen* should according to the present analysis contain a shwa between the phonetically intervocalic consonants when their syllabic structure is determined (cf. Twaddell 1938 p. 223). - Foreign words and names have a deviating phonological structure in this respect as in many others.

The points I wish to illustrate by this digression on German are the following:

(a) The use of syllable boundaries may connect ("explain" in a vague sense) many facts which are apparently quite disparate.

(b) In some cases (e.g. *Grüsse* / grȳs.ə/) the postulated syllable boundary may not coincide with the intuitively felt syllable boundary or with some experimentally established syllable boundary (or better: experimental data may seem to contradict the proposed syllable boundary). This may indicate that the syllable we are dealing with is a more abstract entity than the phonetic syllable, viz. a "phonological syllable". Nevertheless I dare use the term "syllable" since it is an entity which has, in Danish at least, exactly one phonological vowel and whose boundaries can be posited in accordance with some generally recognized principles for syllabification (e.g. principles (A) and (B) below). I should furthermore like to suggest the hypothesis that in the cases where the boundaries of the phonological and phonetic syllable do not coincide, the phonetic syllable boundaries will always be universally less marked than the phonological ones (e.g. if a sequence ... VCV ... has different phonological and phonetic syllable boundaries, the phonetic one will always be before C). This is the case in Danish words like [bæːðə] *bade* where the phonological syllable boundary occurs after [ð], whereas the phonetic boundary (if such a boundary is recognized at all) is before the consonant. In short: a sound chain may be "syllabified" either in the universally unmarked way (in "phonetic syllables", if you like), or in "(phonological) syllables", or (according to principle (C) below) in a way which is sensitive to grammatical boundaries. The three ways of syllabifying a sound chain may of course interact.

2.3.2. Some principles for syllabification in Danish

It has already been made clear that the unit which we try to establish, i.e. the "(phonological) syllable", is one that functions as the domain for several phonological rules, and which furthermore has certain characteristics that are always ascribed to the syllable under any definition. For instance it never contains two or more vowels that are separated by one or more consonants (in Danish there is the stronger requirement that the syllable contains exactly one phonological vowel, and also that in Danish the syllable always follows principles (A) and (B) below).

The following four principles for syllabification in Danish form a sort of hierarchy.

(A) Word boundaries coincide with syllable boundaries.

(B) Syllables always begin with a "full vowel" or with a possible word-initial consonant or consonant cluster, and they always end in a vowel or in a possible word-final consonant or consonant cluster.

(C) If there is an intuitively transparent morpheme boundary between two "full vowels" in the same word, the syllable boundary coincides with this morpheme boundary in so far as it does not thereby violate principle (B).

(D) One intervocalic consonant belongs to the syllable of the preceding vowel if the following vowel is /ə/, and to the syllable of the following vowel if this is a "full vowel" for which a derivation from shwa cannot be postulated (i.e. which is neither /ə/ nor weakly stressed *i* (in the endings *ig*, *isk*, *ik*) nor weakly stressed *e* (in *ing*), see further section 2.3.2.2. below).

I need not emphasize the highly tentative character of these proposals (of which the first ones are of course very well known and the third one extremely vague). It is also clear that further principles are needed for the situation with more than one intervocalic consonant (in some of these cases vacillation may occur, cf. section 2.2. above), cf. the following section.

2.3.2.1. Tentative justification for the proposed principles

The phonological rules which I think can most naturally be formulated with the "phonological syllable" as their domain and which thus constitute evidence for syllabification, may be stated in the following vague form.[7] I do not claim that these are the only such rules.

(1) /o/ is pronounced [o] in open syllables and [ɔ] in closed ones.

(2) Except in the environment of /r/, /a/ is pronounced [α] before a non-coronal consonant belonging to the same syllable, otherwise it is [a].

(3) /g/ is dropped after a nasal belonging to the same syllable, otherwise it is pronounced [γ] in the final part of the syllable and [g] in the initial part.

(4) /d/ is dropped after a sonorant belonging to the same syllable, otherwise it is pronounced [ð] in the final part of the syllable and [d] in the initial part.[8]

[7] As already mentioned I shall not give explicit rule formulations, but in such formulations the syllable should not be mentioned in the environment of the rule (i.e. it should be a property of the rule itself that its domain is the syllable). I certainly do not make the claim that none of these "rules" are instances of the same rule, nor that none of these "rules" contain different rules. Both of these claims would obviously be false.

[8] For some problems in connection with rules (3) and (4) see section 2.3.2.2. below. Furthermore it should be added that /d/ is dropped before a dental stop belonging to the same word, and that /g/ is pronounced [g] in the final part of the syllable before +t if the preceding vowel is shortened. See Rischel 1970a who was the first to state and discuss these problems within a generative framework.

(5) /p, t, k/ are heavily aspirated (and /t/ furthermore affricated) in syllable initial position, but unaspirated in syllable final position. (Furthermore, they may be aspirated in utterance final position.)

(6) /r/ is pronounced [ɒ̯] in the final part of the syllable, [ʁ] in the initial part.[9]

(7) Short /ø/ is lowered before a nasal or a /v/ belonging to the same syllable (see section 4. below for more careful formulations).

(8) /h/ only occurs syllable initially before a vowel (this is in fact not a phonological rule, but a sort of well-formedness condition).

Syllable boundaries are used by Hjelmslev (1951) to account for the different manifestations of /d/ and /g/ (as in rules (3) and (4) above). It is important to realize, however, that our claim is much stronger than Hjelmslev's. Whereas he indicates syllable boundaries everywhere in the underlying representations (his "ideal taxeme notation"), we predict the occurrence of the syllable boundaries by means of general principles, or in other words: the syllable boundaries are inserted by rule.[10]

[9] In this form the rule does not cover very conservative standards where intervocalic word-internal /r/ not occurring before /#/ is most often manifested as a consonant (this manifestation is used for nearly all instances of /r/ except when preceded by shwa in even more conservative standards). For the younger standards it should be added that [ɒ̯] may be substituted for intervocalic [ʁ] before an unstressed vowel. See further footnote 15.

[10] As mentioned in Basbøll 1971 (p. 207f) Hjelmslev misses several generalizations by his way of using syllable boundaries: In all cases where the placement of the syllable boundary has any phonological effect according to him, either the placement of it is predictable, or the phonological effect in question is due to some other independently established factor. (In most cases different placement of the syllable boundary has no phonological consequences at all.)

Notice that as the hypothesis stands, a lot of empirical data could disconfirm the proposed principles for syllabification. Some items of justification will now be given together with types of evidence that would disconfirm the principles.

Ad (A) I know of no cases where the mentioned rules apply across word boundaries. Evidence which would disconfirm (A) would be e.g. if da [da] were pronounced [dɑ] in phrases like da manden kom, or if valg [val?ɣ] were pronounced with final [g] in valg ét. Not the slightest tendencies in this direction can be found.

Ad (B) A form like yngle ['øŋlə] has the syllable boundary before l although the morpheme boundary is after l ([ŋl] is an impossible termination of a Danish monosyllable).

Ad (C) This principle accounts for the fact that the syllable boundary coincides with the juncture # separating the two parts of a compound (e.g. dám#and ['dɑmɒn?] versus dámask ['damasg][11]). Principle (C) also applies to some cases of derivation where (D) would give a different result. For instance a word like skuespillerinde is normally pronounced [sguəsbelɒ'enə] but it has an older pronunciation [sguəsbelɒ'ʁenə].[12] The former

[11] According to the principles put forward in Rischel 1970b, damask should have an underlying geminated m: /dammask/ in order to predict stress on the first syllable. This need not conflict with our principles for syllabification, however, since the rule shortening long (or geminated) word-internal consonants can apply before the syllable boundaries are inserted. I have found no cases where syllabification should apply to geminated consonants.

[12] There is nothing strange in the fact that the latter form seems to correspond to a spelling skuespillerrinde (cf. Rischel 1969 p. 197), since the r-colouring effect applies across syllable boundaries (but not across the juncture #), see section 3. below.

presupposes a syllable boundary after /r/ (coinciding with the morpheme boundary), the latter before /r/ (which is the phonetically unmarked place for it to occur). Similarly in a case like *jøde*, *jødinde* ['jø:ðə, jøð'enə], cf. Rischel 1970b, p. 133f ("additive and replacive suffix insertion"), and section 2.3.2.2. below. Note further that a word like *abusus* is correctly pronounced [ɑb'u:sus], but that a person not being able to analyze it into *ab+usus* will pronounce it [a'bu:sus], as everybody would pronounce some African name *Abutu* [a'bu:tu].

Ad (D) This principle is the crucial one for the present paper, and it is supported by a lot of data, e.g. the following (remember that the term "full vowel" denotes all vowels except shwa and weakly stressed posttonal *i* and *e* occurring in certain endings):

(i) Intervocalic /d, g/ is pronounced [d, g] before full vowel, but [ð, ɣ] before shwa: *Ada*, *saga* versus *node*, *bage*.[13] The sounds [ð, ɣ] never occur before full vowels belonging to the same word.

(ii) Intervocalic /p, t, k/ are heavily aspirated before full vowels, but not before shwa: *kvota*, *ekko* ['k^{h}vo:t^{s}a, 'ɛk^{h}o] versus *otte*, *takke* ['ɔ:də, 't^{s}ɑgə].

(iii) /h/ occurs before full vowels (also unstressed ones), but never before shwa: *Uhu* (a trade mark), *Ahasverus* ['u:hu, ahas've:ʁus] versus *Brahe*, *Brahetrolleborg* ['bʁɑ:ə, bʁɑə't^{s}ʁ̥ʌləbɒ·ʔɣ]. (*h* is not dropped, however, when occurring before a full vowel which is phonetically reduced to shwa, e.g. the first *h* in *kom herhén*!)

[13] In certain northern Jutlandic dialects words like *Ida*, *soda* ['i:da, 'so:da] are pronounced ['i:ðə, 'so:ðə], i.e. with final shwa and therefore intervocalic [ð].

(iv) Evidence from the pronunciation of short /o, a/ has already been given in sections 2.1. and 2.2. above.

As already mentioned, the formulation of (D) has not been given in its full generality. The same principle should be extended to account also for the manifestation of the intervocalic consonant clusters /ng, nd, ld/ by stating that the syllable boundary occurs before the stop when the following vowel is a full vowel, but after the stop, which is therefore deleted, when the following vowel is shwa: *Angus*, *vandal*, *Hulda* [ˈ ɑŋgus, vanˈdæ·ʔl, ˈhulda] versus *bange*, *vande*, *hulde* [ˈbɑŋə, ˈvanə, ˈhulə]. Note especially alternations like *diftong*, *diftongere* [difˈt^{s}ʌŋ, dift^{s}ʌŋˈge·ʔɒ] (and *vand*, *vandig* [vanʔ, vandi], see sections 2.3.2.2. and 4. below). Similarly, the medial clusters /lg, rg/ exhibit the same syllabificational pattern, as shown by the different /g/-manifestations: *Volga*, *ergo* [ˈvʌlga, ˈæɒ̯go] versus *bølge*, *værge* [ˈbølɣə, ˈvæɒ̯ɣə].

As a quite informal experiment, I have tried to syllabify the medial clusters found in native Danish infinitives ending in shwa according to the principle that the border should go as much toward the right as is permitted by principle (B). There are three classes of exceptions where the syllable boundaries thus established fail to predict the correct pronunciation: (i) Where the cluster consists of a sonorant consonant followed by *dr*, the syllable border must be *before d* (e.g. *ændre*, *skildre*, *fordre*, all pronounced with medial [d]).[14]

[14] This is maybe no exception at all, since in the cases in question (viz. the clusters [ndʁ, ldʁ, ʁdʁ]), [d] could possibly be inserted by rule. There are, however, some exceptions in the case of *ldr* (but none in the other two): words like *aldre*, *buldre* etc. have no pronounced [d]. (If there is a /d/ in the underlying form, there are thus some instances of medial /ldr/ with syllable border *after* /d/ (e.g. *aldre* 'ages', related to *ældre*); and if *d* is inserted by rule, this rule has some exceptions.)

(ii) Where the cluster /rd/ is preceded by a short (or shortened) vowel, the syllable border must be before d (see the end of section 2.3.2.2. below). (iii) Two words could not be syllabified at all without violating either principle (B) or the principle of manifestation predictability, viz. the verbs tordne and ordne [ˈtˢoɒ̯dnə, ˈɒːdnə]. It is extremely interesting that both of these verbs are derived from dissyllabic (and quite regular) nouns: torden and orden [ˈtˢoɒ̯dən, ˈɒ.ʔdən], syllabified /tor.dən, ɔrʔ.dən/.[15] The verbs can thus be syllabified /tor.dən.ə, ɔr.dən.ə/ with later loss of their first shwa. This is a striking parallel to the German examples mentioned in footnote 6.

There is no doubt that the reason why the role of syllabification for the determination of e.g. the variants of /o, a/ has not been given full credit in the literature is that native monomorphemic words are generally either monosyllables, or dissyllables with shwa as their second vowel; and in the latter case the first syllable comprises at least one final consonant (if there are any intervocalic consonants, otherwise the distinction between long and short vowel is neutralized in favour of the long one), i.e. the relevant consonantal environment for the first vowel. Occurrences of "unexpected" variants of /o, a/ were then taken as signalling foreign word types. However, in my view the correct way to state the facts is to say that unstressed full vowels in themselves signal foreign word types, whereas all the other facts of pronunciation we have discussed can be deduced directly from the principles of syllabification which are highly sensitive to the distinction between full vowels and shwa.

[15] According to Ordbog over det danske Sprog (ODS), orden is pronounced with a short first vowel and stød on /r/, in contradistinction to words like år which, still according to ODS, is pronounced with a long stød-vowel. Today, however, long as well as short /ɔ/ together with a following /r/ is nearly always pronounced as one long vowel: [ɒ:] or [ɒ.ʔ] (cf. Rischel 1969 p. 194ff). (The /a(:)r/-sequences are pronounced in a similar manner.)

2.3.2.2. Some further problems of syllabification in Danish

It is clear from the preceding discussion of principle (D) that one class of vowels has not been taken into account, viz. those which are neither full vowels nor shwa, i.e. posttonal *i* and *e* in endings like *ig*, *isk*, *ik*, *ing* which can possibly be derived from an underlying shwa with subsequent application of the assimilation rule

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

(These endings are always unstressed, and phonetic shwa is excluded before a velar belonging to the same word, with or without an intervening coronal consonant. The lowering of *i* to *e* before a nasal is regular, see section 4.)

The reason why these endings have been excluded from consideration is that they form a rather complicated picture as regards syllabification, as will be illustrated in this section. Since I do not know how to incorporate the syllabification associated with these endings into an overall description, I shall only briefly state what I think are the main facts.

Consider the following examples:

(i) (a) *Erotik*, *erotisk* [eʁo'it^{s}ig, e'ʁo·ʔt^{s}isg]

(b) *Parodi*, *parodisk* [p^{h}ɑʁo'di·ʔ, p^{h}ɑ'ʁo·ʔdisg]

(c) *Metodik*, *metodisk*, *metode* [met^{s}o'dig, me't^{s}o·ʔðisg,
me't^{s}o:ðə]

(ii) (a) *Oda*, *modig*, *ode* ['o:da, 'mo:ði, 'o:ðə]

(b) *Hulda*, *heldig*, *holde* ['hulda, 'hɛldi, 'hʌlə]

(c) *Gerda*, *færdig*, *færdes* ['gæɐ̯da, 'fæɐ̯di, 'fæɐ̯dəs]

Ad (i) This is evidently a problem of how derivations take place, and the reader is referred to Jørgen Rischel's interesting but brief discussion of examples like these under the heading "additive and replacive suffix insertion" (1970b p. 133f).

Ad (ii) If the distribution of stops and continuants is to be explained by syllabification according to the principles stated earlier, the ending *ig* seems to count as beginning with shwa after one single intervocalic consonant, but as a full vowel-ending after a consonant cluster. This might indicate that syllable boundaries are introduced *before* the rule that raises shwa in cases like *modig* applies, but *after* in cases like *heldig*, *færdig* (with an intervocalic consonant cluster), but this of course does not *explain* anything.

In forms with underlying intervocalic /rd/ before shwa, it looks as if the syllable boundary goes before /d/ if the preceding vowel is short (*færdes*, *hærde*), otherwise after /d/ which is therefore deleted (*på færde*, *jorden*, *Norden*, cf. *jordisk*, *nordisk* with shortened first vowel and pronounced /d/). Compare contrasts like *verden*, *værten* ['væɒ̯den, 'væɒ̯tən], in very conservative standards only, whereas no such contrasts are found where the stressed vowel is phonologically long.

3. r-colouring

Several aspects of the problem of the r-conditioned variants of vowel phonemes have been dealt with elsewhere (e.g. Diderichsen 1957, Rischel 1969, Austin 1971). I shall therefore limit myself to giving some very crude rules accounting for which of the stressed vowels are subjected to "r-colouring".[16]

[16] Austin (1971) gives several complicated rules which seem to me rather unrevealing of the linguistic facts, partly because he uses a distinctive feature system (with "high" and "mid" accounting for vowel height) which in my view obscures the regularity of the processes in question.

The following vowel diagram shows which phonologically distinct vowels are found in the environment of /r/. For reasons which will become clear in a moment, the language used here is the variety of standard Danish spoken by the young Copenhagen generation.

	Long vowels			Short vowels		
After /r/	i:	y:	u:	i	y	u
	ɛ:	œ:	o:	ɛ	œ	ɔ
	æ:[17]	-	ɔ:	æ	-[18]	ʌ
		ɑ:			ɑ	
Before /r/	i:	y:	u:	i	y	y
	e:	ø:	o:	-[20]	-[20]	-[20]
	æ:[19]	ɶ:	ɒ:	æ	ɶ	ɒ
		ɑ:			ɑ	

[17] In words like <u>græde</u>, <u>træde</u> [gʁæ:ðə, t^{s}ʁ̥æ:ðə] where the older generation has [ɛ:].

[18] The vowel [ɶ] occurs before nasals, but there we have no [y], so it is sufficient to posit 2 short rounded front vowel phonemes if partial overlapping is allowed (see section 4. below).

[19] The long /ɛ:/ before /r/ of the conservative standards is regularly lowered to [æ:] in the language of younger people, e.g. <u>bære</u> ['bæ:ɒ] (= <u>bager</u> in this idiolect, conservative ['bæ:ɣɒ]) (Lund and Brink, oral communication).

[20] There is in general no distinction between /i, y, u/ and /e, ø, o/, and the young generation normally uses the narrow manifestations throughout (except for a few words with /o/: <u>sort</u>, <u>hurtig</u>).

The dotted lines in the vowel systems separate those vowels which are r-coloured (in the bottom and, after /r/, left corner of the diagrams) from those which are not.

It is evident from this table that whether a vowel is r-coloured or not depends on whether it comes before or after /r/, but (in the advanced standard Copenhagen dialect) it is independent of vowel length. The rule in this variety of standard Danish can be stated informally in the following way:

$$\begin{bmatrix} \text{V} \\ \text{-high} \end{bmatrix} \ \text{"is r-coloured"}\text{[21]} \quad / \quad \left\{ \begin{matrix} \text{r}\begin{bmatrix} \overline{\text{-back}} \end{bmatrix} \\ \begin{bmatrix} \overline{\text{+low}} \end{bmatrix}\text{r} \end{matrix} \right\}$$

The only important difference between the advanced Copenhagen standard and more conservative norms as to which vowels are subjected to "r-colouring" is that /ɛ:/, and in some standards even /œ:/, is not r-coloured in these latter norms (cf. footnotes 17 and 19). The evolution from conservative to advanced standard in this respect is evidently a kind of rule simplification.

It is interesting to notice that this r-colouring effect applies across syllable boundaries (but not across boundaries marked by the juncture #, including word boundaries). In examples like <u>araber</u> [ɑ'ʁɑ·ʔbɒ] even the first /a/ is r-coloured although the syllable boundary occurs before /r/. (This placement of the syllable boundary is confirmed by the consonantal

[21] Exactly what is implied by a vowel being "r-coloured" is not under investigation here, but roughly speaking it means that the vowel is moved "one degree" in the direction toward the right bottom corner of Jones' vowel diagram. (Note that the /a(:)/ which is input to the rule is not a back vowel.) In more conservative norms the over all degree of r-colouring is smaller than that of the advanced Copenhagen standard. Therefore the phonetic notation used in the vowel diagrams exaggerates the differences between r-coloured and non-r-coloured vowels in the conservative norms. Further, it should be said that in the conservative norms the degree of r-colouring is smaller in the long vowels than in the short vowels (except for /a(:)/).

pronunciation of /r/, but the fact that there are two r-coloured a's but only one /r/ suffices under the present supposition that there are no segment-internal syllable boundaries.) Also compare examples like skuespillerinde [sguəsbelɒˈ(ʁ)enə] and arrest [ɑ'ʁæsd].[22]

4. Short rounded front vowels

As mentioned by Henning Spang-Hanssen (1949 p. 66) there is no environment where more than two contrasting short rounded front vowels are possible, viz. [ø] and [œ] before nasals, [y] and [ø] otherwise. All the vowels in question are subject to r-colouring according to the principles mentioned in the preceding section (cf. Table I below).

Hjelmslev (1951 p. 23) has mentioned derivations like mand - mandig [manʔ, 'mandi] in support of underlying forms like /mand/. The "latent" /d/ (to use his term) explains the stød (which occurs automatically in monosyllables ending in a consonant cluster whose first member is a sonorant), and is pronounced before the derivative ending ig. However, Rischel (1970b p. 129) has proposed that we have long (or geminated) sonorant consonants in such cases (i.e. /mann/), and that d is inserted between long sonorants and the suffix in question by

[22] A small handful of examples like Anders, anderledes, andre, vandre, aldrig ['ɑnɒs, 'ɑnɒˌle.ʔðəs, ɑndʁɒ, vɑndʁɒ, 'ɑldʁi] seem to indicate that r-colouring can apply across intervening consonants. But for the following reasons I think it is preferable to give these words an exceptional phonological form in the lexicon and continue to claim that r-colouring can only affect neighbouring segments: Firstly, there are other words, like klandre, which in the same phonologically relevant environment have the expected [a]; second, this supposed effect does never cross morpheme boundaries: words like vante+r etc. all have [a], although r-colouring normally does: ta', tar etc. [t^sæ·ʔ, t^sɑ·ʔ] ; finally, /a/ would be the only vowel which could undergo this strange rule (e.g. the /ɛ/ of kæntre, ændre etc. does not undergo the slightest r-colouring).

a general rule. Since there are minimal pairs like *skynd* (imperative), *skøn* [sgønʔ, sgœnʔ], which should both end in /-nn/ according to Rischel, he is forced to recognize two different underlying short rounded front vowels before nasals (e.g. *synd* [sønʔ]/synn/ versus *skøn* /sgønn/).

However, we shall follow Hjelmslev more closely and propose an alternative, viz. that there is only one underlying short rounded front vowel before nasals, and that the underlying distinction between *synd* and *skøn* (apart from the prevocalic consonant(s), of course) resides in the final consonant cluster: /sYnd/ versus /sgYnn/. There is then, according to this hypothesis, a regularity (i.e. a redundancy rule or condition) saying that /Y/ is relatively narrow before a nasal followed by a non-nasal consonant, but otherwise relatively open before a nasal. The following facts all speak in favour of this latter hypothesis:[23]

(i) All words which can be shown to have a "latent" /d/ have stød when occurring as monosyllables.

(ii) There are no stød-less monosyllables in [-øn] (cf. *søn* [sœn]).

(iii) Most words having [ø] before *n* can be shown to have a "latent" /d/: *synd*, *fynd*, *ynde*, *kynd(ig)*, *mynd(ig)*, whereas no words having [œ] before *n* can be shown to have /d/.

[23] These arguments are given in fuller form in Basbøll 1972. - It should be borne in mind that monosyllables whose underlying form ends in a sonorant consonant followed by at least one other consonant have stød (/r/ does not count as sonorant in the clusters /rp, rt, rk, rf, rs/, which historically have unvoiced /r/).

(iv) The adjectival derivative endings *ig* and *lig* are synonymous. No words with [ø] before *n* take *lig*, whereas no words with [œ] before *n* take *ig*.[24]

(v) No words ending in *m* can be shown to have a "latent" consonant, say *b*, and there are no words with short [ø] followed by an *m* which is not followed by another consonant.

(vi) There are no words with [œ] followed by [ŋ] (which in turn is derived from /ng/).

