ARIPUC VOL. 12.

ANNUAL REPORT

of the
Institute of Phonetics
University of Copenhagen

COPENHAGEN 1978

ARIPUC VOL. 12.

ANNUAL REPORT

of the
Institute of Phonetics
University of Copenhagen

96, Njalsgade
DK 2300 Copenhagen

COPENHAGEN 1978

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

KØBENHAVNS UNIVERSITET
REPRO. AFDELINGEN

CONTENTS

PERSONNEL OF THE INSTITUTE OF PHONETICS IN 1977

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

Associate professors:

Børge Frøkjær-Jensen, cand.mag. (seconded to the Audiologopedics Research Group)
Jørgen Rischel, dr.phil.
Nina Thorsen, cand.phil.
Oluf Thorsen, cand.mag.

Assistant professors:

Birgit Hutters, cand.mag.
Niels Reinholt Petersen, cand.phil. (temporarily appointed)

Research fellow: Peter Holtse, cand.phil.

Teaching assistants:

Gorm Gabrielsen, cand.stat.
Mimi Jacobsen, stud.mag.
Hans Peter Jørgensen, cand.mag.
Peter Molbæk Hansen, stud.mag.
Ellen Pedersen, cand.mag.

Engineers:

Preben Dømler, B.Sc.
Mogens Møller, M.Sc.

Technician: Svend-Erik Lystlund

Secretary: Else Parkmann

Teachers from other institutes lecturing at the Institute of Phonetics:

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

PUBLICATIONS BY STAFF MEMBERS

Børge Frøkjær-Jensen	"Audiologopædisk uddannelse ved KU", Logos 9
Jørgen Rischel	"The contribution of Louis Hjelmslev", The Nordic Languages and Modern Linguistics 3
Jørgen Rischel and Svend-Erik Lystlund	"The IPUC speech synthesizer", Hamburger Phonetische Beiträge 22
Chr. Berthelsen, I. Kleivan, Fr. Nielsen, R. Petersen and J. Rischel	Ordbogi kalaallisuumit qallunaatuumut (Greenlandic-Danish dictionary), 240 pp.
Nina Thorsen	"On the interpretation of raw fundamental frequency tracings", Preprints from VIII^{èmes} Journées d'Etude sur la Parole, Aix-en-Provence, 25-27 Mai, 1977, p. 175-181.
Nina Thorsen	Talens akustik, skrevet for ikke-teknikere (mimeographed), 120 pp.
Nina Thorsen and Oluf Thorsen	Fonetik for sprogstuderende, 169 pp. (3rd revised edition of Lærebog i fonetik)
Oluf Thorsen, Ole Kongsdal Jensen and Karen Landschultz	Fransk Fonetik, 7th revised edition, 272 pp.

LECTURES AND COURSES IN 1977

1. Elementary courses in general phonetics

One semester courses (two hours a week) in elementary general phonetics (intended for all students of foreign languages except English and French) were given by Peter Molbæk Hansen, Mimi Jacobsen, Hans Peter Jørgensen, Ellen Pedersen, and Nina Thorsen. There was one class in the spring semester and eight parallel classes in the autumn semester.

Courses in general and French phonetics including practical exercises in the language laboratory (three hours a week) were given through 1977 by Oluf Thorsen.

2. Practical exercises in sound perception and transcription

Birgit Hutters gave a course for beginners (two hours a week) in the autumn semester.

Niels Reinholt Petersen and Nina Thorsen gave a course for more advanced students (two hours a week) in the spring semester.

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

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

3. Phonology

Jørgen Rischel and Eli Fischer-Jørgensen gave a two semester course in phonology (two hours a week) through 1977.

Jørgen Rischel gave an introductory course in general linguistics (one hour a week), a course in phonology for advanced students (two hours a week) in the spring semester, and a course in Danish phonetics with emphasis on the phonological aspect (two hours a week) in the autumn semester.

4. The physiology of speech

Birgit Hutters gave a course in instrumental physiological phonetics (two hours a week plus individual exercises) in the spring semester, and a course in the physiology of speech (two hours a week) in the autumn semester.

5. The acoustics of speech

Niels Reinholt Petersen gave a course in instrumental acoustic phonetics (four hours a week plus individual exercises) in the autumn semester.

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

Nina Thorsen and Preben Dømler gave a course in elementary mathematics and electronics in the spring semester.

6. Other courses

Eli Fischer-Jørgensen gave a course in German phonetics (three hours a week) in the spring semester and a course in perceptual phonetics for advanced students (two hours a week) in the autumn semester.

Gorm Gabrielsen gave a course in statistics for advanced students (two hours a week) in the spring semester.

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

Birgit Hutters, Niels Reinholt Petersen and Nina Thorsen presided at a series of seminars for advanced students on topics in experimental phonetics (two hours a week) in the spring semester.

Nina Thorsen gave a course in English phonetics (two hours a week) in the autumn semester.

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

Nina Thorsen and Oluf Thorsen presided at a series of seminars for staff members of the Language Teaching Department of the Danish Refugee Council (Flygtningehjælpens Sprogskole) through January, February, March and October, November.

7. Seminars

Einar Haugen gave a lecture on issues in bilingualism.

Eli Fischer-Jørgensen gave an account of her visits to institutions in Japan and USA.

R.A.W. Bladon lectured on his research in coarticulation

Hajime Hirose gave an account of EMG investigations of laryngeal muscles.

Steen Fibiger lectured on problems concerning hooked wire electrodes for electromyographic investigations.

Victoria Fromkin gave a lecture titled "To err is human" (an account of 'slips of the tongue' in English).

Nina Thorsen gave an account on the identification of intonation contours.

The annual Swedish-Danish phonetics seminar was held at the Institute on November 25-26 with participants from Lund, Odense, Stockholm, Umeå, Uppsala and Copenhagen. The papers read at the seminar are printed in the seminar section of this issue of ARIPUC (p. S-1 - S-112).

8. Participation in congresses, symposia, meetings, etc.

Eli Fischer-Jørgensen visited the following institutions in USA: The Haskins Laboratories, The Bell Telephone Laboratories, M.I.T., University of California at Berkeley, Los Angeles, and Santa Barbara, University of Connecticut, and University of San José, and gave guest lectures at the Bell Telephone Laboratories, at the University of Connecticut, and at the University of San José. She participated in the West Coast Phonetic Symposium at the University of California, Santa Barbara, March 26-28, and gave a lecture on the EMG research of the Institute.

Eli Fischer-Jørgensen gave a guest lecture at the University of Aarhus on problems and methods in perceptual phonetics.