These facts are mere accidents (or better: are quite unconnected) according to Rischel's proposal, whereas they are predictable consequences of our proposal (that there is only one underlying short rounded front vowel before nasals, which shows up as a relatively narrow vowel before a nasal followed by a non-nasal consonant, otherwise as a relatively open vowel[25]), together with independently established suppositions on the stød (on which Rischel agrees).

Table I shows how the different manifestations of the short rounded front vowels can be derived. It should not be taken too seriously, and there is no space here to discuss all the rules mentioned. The language is advanced standard Copenhagen.

[24] I thus consider *ig* and *lig* to be instances of the same formative, the choice between them being determined mainly by phonological environment. A counterexample like *mandig* 'manly' versus *mandlig* 'masculine' is only apparent: the distinction has been lexicalized. The "irregular" (unexpected) form *mand+lig* is probably formed in analogy (whatever that means) with *kvinde+lig* 'feminine' - which is not opposed to anything like *mandig* - where the *lig*-ending is quite regular.

[25] The rule is in fact not restricted to nasals, cf. *fylde*, *fyldig* [fylə, 'fyldi] (and similarly *skylde*, *skyldig*), whereas no derivatives in [-øldi] can be found. Cf. the fact that there are no stød-less monosyllables in [-yl] (but there are in [-øl], e.g. *øl*).

TABLE I

		dyrke	dørke	grynt	grønt	synder	sønner	rytter	bryst	støvle	vrøvle
(1)	UF	y	ø	Y(nt)	Y(nn+t)	Y(nd)	Y(n)	y	ø	Y	Y
(2)	RR	-	œ	y	ø	y	ø	-	-	œ	œ
(3)	NL	-	-	ø	œ	ø	œ	-	-	-	-
(4)	rC	-	Œ	œ	Œ	-	-	-	œ	-	Œ
(5)	PO	y	Œ	œ	Œ	ø	œ	y	œ	œ	Œ

Explication of the abbreviations of Table I:

(1) UF = Underlying form. /Y/ means a short rounded front vowel unspecified as to degree of opening, or the maximally unmarked short rounded front vowel. The symbol /Y/ is used (instead of simply /y/) to indicate that there is no lexical contrast in vowel height.

(2) RR = Redundancy rules, or the like. Including the rule (or condition) discussed in the present section, the rule (or condition) alluded to in footnote 20 that a non-high short vowel before /r/ is low, and the rule (or condition) that a short vowel before /v/ is low.

(3) NL = Nasal lowering. A rule that lowers some short vowels one degree before nasals (the details of this rule are rather dubious and could not be investigated here).

(4) rC = r-colouring. See section 3. above.

(5) PO = Phonetic output.

Note that the ordering of nasal lowering and r-colouring is crucial: the /y/ in *grynt* must be lowered to *ø* before r-colouring can apply to give [gʁœnʔḋ], cf. *rytter* with [y]. This is the unmarked ordering ("feeding order") of the two rules.

References

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

Basbøll, Hans 1969: "The Phoneme System of Advanced Standard Copenhagen", *ARIPUC* 3/1968, p. 33-54.

Basbøll, Hans 1971: "A Commentary on Hjelmslev's Outline of the Danish Expression System (I)", *Acta Linguistica Hafniensia* XIII.2, p. 173-211.

Basbøll, Hans 1972: "Some Remarks Concerning the Stød in a Generative Grammar of Danish", *Proceedings of the KVAL Spring Seminar 1972* (preliminary title only), held in Åbo and Stockholm on April 8th and 9th (forthcoming, to be published by Skriptor, Stockholm).

Diderichsen, Paul 1957: "Om udtalen i dansk Rigssprog", *Danske Studier*, p. 41-79.

Hjelmslev, Louis 1938: "Neue Wege der Experimentalphonetik", *Nordisk Tidsskrift for Tale og Stemme* 2.10, p. 153-194.

Hjelmslev, Louis 1951: "Grundtræk af det danske udtrykssystem med særligt henblik paa stødet", *Selskab for nordisk Filologi. Aarsberetning for 1948-49-50*, p. 12-24. [Reprinted in English translation in Louis Hjelmslev, *Essais linguistiques* II (1972, forthcoming).]

Rischel, Jørgen 1969: "Notes on the Danish Vowel Pattern", *ARIPUC* 3/1968, p. 177-205.

Rischel, Jørgen 1970a: "Consonant Gradation: A Problem in Danish Phonology and Morphology", *The Nordic Languages and Modern Linguistics* (ed. Hreinn Benediktsson), p. 460-480.

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

Spang-Hanssen, H. 1949: "On the Simplicity of Descriptions", *Recherches structurales = TCLC V*, p. 61-70. [Reprinted in Hamp et al. (eds.), *Readings in Linguistics* II (1966), p. 234-241.]

Twaddell, W.F. 1938: "A Phonological Analysis of Intervocalic Consonant Clusters in German", *Actes du quatrième congr. intern. de linguistes* (Copenhagen 1936), p. 218-225.

COMPOUND STRESS IN DANISH WITHOUT A CYCLE

SR: Accentuation

Jørgen Rischel

1. Introduction

The existence or non-existence of cyclic rules in phonology is an important issue in the current debate. Among the processes which are widely assumed to be cyclic, the graded reduction of stresses in compound words probably holds a particularly high rank. The mechanism involved was stated in such terms already by Chomsky, Halle, and Lukoff (1956), and the more recent formulation of it in Chomsky and Halle (1968) largely dominates descriptive approaches these years.

I wish to point to the fact that compound stress in Danish can be described without reference to cyclic rule application (at least in the sense in which "cyclic" is generally taken), and that this approach enables us to abandon the parameter of "degree of stress", which (unlike the abstract dichotomy [± stress]) seems to me fictitious in a description of Danish. Though I do not generalize the results to other languages, the very possibility of describing compound stress in a Germanic language in this way seems to me of interest to phonological theory in general.

The model for generating complex stress patterns which I use below, was actually outlined ten years ago in a paper using English for illustration (Rischel 1964, also cf. the much earlier outline of the hierarchical concept in Fischer-Jørgensen 1948 (1961)). However, the said paper was highly sketchy and moreover contained a number of contentions which are not directly relevant to the present issue. I shall, therefore, confine myself to mentioning the paper rather than referring to it in more detail.

2. The hierarchical model

We assume a basic difference between stressed and unstressed syllables, the distribution of [+stress] and [-stress] being either lexical or introduced by rule.[1] There must then be a device that converts [+stress] into degrees of stress (and a device that converts [-stress] into degrees of stress under certain conditions) if, for the moment, we assume that "degree of stress" is a linguistic parameter (this will be questioned later in the present paper). As shown by Chomsky, Halle, and Lukoff (1956) the grading of [+stress] is closely dependent upon the constituent structure of the syntactic surface representation (though obviously with some adjustments). If degrees of stress that are intermediate between the strongest and the weakest are considered as reduced occurrences of

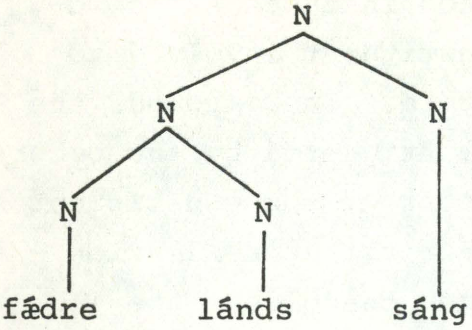
[1] In Rischel (1970) it was shown that the stress placement in Danish formatives (morphemes) is largely predictable from their segmental structure, and that the stress placement in Danish noncompound words is (normally) found by deleting all but the last formative stress. The rules for this simple mechanism of stress assignment were presented without a sufficiently clear statement of the theoretical framework in which they are to be understood (the attempt p. 140 ff to harmonize the approach with that of Chomsky-Halle, must be considered as a failure). The stress assignment rules for formatives are most naturally understood as redundancy rules, whereas the word-stress rule is a process rule, belonging to phonology proper. It must be conceded that not all formative stresses in Danish are predictable; i.e. some formatives are lexically marked for idiosyncratic stress placement, whereas the majority are unmarked (i.e. stressed according to redundancy rules). In a reasonably realistic phonological representation stress may have to be marked more often than suggested in the said paper, but a solution cannot be found until other types of evidence in favour of more or less abstract representations have been investigated, in particular the stød (cf. Basbøll, forthcoming). - As for word stress assignment, I wish to point to the fact that the rule involved requires only a distinction of [+stress] and [-stress], which agrees with the starting-point of the present paper. (The short section on compounds (p. 138) was kept in quite traditional terms and may be disregarded.)

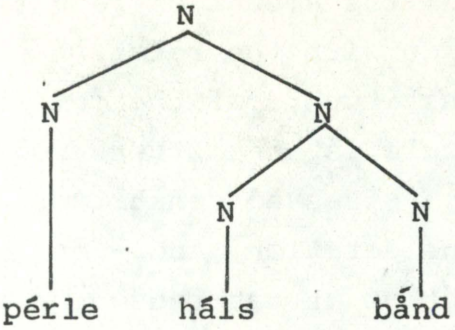
[+stress], the amount of reduction in each instance has something to do with the relation of the constituent involved to other constituents of the complex structure. In compounds the general rule is that stress reduction is triggered by the occurrence of a stressed constituent to the left of the constituent under consideration, and - at least to a first approximation - the effect of stress reduction is stronger the closer the connexion is between the two constituents. This can be taken care of by a cyclic approach, and indeed invites such an approach. However, the concept of cyclicality may be unnecessary here, and hence should be abandoned if it is not required for independent reasons. If stress grading depends on the syntactic "tree-structure", it may be directly deducible from this representation.

In the following I confine myself strictly to compounds with initial "main stress". The exceptions to this general pattern are few (subregularities for these will not be stated here).

Consider compounds like *fædrelandssang* 'patriotic song' (literally: father-land's-song') and *perlehalsbånd* 'pearl necklace'. The internal constituent structures of these compounds (which can be posited no matter whether the compounds are considered to be entirely or partially lexicalized), are obviously different. In the former the primary break is between *fædrelands* and *sang* (with a secondary break between *fædre* and *lands*), in the latter the primary break is between *perle* and *halsbånd* (with a secondary break between *hals* and *bånd*). However, both compounds consist of a sequence of three noun stems: a bisyllabic one followed by two monosyllabic ones. Assuming that each of these gets initial stress by rule (cf. Rischel 1970 p. 119ff) we arrive at structures which can be roughly represented like this:[2]

[2] I assume some readjustment by which, for example, the formative *s* in *fædrelandssang* is incorporated into the second noun.



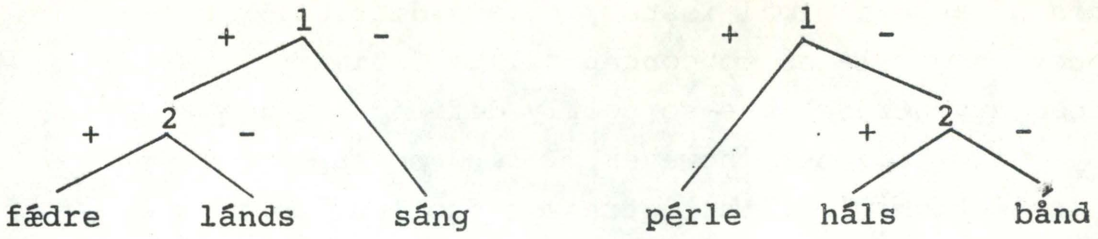


A compound stress rule with associated conventions for stress adjustment may be considered to have the effect of reducing stresses. Under a cyclic application stress reduction would apply twice to *lánds* in the former compound, and to *bǻnd* in the latter, but only once to *sáng* in the former, and *háls* in the latter. This seems plausible enough, since it would not be too difficult to make phoneticians agree that the two forms are stressed like this:[3]

'fædre,lands„sang
'perle„hals,bånd

where [„]indicates a less reduced stress, and [,] a more reduced stress. However, this very accentuation can be read off the tree structure representation, provided that we have a rule saying that left branches are given relatively more prominence than right branches in compounds. If this difference of prominence is indicated by plus versus minus we get the following representations (with omission of the N labels for simplicity):

[3] Alternatively, [,] might be used instead of [„], if the syllables here marked with [,] are assumed to be reduced to "weak stress". A reduction of stresses all the way to "weak stress" might entail a process [+stress]→ [-stress]. I have not considered the occurrence of such a rule in compounds in the present paper.



The degree of stress reduction cannot, of course, be determined solely by counting minusses. The convention involved must refer to the position of the minusses in the hierarchy, e.g. by assigning each minus the number of the node above it. The phonetic stress reduction would then be a function of such numbers. Since this function is unknown[4] I shall represent a nominal amount of stress reduction by simply indicating the hierarchical number, "2" meaning a stronger reduction than "1", etc.

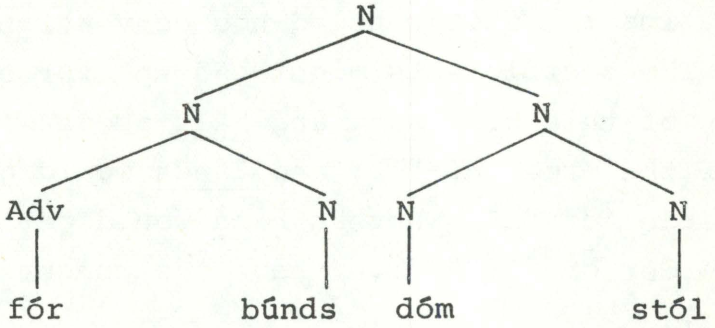
Now one might imagine different ad-hoc conventions for stress reduction. The coefficients might add up, for example. Under a convention of this kind *sáng* and *háls* above would get a reduction of the order of "1", and *lánds* would get a reduction of the order of "2", whereas *bånd* would get a reduction of the order of "1+2" (i.e. have a stronger reduction than *lánds*).

Another possibility would be that the degree of reduction directly reflects the depth of compounding, i.e. that the convention applies to the lowest minus in each case. Under this convention *fædrelandssang* gets stress reductions according to the pattern "0-2-1", and *perlehalsbånd* according to the pattern "0-1-2".

[4] A priori it need not be a linear function at all. Since I shall abandon the use of stress coefficients later in this paper, the problem is of no real interest here.

This is an empirical issue. It is a difficulty that the phonetic correlates to concepts like "primary stress" and "secondary stress" are so poorly defined. As for my own subjective judgment, however, I find no support of the former assumption, i.e. that stress reduction operates on a summation basis. If there is at all a difference between the reduction of stress on *lands* and *bånd* in the examples above, it is rather so that the former is more reduced than the latter.[5] This can be taken care of by a rhythmic convention applying optionally (cf.3.below) if we assume that the basic degree of reduction is the same in both cases.

In compounds like *forbundsdomstol* 'Federal Tribunal' we have a more complex constituent structure which can be represented like this (with some adjustment[6]):



Here we get the strongest stress on *főr*, and the next strongest on *dőm*, under any reasonable convention. However, the weaker stresses on *bűnds* and *stől* would differ crucially depending on the functioning of stress reduction. If it is additive (in some way), we should get less stress on *stől*

[5] This is a highly subjective evaluation. I do not really know how to arrive at a valid criterion for this decision.

[6] Cf. note 2.

than on *bũnds*. Again, this is at variance with my subjective judgment; there is rather a tendency in the opposite direction, which can be taken care of as suggested above.

I would suggest, therefore, that the convention for Danish *reduces each stress of a compound (in relation to the leftmost stress) solely according to the order of (the number assigned to) the lowest node that dominates the constituent in question, and from which it hangs in a branch that is labelled "minus".*

This convention is thus sensitive to depth of compounding and to occurrence in non-initial position under the lowest dominating node.

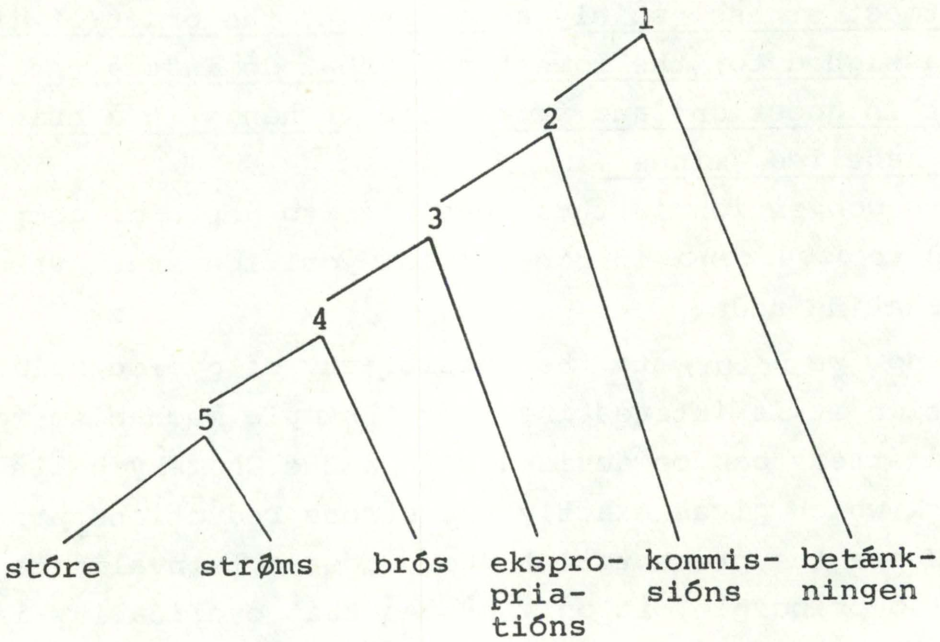
If now we return to the possibility of cyclic rule application it is interesting that a simple mechanism for compound stress can be devised within the Chomsky-Halle framework which gives exactly the stress reductions posited above. Hence the presentation above does not invalidate the cyclic principle; it only argues that cyclicality is unnecessary.

3. Structure Simplification

Both approaches referred to above, cyclic and non-cyclic, can be made to work ad infinitum, though it goes without saying that there is a limit to the degree of stress that are distinguished in actual communication.[7] In Danish it is possible to construct very complex compounds, and if these are entirely right-branching or left-branching the depth of compounding may be considerable. A fancy compound like *storestrømsbroseksspropriationskommissionsbetænkningen* 'the report of the commis-

[7] Cf. Householder's remark to Rischel (1964), *ibid.* p.93, on which I entirely agree.

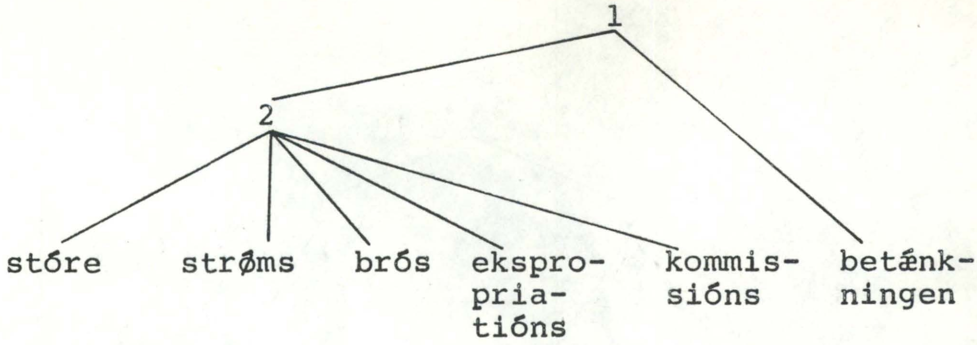
sion for the expropriation for the Storestrøm Bridge (lit.: the bridge of the Great Current)' sounds funny, of course, but is in no way unacceptable from a linguistic point of view. Assume a tree structure like the following:



According to the alleged convention we should get increasing stress reduction from betǽnkningen (1st order) through strøms (5th order). Or put differently: strøms, brós, etc. through betǽnkningen should have increasing stress in the order in which they are spoken. I doubt it that anybody could make a convincing performance of this theoretical stress pattern. There will necessarily be some kind of adjustment reducing the depth.

One possible type of adjustment may be described with reference to a threshold, depending to some extent on tempo and style of speech,[8] below which hierarchical differences vanish. Assuming, for instance, that nodes cannot be of more than second order in casual speech, the structure above would simplify into

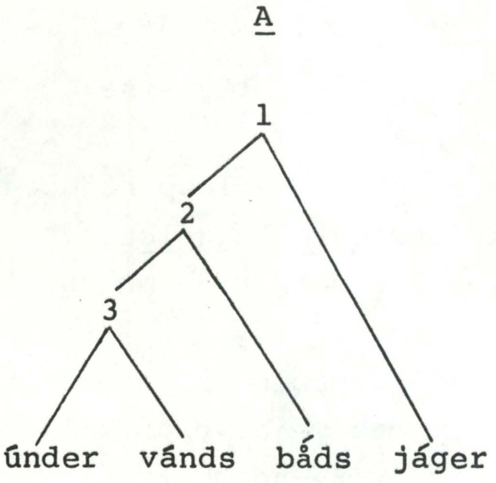
[8] For a device related to this idea of "threshold" cf. Bierwisch (1966) p. 166 ff.

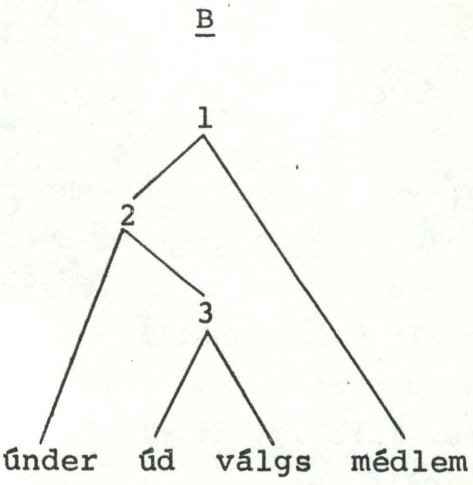


which gives 1st order reduction on betǽnkningen, and 2nd order reduction on the other, non-initial constituents, i.e. in conventional stress notation,

'store+,strøms+,bros+ekspropria,tions+kommis,sions+bé„tænkningen

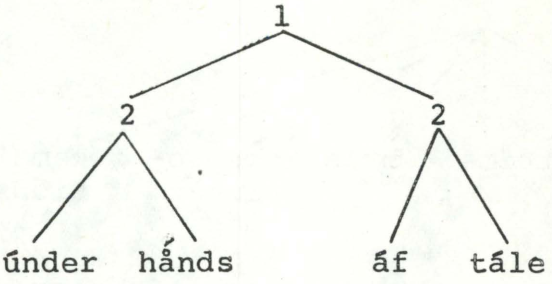
To what extent (under what conditions) such node collapsing actually occurs, could be studied by observing the neutralizations among different hierarchical structures that occur in a given type of speech. With the threshold referred to above we should get a neutralization of the structures in A and B below, whereas A/B would remain distinct from C:[9]





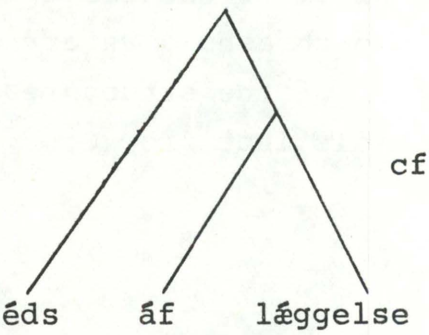
[9] The words mean 'submarine chaser', 'member of a subcommittee', 'private agreement'.

C

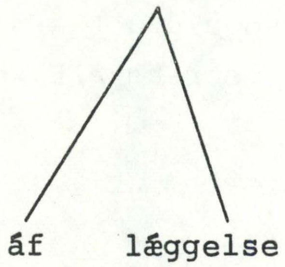


In rapid speech the reduction (and neutralization) may well go even further. Data throwing light on this would deserve close study.

The pattern is complicated by a tendency, in some constructions, to perturbate the relative stresses of constituents. This occurs very clearly in a form like *edsaflæggelse* 'taking the oath', which has the structure

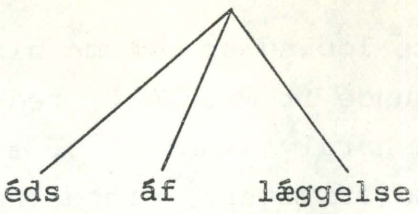


cf.



so that we should expect the stress on *áf* to be less reduced than that on *lǽggelse*. However, the form *edsaflæggelse* can be pronounced with more stress on the ultimate than on the antepenultimate constituent.

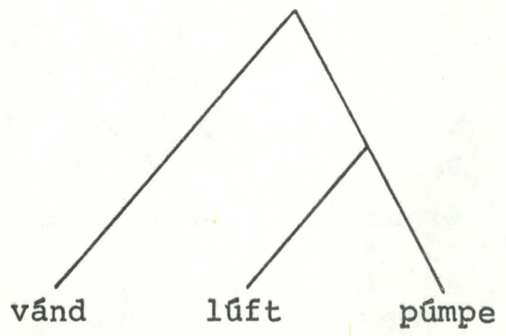
Synchronically, there are several possible explanations of this. Firstly, it may be suggested that we have an optional simplification of the structure to a one-node structure



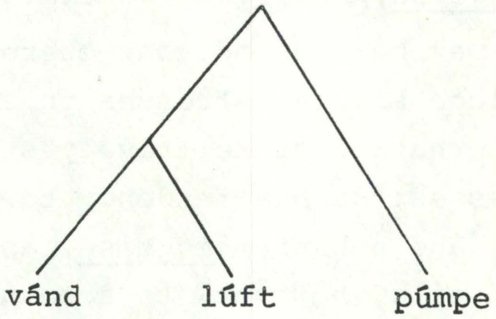
This should give full stress on éds, and evenly reduced stresses on áf and lǽggelse, according to the convention given earlier. Now it may be assumed that there is a phonetic tendency to replace similar stresses on successive constituents by an alternation of relatively stronger and relatively weaker stresses; such a tendency toward contrast would reduce áf, and enhance lǽggelse, as required.

This explanation fails, however, to account for the fact that the tendency to perturbation is not equally strong in all forms, e.g. perlehalsbånd (see above) could hardly occur with stress perturbation. In this respect we are better off if we connect the deviating accentuation with the fact that aflæggelse contains a succession of adverb plus verb, since constructions involving adverb plus verb or verb plus adverb have special prosodic properties anyway. Compounds containing adverb plus verb sometimes have noninitial stress (regularly with +lig, cf. af'tagelig 'detachable') and thus break the most basic rule of initial compound stress. I shall not got further into this here.

There are other cases, however, where it seems to me possible (in my own idiolect, at least) to have stress perturbation though the structure involves no adverb plus verb. Take a technical term like vandluftpumpe. This means 'water jet air pump' i.e. according to meaning criteria the structure should be

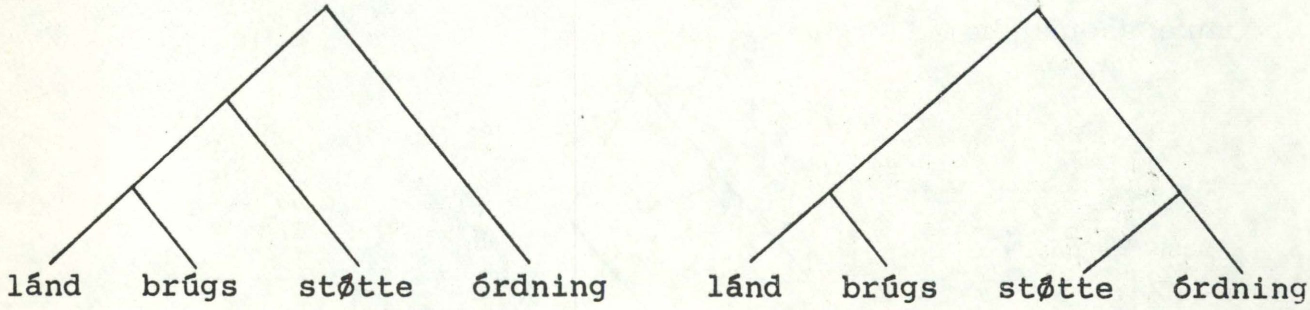


However, even if I have looked up the meaning of the word, I am inclined to pronounce it with more reduction on lúft than on púmpe. This is hardly a matter of simplifying the structure to a one-node structure, since the technicality of the term does not invite casual pronunciation, but rather a straightforward readjustment to



I suggest that there exists a tendency to reinterpret ackward compounds of the former structure as compounds of the latter structure.

Note that the effect of this restatement is that we get an alternation of degrees of stress, rather than a monotonous decrease of stresses. This tendency to avoid stress monotony may also seem to operate in cases where the constituent structure is genuinely ambiguous, e.g. if we take the compound landbrugsstøtteordning which can be read as [landbrugsstøtte][ordning]([arrangement of] [financial support of agriculture]) or as [landbrugs][støtteordning]([arrangement of financial support][for agriculture]), it seems "easier" to pronounce the latter option, viz. structure B below.



Again, the preferred alternative provides rhythmic alternation, and in this case it moreover reduces the depth of compounding.

It is interesting to note that the preferred analysis posited for these forms has the same effect as structure simplification combined with a phonetic tendency toward rhythmic alternation of the stresses on successive constituents (cf. above). It also has the same effect as destressing of adverbs internally in compounds (cf. edsaflæggelse). Thus, until considerably much more is known about the ways in which different compound structures are distinguished or fail to be distinguished phonetically, we cannot decide what is really going on in the forms with superficial stress "perturbation".

4. Abolition of "degree of stress" as a linguistic parameter

Throughout this paper I have referred to "degrees of stress" (disregarding unstressed syllables, i.e. syllables which are not assigned a [+stress]), but I have deliberately avoided any discussion of the meaning of the phonetic label "degree of stress". Thanks to our phonetic tradition it is not difficult to communicate by means of such terms, and I have therefore found it practical to use the terms without any definition whatsoever.

The question of the parameters of stress is crucial the moment we want to give the convention stated in 2. above in an exact form. It may be possible to find a reasonable correlate to stress at the level of speech production, but the signalling of constituent structure is known to be highly complex, at least in English (see Scholes 1971 with references), where it includes both intensity and pitch changes as well as separation in time (i.e.

"disjuncture"). On the basis of the limited data available[10] it may be assumed that pitch jump is an important correlate of stress in Danish, and "disjuncture" is undoubtedly an essential marker of constituent structure also in this language.

Now the question is: do we want our stress rules to give an output in which each syllable is assigned a "degree of stress" represented by a coefficient? Since these coefficients must undergo a highly complex transformation into different parameters before we arrive at anything that can be measured phonetically, the assignment of stress coefficients seems to me warranted only if there is a solid basis for assuming that these coefficients represent a significant level of linguistic specification. As far as Danish is concerned, at least, it does not seem to me intuitively meaningful to specify coefficients of stress the way this is done in Chomsky and Halle (1968) and elsewhere, or for that matter, to specify stress degrees by symbols like ['], [„] and [,], though, as said above, such representations have a communicative value among linguists who know what they refer to.

In my opinion indications of graded stresses are linguistically significant only indirectly, namely by defining types of constructions. Hence it seems to me superfluous to introduce such representations if the constructions themselves contain sufficient information without being transformed into representations with graded stresses. In order to specify the parameters that signal the structure of compounds it would seem appropriate to have recourse to two types of linguistic information, viz. <u>the location of the syllables marked as [+stress]</u> and <u>the location and order of the constituent boundaries</u>. But this is indeed what the adjusted phrase-marker presents after application of the stress redundancy rules.

[10] Eli Fischer-Jørgensen has made som instrumental research on the prosodic characteristics of Danish compounds. Her results indicate that shift of pitch is an essential correlate to "stress". (Personal communication.)

I should prefer, therefore, to replace the convention outlined in 2. above by a convention which specifies more useful phonetic parameters. This means that the expression "reduction of _n_th order" should be replaced, e.g., by information referring to pitch jumps and temporal relations. The pitch change and temporal distance between the stress points of consecutive constituents would be assumed a priori to be lesser the higher the order of the node involved, e.g. in _fædrelandssang_ (see 2. above) the first two constituents should be specified as spoken on almost even pitch and closely adjacent to each other, whereas in _perlehalsbånd_ this would apply to the last two constituents. In so far as a valid set of conventions could be set up, this type of phonetic characterization would seem to me immensely much more satisfactory for Danish than an appeal to fictitious concepts like "stronger" or "weaker" reduced stresses. - The "main" stress of a (normal) compound is simply the leftmost occurrence of the category [+stress]. A "secondary" stress of a (normal) compound is secondary by virtue of _not_ being the leftmost occurrence of [+stress]. The vocal effort or intensity contour of the word may exhibit a peak associated with the first occurrence of [+stress], or a more complex pattern depending on the constituent structure, but this is not inherently the most interesting feature of accentuation though it should be built into the convention, of course, like other parameters. The interesting question is not how to specify degrees of stress as a parameter, but how to choose the phonetic parameters (= instructions to the speech-organs, auditory parameters, or what?) which should be specified by the "stress" convention.

The direct consequence of the contentions stated above is that the output of the phonological component must take the form of a hierarchical representation (i.e. a tree structure or its equivalent: a bracketed representation).

The phonetic conventions, whatever they are, operate on this hierarchical representation. The tree structure is not necessarily congruent with a syntactic surface representation, since various readjustments take place, but surface syntax and lexicon together make it deducible by rule.

It has been argued quite recently by Charles Pyle (1972) that there are no phonological rules (conventions) that replace formative boundaries by boundary markers, i.e. that boundary markers do not exist as phonological units. Pyle argues that the jobs which boundary markers (junctures) do in current formulations, should actually be assigned to the formative boundaries. I agree, since the introduction of boundary markers would be entirely redundant once the adjusted constituent structure (which defines the location of such boundary markers) is present for rules to refer to. There is, however, a problem with lexical items, since "formative" boundaries are marked (i.e. have a phonological effect) in some cases but not in others. If one does not insert boundary markers to indicate the marked boundaries, it is necessary to have rules (triggered, at least in part, by lexical idiosyncracies) which delete boundaries that have no phonological effect. The consequences of such an approach must be investigated.

One is, quite generally faced with the serious question: what kinds of readjustment of boundaries do we have to assume? I think the answer to this question depends on how lexicon is assumed to be organized and how lexical insertion is assumed to take place. Without a theory about lexicon there is no point in discussing whether the constituent structure that is relevant to phonology on different levels has a more or less direct relation to surface syntax.

References

Basbøll, Hans (forthcoming): "Some Remarks Concerning the Stød in a Generative Grammar of Danish", *Proceedings of the KVAL Spring Seminar 1972*, preliminary title only, to be published by Skriptor, Stockholm).

Bierwisch, M. 1966: "Regeln für die Intonation deutscher Sätze", *Studia Grammatica* VII, p. 99-201.

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

Chomsky, Noam and Morris Halle 1968: *The Sound Pattern of English*.

Chomsky, N., M. Halle and F. Lukoff 1956: "On Accent and Juncture in English", *For Roman Jakobson*, p. 65-80.

Fischer-Jørgensen, Eli 1961: "Some Remarks on the Function of Stress with Special Reference to the Germanic Languages", Congr. Intern. Sc. Anthropol. & Ethnol., Comptes-Rendus, IIIe session, Bruxelles 1948, page 86-88.

Pyle, Charles 1972: "On Eliminating BM'S", mimeographed paper (Univ. of Michigan), to appear in *Papers from the Eighth Regional Meeting* Chicago Linguistic Society.

Rischel, Jørgen 1964: "Stress, Juncture, and Syllabicification in Phonemic Description", Proceed. of the IXth Int. Congr. of Linguists 1962, p. 85-93.

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

Scholes, Robert J. 1971: Acoustic Cues for Constituent Structure (Janua Linguarum, Series Minor, 121).

COMPARISON BETWEEN AUDITIVE AND AUDIO-VISUAL PERCEPTION OF PB-WORDS MASKED WITH WHITE NOISE

Carl Ludvigsen

1. Introduction

The basis of this paper is furnished by some experiments carried out two years ago at The State Hearing Center, Bispebjerg Hospital, Copenhagen. The purpose of these experiments was in broad outline to examine the influence on the discrimination score when PB-words were presented to listeners with a normal hearing belonging to three different age-groups, and the words were presented with and without the possibility of seeing the face of the speaker. A report on this experiment is given by Ewertsen et al. (1970).

The aim of the present paper is to study the answers obtained from the subjects under the two modes of presentation and four different signal to noise ratios (S/N).

1.1. The word material

The stimulus material consisted of four phonetically balanced lists (A, B, C, D) each containing 25 words. Due to the phonetic balancing the lists contained only mono- and disyllabic words, all commonly used Danish words. All disyllabic words had a trochaic stress pattern with the first syllable stressed and the second unstressed.

1.2. The presentation

A recording of the lists read by a male speaker whose dialect was close to standard Copenhagen was made on a video tape recorder for use on internal television. The speech signal from the sound track on the video tape was fed through an attenuator and then mixed with noise. The noise level was kept constant throughout the experiment and different signal to noise ratios were obtained by attenuating the speech signal. The noise was approximately "white" within the audible frequency range. The sound signal was presented monaurally through a pair of headphones. The visual signal i.e. a frontal picture of the speaker's face could be presented synchronously with the acoustic signal on a TV-screen in front of the subject.

Each of the four lists was presented four times to every subject. The first presentation was given through the headphones, with no picture on the TV-screen, for the second presentation the TV film was added, and finally approximately one month later these two presentations were repeated.[1] The S/N's were different for the four lists: list A: S/N = -20 dB, list B: S/N = -10 dB, list C: S/N = 0 dB, and list D: S/N = +10 dB.

1.3. The subjects

28 subjects participated in the experiment. For the sake of studying the influence of age on the discrimination score these subjects were selected from different age groups. Thus, 9 subjects were from 20-25 years old, 10 from 45-55 years old, and 9 from 65-75 years old. Audiograms were taken of all subjects, and only subjects with audiograms normal for their group of age participated in the experiment. For the present study, however, it was decided to disregard answers from the oldest group mainly because of the pronounced hearing losses at high

[1] This was done in order to study the effect of retesting.

frequencies which are typical for their age-group. Furthermore two persons from the group 45-55 years were discarded because of hearing losses of respectively 30 and 35 dB for high frequencies. The remaining 17 subjects had audiograms differing less than 25 dB from the normal threshold in the frequency range from 125 Hz to 8000 Hz.

1.4. Experimental procedure

The subject was placed in a quiet room normally used for audiometry. The experimenter was placed in an adjacent room from where he could watch the subject through a window. Near the subject was placed a microphone which allowed the experimenter to hear the replies from the subject. The subject was instructed in the testing routine and was asked to repeat the words as he heard them. The experimenter registered whether the subject was able to repeat the words and if the subject answered incorrectly, i.e. with a word which was not identical to the stimulus, he noted this word on the list.

2. Analysis of answers

As the subjects were not forced to answer, three alternative types of reply are possible: 1) the subject may not have answered or he may have answered 2) with the correct word or 3) with an incorrect word. In the first case no information is obtained about the perception of the stimulus. In the second case some uncertainty exists concerning the cues used for identification of the word. Thus it is possible that a certain feature of a stimulus may not be detected by the subject although a correct answer is given, since the identification may have been based solely upon other features. The most useful source of information about the mechanism of perception seems to be the incorrect answers. Thus a comparison between an incorrect answer and the stimulus word provides information about which cues are detected and which are not.

2.1. Detection of number of syllables

If we assume that the subjects are unable to detect the number of syllables we shall expect the number of syllables found in the incorrect answers to be distributed approximately in the same way as in a normal word material and independently of the number of syllables in the stimulus word. This hypothesis can clearly be rejected from the material: Even at the most unfavourable S/N in the auditive tests the detection of number of syllables is very accurate. This is shown in TABLE 1 below.

TABLE 1

STIMULUS	NUMBER OF WORDS	PRESENTATIONS	NUMBER OF CORR. ANSW.	NO ANSW.	WRONG ANSW.	NUMBER OF WRONG ANSW. WITH 1 SYLL.	2 SYLL.
2 SYLL. WORDS	12	408	52	298	58	0	58
1 SYLL. WORDS	13	442	33	360	49	47	2

TABLE 1. Answers from all 17 subjects pooled (list A, auditive test, S/N = -20 dB)

Table 1 shows that although more than 50 % of the answers are incorrect almost all of these contain the correct number of syllables. This finding also indicates that addition of the visual signal will not improve the detection of syllables appreciably.

2.2. Detection of the unstressed vowels in the disyllabic words

A cursory glance at the incorrect answers tells that the subjects rarely fail to detect the vowel in the second, unstressed syllable of disyllabic words. This impression holds true even for the tests with the smallest S/N and with auditive presentation only. In list A the second syllable of 11 of the 12 disyllabic words were of the type (C)ə(C) (zero or one consonant + schwa + zero or one consonant) and only one ended in a different vowel, viz. ʌ (the word was "tænder" (teeth)). To this word were given 10 answers; six of these were incorrect but all ended in unstressed ʌ. To the remaining 11 words 104 answers were obtained 52 of which were incorrect. 51 of these ended in Cə(C) and only one in a different vowel (unstressed, short i).

This very accurate detection of the unstressed vowels was to be expected: since approximately 90 % of Danish disyllabic words with stressed first syllable have schwa as unstressed vowel, the a priori uncertainty about the identity of the unstressed vowel is relatively small. On the other hand a reduction in intensity or duration of unstressed vowels might be expected to make the identification difficult.[2]

2.3. Confusions between stressed vowels

A larger percentual number of errors occur among the stressed vowels. And, of course, the nature of the errors depends heavily on the mode of presentation, auditive or audio-visual, and the S/N. In order to study the confusions among stressed vowels the words were grouped with respect to the acoustic quality of the stressed vowel. Obviously, this grouping can be done

[2] Intensity curves of the stimulus material showed no reduced intensity for the unstressed vowels.

in more or less detail. For the present purpose the grouping was based on the frequencies of the two lowest formants, F_1 and F_2, as found from a frequency analysis of a tape recording of the stimulus material. The formant frequencies of vowels with marked transitions of F_1 and F_2 were measured at the most steady part of the vowel (generally in the middle) or, if no such part could be found (as in diphtongs), in the first part of the vowel. The letters used for transcription are given below with a few examples.

i	(fine, tit, bi)	[fi:nə, tʰit, bi·ʔ]
e	(dele, fedt, sne)	[de:lə, fet, sne·ʔ]
ɛ	(sæbe, mælk, æg)	[sɛ:bə, mɛlʔk, ɛ·ʔk]
æ	(stave, værst, sal)	[sdæ̈:və, væɒst, sæ̀·ʔl]
a	(nat, trække)	[nat, tʀage]
α	(aften, mig, leg)	[αfdən, mαi, lαiʔ]
ɑ	(varme, brand, barn)	[vɑ:mə, brɑnʔ, bɑ·ʔn]
y	(lyve, ny)	[ly:və, ny·ʔ]
ø	(løbe, dyppe, sø)	[lø:bə, døbə, sø·ʔ]
œ	(køn)	[kœnʔ]
Œ	(gør, tørstig)	[gŒʀ tʰŒʀsdi]
u	(bule, skulder, jul)	[bu:lə, sgulɒ, ju·ʔl]
o	(hoste, sko)	[ho:sdə, sgo·ʔ]
ɔ	(kåbe, stå)	[kʰɔ:bə, sdɔ·ʔ]
ʌ	(slot, øjne)	[slʌt, ʌinə]
ɒ	(går, får)	[gɒ·ʔʁ], [fɒ·ʔʁ]

After the grouping confusion matrices were formed: one for each age-group and mode of presentation. As the confusions were distributed in the same manner for the two age-groups the answers from these were pooled for the subsequent examinations. From these matrices it followed that with a few exceptions the vowels in the incorrect answers were either identical to the stimulus vowel or (especially for a low S/N) with approximately the same first formant frequency. This observation is illustrated in Figs. 1-6.

Results obtained at different S/N cannot immediately be compared since they come from different stimulus words. In order to make such a comparison meaningful, more information about the stimulus material is given in TABLE 2. This table shows for the three lists A, B, and C separately the total number of times a word containing a specific vowel was presented to a subject (e.g. in list A two words have /i/ as stressed vowel; these words are presented twice to seventeen subjects; consequently, the total number of presentations is 68). TABLE 2 also shows the total number of answers containing the correct vowel. The results are given for auditive (upper figures for each vowel) as well as audio-visual presentation (lower figures).

In Figs. 1-6 the number of mistakes between vowels is indicated by different types of lines according to the signature. An arrow on the line points towards the incorrect vowel.

Fig. 1 shows confusions observed at the auditive presentation of list A (S/N = -20 dB). It follows that mistakes occur mainly between vowels with approximately identical first formant frequencies.

Fig. 2 shows that for the same words and the same S/N but with audio-visual presentation no confusions occur between rounded and unrounded vowels. This is in part surprising since no marked difference in lip configurations is present when pronouncing ʌ and α. The explanation seems to be that all words containing ʌ in list A also contain a bilabial stop consonant. This is an easily detected visual cue and therefore the wrong answers observed in the auditive test will not appear in the audio-visual test as they contain no such consonants.