Børge Frøkjær-Jensen participated in the annual meeting of the Danish Audiologic Society, September 12, and gave a paper: "En akademisk overbygning på den audiologopædiske uddannelse".

Birgit Hutters and Børge Frøkjær-Jensen participated in the 17th International Congress of Logopedics and Phoniatrics, Copenhagen, August 15-18.

At a meeting of the Danish Association of Logopedics and Phoniatrics, November 12, Birgit Hutters presided at a panel discussion of problems concerning the clinical use of the Fabre-glottograph.

Jørgen Rischel and Eli Fischer-Jørgensen participated in the International Congress of Linguists, Vienna, August 29 - September 2.

Nina Thorsen and Oluf Thorsen participated in the VIII^{èmes} Journées d'Etude sur la Parole, Aix-en-Provence, May 25-27. Nina Thorsen gave a paper: "On the interpretation of raw fundamental frequency tracings".

Nina Thorsen gave guest lectures on the perception of intonation contours at the University of Lund in June and at the University of Uppsala in October, and lectured on intonation for teachers of the deaf in Copenhagen in May.

Oluf Thorsen participated in a language laboratory seminar at the Institute for Applied and Mathematical Linguistics, University of Copenhagen, April 26-27.

Birgit Hutters and Niels Reinholt Petersen participated in a symposium at the Institute of Linguistics, University of Stockholm, November 2.

The staff and students of the Institute participated in the annual Swedish-Danish Phonetics Seminar, Copenhagen, November 25-26 (cf. p. S-1ff).

INSTRUMENTAL EQUIPMENT OF THE LABORATORY

The following is a list of the instruments that have been purchased or built since January 1, 1977.

1. Instrumentation for speech analysis

1 sonagraph, Kay Elemetrics, type 7029A

2 intensity meters, Fonema

2 fundamental frequency meters, Fonema

2. Tape recorders

2 Semi-professional recorders, Revox, type A77

1 cassette recorder, Tandberg, type TCD 310MK-2

3. Microphones

1 1" microphone, Brüel & Kjær, type 4145

4. Amplifiers

1 measuring amplifier, Brüel & Kjær, type 2607A

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

5. Loudspeakers/headphones

2 loudspeakers, Beovox, type M70

2 headphones, Sennheiser, type 424X

6. Outfit for photography

1 rapidoprint, Agfa, type DD 1437

1 print-drier, Durst, type 400

7. Equipment for EDP

1 disk drive, Plessey, type DD-8/B

ABBREVIATIONS EMPLOYED IN REFERENCES:

AJPs.	American Journal of Psychology
AL	Acta Linguistica
ALH	Acta Linguistica Hafniensia
ARIPUC	Annual Report of the Institute of Phonetics, University of Copenhagen
Folia Ph.	Folia Phoniatrica
FRJ	For Roman Jakobson
F&S	Form and Substance (Akademisk forlag), København 1971
Haskins SR	Status Report on Speech Research, Haskins Laboratories
IJAL	International Journal of American Linguistics
IPO APR	IPO Annual Progress Report
JASA	Journal of the Acoustical Society of America
JL	Journal of Linguistics
JPh.	Journal of Phonetics
JSHD	Journal of Speech and Hearing Disorders
JSHR	Journal of Speech and Hearing Research
Lg.	Language
Ling.	Linguistics
LS	Language and Speech
MIT QPR	M.I.T. Quarterly Progress Report
NTTS	Nordisk Tidsskrift for Tale og Stemme
Proc.Acoust. ..	Proceedings of the ... International Congress on Acoustics
Proc.Ling. ...	Proceedings of the ... International Congress of Linguists
Proc.Phon. ...	Proceedings of the ... International Congress of Phonetic Sciences

STL-QPSR	Speech Transmission Laboratory, Quarterly Progress and Status Report, Royal Institute of Technology, Stockholm
SL	Studia Linguistica
SPE	The Sound Pattern of English, Chomsky & Halle, 1968
TCLC	Travaux du Cercle Linguistique de Copenhague
TCLP	Travaux du Cercle Linguistique de Prague
UCLA WPP	Working Papers in Phonetics, University of California, Los Angeles
Zs.f.Ph.	Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung.

The annual Swedish-Danish Phonetics Seminar was held in Copenhagen on November 25-26, 1977. The seminar was a continuation of similar seminars held at Lund, Stockholm, Uppsala, and Umeå. The aim of the seminars is to provide an opportunity for advanced students and graduate students to present and discuss their current research projects. On the following pages papers given at the seminar are published.

(ARIPUC 12, 1978)

ON SLIPS OF THE PEN

Kerstin Nauclér[1]

Definition

The easiest way to define a slip of the pen is to describe what it is not. A slip of the pen is not a spelling error, caused by deficient knowledge of spelling rules, nor is it a lexical error, i.e. it does not stem from an incorrect interpretation of the meaning or the origin of a morpheme. Spelling and vocabulary belong to the competence, slips belong to the performance. Boomer and Laver (1968) have defined a slip of the tongue as "an involuntary deviation in performance from the speaker's current phonological, grammatical or lexical intention". This definition is applicable to a slip of the pen too. A slip arises somewhere in the program of a linguistic achievement, spoken or written.

Purpose

The purpose of the present study is to find out to what extent, if at all, the writer uses the same linguistic units as the speaker when programming his linguistic performance. In the debate on reading and reading processes there is a discussion about the existence and, if so, the necessity of speech recoding as a mediating stage between visual input and meaning analyses in reading. Recent results (e.g. Kleiman, 1975, and Levy, 1977) support the existence of a speech recoding stage which occurs after lexical access, i.e. word comprehension, and facilitates the temporary storage of words necessary for sentence comprehension. Earlier experiments on reading at the Haskins Laboratories

---

[1] Department of Linguistics, Phonetics Laboratory, Lund university, Sweden.

by Erickson et al. (1973) also point to a phonetic-phonological code as appropriate for temporary storage. (See Nauclér, 1975, for further discussion.) Also in writing it is necessary to keep a number of linguistic units in temporary store while coding them into graphomotoric commands. It is therefore reasonable to assume some kind of phonetic-phonological mediation for writing as well.

From analyses of slips of the tongue it is evident that different kinds of linguistic units are involved on the phonological level, indicating the discreteness of phonological units of different sizes, all of which can play a role in production. (For a survey, see Fromkin, 1973.) Consequently, my first question is: Do slips of the pen support the assumption that we use the same linguistic units to program both speech and writing?