TABLE 2

	LIST A (S/N = -20 dB)				LIST B (S/N = -10 dB)				LIST C (S/N = 0 dB)			
	NUMBER OF:		NUMBER OF:		NUMBER OF:		NUMBER OF:		NUMBER OF:		NUMBER OF:	
	PRES.	ANSW.	CORR. ANSW.	CORR. VOWEL	PRES.	ANSW.	CORR. ANSW.	CORR. VOWEL	PRES.	ANSW.	CORR. ANSW.	CORR. VOWEL
i	68 68	8 50	1 22	6 48	34 34	16 24	1 6	7 24	68 68	66 68	58 68	65 68
e	68 68	19 55	8 37	15 53	136 136	99 131	55 114	77 127	136 136	114 131	67 114	98 122
ε	170 170	34 144	14 124	23 139	136 136	84 127	42 108	62 124	102 102	97 102	87 101	91 102
æ	68 68	17 58	5 40	14 55	34 34	21 30	1 11	20 20	68 68	63 67	50 58	63 67
a	34 34	5 24	1 4	4 16	68 68	39 62	7 54	20 58	34 34	30 34	29 34	29 34
α	102 102	35 96	29 92	32 94	34 34	29 28	6 15	17 28	68 68	65 68	62 66	65 68
ɑ	34 34	3 16	2 5	2 6	34 34	22 31	17 30	22 31	68 68	67 68	62 68	66 68
y	34 34	8 25	2 10	3 14	- -	- -	- -	- -	34 34	34 34	34 34	34 34
ø	- -	- -	- -	- -	34 34	20 30	6 23	15 30	- -	- -	- -	- -
œ	- -	- -	- -	- -	- -	- -	- -	- -	34 34	31 33	28 32	28 32
Œ	- -	- -	- -	- -	34 34	27 33	16 30	24 33	- -	- -	- -	- -
u	68 68	5 38	0 27	0 31	68 68	53 66	46 60	48 61	34 34	23 32	17 31	19 32
o	34 34	8 28	2 21	5 25	68 68	59 67	47 62	57 67	34 34	33 34	31 34	33 34
ɔ	34 34	16 29	5 19	12 28	34 34	30 34	23 34	30 34	34 34	32 32	21 27	30 31
ʌ	102 102	26 94	13 76	18 93	102 102	74 101	58 96	67 101	102 102	99 99	81 91	90 99
ɒ	34 34	13 28	8 22	12 27	34 34	29 32	21 31	26 32	34 34	33 34	33 34	33 34

LIST A (S/N=-20 dB)

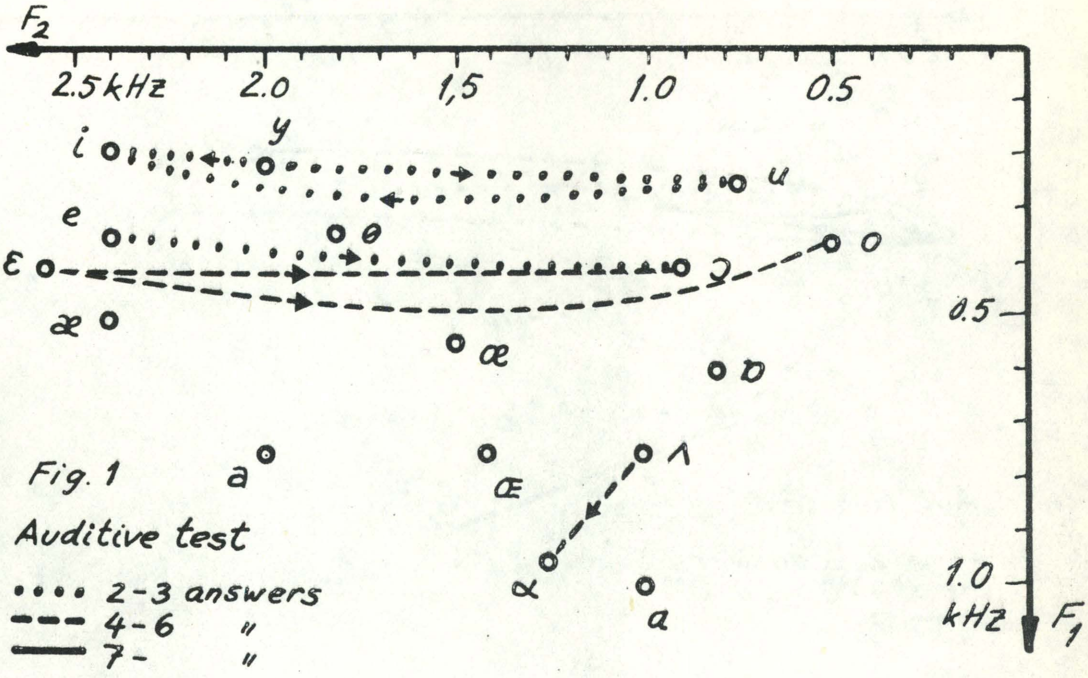


Fig. 1

Auditive test

LIST A (S/N=-20 dB)

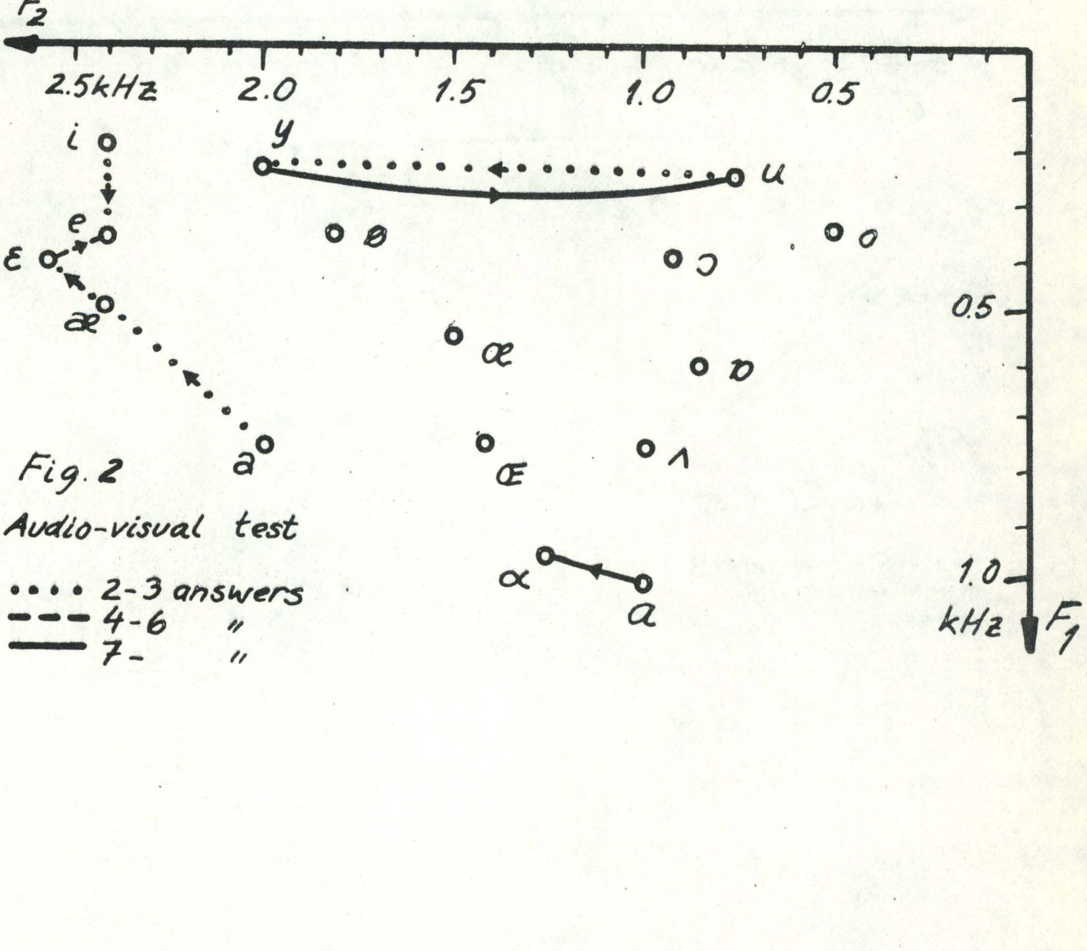


Fig. 2

Audio-visual test

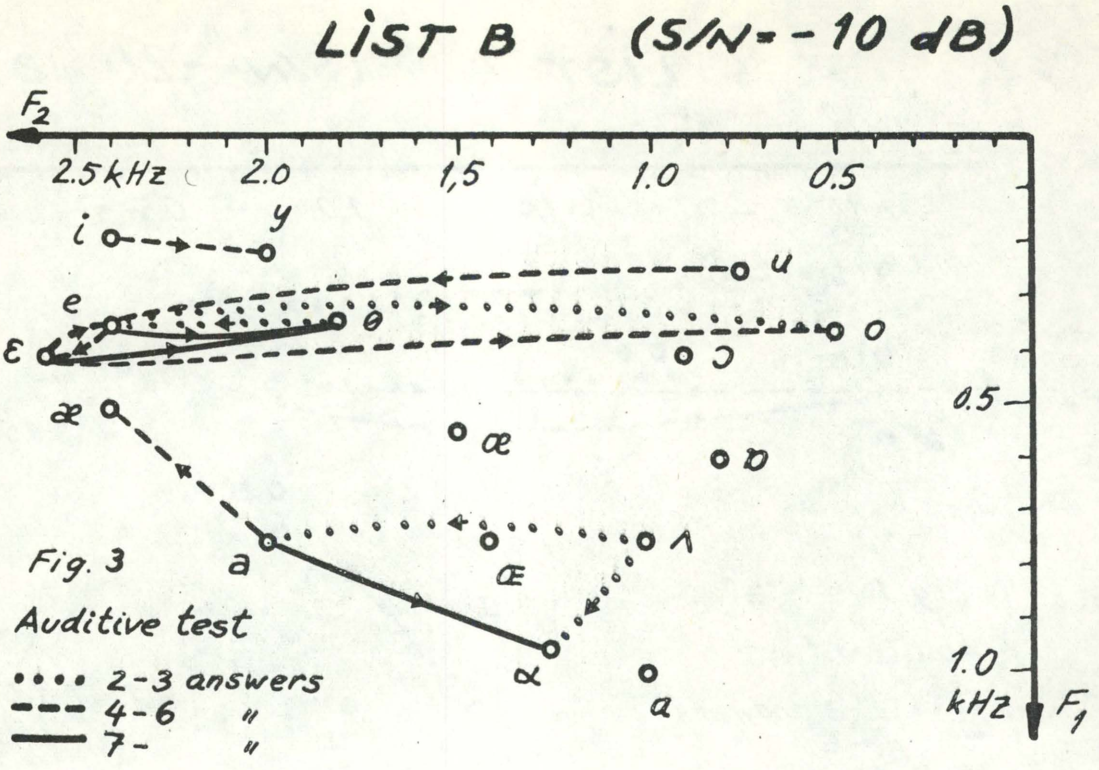


Fig. 3

Auditive test

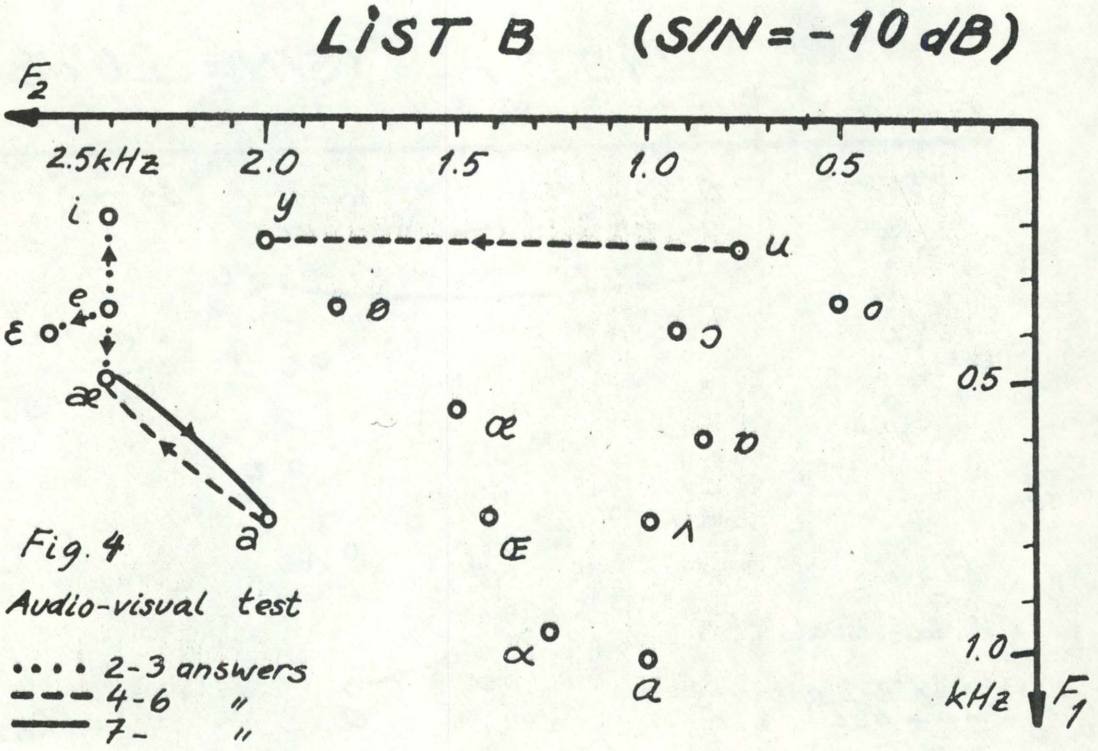


Fig. 4

Audio-visual test

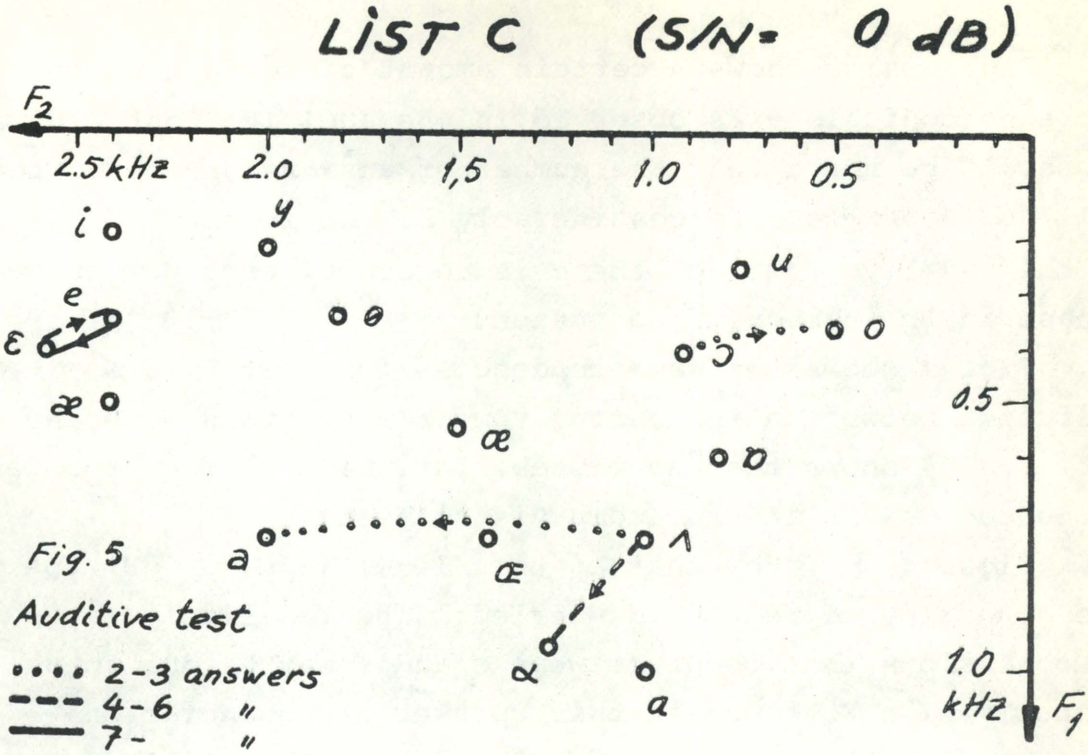


Fig. 5

Auditive test

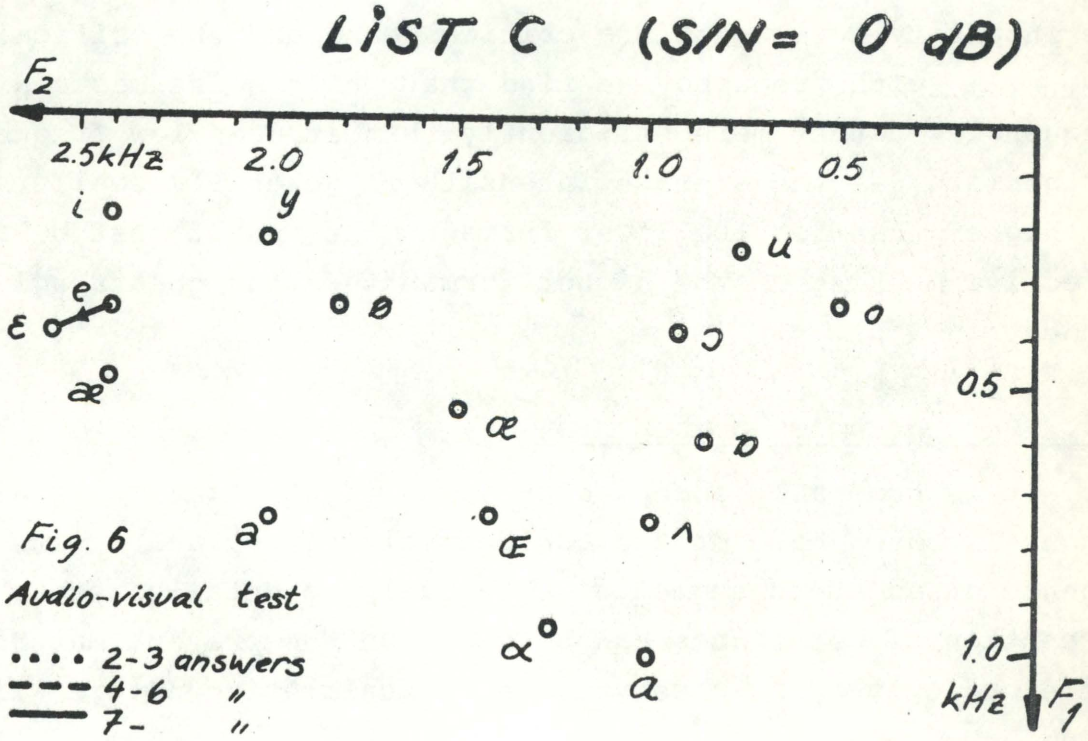


Fig. 6

Audio-visual test

Fig. 2 also shows a certain amount of confusion among vowels where no confusion was observed in the auditive test. However, it should be noted that the number of answers given (correct as well as incorrect) are considerably higher in the A-V test than in the auditive test and there is no significant percentual increase in the number of confusions.

Fig. 3 shows the same tendencies for list B (S/N = -10 dB): confusion between neighbouring vowels and vowels with the same F1.

Fig. 4 shows that no mistakes are made between rounded and unrounded vowels in the audio-visual test.

Figs. 5-6 show that at this level (S/N = 0 dB) the vowels are generally correctly identified. The relatively large number of confusions that occur between e and ε may be due to an exceptionally high F_1 found in this speaker's pronunciation of e.

No diagrams are shown for the test with list D (S/N = +10 dB) since no confusion between vowels was observed.

The results obtained here fit with the well-known masking proporties of white noise. As masking is roughly determined by the intensity level within a critical band and the critical bandwidth grows with frequency, we find that white noise masks high frequencies rather more efficiently than it does low frequencies. Furthermore, as the average intensity is generally smaller for the higher than for the lower formants the result must be an effective masking of the higher formants (although strongly dependent on S/N).

2.4. Perception of consonants

It is generally accepted that the identification of a consonant is based both on the consonantal segment and on its influence on adjacent segments. Obviously no detailed study of the perception of consonants can be based on the present material. Therefore, only a few observations of qualitative nature will be given here.

2.4.1. Labials

The most notable observation is the difference in discrimination of bilabials (p, b, m) as well as labio-dentals (f, v) found for the two types of presentation. While the discrimination of these consonants (especially that of f and notably for S/N $\leqq$ -10 dB) is remarkably poor in the auditive tests, the discrimination in the audio-visual tests is very high, even for S/N = -20 dB. Generally the audio-visual detection of f, v, and m (word initially) is almost perfect, i.e. even in the incorrect answers these consonants are always found in the correct positions when they were present in the stimulus word. The consonants p and b are often mutually confused for S/N $\leqq$ -10 dB but if a bilabial stop occurs in the stimulus word then a bilabial stop will be found even in incorrect answers.

This finding agrees well with earlier observations. The auditive discrimination between voiced and voiceless consonants in white noise is rather good (see e.g. Miller and Nicely 1954) and so is the visual detection of bilabials and labiodentals (see e.g. Woodward and Barber 1960). A generalization of these findings to audio-visual perception agrees with the above mentioned observations (remembering that Danish b is voiceless).

2.4.2. Voiceless fricatives

Another interesting observation is that voiceless fricatives are very hard to detect in white noise. This is not very surprising, but furthermore voiceless fricatives (especially s) as well as the affricated stop t occur in the incorrect answers where the stimulus word had no such consonants. This must be due to the pronounced similarity in acoustic quality between white noise and fricative sounds.

2.4.3. Voiced consonants

At a S/N = -20 dB wrong answers to stimulus words containing voiced consonants (m, n, ŋ, l, v, j, ð, r, ɣ) generally contain voiced consonants but with several confusions between them. At more favourable S/N the consonants are discriminated more accurately. The material gives no support to the theory that nasals are detected as a separate group as found by Miller and Nicely (1955). One reason for this could be that in American English vowels are generally strongly nasalized in nasal surroundings.

2.4.5. Stops

The stops in Danish are all voiceless and the difference between b, d, g and p, t, k is mainly one of aspiration (t is also somewhat affricated). The material from the auditive tests gives no indication that confusion within these groups[3] should be more likely than confusion between the groups or even with other consonants.

3. Conclusion

The results obtained in this paper show in agreement with earlier findings (see e.g. O'Neill 1954, Sumby and Pollack 1954) that the visual signal of a speaker's face considerably improves the detection of certain speech segments especially when the signal to noise ratio is unfavourable. The improvement is particularly conspicuous in the detection of bilabials and labio-dentals but also in separating rounded vowels from unrounded. The results are obtained from a discrimination test of isolated words and it may be expected that the influence of the visual signal is less pronounced in the perception of running speech, since e.g. the syntactical structure of preceding strings will make detection of certain segments redundant. And furthermore the articulation will generally be less distinct.

[3] viz. the group b, d, g and the group p, t, k.

It is also shown that the masking of white noise is not uniformly distributed over the frequency range and that the masking is one of higher frequency components mainly. Thus if the background noise is to give approximately the same decrease in redundancy for all components of the speech signal, another type of masking sound with less intensity in the higher frequency region must be used.

References

Ewertsen, H.W., Nielsen, H.B., and Nielsen, S.S. 1970: "Audio-visual Speech Perception", Acta Otolaryngologica 263, p. 229-230.

O'Neill, John J. 1954: "Contributions of the Visual Components of Oral Symbols to Speech Comprehension", Journal of Speech and Hearing Disorders 19, p. 429-439.

Miller, G.A. and Nicely, P.E. 1955: "An Analysis of Perceptual Confusions Among Some English Consonants", Journal of the Acoustical Society of America 27, p. 338-352.

Sumby, W.H. and Pollack, I. 1954: "Visual Contribution to Speech Intelligibility in Noise", Journal of the Acoustical Society of America 26, p. 212-215.

Woodward, M.F. and Barber, C.G. 1960: "Phoneme Perception in Lipreading", Journal of Speech and Hearing Research 3, p. 212-222.

SOME REMARKS ON ACOUSTIC PARAMETERS IN SPEECH DISORDERS

Nils H. Buch[1]
Børge Frøkjær-Jensen

1. Introduction

This paper and the following one deal with automatic sampling of acoustic data derived from speech. The actual project is in a preliminary stage.

The first paper presents some reflections about the extraction of the acoustic parameters and the applications to different voice disorders. The next paper describes the data collecting system for the investigation.

1.1. Diagnostics based upon recordable criteria?

Within the fields of hearing pathology the diagnostics has for several years been based on so-called objective criteria - first of all on audiometry. Otologists have in this way been fortunate in possessing methods that could be used for testing the patients, e.g. by means of the patients' responses to given stimuli. Audiometry has been used in its present form for more than 20 years and is now a well-known and well-established method for routine testing of hearing.

In the last few years this has been supplemented with a great deal of investigations on the distinct characteristics of different functions of the ear. Measurements of the impedance in the tympanic membrane, the airpressure, tone-decay, recruitment, etc., and especially the development within the

[1] Dr. med. Nils H. Buch is in charge of the Phoniatrics Laboratory, ENT Department, of the University Hospital, Copenhagen, and lectures on speech disorders at the Institute of Phonetics.

field of the ERA-audiometry[2] has placed audiology at a central position in modern research.

Within the field of speech pathology it is much more difficult to set up recordable criteria for the defects. The phoniatrics and logopedics still base their diagnosis mainly on subjective estimates combined with laryngoscopic examination by means of the larynx mirror - in recent years to some extent combined with a stroboscopic light source. This poses few problems for the phoniatric doctors who have trained their ears for years, but the method is very uncertain for doctors or logopedics without such experience and without knowledge of the auditive impression of the different speech disorders.

In fact, typical changes in the acoustic speech spectra are often the sole symptoms of many different voice or laryngeal diseases - at least in the initial stage.

These acoustic changes of the speech spectrum have given rise to a very confusing terminology, and such denotions as weak, breathy, noisy, harsh, husky, hoarse, shrill, dull, etc. have been used in the literature for different voice qualities, in most cases without any definition of the words. Furthermore, these characteristics are mostly based upon subjective descriptions of the human voice qualities. The terms do not refer to well-defined standards such as synthetic vowel spectra, which could be very useful in this case. However, in the modern phoniatric examination etiological diagnosis is used in combination with the acoustic/auditive description.

It seems that an international terminology is needed (Sonninen and Damsté 1971, Wendahl 1963, 1966, Isshiki 1966), and that this terminology must be based on recordable characteristics in the acoustic spectrum. This demands a basic

[2] The ERA-audiometry is a recordable and objective registration consisting of a summation of the cerebral electric potentials which occur after repeated acoustic stimulation of the ears.

research connecting the different morbi acoustically. It is notorious that this could be done, as proved by the fact that the experienced doctor or speech therapist is able to base much of his diagnosis upon the auditory impression. Moreover, a diagnosis based on analyses of the speech spectrum would be easy to carry out for the nurse and would not be uncomfortable for the patients.

1.2. Phoniatric and logopedic research based on phonetics

The methods and instrumentation of modern experimental phonetics in many respects provide a good frame for applied phoniatric and logopedic research. The phonetician deals with basic research on normal and healthy voices, and he is therefore able to supply the research on voice disorders with his experience and with the necessary reference material for pathologic research. In order to get a sufficiently good reference frame, a great deal of acoustic data of normal speech are needed before it is possible to deal with the acoustics of dysfunctions. It seems to be one of the tasks for the phoneticians to provide doctors and speech therapists with this normative material.

It is obvious that the methods and instrumentation which are used in acoustic analyses of normal voices could be very useful in the analysis of the pathological voice quality as well. During the last years such analyses have been made (Smith 1961, Wendahl 1963, Lauritzen and Frøkjær-Jensen 1970). However, as the measurements normally have been made by hand from curves such as sonagrams and mingograms, the analyses have been of very little use in the clinical situation.

The first demand on a phonetic/acoustic analysis used for the clinical diagnostics must be that the analysis can be carried out automatically as a routine investigation.