Since both competence in oral language and special writing rules are required for the acquisition and mastering of written language, mastering written language may be regarded as an augmented competence (Weigl, 1972). This augmented competence implies among other things close connections between linguistic expression and graphic performance, groups of graphemes (graphic units) being related to groups of segments by means of correspondence rules in an almost automatic way (Bierwisch, 1972). But before such rules are internalized, beginners have to rely only on their competence of oral language. They must "sound out" the words to be able to write them down. This is apparent in young people's spelling and lexical errors.

Thus, we must assume two programs for writing, one for coding the message into a phonetic-phonological code (available but not necessarily used) and one for converting it into graphomotoric units. A slip of the pen can therefore arise either in the linguistic or in the graphic program. A comparison between linguistic and graphic slips can provide interesting information as to the serial ordering of units of different programs. However, graphic slips are not discussed in this study. As far as possible they have been separated from the linguistic slips. Instead, a comparison between two groups of writers, differing in age, was carried out for the following reasons: On one hand, young people can be expected to make more linguistic slips than adults as they have to depend more on speech recoding. On the other hand, adults plan longer sentences and need all the storing capacity of their short-term memory even if they have internalized the correspondence rules. But since the short-term memory may contain about seven items at a time regardless of size, the young people need the capacity of the short-term memory, too. They fill it with items of smaller linguistic size. Therefore, a difference between skilled and less skilled writers is not likely to show up as a difference in frequency between linguistic and graphic slips. It is more reasonable to assume that the difference in size of linguistic units used for the graphomotoric program can influence the slips of the two groups. My next question is therefore: Is the distance between a slip and its "trigger" shorter for slips made by young people than by adults, the former presumably storing units of smaller size in their short-term memory?

Corpus

The material used for this study consists of compositions written by students of different levels. 190 compositions were written by students in the 4th grade (aged 10-11 years), 190 by the same students when in the 6th grade (now 12-13 years), and 150 by upper secondary school students (12th grade) as a school-leave examination (aged 18-19 years). In this report the two lower grades are treated as one group, called "less skilled writers". The material has certain drawbacks. The writers have obviously corrected all slips they noticed. The slips still remaining are those which are difficult to discover when proof-reading. Not only the number but also the variety of slips has decreased as a consequence. Furthermore, a slip might look just like a spelling error or a lexical error, and since it is impossible now to ask the subjects about their errors, there is no way to be sure of the right category. It seems, however, as if my classifications were made correctly on the whole, as the number of slips does not decrease much from grade 4 to grade 6 (in contrast to the number of spelling and lexical errors). Learning does not affect this type of performance.

Classifications

The material was first divided into deletions, substitutions and additions. Only very few examples of metathesis, corresponding to the classic spoonerism in speech, were found. Deletions are by far the most common type of slip of the pen, four times as frequent as each of the two other types.

Results

The smallest linguistic unit involved in slips of the pen, i.e. the feature, is mainly found in the substitutions. When one grapheme is written for another, there are two possible explanations for this. Either the substitution is graphic, i.e. the graphomotoric commands given were wrong, resulting, for instance, in an upstroke instead of a downstroke. For this to happen the graphemes substituting each other must look alike, as p-b, g-d, d-b, and so on. Or the substitution occurs before the motor programming. In that case the slip is regarded as linguistic and the segments corresponding to the graphemes substituting each other must have linguistic features in common.

As can be seen from Table 1, similar graphemes are not the most frequent substitutes. But if the two segments, corresponding to the graphemes involved, are compared, they turn out to contain the same features but one in about 70% of the substitutions. In the rest of the cases all features but two are shared by the two segments involved. This is in consistency with Fromkin's results from slips of the tongue (personal communication).

Table 1

Numbers of obstruents substituting each other

Grade	p-b	t-d	k-g	b-d	t-k	g-d	k-p	g-t	k-d	v-f	p-f	p-v
4	2	6	2	2	2				1			
6		4	1					1		2	1	
12	3	2	4			1	1					1
	5	12	7	2	2	1	1	1	1	2	1	1

What is the reason for writing another grapheme than the intended one? It seems as if the substitutions are influenced mostly by a similar segment in the immediate context (syntagmatic assimilation). This sometimes leads to partial identity between the segments

(1) P har under senare tid tagits upp i diskussionen → takits
(P has lately been brought up in the discussion)

but in most cases to complete identity

(2) ej bör vänta med att ingå äktenskapet → äktenspapet
(should not wait to get married)

(3) nu styrde maskinen ut ur folkhopen → folkhofen
(now the machine steered out of the crowd)

(4) vi fick ta upp den igen → vick
(we had to pick it up again)

The "trigger" is not necessarily found in the syntagm. There are many examples of paradigmatic influence

(5) Äktenskapet bygger till 75% på → by<u>ck</u>er[1)]
(Marriage is based on ... to 75%)

(6) Alla fyra grupperna har följande uppgifter → fö<u>r</u>jande
(All four groups have the following tasks)

It can be argued, of course, that in (2), (3) and (4) the substitutions are graphic anticipations and duplications, but if that is so, there is no sufficient explanation for the fact that the substituted segments have all the phonological features but one in common with the segment substituting them and no striking graphic similarity. It should also be emphasized that vowels and consonants do not substitute each other, a fact which can only be accounted for linguistically.

Features expressing manner of articulation are more frequently exchanged than features expressing place of articulation by both groups. But if the feature VOICED is disregarded the skilled writers substitute fewer "places of articulation" than "manners of articulation". The opposite is true for the less skilled writers, a somewhat puzzling finding.

The voicing feature is the most common one substituted among the paradigmatic slips. This is not the case among the syntagmatic ones. The exchange of value of the feature VOICED might as well be a lexical error as a slip, caused by deficient knowledge of the morpheme in question, since it has been shown, i.a. by Simon (1975) that the distinction voiced/voiceless is mastered only late in the acquisition of language. However, as the feature VOICED is the most frequent one substituted paradigmatically by both groups of writers in this study, there is hardly any reason to doubt its classification as a slip.

In a few cases the substitutions are not influenced by a single segment but by a sequence of segments in the paradigm similar to the intended one

---

1) Swedish spelling rules do not permit ⟨kk⟩ . The rule ⟨k⟩ → ⟨c⟩/_⟨k⟩ has been applied after the shift from VOICED to VOICELESS.

(7) anpassningsproblem med den → anpassningsproblev
(adaptional problems to the) "blev" is the past tense
of a modal auxiliary

A single segment turns out to be most often involved in additions. The following examples are regarded as linguistic slips due to the context in which the graphemes are added

(8) en svag lukt → svagt [svɑːkt lukt]
(a slight smell) 't' is an inflection suffix

(9) naturligtvis [natɯrliktvis] har ökat → öktat
(naturally have increased)

(10) smärtorna på Bosse → spå
(the pains on Bosse) 'Bosse' is a proper name. The syntactic structure of the phrase is as bad in Swedish as in English.

In some cases of additions where the linguistic influence is not obvious it may perhaps be more appropriate to consider the slip as graphic, i.e. a graphomotor command has been released twice

(11) rätt nöjd faktiskt → fakstiskt
(quite satisfied, as a matter of fact)

Also among the additions are examples of external influence from similar sequences, syntagmatic as well as paradigmatic

(12) en gräsklippare berättar → berättare
(a lawn-mower tells) '-are' is a derivative suffix meaning "the one that/who"

(13) började det plötsligt osa bränt → dosa
(it suddenly started to smell of smoke), "osa" is a verb, "dosa" is a noun meaning "box"

Less skilled writers have a certain tendency to simplify the structure of the word by inserting a vowel in a consonant cluster. The inserted vowel is an anticipation of the following one

(14) borta bra men hemma bäst → bara
(proverb: there is no place like home)

(15) sen dess tycker jag illa om piggsvin → piggsivin
(ever since I dislike porcupines)

The same effect, i.e. a less complicated consonant structure, is obtained when a consonant is deleted in a cluster, a phenomenon which dominates the deletions. Whenever a consonant is left out, it is deleted from a cluster. The only exception to this "rule" is single consonants in word final positions. Deletions in clusters can be regarded as a way to make sequences of linguistically similar segments (consonants) simpler. But it can also be ascribed to the fact that consonants in clusters very often share the same place of articulation. In this case, the similarity leads to deletion and not to assimilation as was the case with the substitutions.

As can be seen from Table 2, both skilled and less skilled writers delete more than half of all consonants in medial clusters. Only the less skilled writers delete in initial clusters, whereas the skilled ones are more inclined to do so in final clusters.

Table 2

Deletion of consonants in clusters (%)

Grade	Position of cluster init.	med.	final
4	30	52	18
6	14	51	35
12	2	55	43

Emphasizing some differences between the two groups we must try to explain them. The fact that less skilled writers unlike skilled ones delete consonants in initial clusters has probably to do with the graphic units stored by the skilled writers. Such units are certainly more easily acquired for the first part of a word than for any other part, the structure of the first part of a word being constant in contrast to the final part, where the structure is changed by inflectional suffixes.

When a skilled writer deletes a single consonant at the end of a word (i.e. the consonant is not part of a cluster), the consonant is in most cases an inflection, but when deleted by a less skilled writer it is more often part of the stem. Consequently, it seems to be merely a matter of position when a less skilled person deletes a consonant at the end of a word, whereas in the case of a skilled writer one might assume a strategy implying a programming of lexical and grammatical morphemes separately. The latter morphemes, being more redundant, may sometimes be overlooked.

It is apparent from what has been said about deletions in consonant clusters that, in contrast to substitutions and additions, no segments (or sequences of segments) of similar phonological quality triggering the deletions are found in the context. The occurrence of two or more consonants in sequence seems to be enough to cause a deletion. This is one of the few cases where there is not a certain distance between the trigger and the slip. In the substitutions and additions there are some examples of the trigger and the slip being separated by a word boundary only. In the group of the less skilled writers there are also a few examples of trigger and slip in the same syllable, the distance between them amounting to no more than one or two segments.

On the whole, both groups favour the distance of one syllable between the trigger and the slip. Less skilled writers prefer the syllables involved to be part of the same word. No distance of more than six syllables is observed, once in each of the three grades.

Conclusions

Since it is apparent from the slips of the pen examined in this study that both features, segments and morphemes are involved when comparing errors, intentions and triggers, the assumption that we use the same linguistic units when programming speech and writing seems to be correct. The hypothesis that the distance between the trigger and the slip ought to be greater when the slip is made by a skilled writer, was confirmed inasmuch as the trigger and the slip were never found in the same syllable in errors made by the skilled writers. This was sometimes the case with errors made by less skilled writers. Moreover, in most cases concerning less skilled writers the trigger and the slip were found in the same word. For the skilled writers they were more often found in different words.

Summary

A phonetic-phonological coding is used, not only by less skilled writers but also by skilled ones. The difference between the two groups of writers lies mainly in the size of the linguistic units used as input to the graphomotor program and in the number and size of the graphic units internalized.

References

Bierwisch, M. 1972:	"Schriftstruktur und Phonologie", Probleme und Ergebnisse der Psychologie 43, p. 21-44
Boomer, D.S. and J.D.M. Laver 1968:	"Slips of the tongue", British Journal of Disorders of Communication 3 (1)
Erickson, D., I. Mattingly and M. Turvey 1973:	"Phonetic activity in reading: An experiment with kanji". Haskins SR-33
Fromkin, V. 1973:	Introduction. In Fromkin (ed.), Speech errors as linguistic evidence.
Kleiman, G.M. 1975:	"Speech recoding in reading", Journal of Verbal Learning and Verbal Behaviour 14, p. 323-339
Levy, B.A. 1977:	"Reading: Speech and meaning processes", Journal of Verbal Learning and Verbal Behaviour 16, p. 623-638

Nauclér, K. 1975: "Some thoughts on reading and writing", *Phonetics Laboratory, Department of General Linguistics. Lund University. Working Papers* 10, p. 129-138

Simon, C. 1975: "The development of some speech patterns in children", *Paper presented at the VIIIth International Congress of Phonetic Sciences*. Leeds

Weigl, E. 1972: "Zur Schriftsprache und ihrem Erwerb: Neuropsychologische und psycholinguistische Betrachtungen", *Probleme und Ergebnisse der Psychologie* 43, p. 45-105.

(ARIPUC 12, 1978)

STØD AND SYLLABICITY IN A JUTLANDIC DIALECT

Peter Molbæk Hansen[1]

Abstract: The paper is concerned with the Himmerlandic dialect in North East Jutland. Certain phonological structures which are ambiguous with regard to occurrence of syllabicity, namely long vocoids and certain sequences of sonorant segments, are discussed. It is argued - on the basis of certain prosodic phenomena (the behaviour of stød and stress) - that from a functional point of view these structures are best described as disyllabic. A brief discussion of the possible connection between the Jutlandic apocope and the behaviour of these structures is given.