The second demand must be that the results must be available shortly after the microphone recordings have been made. Furthermore, it would be an advantage if the results are presented in a form which is easy to handle in the total clinical estimate.

1.3. Collection of data for the comparative analysis

As mentioned above it is necessary to collect a normal material which is adequate for the comparisons between pathologic and normal voices. On this basis it will be possible to compare the acoustic parameters collected from the different voice disorders with those of normal voices.

With reference to the above-mentioned paper (Lauritzen and Frøkjær-Jensen 1970) it is furthermore of practical interest for the value of logopedic therapy to compare the results before and after voice training.

2. The change in voice quality for some typical disorders

The planning of this project has been based upon the following specific speech disorders: vocal mutation, laryngeal paralysis, voice disturbances after androgen therapy (treatment of women with masculine hormons), vocal nodules, cronical laryngitis, and psychogenic dysphonia. The treatment of these disorders can be either pedagogical, medical or operative.

2.1. Vocal mutation

In this physiologically conditioned type of dysphonia the intonation range is normally reduced, and shifts between the voice registers happen very often. The phonation is in many cases pneumophonic.

2.2. Laryngeal paralysis

The voice of patients with a paralysis, most commonly caused by damage of the recurrens nerve, sounds airfilled, noisy, soft, weak, and hoarse. Occasionally the voice is unstable with a tendency to diplophonia. If such a voice is trained the quality will change to a brighter, more modulated, and less noisy one.

2.3. Androgen damages of female voices

This type of artificial dysphonia causes a lowering of the mean fundamental frequency, reduces the intonation range, and makes the voice sound rough and coarse like a man's voice. Observation of a glottal transverse insufficiency - just as in boys with mutation dysphonia - is characteristic of this disorder.

2.4. Vocal nodules

Bilateral benign tumours on the vocal folds are in most cases observed in children. In adult voices it most commonly is found in connection with hypercompression, - especially in the voices of singers and professional speakers.

The normal symptom is a gruff voice quality which is poorly equalized. Special difficulties arise during phonation in the middle register.

2.5. Chronical laryngitis

The symptoms of this group of voice disorders are mostly differentiated and cause difficulties when making the precise differential diagnosis.

In some cases the vocal folds are more or less oedematous and injected, now and then with hyperkeratosis and uneven edges. The voice quality changes very much and is often dominated by variations in compression, combined with pneumophonia and bad equalization.

2.6. Psychogenic dysphonia

In order to supplement the analysis with an examination of a characteristic non-organic type of dysphonia we have chosen the psychogenic dysphonia (functional dysphonia).

There is never any agreement between the laryngoscopic observation and the distinct changes in voice quality. It is typical for the psychogenic behaviour of these patients that they are not motivated for the clinical functional tests. It is therefore of great interest to include this group in the project. This means that the material thus enables us to compare the parameters from the normal voices with both organic and non-organic voice disorders.

3. Acoustic parameters for different voice qualities

As mentioned in 1.1. the auditive terminology of the different voice qualities is rather confusing. However, it seems that two or three main groups can be separated.

3.1. The weak, noisy, and breathy voice

At the physiological level this phonation is characterized by a more or less continuous airflow during the entire vibratory cycle, and no glottal closing phase can be observed. This is undoubtedly a result of either an insufficient medial compression of the arytenoid cartilages or a lateral excursion of the vocal folds.

At the acoustic level the escaping air generates noise especially at frequencies above 3000 Hz, and due to the lack of a glottal closing phase the acoustic vowel spectrum has very weak harmonics which reduce the levels of the higher formants. The first harmonic is very prominent.

The auditive impression is a breathy and noisy voice quality with a dull and dark timbre mainly caused by the lack of higher harmonics, but in some cases also caused by a lowering of the fundamental frequency. The voice is weak and the intelligibility is bad.

It seems quite natural that this voice quality appears particularly in paralyzed voices.

3.2. The harsh and rough voice quality

This voice quality has its origin in a disturbed vibration of the vocal folds. According to recent investigations (Wendahl 1963 and 1966, Isshiki et al. 1969, Lieberman 1963) the duration of the glottal vibrations changes from period to period even for a vowel phonated at a constant fundamental frequency. This irregularity of the vocal fold vibration is often due to excessive tension of the folds (Zemlin 1969). Furthermore, the harsh quality is combined with frequent use of glottal attack and often with a slightly lowered fundamental frequency.

The harsh or rough voice quality seems to dominate in voice disorders such as the different forms of vocal nodules and in some forms of laryngitis.

3.3. The hoarse voice quality

According to Isshiki et al. (1969) the auditively based term "hoarseness" comprises at least four auditive parameters, including breathiness and harshness.

Acoustically it manifests itself as a combination of noise and lack of harmonics in the upper part of the spectrum, and as a continuous fluctuation in periodicity (in extreme cases no pitch can be registered at all).

Physiological changes which occur (among other things) during the mutation, or changes caused by neoplasms or by vocal nodules and laryngitis, are the main causes for this change in vocal quality.

4. The temporal relations in pathologic speech

Apparently very little research has been made on the altered temporal relations in pathologic speech. It is a clinical experience that both the length of vowels and consonants and the length and number of the inspiratory pauses are extended in different voice disorders.

Normally the laryngeal paralyses will result in a greater consumption of air caused by the insufficient glottal closure. This means that the respiratory frequency is increased which in turn reduces the length of the voiced passages (phonation groups). An increased consumption of air may also be registered e.g. in laryngitis and often in psychogenic dysphonia. The function of the glottis when affected by recurrent paralysis will cause a lengthening of the transition time between the sound segments.

It is obvious that the changes in the temporal relations could be used as a tool for describing the different voice disorders, too. In this first trial we have therefore decided to use the length of the continuous voicing as a parameter (see 6.3.).

5. The most important acoustic parameters

5.1. The fundamental frequency

Information about the fundamental frequency may be obtained by an automatic scanning of the output voltage from a fundamental frequency detector. The collected data contain information about the lowest, the highest, and the mean fundamental frequency, the intonation range, the frequency and the range of falling/rising changes in the intonation, and the way in which these changes occur (e.g. continuously or in jumps).

If the automatic scanning is synchronized with the vocal fold vibrations, the parameter also contains information about variations in periodicity.

5.2. The relative intensity of the higher harmonics

An automatic scanning of the relative vowel intensity above 1000 Hz seems to be a good parameter for the amount of energy in the higher part of the spectrum. ("Relative" means relative in relation to the total energy at the moment of measurement.) A reduced energy of the harmonics in the F_2-F_3 region in relation to the F_1 region will result in reduced intelligibility. (I. Lehiste and G. Peterson 1959, S. Smith 1961).

The parameter contains information about the strongest and weakest relative intensity, the mean relative intensity, and the dispersion of relative energy above 1000 Hz for all the voiced sounds in a spoken text.

5.3. Duration of the phonation groups

We define a phonation group as a group of voiced segments, i.e. the phonation groups are interrupted by pauses and unvoiced consonants. As previously mentioned several phonation disorders are characterized by too short phonation groups caused by an insufficient closure of the vocal folds or lack of ability to close the folds. This parameter thus gives very valuable information about the ability to close the vocal folds and to keep the folds in the phonatory position. Indirectly, it gives information about the respiratory frequency.

Information derived from that parameter consists of measurements in milliseconds for the longest, the shortest, and the mean duration of the phonation groups together with details concerning the distribution of the durations in a spoken text.

6. How to extract the parameters

6.1. The fundamental frequency

A purely automatic scanning of the fundamental frequency will register a frequency of 0 Hz in all pauses which is of no interest in the further treatment of the data. Therefore the data collecting system should be stopped when the voicing stops. The data collection is under supervision of a "voice indicator" (see the next paper 3.1.) which starts the data collecting procedure 25 ms after the voicing starts, and stops it immediately when the voicing stops. The "voice indicator" is operated by the intensity level of the fundamental frequency. The speed of the data collection is set by a "pulse generator" (see next paper 3.2.) which is normally triggered by the mains supply at 50 Hz. The data are either stored on a punched paper tape (see next paper 3.3. and 3.4.) or fed on-line into a computer. The punched paper tape is an appropriate way of storing the data if the system is to be used in a phoniatric clinic without EDP-possibilities. The on-line processing calls for a small computer[3] connected with the data collecting system. In this way the EDP-output will appear immediately after the patient has been recorded, which is a great advantage in a routine analysis. By means of appropriate software the EDP-system will handle the data, and the output will primarily appear as histograms with standard deviations showing the distribution of the different fundamental frequencies in a spoken text.

6.2. The relative intensity of the higher frequencies

The relative measure of the intensity above 1000 Hz calls for two intensity registrations: (1) total intensity and (2) intensity above 1000 Hz.

[3] Our laboratory has just ordered a computer, which will be used for on-line run of this project, as well as for other purposes.

The relative intensity above 1000 Hz in relation to the total spectral energy is expressed as:

$$\frac{I_{hp1000}}{I_{total}}$$

Electronically these calculations can be made by substracting the logarithmic value of the denominator from that of the numerator:

$$\log \frac{I_{hp1000}}{I_{total}} = \log I_{hp1000} - \log I_{total}$$

If, therefore, the two intensity measures are represented as logarithmic voltages (i.e. in dB), these voltages can be subtracted from each other in a differential amplifier, the output of which thus represents the relative intensity above 1000 Hz measured in dB.

The further treatment equals the treatment of the output voltage from the fundamental frequency meter (see 4.1. in this paper and 3.2. in the next paper). Ideally the fundamental frequency and the intensity should be registered simultaneously. However, due to the electronic circuitry there will be a short delay of 0.45 msec. between the registration of the two parameters (see Fig. 2 in the next paper). On the punched paper tape each fundamental frequency measurement will be followed by an intensity measurement.

The final representation in the EDP-output equals the representation of the fundamental frequency parameter, i.e. as histograms showing the distribution of the energy above 1000 Hz.

6.3. Duration of the phonation groups

The above-mentioned "voice indicator" registers when voicing is present, and in the "voice duration meter" (see next paper 3.3.) the duration of the phonation groups is converted to an electrical voltage, which is fed via the multiplexer and the A/D-converter to the puncher. The system then enables us to register 120 different durations (see next paper 2.3.).

On the punched tape the duration data are separated from the two other parameters by means of "boundary markers" generated by the "voice duration tape coder" (see next paper 3.4.).

The EDP-output represents the durations in the form of easily comparable histograms.

6.4. Resolution of the parameters

Considerations about the resolution of the parameters indicate that the registration of the fundamental frequency (6.1.) calls for about 100 decimal values per collected data and the resolution of the relative intensity registration (6.2.) requires about 60 values per data. The resolution of the time parameter (6.3.) depends very much upon the voice disorder considered, but again 100 values seem to satisfy our demands.

7. The value of non-acoustic tests

The present lack of empirical data concerning the acoustic changes in voice quality for different voice disorders justify a procedure which compares the above-mentioned acoustic parameters with the results of examinations made by the classical methods. We have planned to make complete diagnostic investigations at the University Hospital or at the Institute of Speech Disorders for all voices recorded, normal as well as pathological. In this connection it is our plan to make laryngoscopy and stroboscopy, to establish the total phonation range and the maximum duration of phonation, and according to requirements to supplement the results with the necessary examinations by means of X-ray, glottography, blood and hormon tests, etc. A profound anamnesis will be made of all patients in order to obtain the best possible classification. Only in this way it will be possible to get hold of and to get experience about the specific characterization of the acoustic parameters.

8. Some further acoustic parameters

This paper has been called "Some remarks ...". It is only meant as a description of a project which has just started. The three parameters which have been described in the paper seem, after pilot tests, to be the ones that contain the most relevant information. Naturally several other possibilities exist, and as the project proceeds we will test the utility of other parameters. Among these we only want to draw the attention to:

(1) the duration of the pauses in a spoken text,

(2) the relative intensity of the unvoiced consonants in relation to the vowels,

(3) separate analyses of the irregularities in periodicity, and

(4) correlation analyses of the intensity variations for the glottal cycles (Koike 1968).

References

Isshiki, Okamura, Tonabe, and Morimoto 1969: "Differential Diagnosis of Hoarseness", Folia Phoniatrica 21, 1.

Koike, Y. 1968: "Vowel Amplitude Modulations in Patients with Laryngeal Diseases", JASA 45,4, p. 839-844.

Lauritzen, K. and B. Frøkjær-Jensen 1970: "Comparative phonetic-acoustic analysis before and after speech therapy of voices suffering from recurrens paresis", ARIPUC 4, p. 145-165.

Lehiste, I. and G. Peterson 1959: "The Identification of Filtered Vowels", Phonetica 4, p. 161-177.

Lieberman, P. 1963: "Some Acoustic Measures of the Fundamental Periodicity of Normal and Pathologic Larynges", JASA 35,3, p. 344-353.

Smith, Svend 1961: "On Artificial Voice Production", Proc. of the IV. Intern. Congr. of Phonetic Sciences, Helsinki, p. 96-110.

Sonninen, A.,and P.H. Damsté 1971: "An International Terminology in the Field of Logopedics and Phoniatrics", Folia Phoniatrica 23,1, p. 1-80.

Wendahl, R.W. 1963: "Laryngeal Analog Synthesis of Harsh Voice Quality", Folia Phoniatrica, 15, p. 241-250.

Wendahl, R.W. 1966: "Laryngeal Analog Synthesis of Jitter and Shimmer, Auditory Parameter of Harshness", Folia Phoniatrica 18,2, p. 98-108.

Zemlin, W.R. 1969: Speech and Hearing Science.

CONSTRUCTION OF AN AUTOMATIC DATA COLLECTING SYSTEM

Poul Thorvaldsen

1. Introduction

In order to carry out the investigation described in the previous paper "Some Remarks on Acoustic Parameter in Speech Disorders" it is necessary to collect a great amount of data. Since the measurements evidently cannot be carried out manually with sufficient speed, the need for some automatic data collecting equipment arose.

As the development went on, it became clear that the system had to be generally applicable in order to comply with a wish for a "general-purpose" automatic data collecting system.

The system now consists of a basic general-purpose unit, to which are added module units designed for the above-mentioned investigation. Furthermore, there is a possibility of adding units for special investigations.

The system at its present stage is shown in the block-diagram (Fig. 1). The general-purpose system consists of the blocks fully drawn. The blocks drawn with dotted lines show the special units developed for the acoustic analyses of speech disorders.

2. The general-purpose system

2.1. Multiplexer, A/D-converter and tape-puncher

The general-purpose system is constructed around a multiplexer, an analog/digital converter and a tape puncher.

The multiplexer is manufactured by Analogic (type MUXPAC MP 4108) with 8 single or 4 differential inputs and a settling time of max. 5 μsec.

BLOCK - DIAGRAM.

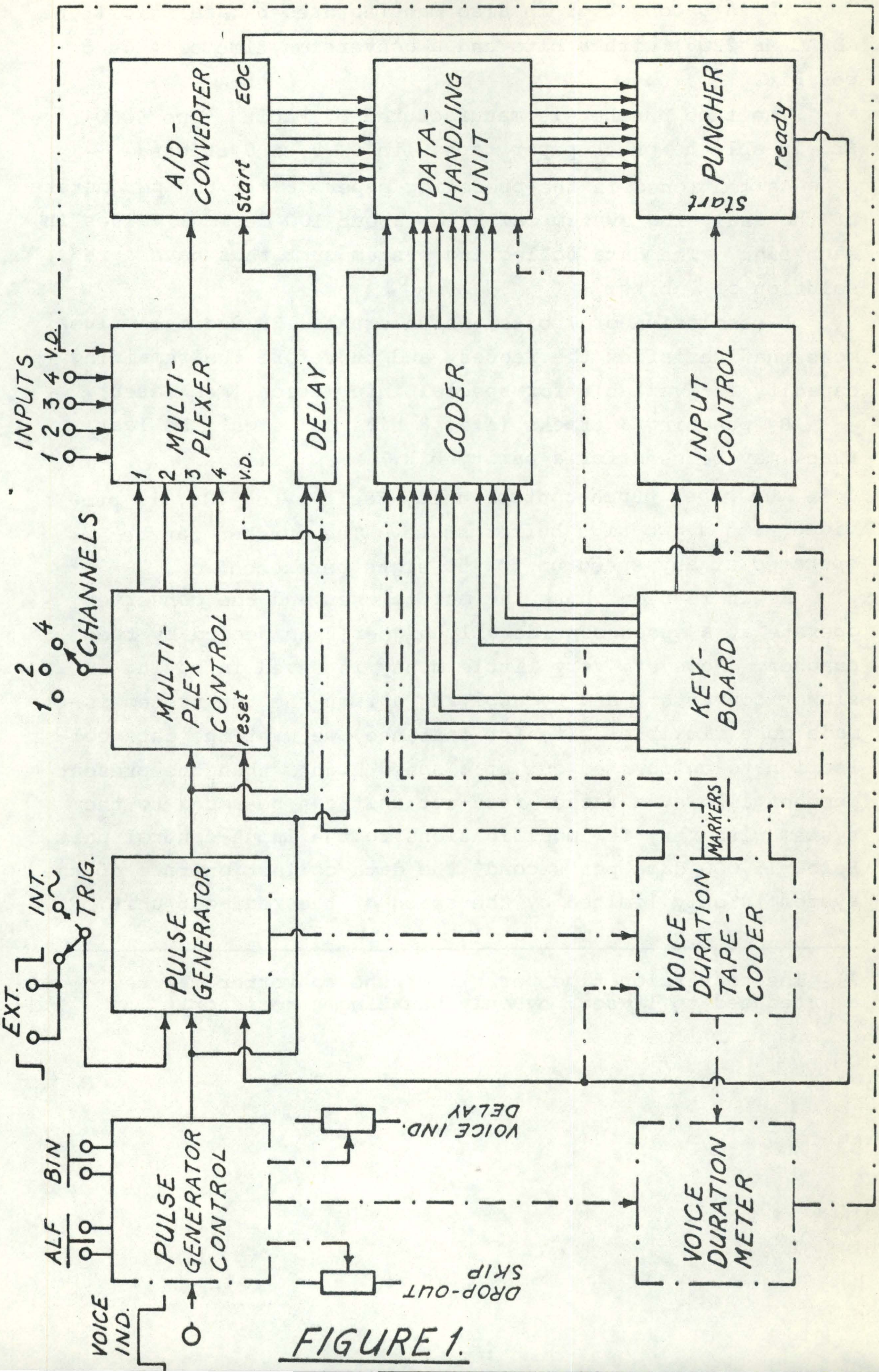


FIGURE 1.

The A/D-converter is also manufactured by Analogic (type ADPAC MP 2208) with 8 bits and a conversion time of 4 µsec. per bit.

The tape puncher is manufactured by Facit (type 4060) and is able to punch paper tape with 6, 7 or 8 tracks.

As mentioned in the preceding paper (6.4.) the parameters of the analyzing system calls for about 100 decimal values in each sign. The data collecting system must thus have a resolution of 7 bits.

A resolution of 7 bits, which equals 128 decimal values, more than satisfies the request and therefore the remaining capacity is available for special information (see later).

By punching 8 tracks (i.e. 8 bits per sign) the last track may be used for a parity bit.

The paper punch control unit, Facit type 5106, is provided with a two-sign buffer store. The puncher can be operated at any speed up to 150 signs per second.

As it is seen, both the multiplexer and the converter operate at a speed considerably higher than needed by the puncher. However, very little money is saved in buying a slower converter, and by applying a fast one the system is made more flexible. If, for instance, we want the data collection to be carried out at a speed higher than the present puncher allows, a faster read-out unit can be added to the system with very few modifications to the input-control unit. Below 75,000 data per second[1] the data collection rate of the system is only limited by the speed of the read-out unit.

[1] The conversion time per bit of the converter may be reduced to 1 µsec. by only shunting a resistor.

2.2. Keyboard and coder

When dealing with different kinds of data, it is useful to be able to put comments to the different data sets. To this end the system has been provided with a keyboard. From here it is possible to put the normally used alphanumerical signs on the tape.

The keyboard is combined with a coder which transforms the signs into a seven-bit code. With reference to the computer[2] which is to be used for the data processing an ASCII code has been chosen.

Furthermore the keyboard acts as a control panel. By means of the keys "ALF" and "BIN" it is decided whether the data-handling unit shall allow the alphanumerical data from the keyboard or the binary data from the A/D-converter to be fed to the puncher.

2.3. Data-handling unit

Besides transferring the signals from keyboard and converter to the puncher, the data-handling unit acts as a generator of special signs and parity bits.

Only data having values between 1 and 119 are passed directly from the A/D-converter. If the value is 0 or negative the data-handling unit will generate a sign of "underflow". Similarly, if the value is greater than 119 a sign of "overflow" will be generated.

[2] A Digital Equipment computer type PDP8/e with an 8 K central processing unit will be installed at the laboratory during the autumn 1972.

In order to facilitate the software work, the area between 120 and 127 is reserved for special signs. In addition to "underflow" and "overflow", the signs "ALF" and "BIN" and the later explained "voice duration boundary markers" are placed here.

As mentioned above a parity generator adds an eighth bit to the data before it is transferred to the puncher. Odd or even parity can be chosen at discretion.

2.4. Additional units

The general-purpose function of the remaining units is quite trivial.

When the pulse generator control is set to binary-mode (automatic data collection), the pulse generator generates a pulse for every positive or negative triggering edge applied to the "EXT. TRIG." input. This pulse is generated on condition that a "ready"-signal has been received from the puncher. This happens when the average frequency of every 3 triggerings does not exceed 150 per second.

The output from the pulse generator is fed to the multiplex control, which chooses among the multiplexer inputs which one should be fed through. The possibilities are: input 1 alone, input 1 and 2 alternately, and input 1, 2, 3 and 4 successively. The "RESET" input of the multiplex control ensures that input 1 is always the first one to be activated when the measuring-mode has been started.

The output of the pulse generator also starts the A/D-converter. This is done via a delay corresponding to the settling time of the multiplexer. When the conversion has been performed an EOC (end of conversion) signal is sent to the input control, which starts the puncher.

3. The special-purpose system

3.1. Pulse generator control

In the special-purpose mode the pulse generator control is managed by a "voice-indicator", in fact a voltage which is a logical "1" for voiced segments and a "0" for voiceless segments. This management means that the pulse generator is running only when the voice indicator is a logical "1".

In order to avoid that short drop-outs in the voice indicator may interrupt the pulse generator, the pulse generator control is provided with a "drop-out skip" facility. Such drop-outs in the voicing may be caused by articulatory reasons as for instance a rolled "r" or a flapped "d". The maximum duration of the drop-outs to be skipped is adjustable in the range 5-50 msec.

As it is of no interest to make measurements in the beginning of a voiced segment, e.g. during a voiced stop consonant, the onset of the pulse generator is delayed. The delay is adjustable.

3.2. Pulse generator

The internal triggering of the pulse generator is operated from the mains supply (50 Hz). In order to make nearly synchronous measurements of fundamental frequency (channel 1) and intensity (channel 2), two pulses for the multiplex control are generated for every triggering. The duration of every pulse is 300 µsec. and the interval between them 150 µsec. It is impossible to make more than two nearly synchronous measurements at a time, as the average frequency of every three triggerings of the puncher must not exceed 150 per second.

Figure 2 shows how the sampling of the inputs is organized in the mode of 1, 2 or 4 sampled channels.

SAMPLING OF INPUTS.