1. Introduction

This paper is concerned with some problems connected with the identification of syllabicity in a dialect spoken in Himmerland, a region in North East Jutland. In the following I shall refer to this dialect - which is my own first language - as H. The main problem is the following: in most cases there is general agreement among native speakers of H (including myself), and also among dialectologists who have dealt with H or related dialects, on the number of syllables in a given word. Thus, words such as [nɛi̯’] 'no', [sdɛi̯] 'place', [hɑu̯] 'sea', [han] 'he', [u̯an’] 'water', and [lan’] 'country' are considered by everybody to be monosyllabic; and words such as [nóːln̩] 'the needle', [nó·ln̩] 'the needles', [sɑ́mlɒ] 'gathers', [kɛ́lɒ] 'basement', [kánɒ] 'jugs', and [sdœ́u̯ɒ] 'spreads dust' are considered by everybody to be disyllabic. This is in fact implied by the above transcriptions in which nonsyllabic vocoids and syllabic contoids are marked off in accordance with the usual IPA transcriptional

---

[1] Institute of Phonetics, University of Copenhagen.

practice. There are, however, certain structures which may be called ambiguous as regards the number of occurrences of syllabicity. These structures may be grouped as follows:

$$\left.\begin{array}{ll} \text{a)} & V{:} \\ \text{b)} & V\cdot\text{ɒ} \\ \text{c)} & V_h\text{ə}\cdot \\ \text{d)} & V_h\cdot\text{ə} \\ \text{e)} & V_h\text{ə}C_s \\ \text{f)} & V_h\cdot C_s \\ \text{g)} & VV_hC_s \\ \text{h)} & VC_sC_s \end{array}\right\} \quad \underline{\text{followed by pause or by a nonsyllabic segment}}$$

where V = vocoid; V_h = high vocoid; C_s = sonorant contoid.

Some examples of words containing these structures are given below; the sequences corresponding to the syllabically ambiguous structure types above are underlined. $\frown$ denotes a particular rising-falling tonal movement (weak stød, cf. below).

a)	$[\text{b}\underline{\overset{\frown}{\text{o}}\text{:}}\text{s}]$ 'stall'	$[\text{l}\underline{\text{ø:}}\text{s}]$ 'read'	$[\text{it}\underline{\overset{\frown}{\text{a:}}}\text{li̭ən}]$ 'Italy'
b)	$[\text{g}\underline{\text{o·ɒ}}]$ 'farms'	$[\text{t}\underline{\text{u·ɒ}}]$ 'tours'	$[\text{f}\underline{\text{i·ɒ}}]$ 'four'
c)	$[\text{h}\underline{\text{iə·}}]$ 'heath'	$[\text{h}\underline{\text{uə·}}\text{s}]$ 'sock'	$[\text{g}\underline{\text{yə·}}]$ 'fertilizer'
d)	$[\text{m}\underline{\overset{\frown}{\text{i·ə}}}]$ (personal name)	$[\text{tr}\underline{\text{u·ə}}\text{s}]$ 'is threatened'	
e)	$[\text{d}\underline{\overset{\frown}{\text{iəl}}}]$ 'part'	$[\text{s}\underline{\overset{\frown}{\text{uəl}}}]$ 'sun'	
f)	$[\text{b}\underline{\overset{\frown}{\text{i·l}}}]$ 'car'	$[\text{s}\underline{\overset{\frown}{\text{u·l}}}]$ 'meat'	$[\text{p}\underline{\overset{\frown}{\text{i·l}}}]$ 'arrow'
g)	$[\text{sg}\underline{\overset{\frown}{\text{ɔul}}}]$ 'shovel'	$[\text{t}\underline{\text{ɛin}}]$ 'draw'	
h)	$[\underline{\text{ɛlm}}]$ 'elms'	$[\text{s}\underline{\text{alm}}]$ 'hymn'	$[\text{f}\underline{\text{ɑmlg̥}}]$ 'fumbled'

The problem with these forms is that native speakers of H (including myself) do not seem to have any firm intuition concerning the number of syllables in these particular types, cf. above. Personally, I tend to perceive them as disyllabic, but I am not sure that I am a reliable informant: during my work with my own native dialect I have become increasingly aware that my perception of these things is somewhat influenced by my phonological analysis. Unfortunately, it is not very easy, either, to elicit reliable information on the perception of syllabicity in these particular structures from other (linguistically more naïve) speakers of H. In conversations with informants I have often experienced that one and the same person stated, within less than half an hour, that 1) the words degn 'parish clerk' and dejen 'the dough' (rendered here in normal Danish spelling) sound exactly the same,[1] 2) that the word degn is monosyllabic, and 3) that the word dejen is disyllabic. This does not, of course, show that the syllable is irrelevant, but rather that the term 'syllable' is too tied up with the orthographic tradition (of another dialect, namely Standard Danish (SD)), and that spelling rules learned in school are applied in such situations.

In phonetics textbooks one often finds statements to the effect that native speakers of a language can, as a rule, "count syllables", i.e. they can identify the number of syllables in a given utterance. Although there has been a tendency for it to be axiomatized, it should be remembered that this claim is, in principle, an empirical hypothesis (unfortunately, little work has been done to test this hypothesis; Bell, 1975, may be mentioned as an example; note that his results cannot be said to support the hypothesis). It must be remembered, too, that the very content of this hypothesis is open to several interpretations, i.e. the whole issue depends crucially upon what is meant by the term 'syllable'.

2. The phonological behaviour of syllabically ambiguous structures

Whatever the perceptual status of the structures a) through h), there are good arguments for describing them as functionally

---

[1] This seems beyond dispute; in my raw transcription they would both be rendered [dɛ͡in].

disyllabic. This may best be illustrated by the behaviour of stød in H. Like many other Danish dialects H is a stød-dialect, i.e. the stressed part of some words is characterized by a distinctive glottal modification (the stød). In SD the stød is not normally considered to have systematic allophonic variants, whereas in H two clearly distinguishable types of glottal modification are in complementary distribution within the simple word, cf. below. I shall refer to these two stød-types as strong and weak stød. Impressionistically, the strong stød is very similar to the SD stød, whereas the weak stød is more aptly described as a rising-falling tonal movement, the last portion of which may be more or less glottalized. In my transcriptions I shall designate the weak stød by the symbol ⌢ over the sequence of segments modified by it (cf. the transcriptions in section 1.). The strong stød I shall designate by the symbol ' after the segment that is most clearly glottalized (this is the traditional way of transcribing the SD stød).

Morphophonemically, the strong stød and the weak stød are closely related, cf. e.g. [$\overset{\frown}{\text{no:l}}$] - [no'ln̩] 'needle' - 'the needle', and from a generative viewpoint[1] the occurrence of one or the other of the two types is entirely predictable if they are identified on an abstract level (whether this abstract stød be underlying or inserted by rule). However, I shall be concerned here with the phonological behaviour of the two types of stød in the surface phonological structure of short utterances (simple words). From this point of view the two types of stød are in complementary distribution; and if, as seems natural, the strong and weak stød are identified *phonemically*, their distribution must be stated in terms of *phonemic* conditions. However, the formulation of these conditions may vary drastically depending on whether or not reference to syllabification is allowed, and depending upon whether the structures a) through h) are interpreted as mono- or disyllabic. Let us assume first that the structures a) through h) are monosyllabic. The necessary conditions for the two types

---

1) I.e., from the point of view of the predictability of the phonetic realization of a string of morphophonemic entities, including grammatical boundaries.

of stød to occur may then be formulated like this: 1) The weak stød may occur on one of the structures a) through h) or on one of the disyllabic structures $V\underset{\wedge}{V}V$ and VC_sV, provided that none of the structures is followed by a syllabic segment. 2) The strong stød may occur on one of the structures V·, $V\underset{\wedge}{V}$, and VC_s, provided that these structures do not meet the conditions under which the weak stød is possible. Thus stød-words like [nɛi̯’] 'no', [lɛi̯’f] (a personal name), [han’sn̩] (a surname) have the strong stød (the conditions for the occurrence of weak stød are not met); stød-words like [ma͡nl] 'almond', [ita͡:li̯ən] 'Italy', [o͡:lsn̩] (a surname), [brɔ͡u̯n] 'the bridge' 'brown', [la͡nə] 'the country', [hɑ͡u̯ə] 'the sea', on the other hand, have the weak stød (although the sequences V·, $V\underset{\wedge}{V}$, and VC_s occur, the conditions for the weak stød to occur are met; therefore, the strong stød is excluded); finally, stød-words like [hɛi̯’nə] 'the fence', [bi:’ln̩] 'the car' have the strong stød (the structures which might otherwise carry the weak stød are followed by a syllabic segment in these examples). These formulations, although observationally correct, are obviously quite unrevealing and unnatural from a phonological point of view: for one thing, if, as assumed here, the structures a) through h) are monosyllabic, it is strange that they should be equivalent (in relation to the stød) to the structures $V\underset{\wedge}{V}V$ and VC_sV which are undoubtedly disyllabic, cf. above. This might suggest that the syllable is irrelevant to the manifestation of the stød, but that seems to be contradicted by the fact that the occurrence of a <u>syllabic</u> segment after the structures in question is crucial for the manifestation of the stød.

However, a close inspection of the structures a) through h) and the structures $V\underset{\wedge}{V}V$ and VC_sV, i.e. the structures which may carry the weak stød if they are not followed by a syllabic segment, will reveal that a long vowel is equivalent to a short vowel followed by two sonorant segments, and that a half-long vowel is equivalent to a short vowel followed by one sonorant segment in relation to the weak stød; the sequences which may carry the strong stød, namely V·, $V\underset{\wedge}{V}$, and VC_s, show that in relation to the strong stød a half-long vowel is equivalent to a short vowel followed by one sonorant segment. This suggests an analytical interpretation of length, that is V· = $V\underset{\wedge}{V}$, and V: = $V\underset{\wedge}{V}\underset{\wedge}{V}$ (if monosyllabic), or V: = $V\underset{\wedge}{V}V$ (if disyllabic, cf. below).

Under this interpretation the formulation of the stød-conditions will have to run like this (on the assumption, still, that the structures a) through h) are monosyllabic): 1) the stød requires for its occurrence a sequence of sonorants, the first of which must be a vowel; this sequence must contain at least two segments. 2) if the sequence of sonorants contains only two segments, the stød will be of the strong type. 3) if the sequence of sonorants contains three segments, the stød will be of the weak type. 4) if the sequence of sonorants contains more than three segments, the stød will be of the weak type provided that the fourth segment is nonsyllabic; otherwise the stød will be of the strong type. This formulation, too, fails to account for the fact that disyllabic structures like $V\underset{\wedge}{V}V$ and VC_sV behave like the allegedly monosyllabic structures a) through h), although it seems somewhat less unnatural than the first formulation.

If, instead, we assume 1) that the structures a) through h) are all disyllabic, and 2) that syllabification is relevant to the distribution of the two types of stød, the utterances under investigation must of course be syllabified on independent grounds. I shall not discuss the general problems connected with syllabification (for a detailed discussion of such problems, see Basbøll, 1974); suffice it to mention that there are at least two types of criteria which have generally been considered important in this respect, namely 1) universal (phonetically oriented) tendencies (e.g. two intervocalic consonants the first of which is sonorant, are normally heterosyllabic), and 2) language specific distributional criteria (above all: syllable initial and syllable final segment combinations should correspond to (structurally) possible word initial and word final segment combinations, respectively) (see also Pulgram, 1970).

Now, according to both types of criteria the structures which may carry the strong stød show a striking similarity: these words can all be syllabified in such a way that the structure which carries the strong stød (a vowel + one sonorant, nonsyllabic segment) is homosyllabic, cf. /hɛi̭'-nə/, /nɛi̭'/, etc. If, as suggested above, the structures a) through h) are interpreted as disyllabic and if, in addition, the analytical interpretation of vowel length is maintained, these structures can be brought together with the undoubtedly disyllabic structures VV̭V and VC_sV under the common formula V́CV where V designates any syllabic segment, and C designates any nonsyllabic segment. It may then be stated that the structures which may carry the weak stød share the property of having no well defined internal syllable boundary. In languages (like the Germanic ones) with heavily stressed syllables contiguous with unstressed syllables, single intervocalic consonants are often described as ambisyllabic, and this interpretation can be applied to the dialect in question, cp. [la͡nə] and, e.g., English words like <u>bitter</u>; the same is true of single consonants separating a stressed vowel and a syllabic consonant, cp. [ma͡nl] and, e.g., English words like <u>little</u>. In the case of long vocoids forming two syllables (hiatus), cp. [bo͡:s], there is no well defined internal syllable boundary either. We may speak of overlapping syllables in all such cases (cf. Rischel, 1964; Pike, 1947, p. 65, 90). Under the disyllabic interpretation of the structures which may carry the weak stød, the latter may be said to be the manifestation of the stød in cases of overlap between the stressed syllable and the following unstressed one. It is not clear to me whether such a mechanism can be said to be natural; anyway, the very possibility of bringing the structures carrying the weak stød together under a common and typologically plausible structural description seems to me of interest. This could, of course, also be done by claiming that the structures which have hitherto been transcribed VV̭V and VC_sV are functionally monosyllabic; as I have repeatedly mentioned, however, they are undoubtedly phonetically (perceptually) disyllabic; this is further supported by typological considerations: to say that a word such as [la͡nə] is monosyllabic amounts to accepting that homosyllabic vocoids may be separated by a contoid, and this must be considered implausible from a typological point of view: such an interpretation should only be accepted if there is strong language specific evidence for it. I have not found such evidence.