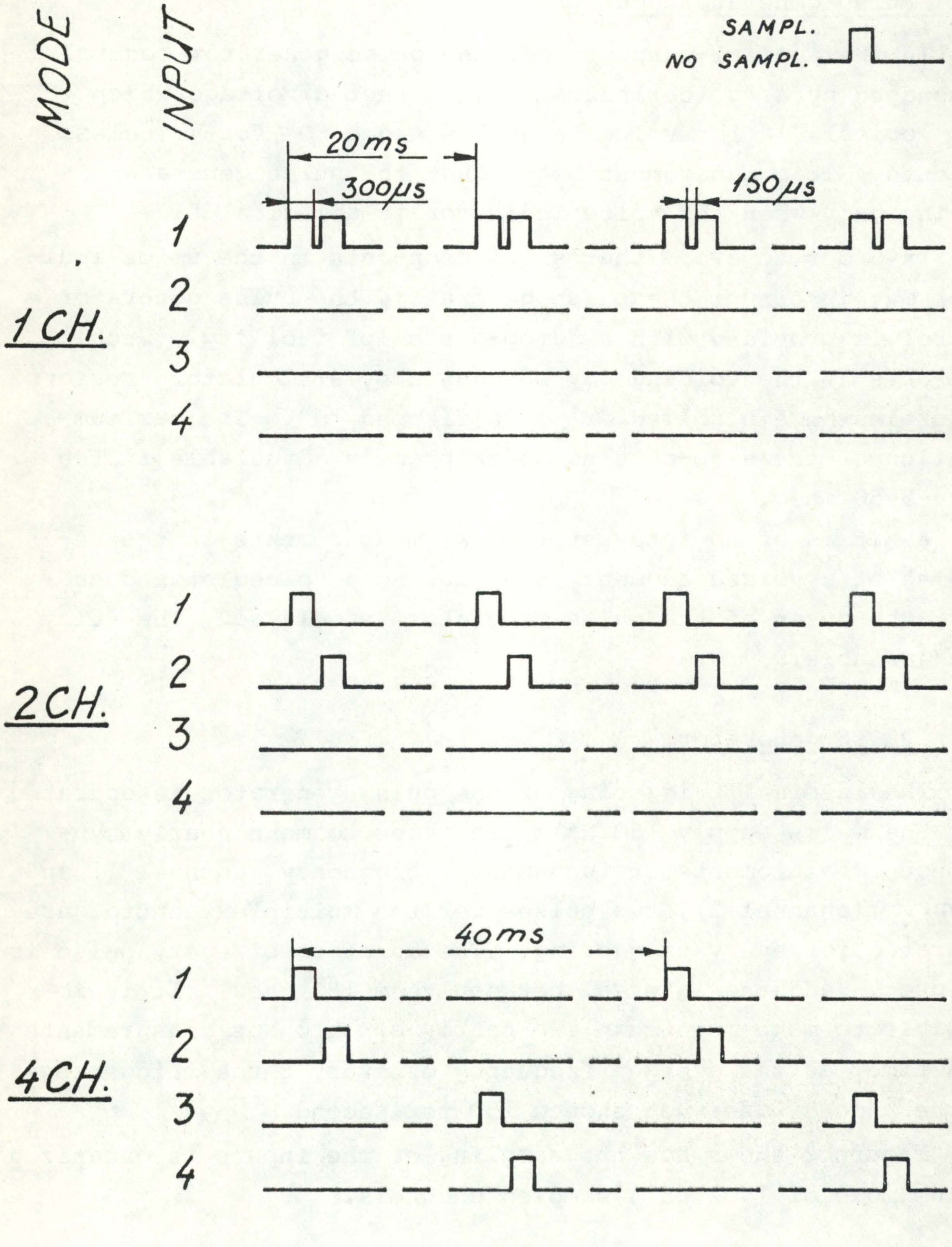


FIGURE 2.

3.3. Voice duration meter

As length of phonation is a significant parameter of the present investigation, special circuitry has been constructed to measure and read out this quantity.

The voice duration meter contains a controlled integrator charged by a step voltage synchronized with the voice indicator. So the output is a voltage proportional to the length of phonation. The integrator is started when phonation starts and is reset when a signal from the voice duration tape-coder indicates that the output has been read out. The signal is fed through a fifth channel of the multiplexer controlled by the voice duration tape-coder.

3.4. Voice duration tape-coder

Since voice duration data by itself in no way differ from other data on the tape, it is necessary to mark them in some way. This is done by the voice duration tape-coder, which puts a "boundary marker" in front of and behind the measurement. The boundary markers are special, easily recognizable signs placed in the area between 120 and 127 as previously mentioned.

The voice duration tape-coder is controlled by an "end of voicing"-detector. When the pulse generator has been idle for a period slightly longer than 20 msec. the detector causes the first boundary marker to be generated. The boundary marker is followed by a signal permitting the output of the voice duration meter to pass the fifth channel of the multiplexer (V.D. in the block-diagram), and then the generator of the second boundary marker is triggered by the next ready-signal from the puncher.

As a part of the procedure the data-handling unit and the input control are of course manipulated to select the proper inputs.

This pretty troublesome procedure is necessitated by the wish for a fast read-out routine in conjunction with the demand of ensuring that both data and markers are accepted by the puncher.

Acknowledgement

The work described above has been carried out in cooperation between the Institute of Phonetics, The Phoniatric Laboratory of the University Hospital, and the State Institute of Speech Disorders. The development has been financed by Grants from The Danish Research Council for the Humanities (Statens humanistiske Forskningsråd).

ON BRITISH ENGLISH p-, b-, sp- and -s + b-

Henry Petersen

1. Introduction

1.1. The problem

The problems connected with the clusters consisting of s + stops in English must be intriguing indeed, since so many linguists (and educationalists, for that matter) have done an almost incredible amount of work to solve them. The attention of the present writer was also engaged, and the result is the article below, which, it is hoped, will throw some light on the articulatory conditions of p-, b-, sp- and -s + b-.

One of the pedagogical questions facing Danish teachers and learners of English is whether our pronunciation of written sp- [sb̥] can be transferred to English without modification. Without unduly anticipating the conclusion of this article it may be said that the answer must be in the negative.

Otto Jespersen (1958, p. 51) states "that the English sounds [p], [t], [k] remain practically unaltered when preceded by [s]." His view is supported by Daniel Jones (1962, p. 139), who says that "Scandinavians also have a tendency to replace p by b̥ when it occurs ... after s as in spend ...".

A.C. Gimson (1964, p. 146) finds that "when /s/ precedes /p, t, k/ initially in a syllable, there is practically no aspiration, even when the sýllable carries a strong accent, cf. pin [p^{h}In] and spin [spIn] ..." and (not quite consistently, perhaps, p. 48) that "when /p, t, k/ follow an initial /s/, ... they are realized with no aspiration even when stressed". These phonetic (auditory?) facts are interpreted to the effect that, phonemically, the words spin, steam, scum may be trancribed either as /spIn, sti:m, skʌm/ or as /sbIn, sdi:m, sgʌm/ without ambiguity. Providing that Gimson's phonetic basis is sufficiently broad, the latter transcription is, phonologically, as sound as the former, of course.

The first incentive to inquire into these things came 5 or 6 years ago from a discussion with Professor Poul Steller, of Danmarks Lærerhøjskole. He maintained the views of Jespersen and Jones, perhaps with some modification. I agreed.

The second impetus was given by N. Davidsen-Nielsen (1969a 1969b),whose experimental procedures and the phonological results derived from the experiments appeared to me to be methodically unsound, involving, it would seem, some confusion of the phonetic and phonemic levels.

The achievement of the object outlined above rendered necessary not only an experimental investigation of the articulation of *p*-, *b*-, *sp*- and -*s* + *b*-, but also a statistical analysis of the varying lengths of these sounds.

1.2. Linguistic material

The linguistic material consisted of two groups of sentences called Ia, Ib, Ic, Id (40 sentences) and II, 1a; II, 1b; II, 1c; II, 1d; II, 2a, etc. (440 sentences). The number 480 was not attained, however, partly owing to miscounts, partly on account of inferior quality of a few of the recordings.

Group I comprised four sentences with the intonation indicated, each spoken ten times in succession:

Ia I ˈswallow a ˈbitter ˎpill today
Ib I ˌswallow a ˋbitter ˌpill today
1c I ˌswallow a ˌbitter ˋpill today
1d I ˌswallow a ˌbitter ˏpill?

The object of the analysis of group I was to find out whether the realizations of *p* and *b*, respectively, in the same phonetic surroundings, but in different tonetic groups differ in respect of measurable characteristics to such a degree that they must be considered as belonging to the same/different statistical population(s). Furthermore - as in Frøkjær-Jensen et al. (1971) - there is an indication of the position in time of the plosion relatively to the glottal gesture, if any. In this way it was hoped that an exact assessment of the occurrence of aspiration, if any, especially in the sp- cluster, would be rendered possible.

Group II comprised 11 sub-groups, each of 4 sentences. The sounds under investigation were said within the frame of a carrier sentence, word-initially, and followed by the eleven "pure" English vowels, in the following order:

II, 1a I say ˋpeak today II, 1b I say ˋspeak today
II, 1c I say ˋbeak today II, 1d 'Jack's 'beak is ˎred

The other words were: pill, spill, bill, Jack's bill (sub-group 2), pell, etc. (sub-group 3), pan (sub-group 4), par (sub-group 5), pore (sub-group 6) pot (sub-group 7), pun (sub-group 8), purr (sub-group 9), pull (sub-group 10); in this group was included the only nonsense word used: spull [spʉɫ]), poon (sub-group 11). Each sentence was spoken 10 times in succession in the order given above.

From Group II were selected 3 sub-groups, 1, 2 and 9, with the object of elucidating the articulation of p-, sp-, b- and s- + b- (114 sentences). Furthermore, sentences from sub-group 5 were selected and compared with sentences from sub-group 1 with a view to finding out whether in comparable phonetic and tonetic surroundings the mean length of consonants preceding vowels with maximum difference in the degree of opening is influenced significantly by this difference.

1.3. The subject

The subject was an Englishman, who gave the following account of his accent and educational background: 26 years of age, B. A. English from Cambridge. M. A. Applied Linguistics - Essex University. At present completing Ph. D. in linguistics.

Accent: Near R. P. Southern English imposed upon South Wales English. Very little remaining of previous dialect, except in the case of, e. g., orthographic <u>flew</u>, <u>blew</u> [flIu, blIu].

According to this information the subject - hereafter referred to as P. M. - is assumed to be representative of his linguistic community. Obviously, the results of investigations of such limited scope as the present experiment are in principle of highly limited validity because the material offers no possibility of estimating several important factors. It is hoped, nevertheless, that inferences drawn from the sentences recorded may have a certain general application as regards the (social) group of speakers of R. P.

1.4. Instrumentation and comments on the set-up

The output of the Fabre Glottograph was tapped with a time constant of 4 seconds. The input to the transpitchmeter (settings: LP filter 120 Hz, HP filter 70 Hz) was taken from the Fabre Glottograph, consequently the output curves were physiological. The transducer of the Photoglottograph was inserted through the subject's nose and fixed in the oesophagus. An aerometer was placed before the mouth, and a Revox Recorder (tape speed 7.5"/sec.) was connected to the microphone while Group II was being recorded.

The speed of the recording paper was in all cases preset at 10 cm/sec. As, however, the speed of the mingograph was not exactly 10 cm/sec., but 10,25 cm/sec., all figures are divided by 1.025, the results thus obtained being very near approximations to the true time.

No doubt one would be to some extent justified in reviving the objections of the classical phoneticians against experimental phonetics (Eli Fischer-Jørgensen (1962), p. 115 and (1957), p. 435). Though P. M. as a subject was all that could be desired, unaffected psychologically, calm and patient, even his speech was influenced by the apparatus to such an extent that it cannot be said to be natural. This is particularly the case with the nasals, which sound as if at the time of recording the subject had a cold in his head.

1.5. Introductory remarks on the results from group I

p and b were chosen as objects of investigation for the same reasons as adduced in the above-mentioned article by Frøkjær-Jensen et al.: "The study is limited to labials, partly because this simplifies the problems since there is no affrication present, and partly because the choice of labials minimizes the risk of artifacts in the glottogram stemming from tongue movements" (p. 123).

Artifacts occur, nevertheless, on account of the movements of the back of the tongue, most perceptibly in open back vowels. Here the articulatory movements may cut off the light from the Photoglottograph to such an extent that glottal gestures leave no traces in the mingogram. Besides, there is a very marked artifact in most of the n´s.

Both p and b are word-initial and intervocalic. Their surroundings are chosen so as to be very nearly identical, viz. allophones of [ə] and [ɪ], whereas the intonation of the sentences varies. In these circumstances, such significant differences in length as might be established between the means of the four groups of b´s and the four groups of p´s respectively must be ascribable to varying stress and intonation.

This assumption is the basic idea of this article.

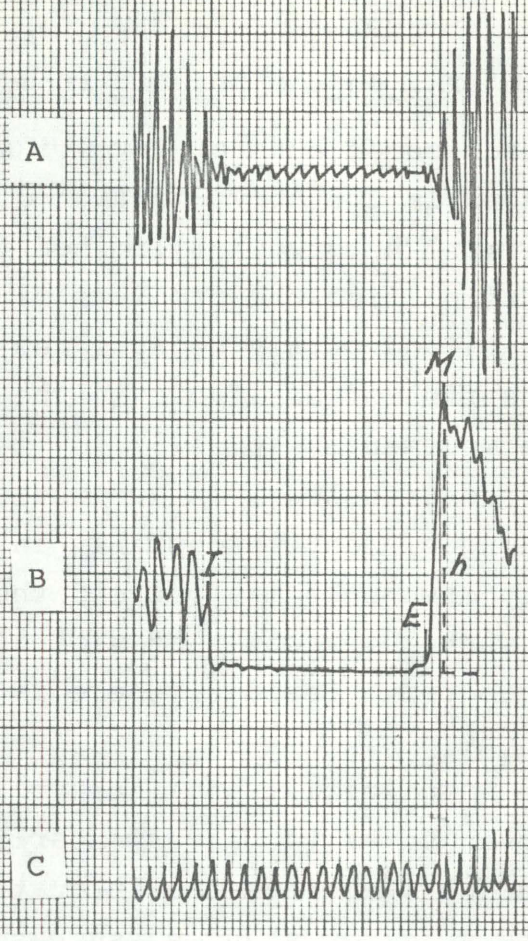
2. Measurements of b

2.1. Definition of the quantities measured in the sound b

Figure 1 illustrates the Photoglottograph and aerometer curves and the microphone oscillogram.

The "dip" of the curve between the points I and M is hereafter called the trough (also in p).

Fig. 1: parameters in b



A: Microphone oscillogram

B: Aerometer curve

C: Photoglottogram

It will be seen from Fig. 1 that the aerometer curve contains several salient points of interest for the measurement of the total duration of *b*. Point I is chosen at the vertex of the last well-defined vibration of the vocal chords in [ə]. Point M indicates the maximum deflection of the aerometer curve after the plosion. On close inspection of the material of group I it was found that in all partly devoiced specimens of *b* M coincided very neatly with the onset of vowellike vibrations in [I], whereas deviations were sometimes noted in fully voiced *b*'s.

In the bulk of these cases there was perfect coincidence; but in some the onset of the vowel occurred from 1.5 cs to less than 0.5 cs before M.

Consequently, the total duration of *b* is measured along the time axis from I to M, and, when necessary, this measurement is adjusted in accordance with the distance between the vowel segments in the microphone oscillogram in order to obtain the best possible accuracy.

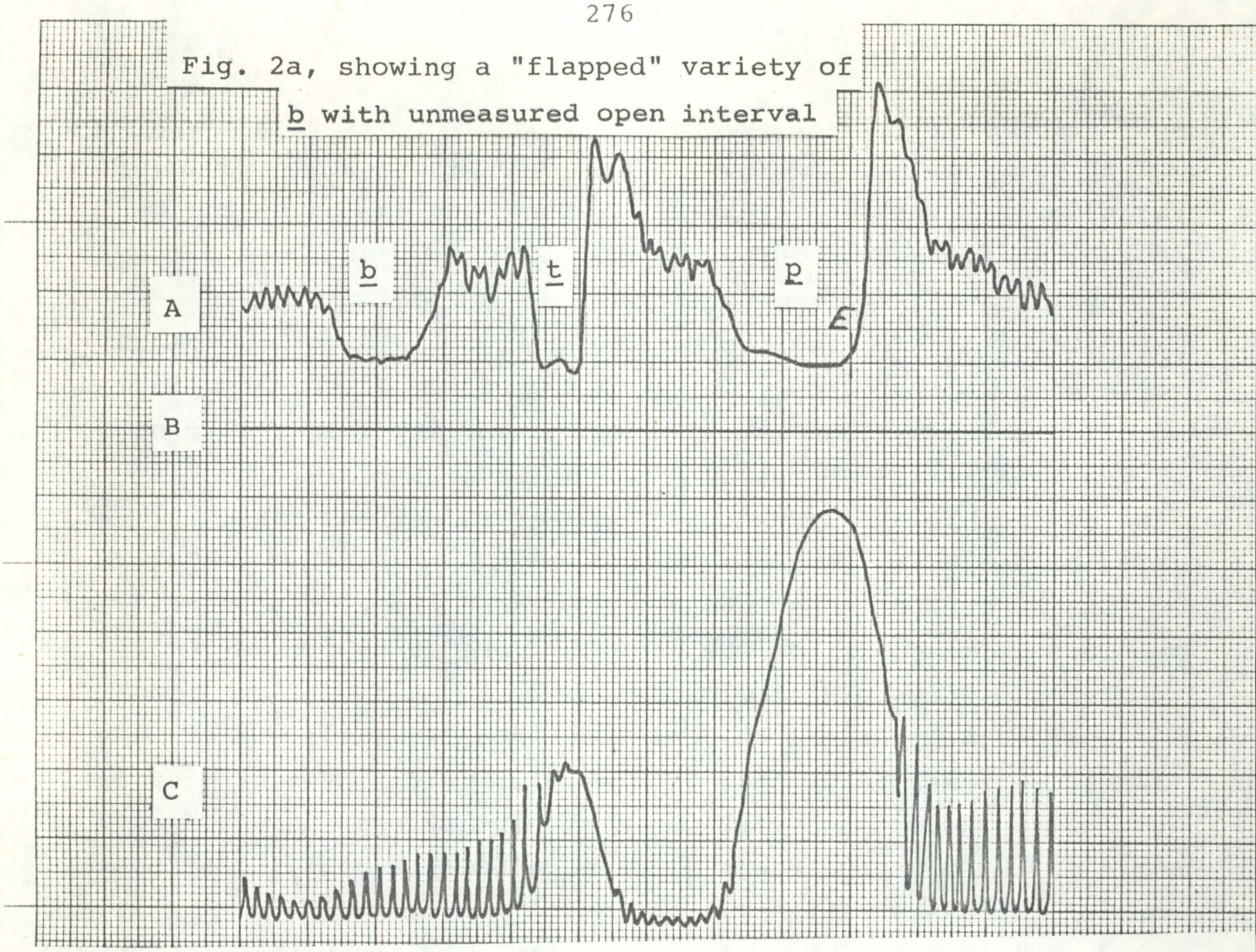
E indicates the time of the plosion, EM (along the time axis) the duration of the open interval, while IM - EM is an expression of the closure stage in a wide sense: transition + hold stage.

On account of the "floating" zero-line (see Figures 2a and 2b) the exact duration of EM cannot be calculated in most of the fully voiced *b*'s, many of which seem to be fricative or "flapped" varieties of the sound (Gimson 1964, p. 154). It naturally follows that the mean values of EM given below are *not* representative.

In the partly devoiced *b*'s and in some of the fully voiced ones the duration of EM was calculated as shown in Fig. 2b.

The ordinate h is a relative measure of the maximum airflow in M, measured in a cubic unit, e.g. cm^3/sec. Any ordinate between E and M will be a measure of the relative airflow at the given moment, whereas the slope of the curve indicates the increase in airflow at that moment. The slope of the curve is

Fig. 2a, showing a "flapped" variety of b with unmeasured open interval

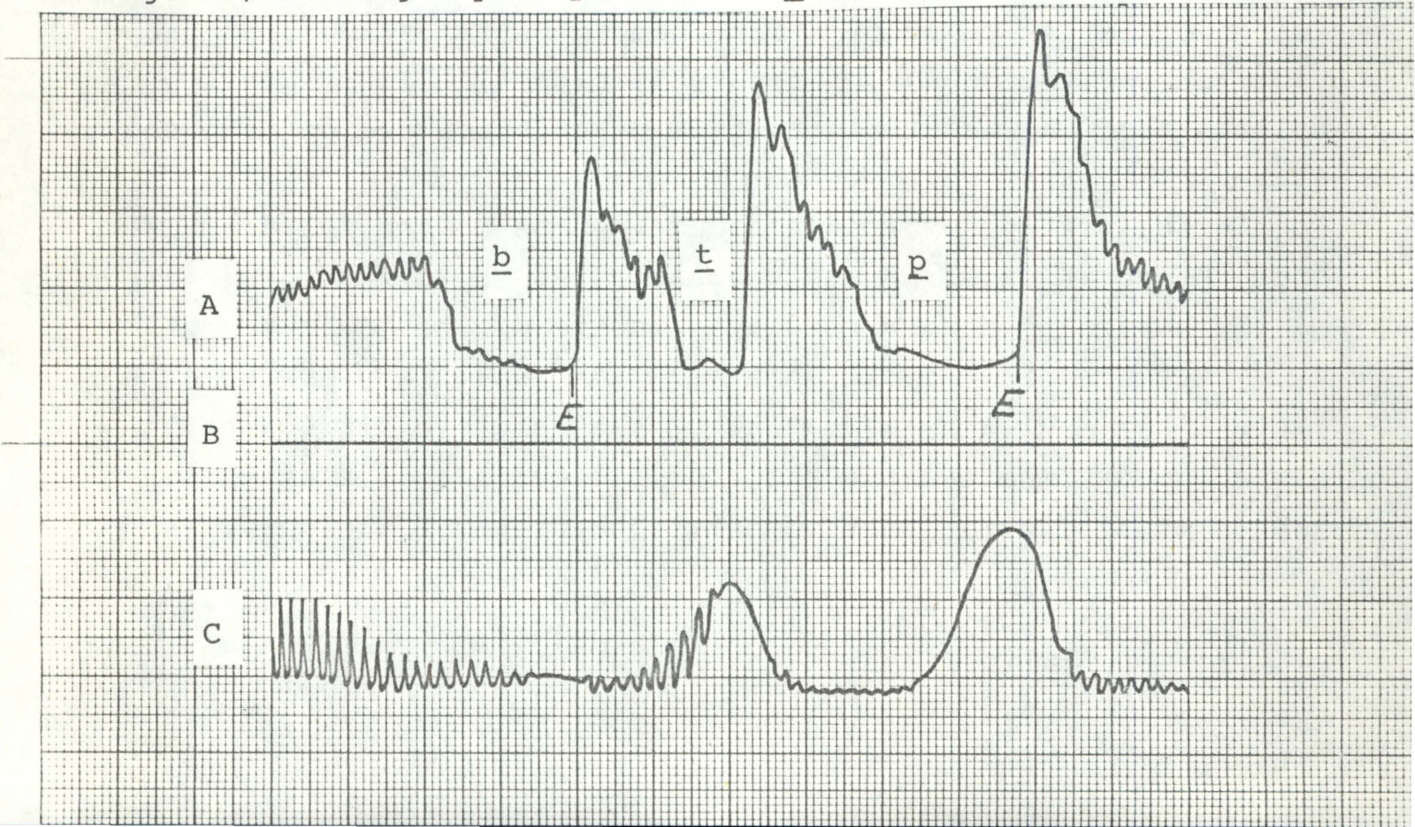


A: Aerometer curve

B: Zero line in pauses between sentences

C: Photoglottogram

Fig. 2b, showing a partly devoiced b with measurable open interval



understood to be tan. of the angle between the abscissa and the tangent touching the curve at the point investigated. When, as was the case in the present experiment, the apparatus is not calibrated, the slope indicates the relative increase in airflow per unit of time.

As the curve between E and M may be taken to be roughly linear, it follows that its slope is indicative of the relative increase in airflow per second at any time of the interval EM.

This assumption underlies the following measurements of relative increase in airflow.

2.2. Variations in the duration of word initial b

As far as the duration of word-initial b is concerned, it is important to decide whether samples of b's drawn at random and in random numbers from uniform phonetic, but tonetically varying groups - like those established above - constitute a homogeneous material from which the average total duration of b could be calculated with confidence. This would not seem to be a likely proposition (Eli Fischer-Jørgensen 1970, p. 37 ff), and the figures below also suggest that the answer must be in the negative.

The average durations of b in cs in Ia through Id are:[1]

Ia	ˈbitter	9.4	***	*
Ib	ˋbitter	11.2		
Ic	ˌbitter	7.6	*?	
Id	ˌbitter	8.6		

The difference between the means of b in a stressed syllable with (relatively) high and static tone (Ia) and b in nuclear syllable with High Fall (Ib) is highly significant (exceeding the

[1] One asterisk indicates significance at the 95% level (or better), two and three asterisks at the 99% and 99.5% levels respectively. o = not significant.

99.9% level), which must very nearly rule out the possibility that the difference is accidental. As, however, the two sounds occur in the same phonetic environment (with the possibly best approximation), it would seem safe to conclude that the difference in total length must be ascribed to membership of different tonetic groups (this does not answer the question whether b's in the same tonetic, but in different phonetic surroundings constitute distinct statistical populations).

The other two analyses (a b in a "normal" nuclear syllable does not, unfortunately, occur in the material) result in a similar, but perhaps a little more non-committal no to the question posed, the significance level in the case of Ia and Id exceeding the 95%, but not the 97.5% level, and in that of Ic and Id the t value being not quite, but very nearly equal to the t value of the 95 % level (this is suggested by the ? after the asterisk).

The conclusion of the above considerations must be that the duration of word-initial b in intervocalic surroundings cannot be calculated without due allowance being made for rhythmic and tonetic groupings.

2.3. Degree of voicing in b

There are rather large discrepancies between the amount of voicing shown by the Photoglottograph and the amount shown by the physiological duplex oscillogram (tapped from the Fabre Glottograph) - the latter showing longer intervals of voicelessness than the former.

It was therefore decided to measure the amount of devoicing from the duration of the glottal gesture in the Photoglottograph curve.

The figures are shown in table I below.

The average duration of all partly devoiced b's and the ratio of devoicing to total duration are given in parentheses, the average duration of all fully voiced items in square brackets.

As the numbers in question are too small, no conclusions can be drawn, but the figures may give interesting hints.

TABLE I

	fully voiced items	partly devoiced items	average duration of b in cs	average devoicing in cs	devoicing in %
Ia ˈbitter	3	6	9.4 (10.0) [8.4]	1.4	14.9 (20.8)
Ib ˋbitter	2	8	11.2 (11.3)[10.9]	2.4	21.4 (26.4)
Ic ˌbitter	7	2	7.6 (9.9) [7.0]	0.4	5.3 (19.3)
Id ˌbitter	5	5	8.6 (9.1) [8.0]	1.1	12.8 (24.1)

The above ratios suggest that the voicing of b is to some extent dependent on the total duration of the sound: the shorter it is, the smaller the amount of devoicing in it, and the greater the probability of its being fully voiced.

In group I all b's with a total duration of 8.4 cs or less were fully voiced.

Allowance being made for the small populations of the present material, the general conclusion would seem to be warranted that the duration of b is dependent on the rhythmic/tonetic conditions of the words in which the sound occurs and also (as will appear later) on the length of the sentence.

Consequently, the maximum probability of intervocalic b being fully voiced will occur in weakly stressed words in (relatively) long sentences.

2.4. Survey of b in group I

TABLE II

		N	s	M in cs	open interval in cs	closure in cs	h	V (slope)	tan. V
Ia	'bitter	9	1.1	9.4	2.2	7.7	12.5	$68^{\circ}-84^{\circ}$	5.9994
Ib	`bitter	10	1.2	11.2	2.4	9.0	14.1	$71^{\circ}-84^{\circ}$	6.5341
Ic	ˌbitter	9	1.7	7.6	2.3	7.3	10.2	$69^{\circ}-83^{\circ}$	4.3227
Id	ˌbitter	10	0.8	8.6	2.0	6.8	10.7	$68^{\circ}-83^{\circ}$	6.1893

these figures are not representative (see 2.1.)

3. Measurements in p

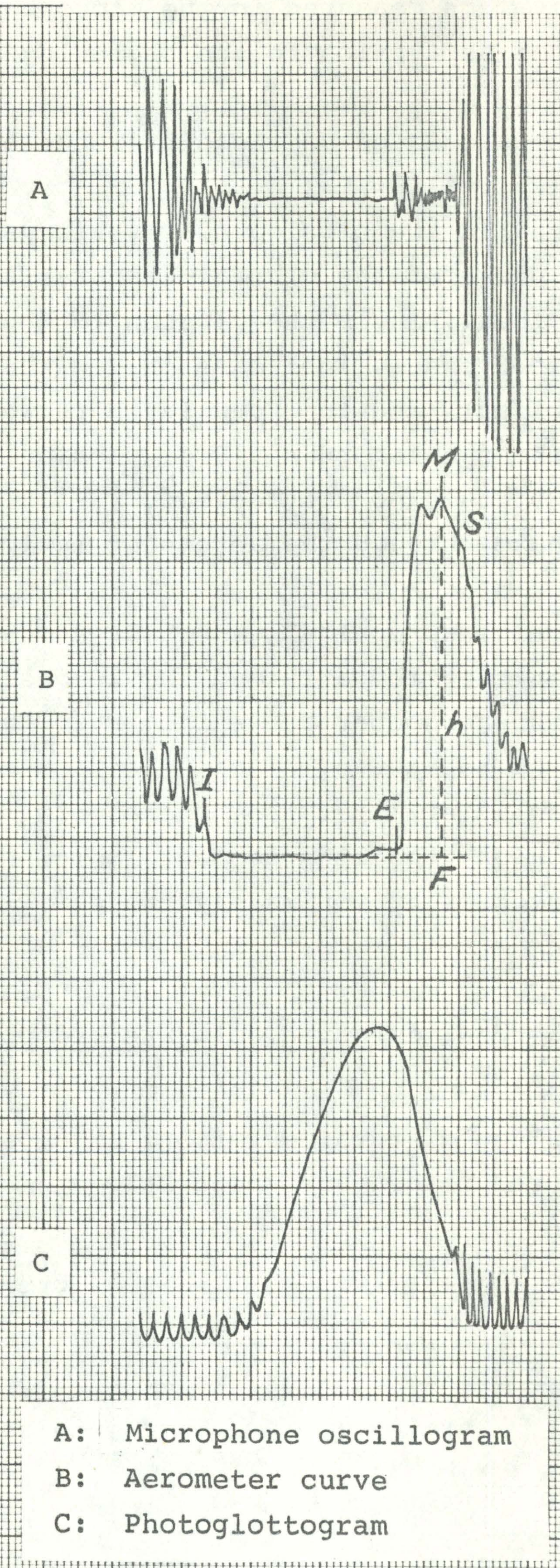
3.1. Definition of the quantities measured in the sound p

The parameters of the sound p are shown in Fig. 3.

The total duration of the sound p is measured along the time axis from I to S, and the measurement is compared with the corresponding distance between the vowel segments in the microphone oscillogram in order to obtain the best possible accuracy. Point I is placed at the vertex of the last well-defined vibration of the vocal chords in [ə] and S at the starting point of the first regular vibration of [I] after the aspiration phase (in the microphone oscillogram).

Very frequently point E - the lowest point in the aerometer curve - did not coincide with the plosion, so that in many cases the airflow through the instrument had started before the plosion. The cause of this is probably movements of the lips preceding the plosion proper.

Fig. 3: parameters in p



A: Microphone oscillogram

B: Aerometer curve

C: Photoglottogram

The open interval is measured in the microphone oscillogram from the onset of the aspiration to the beginning of vowellike vibrations.

The closure is calculated as the difference between total duration and open interval.

In groups Ic and Id the plosion nearly always took place during the opening phase of the glottal gesture (2 instances in Ic could not be decided), in Ib the plosion took place during the closing phase in 6 cases, in 1 case during the opening phase (3 could not be decided), and in group Ia there were 2 instances of plosion during the opening phase, 2 of plosion during the closing phase, and 3 could not be decided.

h is the ordinate through the aerometer maximum normal to the axis in F.

The angle V is an expression (generally somewhat "idealized") of the angle between the time axis and the longest nearly rectilineal segment of the aerometer curve from E to the point of maximum deflection of the curve.

As will be seen later, in (s)p-, where point E indicates the time of the plosion, the line segment IS (along the time axis) is in most cases of approximately the same length as IF, so that ES≈EF.

3.2. Variations in total duration, closure and open interval in word initial p

The average values of total duration, of closure and open interval (all measured in cs) are shown below.

		total duration		closure		open interval	
Ia	ˎpill	14.5		10.4		4.2	
			***		O		***
Ib	ˈpill	13.2	***	10.0	**	3.1	O
			*** O		*** O		O O
Ic	`pill	15.6		12.0		3.6	*
			*		*		O
Id	ˏpill	14.8		10.9		3.9	

It will be seen that the total duration of intervocalic word-initial p is highly sensitive to shift of stress and tone. Also the closure of p is affected, but not to the same degree, perhaps.

The open interval seems to be the stablest component of the sound: only the short open interval of Ib (unstressed, low and level tone) differs significantly from Ia and Id. In all other cases there is no significant difference.

The figures seem to indicate that the open interval of p in nuclei with Low Fall and Low Rise is comparatively long. It may be a little surprising that the duration of the open interval in the High Fall group (Ic) does not differ significantly from that of the other groups.

The difference between M in Ia and Id is not significant, which means that p's in Low Fall and Low Rise groups constitute one statistical population. The difference in average duration may be supposed to be accidental and not attributable to membership of different tonetic groups. - N.B.! The length of the sentences is not the same, so that the validity of the conclusion may be contestable.

In all other cases the difference between the means is significant or highly significant, for which reason the groups must be considered as belonging to different statistical populations.

The conclusion of the above is that in uniform phonetic surroundings the total duration of p varies significantly with membership of different tonetic groups (with the said exception). The stronger the stress and the longer the tonetic glide, the greater the total duration.

On the whole, the standard deviations for the four groups of p's are smaller than those for the four groups of b's. Statistically, the duration of p must therefore be said to be a relatively more stable quantity than that of b. The reason is not far to seek. Phonetically, the b's are two different sounds:

a partly devoiced and statistically rather stable sound and a fully voiced and comparatively unstable sound, which in some cases, it would seem, may be "flapped" or even fricative. As an example of the unstableness of *b* it may be mentioned that the standard deviation for *b* in group Ic (ˌbitter) is 1.7 for the whole population; the range is 5.4 cs. The range of the fully voiced *b*'s is 4.2 cs, of the partly devoiced items it is 0.9 cs (there are only 2 such sounds in the population). In *p* in Ib (ˌpill) the standard deviation is 0.7, and the range is 2.4 cs.

3.3. IS and IF values

It is of importance to remember that in initial *p* total duration (IS) is always distinct from the distance in time between the implosion and the moment of maximum airflow, so that IS > IF, whereas in *p* preceded by *s* the difference between the two values tends towards zero.

3.4. Increase in airflow on release of p (tan. V values)

It will be seen that the tan. V measurements show a surprising amount of uniformity, the four groups only differing a few degrees from one another. As some uncertainty is inherent in the method of measurement, a certain caution should be exercised in the interpretation of the differences between the tan. V values. No doubt, however, it would be safe to suppose that of the four groups Ic has the relatively greatest increase in airflow per unit of time, and that this quantity is dependent on the extent of the tonetic glide (and the stress?).

3.5. Comparisons between the lengths in groups I and II

Some statistical calculations were made in order to compare lengths of certain segments selected from groups I and II.

a. Ic total duration of p in ˋpill and II, 2a total duration of p in ˋpill

The difference between the two means is significant (exceeding the 99.5 % level). Both occur in intervocalic positions in High Fall groups, though not between the same vowels.

Alternative interpretations present themselves:

(i) the difference between the means is either ascribable to the fact that the phonetic surroundings of the two p's are not quite the same, [ə-p-I] in Ic and [ei-p-I] in II, 2 a, or to

(ii) the rhythmic-tonetic differences between the sentences.

Ic has 9 syllables, of which - besides the nucleus - at any rate two (Nos. 2 and 5) carry a certain amount of stress, whereas II, 2a has only one stress, viz. on the nuclear syllable.

It is hardly possible from the material to make out a strong case for the rejection of (i). However, two statistical calculations speak in favour of this.

Firstly, the difference between the total duration of p in II, 2a and in II, 3a is not significant (below the 60 % level): they belong to the same statistical population. These two sentences are rhythmically and tonetically identical, whereas phonetically they are "unilaterally" identical. Thus the difference in degree of opening/place of articulation for [I] and [e] does not bring about a significant difference between the p's in these phonetically different surroundings.

Secondly, the word pill in the sentences Ic and II, 2a was segmented into [p] and [Iɫ] on the aerometer curve (according to the usual criteria).

The mean of the total length of the latter segment was in Ic 19.9 cs, s = 2.1. The corresponding figures were for II, 2a 26.1 cs and 1.2. The difference between these means is highly significant (exceeding the 99.95 % level). Thus the two segments belong to separate statistical populations. As the phonetic surroundings of [Iɫ] in the two words are identical, there is every probability that the difference arises from tonetic conditions.

Moreover, the difference between the mean of the word pill (total length) in Ic (M = 35.5 cs, s = 2.0) and the mean of the word pill in II, 2a (M = 43.9 cs, s = 1.4) is also highly significant (exceeding the 99.95 % level).

On the strength of this evidence the first sentence of the alternative (i) is rejected, and the difference between the means is supposed to arise from rhythmic-tonetic "causes".

b. Possible influence of following vowel

Though the difference between the means of p in II, 2a and II, 3a is not significant, the possibility remains that significant differences between the means of the two sounds might be concomitant to greater differences in degree of opening/place of articulation of the vowels following p/b.

Comparisons were only made between total duration of p in II, la (peak) and of p in II, 5a (par) and between total duration of b of the same sub-groups (beak, bar). For the p's the difference between the means was highly significant (exceeding the 99.95 % level), and for the b's it was very probably significant (exceeding the 97.5, but not the 99 % level). As the rhythmic-tonetic conditions of these sentences are identical in pairs, it is hereafter taken for granted that at any rate extreme differences in degree of opening/place of articulation may influence the total length of p/b significantly. (N.B. perhaps the difference between open/closed syllable should not be disregarded, but further experiments are required to verify/falsify this proposition).

In consequence of the facts established above it was considered safer to make comparisons between items from within the same group and not between items from different groups.

Though the influence of the degree of opening/place of articulation has not been sufficiently clarified, there is, on the whole, no doubt that if reliable statistics of quantity in English are to be established, due allowance must be made not only for this problem, but also for rhythmic-tonetic phenomena.

Failing this, one runs the risk of uncritically confusing quantities from distinct statistical populations.

3.6 Survey of p in group I

TABLE III

		N	s	total duration in cs	open interval in cs	closure in cs	IF in cs	h	V (slope)	tan. y
Ia	ˎpill	7	0.7	14.5	4.2	10.4	13.4	19.2	82°-85°	9.5233
Ib	ˌpill	10	0.7	13.2	3.1	10.0	11.9	16.4	82°-85°	8.9726
Ic	ˋpill	9	0.7	15.6	3.6	12.0	14.3	20.3	85°-86°	12.7060
Id	ˏpill	10	1.1	14.8	3.9	10.9	13.2	18.6	84°-86°	11.1010

4. p-, sp-, b-, -s + b-

4.1. General remarks

Groups II,1; II,2; II,9 were selected for the investigation of the relationships between the above-mentioned sounds (see 1.2.).

Some calculations were made to establish whether the p's and the b's of the 3 groups differ significantly or not.

The results were as follows:

p

		total duration cs			open interval cs			closure cs		
II, 1a	peak	17.3			3.9			13.4		
			o			*			o	
II, 2a	pill	17.8		***	4.4		**	13.4		o
			*			*?			o	
II, 9a	purr	18.8			4.8			13.9		

b

		total duration cs		
II, 1c	beak	15.o		
			***	
II, 2c	bill	13.5		o
			***	
II, 9c	burr	14.9		

Accordingly it was decided to make comparisons only between members of sub-groups within the same group (see 3.5.).

4.2. Comparisons in group II,1

a. p-/(s)p-

The segmentation of this cluster presents a problem on account of the co-articulation of the two sounds. However, there is always a clear "hump" in the aerometer curve immediately followed by the sudden drop of the p transition. It was decided to place the dividing line between s and p just after the vertex of this hump. At this point the s noise in the microphone oscillogram seems to have weakened very much or to have stopped altogether.

The difference between the means of the total duration of p (IS) in peak and speak is highly significant (at the 99.9 % level). Statistically, the two stops, which occur in comparable rhythmic-tonetic groups, do not belong to the same population. The difference must be ascribed to phonetic "causes": s precedes p.

The duration of the open interval of (s)p in this group is only about 1/3 of that in p. Besides the intensity of the aspiration is slight, it may even be inaudible. It is, however, in all cases a stable physiological component of the "weakened" p. Also the means of the closure (10.4 cs and 13.4 cs respectively) differ significantly (at the 99.5 % level).

The h values show that, on the average, the relative maximum airflow in (s)p- is only about 2/3 of that in p. Similarly, the relative increase in airflow per unit of time is smaller in the former than in the latter (see table IV, section 4.5 below).

The most important difference, probably, between the two sounds (and very likely the one from which all other differences derive) is found in the glottal gesture: p- has its own gesture with a duration of about 17 cs, whereas s and p in sp-, as it were, share a glottal gesture of somewhat longer duration, 23 - 24 cs. The outstanding feature is that s always occupies the whole of the opening phase and more or less of the closing phase of the glottal cycle, so that the implosion of p begins only during the latter phase.

The plosion (point E) takes place late, in fact only when the closing has been accomplished (the curve of the Photoglottograph has become "horizontal" some cs before E), from which it will be understood that the open interval is short, and the aspiration weak: at the time E the glottis is being adjusted for phonation. This adjustment is achieved when - after the release - the supraglottal pressure has dropped sufficiently. But in E this has not yet taken place. - Kim (1970 p. 114) states that "by the time /p/ is released, the glottis will already have become so narrow that the voicing for the following vowel will immediately start, and then we have an unaspirated /p/ after /s/". If this is to be taken as a general rule, I disagree. My photoglottograms show very clearly that as far as the onset of voicing is concerned, the statement does not hold generally, at least not for my (British English) informant.

Everywhere in group II, lb it has been possible to draw an acceptable and consistent dividing line between the s and p segments.

b. p in sp- and b in -s + b-

It has been demonstrated in section a. above that the characteristics of p when preceded by s are altered radically. Phonetically, the sound comes closer to b in many respects. The question is whether it is identical with b from an articulatory point of view.

The difference between the means of total duration of p in speak (11.8 cs) and of b in Jack's beak (12.8 cs) is significant at the 99.5 % level.

Now the previous investigations have shown that quantity in phonetically comparable groups is greatly influenced by differences in the rhythmic-tonetic structure of the sentences. Unfortunately, at the outset I was unaware of the exact nature of these facts (but had a suspicion of them). It follows that the appropriate test sentences were not made.

However, from a durational point of view (s)p- is neither p nor b, total duration of p-, (s)p-, b and (-s+)b- being 17.3, 11,8, 15,0 and 12,8 cs respectively.

As regards aspiration and voicing, the two sounds also differ from each other: in (s)p- aspiration is a stable feature, whereas in (-s +) b - it is mainly unstable or missing. Moreover, there is a somewhat unstable, but nevertheless unmistakable tendency towards voicing in the EM phase in (-s +)b-. In (s)p-this tendency is entirely lacking during the release stage.

Only in 7 cases out of 9 did the Photoglottograph show a glottal gesture underlying the *k*-*s*-*b* segments (and even these seven were not all very clear).

There seems to be no evidence of any particular artefacts (e.g. that the tongue or the epiglottis may cut off the light from the light source) associated with the photo-glottography of sequences like *s* + labial + high front vowel (Frøkjær-Jensen et al. 1971). Hence a tentative interpretation of the timing of *b* relatively to the glottal cycle may be ventured: *b* only occupies a small part of the glottal gesture, if any, in the articulation of *k*-*s*-*b*. As far as can be ascertained, its implosion falls very late in the closing phase, and no separate gesture underlies the hold and release stages, during which the curve is "horizontal". This feature, too, distinguishes (-*s* +)*b*- from (*s*)*p*-. The fact that the sounds of the *sp*- cluster share a glottal gesture, whereas *b*- is located outside or "at the edge of" the glottal gesture for -*k*-*s* might, perhaps, be interpreted to the effect that there is a juncture effecting a less intimate connection between the sounds in -*s* + *b*- than in *sp*-.

The "two-humped" shape of the few tolerably clear photoglottograms of the series II, 1d (Nos. 1, 4, 5, 6) leads one to suppose that -*k*-*s* are less intimately connected (are there 2 commands?) than *sp*-, which only seems to have one command, and that *b*- has no command at all.

In this sub-group there is no voicing during the hold stage.

c. b in beak and b in Jack's beak

The difference between the means of the total duration of the two *b*'s is very probably significant (exceeding the 95 %, but not the 97.5 % level). This agrees well with previous findings.

d. Total duration of sp- in speak and -s+ b- in Jack's beak

As they occur in separate tonetic groups, and as the rhythmical structure of the sentences is not the same, these two values are not directly comparable. Statistically, the latter (s = 0.8) is a more homogeneous group than the former (s = 1.6). The difference between their means is highly significant (at the 99.5 % level).

4.3. Comparisons in group II, 2

a. p-/(s)p-

Here, too, the difference between the total duration is highly significant (at the 99.9 % level). It follows that the two p's belong to separate statistical populations.

The aspiration in II, 2b is more conspicuous in the microphone oscillogram and makes up a larger proportion of the aspiration in II, 2a than is the case in II, 1b compared with II, 1a.

On the whole, (s)p- in this group is more like a p- than the corresponding sound in II, 1b; thus the values of maximum airflow and relative increase in airflow per unit of time are greater and approximate or exceed the values of b- in II, 2c - in contradistinction to II, 1, in which the values of b- exceed those of (s)p- in either respect.

b. p in sp- and b in -s + b-

A glottal gesture for k-s-b was only visible in a few cases in the photoglottogram (always a minor one); in the remaining cases the glottogram suggested "closed glottis" without voice.

In the aerometer curve for Nos. 6 and 7 (II, 2d) the k-s segments were fused in such a manner that segmentation was rendered impossible. This procedure was, however, very difficult in several items of this series. Another peculiarity was a relatively heavy aspiration sometimes occurring side by side with beginning voice in the release stage. Thus this *b* with its great h and tan. V values is to some extent suggestive of a *p*-.

For significance, etc., see 4.2 a.

c. Total duration of sp- in spill and -s + b- in Jack's bill

The difference between the two sets of means is highly significant (99.95 % level).

4.4. Comparisons in group II, 9

a. p-/(s)p-

As in the previous two groups, the difference between the means of total duration is highly significant (at the 99.5 % level).

The photoglottograms for *p*-, whose total duration is the greatest in the 3 groups, show outlines differing much from the usual, rather regular conic section. Three (Nos. 1, 2, 10) show a marked plateau, several cs in duration, instead of the usual "peak", three (Nos. 7, 8, 9) are distinctly "two-humped", whereas only one (No. 6) is of normal shape.

The two-humped shape is very likely the most interesting, as it is difficult to explain from the assumption of only one command per *p*. How are the movements maximum glottal opening → diminishing glottal opening → maximum glottal opening to be visualized from the assumption of only one command?

It is impossible to take a decision on the question whether the apparatus functioned abnormally (small displacements of the light source?), as the effect on the glottogram, if any, of such displacements is unknown.

The total duration of *p*- in II, 9a was the greatest of the three groups under consideration. Its relationships in this respect to the other two *p*'s is seen in 4.1.).

The outline of the glottal opening for the *sp*- cluster is decidedly asymmetric, even more so, perhaps, than in the other two groups, so that the opening phase of the cycle is clearly shorter than the closing phase.

b. p- in sp- and b- in -s + b-

As the b's of II, 9d - alone among the b's of the 3 groups - show a marked tendency towards voicing in the IE phase, they deserve special mention. This applies to Nos. 2, 4, 6, 8, 9, possibly also to 1 and 10. This fact is remarkable as it demonstrates that b preceded by s (in open syllables?) is not necessarily devoiced in its entire length, even during the hold stage, where onset of voice occurs in the latter half of the closure.

This may also provide an answer to the important question: what is the fundamental difference between p and b in English.

On the evidence of the photoglottograms investigated for this article my tentative answer to the question would be like this: p is, as it were, "tied" to a glottal gesture in all positions (with potential aspiration as a consequence), whereas b is "free" in this respect. The supraglottal pressure having risen to the critical level, phonation ceases in b, but the vocal chords do not part, and so phonation is resumed whenever permitted by a sufficient drop in the oral pressure.

This is seen clearly in the case of (s)p- and (-s +)b- in groups 9b and 9c. In the latter, phonation is frequently resumed during the hold stage. This phenomenon is never seen in (s)p-: in this cluster p is also "tied" to a glottal gesture shared with s.

With some confidence I therefore venture to say that a "p" is a [p] also when preceded by s, and a "b" is a [b].

The difference between the means of *b* and *p* is not significant.

c. b in burr and b in Jack's burr

The difference between the means of the two sounds is highly significant (exceeding the 99.5 % level).

d. Total duration of sp- in spur and -s + b- in Jack's burr

The difference between the two sets of means is highly significant (exceeding the 99.5 % level).