The disyllabic status of the structures a)-h), on the other hand, is typologically and phonetically plausible, since it is in good agreement with the sonority principle, which is known to play a role in the syllable structure of many languages. Moreover, there are some language specific arguments in favour of this interpretation. For reasons of space I shall only discuss one (probably the strongest) of these arguments here. In prepositional constructions (preposition + noun), clearly monosyllabic prepositions like [i] 'in' and [te] 'to' are always[1] unstressed if the first syllable of the following noun is stressed, and they are stressed if the first syllable of the following noun is unstressed, cp. [íhobʁɔ́ṷʼ] 'in Hobro' (name of town) and [iʁɑ́nɒs] 'in Randers' (name of town). Prepositions like [œ͡ṷɒ] 'over', [u͡nɒ] 'under', [uðn] 'without', and [o:] 'off (with locative meaning)' do not, however, obey this rule: they are always stressed, irrespective of the stress contour of the following noun, cp. [œ́͡ṷɒʁɑ́nɒs] 'over Randers' and [ó:ṷɔ́ṷʼnn̩] 'off the wagon'; this behaviour is also characteristic of prepositions of the structure $VC_{obs}V$, e.g. [ɛ́tɒ] 'after'. This clearly points to a disyllabic status of the structures a)-h): if [o:] were monosyllabic, it would form an exception to an otherwise quite general rule.[2]

---

1) Except for special cases of emphasis.

2) The stress pattern of prepositional constructions seems to be a special case of a more general tendency towards a trochaic phrase rhythm in H.

3. The Jutlandic apocope

Like all Jutlandic dialects, H is - from a dialectological point of view - characterized by the Jutlandic apocope, i.e., in these dialects the unstressed vowels of Old Danish have been deleted in word final position. In Eastern Danish dialects, upon which SD is mainly based, these vowels have been reduced to schwa instead, cp. the following correspondences between SD and H (the qualitative differences between mutually corresponding segments are irrelevant in this connection):

	SD		H
(1)	[bɔ́:sə]	'stalls' (sb. pl.)	[bo:s]
	[dái̭nə]	'parish clerks'	[dɛin]
	[ɛ́lmə]	'elms'	[ɛlm]
(2)	[lánə]	'countries'	[lan]1)
	[tɑ́g̊ə]	'thank' (vb. inf.)	[tɑg̊]2)

It has often been assumed (more or less implicitly) by (some) Danish dialectologists that the loss of a vowel entailed the loss of a syllable. This is undoubtedly true of words like those in (2) above. If, however, as I have suggested, H words like those in (1) above are disyllabic, two hypotheses concerning the historical development suggest themselves: 1) such words remained disyllabic after the apocope; 2) such words became monosyllabic as a result of the apocope but were later reinter-

---

1) Aged speakers of the (eastern) variety of Himmerlandic, here referred to as H, may distinguish words such as [kan] 'can' [kan·] 'jug'. Both these words are distinct from [kann] (in a narrow transcription perhaps [kan:]). It is hardly to be doubted that the latter word is perceived as disyllabic, whereas the two former words are both perceived as monosyllabic. In my own speech the two former words cannot be distinguished.

2) In the Western varieties of Himmerlandic, such words have the so-called West Jutlandic stød, cf. Ringgaard 1960.

preted as disyllabic. It is interesting that stød words with a segmental structure like that of the stød-less words in (1) may be disyllabic (since e.g. [ɛ͡lm] and [ɛlm] are only distinguished by presence vs. absence of weak stød, they must be equisyllabic; it would seem far-fetched to claim in a *synchronic* phonology of modern H that presence of (weak) stød is a "phonetic manifestation of monosyllabicity" or the like. Thus, if the above words are disyllabic, the corresponding stød words [bo͡ːs] 'stall' (sb. sg.), [dɛ͡in] 'parish clerk' and [ɛ͡lm] 'elm' are also disyllabic (I am not concerned here with the problem of how to transcribe these forms adequately)). These stød words were undoubtedly monosyllabic prior to the apocope. (On the apocope, see Ringgaard, 1963.) If hypothesis 1) above is correct, the result of the apocope may have been almost the opposite of what is normally assumed for H words which today have a segmental structure like those in (1) above: it may be imagined that the apocope did not cause originally disyllabic words *of this type* to become monosyllabic; rather it may have been an indirect result of the apocope that originally monosyllabic words like [bo͡ːs] etc. were reinterpreted as disyllabic.

References:

Basbøll, Hans 1974:	"The phonological syllable with special reference to Danish", *ARIPUC* VIII, p. 39-128
Bell, A. 1975:	*If speakers can't count syllables, what can they do?* (Indiana University Linguistics Club)
Pike, K.L. 1947:	*Phonemics* (Ann Arbor)
Pulgram, Ernst 1970:	*Syllable, word, nexus, cursus* (Janua Linguarum, Series Minor. 81. The Hague - Paris)
Ringgaard, K. 1960:	*Vestjysk stød* (Århus)
Ringgaard, K. 1963:	"The apocope of disyllables", *Phonetica* 10, 3-4, p. 222-229
Rischel, J. 1964:	"Stress, juncture and syllabification in phonemic description", *Proc. 9 int. cong. ling.*, p. 85-93

(ARIPUC 12, 1978)

"STØD" AND UNEXPECTED SUBSTITUTIONS IN THE SPEECH OF DANISH 2-YEAR-OLD CHILDREN

John Jørgensen[1]

1. Introduction

The aim of the present paper is to add new data - in the form of observations - to the study of child language. Danish being the maternal language of the children, it seemed natural to pay special attention to a phenomenon which is characteristic of Danish, viz. the "stød". Danish "stød" can be characterized as a sort of phonemic accent with the syllable as its domain. A phonological description of "stød" is given by e.g. H. Basbøll (1969). It is well known (to Danes at least) that foreigners learning Danish have great difficulties both in producing the "stød" and in placing it correctly. But is that also the case with Danish children learning their mother tongue?