4.4. Conclusion

In the preceding paragraphs an attempt has been made to find an answer to the question: can (*s*)*p*- be looked upon as a *b* on articulatory grounds, phonetically *and* phonologically? In other words, is the notation /sbi:k/ "as good as" the notation /spi:k/?

The answer must be in the negative on grounds deducible from the phonetic facts brought to light by this investigation.

Admittedly, the total duration of *p* is shortened drastically in the *sp*- cluster, which brings (*s*)*p*- nearer to *b*. But in spite of any such alterations (*s*)*p*- potentially retains its aspiration actualized in a measurable quantity in 28 of the 29 specimens investigated. Moreover, (*s*)*p*-'s relationship to the underlying glottal gesture is different from the conditions obtaining in both *b*- and -*s* + *b*-. This is most important.

In a similar way *b* is changed after *s* (devoicing, tendency towards aspiration in the open interval), but the sound potentially retains voicing, frequently actualized in the open interval and even during the closure (group II, 9 d). Furthermore, it may be supposed that the total length of *b* is not materially shortened after *s*. Finally, that unlike *p* it has no separate command.

But perhaps the question itself is posed the wrong way, and the answer consequently inadequate or inconclusive.

Perhaps one must say that under such and such conditions a given sound will be perceived as *p* (in English), whereas the identical sound under other conditions would be perceived as *b*.

This might be called a relational approach. As regards labial stops it implies that what is heard not only depends on what is actually said, i.e., what can be said about the utterance in physiological and acoustic terms, but also on the ears that hear, and on the relational conditions under which the utterance is said *and* heard (perceived).

And linguistic "relativity", it would seem, not only applies to the total duration of stops, but to any speech sound.

It was demonstrated in 3.5. that the total length of pill in Ic and pill in II, 2a varies significantly with the rhythmic-tonetic conditions obtaining in the sentences. The same is the case with the [I] segment in the same two groups. But even vowels seem to be subject to "relativity". The length of the [I] segment in bitter (group I) can easily be measured in the aerometer curve. The means of the four groups are respectively a: 5.1 cs; b: 6.2; c: 4.9; d: 5.0. The difference between the means of a and b is significant (at the 99 % level), s = 0.7 (for a) and = 1.0 (for b). Other calculations were not made.

Seen from a relational point of view the question posed above concerning the identity of *p* in the *sp* cluster actually becomes a pseudo-problem.

If in a given relation (or context), e.g. in the *sp* cluster, the sound is perceived as a *p*, it must be accepted that the sound *is* a *p* in that context, even if a certain amount of physiological-acoustic evidence could be adduced to the contrary. Which is hardly possible in this case (see above).

Consequently, if such a p is isolated from its relations, by removal of the s noise, e.g., one might very well anticipate b perceptions (in English). From a relational point of view it would seem possible to predict this kind of perception with appreciable accuracy on condition that one knew the rhythmic-tonetic conditions under which the word was spoken. But the amputation of s brings about new contexts, new significant relationships, probably. And what importance can then be attached to the result of such an operation?

If this is so, the ground is cut from under the phonological discussion in favour of the interpretation [sp] ⟶ /sb/.

4.5. Survey of group II

TABLE IV

a) subgroups 1a, 2a, 9a

p in	N	s	total duration in cs	open interval in cs	closure	IF	h	V (slope)	tan. V
ˋpeak	9	1.3	17.3	3.9	13.4	15.7	22.8	85°-87°	14.7252
ˋpill	10	0.8	17.8	4.4	13.4	16.4	24.6	86°-87°	17.1690
ˋpurr	10	1.0	18.8	4.8	13.9	17.2	25.4	80°-87°	13.3884

b) subgroups 1b, 2b, 9b

sp- in	N	duration of s	total duration of p	s	open interval	closure	EF	h	V (slope)	tan. V	s + p
ˋspeak	9	9.7	11.8	0.6	1.4	10.4	1.5	15.3	83°-87°	11.4893	21.5
ˋspill	10	11.6	12.3	1.2	1.5	10.8	1.6	17.2	85°-87°	14.9699	23.8
ˋspur	10	11.3	13.0	0.9	1.8	11.1	1.9	17.8	83°-86°	11.6756	24.2

Of a total of 29 p's 10 items had relatively strong aspiration, 18 items had weak aspiration, and 1 could not be decided.

c) subgroups 1c, 2c, 9c

b in	N	s	total duration	open interval	closure	h	V (slope)	tan. V	fully voiced
ˋbeak	9	0.9	15.0	1.5	13.5	16.3	84°-87°	13.3433	2
ˋbill	10	0.9	13.5	1.3	12.2	17.8	85°-86°	12.8655	4
ˋburr	10	1.2	14.9	1.4	13.5	18.0	82°-87°	13.5395	6

d) subgroups 1d, 2d, 9d

-s + b in		N	duration of s	total duration of b	s	open interval	closure	s + b	h	V (slope)	tan. V
'Jack's	'beak	9	3.3	12.8	0.7	2.0	10.8	16.1	15.5	83°-87°	11.3827
' -	'bill	8	4.3	12.8	0.7	1.6	11.2	17.1	17.2	85°-87°	14.4194
' -	'burr	10	5.0	13.4	0.5	1.8	11.6	18.4	17.7	80°-87°	11.0429

5. Niels Davidsen-Nielsen: English Stops After Initial /s/ (1969a, 1969b)

As stated in 1.1, one of the impulses to investigate English labial stops came from the above-mentioned article (1969a), whose approach seemed to me to be open to question.

It says (p. 56): "The words were inserted into sentences and recorded by four persons (two English and two American) ... The test words from column I (among them the word spear, H. P.) were then cut out of their environments. In the six words thus isolated the initial [s] was removed ...

The perceptory experiment consisted in letting 32 test persons (24 English and 8 American) identify 52 recordings of these truncated words. Each of these words, which had been randomized, was played twice to the test persons, who were then asked to write down the English word they thought they heard".

The phonetic surroundings of the 6 test words are not identical (1969b p. 324), and it may be questioned whether in all cases the rhythmic-tonetic patterns are the same. This may have had some (perhaps negligible) influence on the total duration of the stops. More important, however, are the implications of cutting experiments as such.

(a) The amputation of s, e.g. in the word spear, interferes with the inner relational conditions of the word and thereby creates new relations, which as likely as not do not exist in identical form in the English language, and which at any rate did not exist in the word before the amputation.

(b) By the removal of s a new initial stop, unknown to the English linguistic system, comes into existence. No matter how natives may perceive that sound - as b or p - this perception says nothing about (s)p-. For the new labial stop is non-existent in English, so to speak. And units being non-existent in a language must be rejected as elements in linguistic reasoning bearing on that language.

(c) If nevertheless the new stop is presented as belonging to the phonetic pattern of English it can be predicted with good certainty that it will be perceived as b since, in the main, no alternative is left open.

In order to substantiate the assertions made above the words purr, spur, burr were selected. They were said in the carrier sentence I say ˋ... today. It will be seen that the 3 words occur in identical surroundings, phonetically, tonetically and rhythmically. Besides, the relations within this group do not differ much from the ones that would obtain in the group pier, spear, beer of Davidsen-Nielsen's experiment if these words had occurred in a similar sentence.

The average total duration of the words purr, spur, burr is 45.6 cs, 45.2 cs, 51.9 cs respectively. If s is cut away from spur, it leaves a "truncated" word with a total duration of 40.6 cs. This word is called t. The relations obtaining in the 4 words are as follows:

$$\frac{\text{p}}{\text{purr}} = 0.41; \quad \frac{\text{b}}{\text{burr}} = 0.33; \quad \frac{\text{p}}{\text{spur}} = 0.25; \quad \frac{\text{p}}{\text{t}} = 0.32;$$

$$\frac{\text{closure}}{\text{purr}} = 0.31; \quad \frac{\text{closure}}{\text{burr}} = 0.30; \quad \frac{\text{closure}}{\text{spur}} = 0.21; \quad \frac{\text{closure}}{\text{t}} = 0.27;$$

(see 4.5)

It will be seen

(a) that the amputation of s has created new relations in t which did not exist in spur before the amputation

(b) that a new initial stop has come into existence

(c) that the relations obtaining in t are much nearer to those obtaining in burr than to those obtaining in purr.

As voicing/voicelessness is of minor importance for the differentiation of /b/ - /p/ initially, and as the relations between the durations in the truncated word agree pretty closely with those in burr (but differ widely from those in purr), it is possible - in so far as these relations are decisive for the perception - to predict with a high degree of probability that the number of b perceptions will be great compared with the number of p perceptions.

To this must be added that the open interval of p in spur (1.8 cs) does not differ much from that of b in burr (1.2 cs), but differs widely from that of p in purr (4.8 cs). Even in isolation this fact should greatly increase the probability of b perceptions in the truncated word (Lisker & Abramson 1964).

The conclusion to be drawn from the above and from Davidsen-Nielsen's experiments is that an initial labial stop characterized by relations elsewhere unknown to the linguistic system of the English language, but whose characteristics are similar to those of b in the same position, will be perceived as b with very great probability. For it is a well-known fact that when people are faced with linguistic relations unknown from their mother tongue, they will perceive a given foreign speech sound as "a something" whose relations are nearest to those of a speech sound known from the mother tongue (cp. that no doubt most Danes without phonetic training would spontaneously perceive b in the local pronunciation of the word Palermo; but this is neither a phonetic nor a phonological proof that the initial sound of that place name really is a b)

This, however, says nothing about the natural relations of p in the sp cluster and, consequently, nothing about the phonetic characteristics and phonemic status of (s)p-.

On the whole, amputation experiments seem to me to be methodically inadmissible as evidence in phonological discussions. On the other hand, they are of course quite legitimate in perceptory experiments whose sole object is to find the cues requisite for a given perception.

Acknowledgements

I wish to thank Professor Knud Sørensen, of Aarhus University, who has revised the draft of this article; my son-in-law Niels Glerup Sørensen, who has drawn the figures; the staff of the Institute of Phonetics, especially Mr. Jørgen Rischel, for their kind assistance.

References

Abramson, A.S., L. Lisker and F.S. Cooper 1965: "Laryngeal Activity in Stop Consonants", Speech Research Nov. 1965, p. 6.1.-6.13.

Chomsky, N. & M. Halle 1968: The Sound Pattern of English New York.

Davidsen-Nielsen, N. 1969 a: "English Stops After Initial /s/", Aripuc 3/1968, p. 55-62.

Davidsen-Nielsen, N. 1969 b: "English Stops After Initial /s/", English Studies, Vol L,4, p. 1-19.

Davidsen-Nielsen, N. 1970: Engelsk Fonetik, Copenhagen.

Denes, P. 1955: "Effect of Duration on the Perception of Voicing", JASA, Vol. 27,4., p. 761-764.

Fischer-Jørgensen, E., 1957: "What Can the New Techniques of Acoustic Phonetics Contribute to Linguistics", Proc. VIII Intern. Congress of Ling., Oslo, p. 433-479.

Fischer-Jørgensen, E., 1962: Almen Fonetik, Copenhagen.

Fischer-Jørgensen, E., 1970: Indledning til Experimentalfonetisk Kursus, Copenhagen.

Frøkjær-Jensen, B., C. Ludvigsen and J. Rischel, 1971: "A glottographic study of some Danish consonants", Form & Substance, Copenhagen, p. 123-140.

Gimson, A.C., 1964: *An Introduction to the Pronunciation of English*, London.

Jespersen, O., 1958: *English Phonetics*, Copenhagen.

Jones, D., 1962: *An Outline of English Phonetics*, Cambridge.

Kim, Chin-Wu, 1970: "A Theory of Aspiration", *Phonetica* 21,2.

Kingdon, R., 1966: *The Groundwork of English Intonation*, London.

Lisker, L., 1957: "Closure Duration and the Intervocalic Voiced-Voiceless Distinction in English", *Language* 33, p. 42-49.

Lisker, L. and A.S. Abramson 1964: "A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurement", *Word* 20,3, p. 384-422.

O´Connor, J.D. 1963: *A Course of English Intonation*, Stockholm.

O´Connor, J.D. & G.F. Arnold 1966: "*Intonation of Colloquial English*, London.

Thorsen, N.G., 1971: "Voicing in British English *t* and *d* in Contact with other Consonants", *Aripuc 5*, p. 1-39.

A REPLY TO HENRY PETERSEN

Niels Davidsen-Nielsen

1. The editors of ARIPUC have given me permission to reply briefly to Mr. Henry Petersen's phonetic and phonemic criticism of my paper *English Stops After Initial /s/* (1969). I will therefore procede to consider the objections which have been raised one by one.

2. H.P. claims that the phonetic surroundings of my test words are not identical. This is naturally correct if 'identical' is understood to mean 'exactly alike', 'agreeing in every way', but the environments are certainly greatly similar, as they should be. In all the phrases, the test words are placed post-initially and stressed. They are also preceded by an unstressed syllable ending in a vowel and followed by an unstressed syllable. Furthermore the phrases are constructed in such a way that any emphatic rendering of them is avoided, and their rhythmic-tonetic pattern is in all cases ∪-∪∪-(∪). That the presence or absence of a final unstressed syllable, as in *The steam from the chimney* and *A sty to be cleaned* respectively, should be relevant to the pronunciation of the post-initial word seems quite unlikely to me. In short, all the elementary precautions necessary in investigations of this type have been taken.

H.P. himself prefers the carrier frame 'I say ____ today', and in this case he is of the opinion that the surroundings qualify for the term 'identical'. However, this is certainly not true in the narrow sense of the word mentioned above, since there will always be variations, e.g. in stressing, from one pronunciation to the next. When I decided against one particular carrier frame it was because I wanted the test words to be pronounced as naturally as possible, rather than like words in quotation marks. When the recording of the phrases had been completed, my test persons told me that they did not know what words were being investigated, and this would obviously not have been the case under the alternative approach.

Clearly each of these two methods has its advantages, and I consider the one chosen in my particular investigation perfectly appropriate.

3. According to H.P. it is predictable that *b*, *d*, *g* are perceived in the truncated words, and he therefore considers the perceptory experiment virtually superfluous. I do not think that it is at all justifiable *a priori* to rule out other perceptions. H.P. bases his opinion on duration factors, but

he concedes himself that voicing is of some importance, and it is quite likely that a number of other factors could be relevant (release burst, transitions, and what not). When I excised the *s*-sounds I was by no means certain that the truncated words would be perceived the way they were.

4. Whereas H.P.'s phonetic objections do not invalidate my investigation I can at least understand his doubts as to the relevance of amputation experiments to phonological analysis. This approach is indeed somewhat reminiscent of Hjelmslev's proposal for "experimental commutation" (1937), which has been criticized by e.g. Fischer-Jørgensen (1949). I therefore agree that the tape cutting method is not applicable to phonemic analysis in any completely general sense. In this particular case, however, I can see no synchronic criteria by which to choose between the interpretation /sp-, st-, sk-/ and /sb-, sd-, sg-/ except phonetic similarity (I have discussed the symmetry argument advanced by Hockett (1955) in my paper). I am therefore here willing to "hug the phonetic ground" and propose the latter solution on the grounds of greater acoustic similarity, and also greater perceptory similarity, as demonstrated not only by the tape cutting experiment, but also by the test with the phrase *Thanks, Stan, that'll be all*, where the proper name was perceived as *Dan* in nearly half of the cases. It might be added that if /sp-, st-, sk-/ were chosen, one would have to accept overlapping manifestation of phonemes without gaining any compensatory advantages.

5. For these reasons it is my opinion that H.P.'s criticism of the phonetic part of my investigation can be refuted completely, and that my phonological conclusion, which like all such interpretations is certainly open to discussion, can in no way be dismissed.

References

Davidsen-Nielsen, N. (1969), "English Stops After Initial /s/", *English Studies*, vol. L. No. 4, p. 321-339.

Fischer-Jørgensen, E. (1949),"Remarques sur les principes de l'analyse phonémique", *Recherches structurales TCLC* V, p. 214-234.

Hjelmslev, L. (1937), "Neue Wege der Experimentalphonetik", *Nordisk Tidsskrift for Tale og Stemme*, p. 154-194.

Hockett, C.F. (1955), *A Manual of Phonology* (Baltimore).

INVESTIGATING THE INFLUENCE OF BLOWING TECHNIQUE ON PITCH AND TONE IN RECORDER PLAYING

Niels Bak

1. Experimental conditions

The investigations to be reported here are concentrated about two main aspects:

1) analysis of the acoustical properties of the tone

2) the physical/physiological aspects, or the way the player blows his instrument.

These two aspects make different demands on room conditions and experimental setup.

The acoustical investigations - based on microphone signals - make special demands on the properties of the recording room as well as on the quality of microphone and tape recorder. All recordings of this kind were made in a sound treated room.

These demands are of secondary importance only during the physical/physiological investigations, which aim at an exact and detailed registration of variations in blowing pressure synchronized with continuous X-ray recordings of the pattern of movements in the mouth and pharynx during play. Internal television equipment was used for the latter recordings.

2. Pilot experiments

Partly in order to gain some experience in the use of unfamiliar equipment, and partly in order to test the validity of information found in current scientific and pedagogical literature, some pilot experiments were first carried out. These experiments confirmed that

1) the humidity of the air, which during natural blowing condenses to water on the inside of the bore of the recorder, has a strong influence on the tone of the recorder. After a very short period of blowing a "dry" recorder the tonal spectra of the notes show an increasing intensity of higher harmonics (keeping blowing pressure and other factors constant).

2) an increase in blowing pressure increases the relative intensity of higher harmonics.

3) when a recorder player repeats a note choosing a new vowel position for every new attack, the blowing pressure is in almost all cases involuntarily changed, even if he takes pains to keep it constant. This applies even to experienced professional recorder players. Thus one subject who was given the opportunity to read the frequency from the frequency counter while playing was unable to prevent the frequency from rising during the articulation of [i] and falling during the articulation of [u]. Simultaneous pressure recordings show that the changes in frequency were due to variations in blowing pressure (cf. Bak 1970). However the variations seem to depend to a considerable extent on the schools and systems employed.

It should be noted that for technical reasons the blowing pressure quoted in Bak (1970) is defined as the air pressure in the mouth cavity just in front of the mouthpiece when the lip opening is not narrowed to a degree which might offer noticeable resistance to the air stream. This reservation must be made since lip resistance - which forces the player to increase the air pressure in the mouth cavity if the pitch is to be kept constant - is a parameter that is difficult to control.

However, since the lip function with the professional recorder player appears to be a more important factor than has so far been assumed, coming investigators are recommended to measure simultaneously the pressure in the mouth cavity and in the windway of the recorder.[1]

3. First series of experiments: Investigation of the influence from mouth and throat cavities on tonal quality and pitch

Much could be said for the use of natural blowing of the recorder, meaning that the recorder is blown by a subject. Thus a number of pedagogues (with whom on the long view a dialogue would be desirable) are notoriously suspicious of how far mechanical blowing provides an adequate substitute to the assistance of a musician. However, the pilot experiments had shown that natural blowing entails a number of problems in technique of measurement - problems which are of less prominence if the recorder is blown mechanically. The latter procedure was therefore used in all cases except in the two first experiments of the series. The whole series of experiments was arranged on the assumption that if, as is commonly held to be true, the musician may influence tone and pitch appreciably by changing the resonances of the mouth cavity, it should be possible to register this same tendency by using a mechanical blowing technique via an artificial mouth. In the first experiment the subject played the same notes three times with changing vowel

[1] Information on the pressure in the windway can be used to calculate the particle velocity of the air stream which is sent towards the lip of the recorder. This is of great importance to the understanding of the acoustical system of the recorder; and by comparing the pressure differences between mouth and windway much information can be gained on the activity of the lips during play, since the fall in air pressure from mouth to windway will indicate the degree of narrowing in the lip opening.

positions, namely [a], [i], and [u]. In this case the subject succeeded in keeping the blowing pressure relatively constant. In the second experiment the resonances of the mouth cavity were changed during every other note by alternatively playing with and without a wad of cotton-wool in the front part of the mouth. The following experiments were all carried out with an artificial mouth.

It would lead too far to describe here all the changes that were made on the artificial mouth during the following (futile) endeavours to affect tone and pitch by varying the resonance conditions of the artificial mouth. By way of examples may be mentioned: variations in the shape of the "mouth" with an artificial tongue made from plasticine; choice of different degrees of resistance between air tube and "throat-mouth" cavity, i.e. at the "glottis"; use of three different mouth cavities, the dimensions of which (see G. Fant 1960) corresponded to the articulation of [a], [i], and [u]. However, it turned out that none of these manipulations had any effect which could be said to support the "resonance theory". Consequently, this series was terminated with two experiments in which the artificial mouth as a resonator had the highest possible effect: With an artificial mouth, having a cylindrical shape, plane end plates, and a length of 150 mm, the note c''' was blown. This note was chosen because its wavelength was twice the length of the "mouth". An extra resonator with an adjustable volume was coupled to the artificial mouth, and the tonal spectrum inside the mouth was monitored via a probe microphone.[2] With the help of the extra resonator the tonal spectrum in the "mouth" could be changed, since the amplitude of a given harmonic could be varied with the setting of the adjustable resonator.

[2] The probe microphone was kindly placed at our disposal by Brüel and Kjær, Nærum, Denmark.

It was found that the pitch and tonal spectrum of the recorder were unaffected by the changes in the tonal spectrum inside the artificial mouth (as conditioned by the adjustable extra resonator) no matter the size of these changes.

A similar experiment was carried out with the note d″. In order to adjust the length of the "mouth-throat" cavity to the wavelength of this note a "mouth" with the extraordinary length of 298 mm (= 1/2 wavelength) was used. The probe microphone showed that extensive modifications could be made in the resonance conditions of the artificial mouth. Thus, inside the "mouth" the fundamental of the tone could reach a sound pressure level of 132.6 dB, and by altering the volume of the extra resonater the SPL of the fundamental could be reduced to 93.6 dB. Even this reduction in the energy of the fundamental had no measurable influence on the tonal spectrum and pitch of the recorder itself.

The experiments indicate that it is not possible to affect the playing of a recorder through variations in the resonance of the mouth cavity.

4. Second series of experiments: Investigation of the activities in mouth and throat during play, using X-ray equipment in combination with recordings of tone and blowing pressure

The experiments were carried out at the X-ray department, Copenhagen Dental Hospital, in the middle of January 1972. Eight recorder players participated, representing all stages of experience from the beginner to the artistic player. While the subjects played a previously determined program a continuous X-ray recording of the movements of the mouth, tongue, and throat was made on a video tape in synchrony with a microphone recording of the play. Furthermore, blowing pressure and microphone signal were recorded with the help of equipment borrowed from the Institute of Phonetics.

As yet, all the recorded material has not been thoroughly analysed, but even now it is evident that these recordings offer exceptional possibilities of analysing recorder playing from quite a new angle.

5. Conclusion

From the results of the experiments it can be concluded that:

1) The "vowel theory" claiming that the mouth cavity has a relatively large volume during the play of deep notes and a reduced volume at high notes (articulation of [u] and [i]/[y] respectively) must be rejected.

2) The resonance conditions of the mouth cavity have no appreciable effect on the pitch of the recorder, just as the tonal spectra recorded in the sound-treated room of the Institute of Phonetics reveal no influence from the resonance of the mouth cavity on the quality of the tone.

3) The variations in blowing pressure during play are of paramount importance to the quality of the play. Not only is vibrato caused by variations in pressure, but even the purity of the play depends on how far every note is intoned with exactly the blowing pressure which is characteristic of it. Especially highly skilled recorder players will change the blowing pressure in a meaningful way with an astounding precision - even during rapid passages.

References

Bak, Niels 1970:	"Pitch, Temperature and Blowing Pressure in Recorder-Playing", Acustica 22 (1969/70), p. 295-299.
Fant, Gunnar 1960:	Acoustic Theory of Speech Production.

Acknowledgements

The research reported in this paper is supported by the Danish Research Council for the Humanities (Statens humanistiske forskningsråd).

Professor Arne Björk at the Copenhagen Dental Hospital has kindly put his X-ray facilities at my disposal, and his photographer has given me highly valuable assistance with the X-ray recordings. The staff at the laboratory of the Institute of Phonetics has been of much help to me in my experimental work. I am particularly indebted to Mr. Carl Ludvigsen for his valuable assistance with the experiments involving use of the artificial mouth. Finally, I wish to thank the recorder players who have kindly allowed me to make X-ray recordings of their play.

6-72