Another feature which attracted my attention during my survey of the material was the children's substitutions of other sounds for certain sounds of the adult language. The substitutions were unexpected in the sense that they did not agree very well with what might be predicted from the hypotheses proposed by Roman Jakobson in his fundamental work on this topic (1941).

According to Jakobson, stops are acquired before fricatives (and therefore stops are substituted for fricatives, whereas the reverse substitution is not possible). Among the fricatives /s/ is claimed to be the most fundamental and, accordingly, it should not be possible to find any phonemic system where /s/ is lacking.

---

[1] Institute of Phonetics, University of Copenhagen.

Jakobson does not directly mention "stød" in his work but, using nasalized vowels as an example, he states that phonemic phenomena which are found only in few languages, are acquired very late by children.

2. Material[1]

The material on which my observations are based consisted of tape recordings of four two-year-old children's non-imitated speech. About 125 phrases (containing 2-3 syllables on an average) were recorded for each child.

All recordings were transcribed in relatively narrow phonetic transcription. About one fifth of the phrases were transcribed in very narrow phonetic transcription.

3. Results

3.1 "stød"

Out of a total of at least 1000 syllables "stød" was incorrectly treated by the children in only 41 instances, i.e. less than 5 percent. In order to decide whether this surprisingly good result was caused by the fact that the children had avoided phrases containing "stød" in the target language, the percentages of phrases with and without "stød" was calculated. 40 per cent of all phrases contained one or more occurrences of "stød" and 60 per cent contained no "stød". These figures agree quite well with what is found in adult language. The distribution of "stød"-mistakes among the subjects was relatively uniform, viz. MI 11, KR 12, NO 11, and TI 7.

A quantification of the mistakes into 3 groups gave the results shown in table 1.

---

1) A detailed description of the material and how it was collected will appear in the next issue of ARIPUC.

Table 1

Distribution of "stød" errors

Lack of "stød" on a syllable which should have the "stød"	25
"Stød" on a syllable which should have none	10
"Stød" on wrong segment of the syllable	6

It should be added that the syllables with "stød"-mistakes seemed to have no specific phonological structure in common.

3.2 Substitutions

Only two types of substitutions will be dealt with here. One concerns the phonetic realisation of the phoneme /s/, the other the realisation of the phoneme /k/. With a very few exceptions the /s/ phoneme was realised by all four children as a non-sibilant fricative with its place of articulation somewhere between the dental and pre-palatal regions. As no palatograms or other physiological data were acquired, the above articulatory description can only be a qualified guess. It has, however, been confirmed by other trained phoneticians who listened to the tapes. The sounds produced by the children may be transcribed [θ] or [ç$^{+}$].

As for the realisation of the /k/ phoneme there seemed to be very little, if any, consistency among the subjects. One child correctly realised /k/ as [g̊^{h}]. Another substituted a dental stop for /k/. Still another substituted the fricative [x] for /k/. And the fourth child realised /k/ as a retracted [g̥$^{h-}$] or a relatively advanced [g̊$^{h+}$].

4. Discussion

I shall not attempt to give a conclusive explanation of the discrepancies between Jakobson's predictions and my observations on the linguistic behaviour of a very limited number of Danish children. A few points will be mentioned, however, which may deserve being taken into consideration in discussions of Jakobson's hypotheses:

a) One possible explanation of the early acquisition of the "stød" may be that the "stød" is treated as a prosodic phenomenon by the children. It is well known that prosodic phenomena (e.g. intonation) are acquired very early.

b) The [s] — [θ] substitution may reflect the children's inability to produce an [s] with its characteristic and complex tongue shape. Thus the substitution may be considered a purely phonetic one, and it is worth noticing that none of the children had a phonemic opposition between /s/ and /θ/.

c) The substitution of a fricative for /k/ may be explained partly by the fact that Danish /k/ is strongly aspirated (and affricated), and partly by the fact that it is more difficult to make fine adjustments with the back of the tongue than with the tip and blade.

5. Final remarks

It should be stressed that the value of the above considerations is strongly limited by the small amount of data on which they are based. Moreover, there is an unfortunate scarcity of data from other studies to compare with. Especially the very first steps in language acquisition are almost unexplored.

References

Basbøll, Hans 1969:	"The phoneme system of advanced standard Copenhagen", ARIPUC 3, p. 33-54
Jakobson, Roman 1941:	Kindersprache, Aphasie und allgemeine Lautgesetze

(ARIPUC 12, 1978)

ON THE PRODUCTION OF THE SWEDISH TJE-SOUND

Per Lindblad[1]

Introduction

The Swedish tje-phoneme in words such as *kivas* and *tjuta* have two main variants; being, firstly, an affricated pronunciation that can be written as [tç] and, secondly, a fricative pronunciation usually transcribed as [ç]. Both variants have a certain latitude of pronunciation. The affricate is the normal pronunciation in Fenno-Swedish and is found to some extent in a number of dialects in Sweden. The affricate was the dominant pronunciation in Sweden, even up to the early 1900's, but today nearly all Swedes use the fricative variant.

The phonetic value of the symbol [ç] is exemplified in IPA as the German *ich*. The normal Swedish tje-phoneme and the typical German [ç] are, however, distinctly different, while the Polish *ś* and the Swedish tje-phoneme are very similar. The IPA symbol of the Polish *ś*-sound is [ɕ]. In the following paragraphs I will use the IPA symbol [ɕ] to clearly accentuate the special pronunciation of the Swedish tje-sound and also to establish connections with IPA.

Both [ɕ] and the palato-alveolar [ʂ, ʃ] can be produced, as is known, by different articulatory gestures. It is possible that the specific articulations for [ɕ] also vary with different speakers, perhaps because, just as for [s] and [ʂ, ʃ], the distinctive high frequency components of [ɕ] are determined much more by the dimensions of the front cavity than by the maximum contact area. Since the shape and size of the tongue, tooth ridge, teeth and lips may vary from speaker to speaker, a given size of the front cavity may be obtained by different gestures in different speakers. Another important quality of the sounds

---

[1] Department of Linguistics, University of Gothenburg, Sweden.

in question, their sharpness, is produced primarily by a concentrated airstream directed by the tongue against the teeth. Speakers with different shapes of the anterior part of their vocal tracts would achieve this effect in different ways.

Despite the variation between different speakers in the pronunciation of [s] and [ɕ, ʃ] it seems reasonable to assume that articulatory gestures can be grouped into a limited number of gesture types, and that a certain gesture type would be the norm for a certain sound within a language, and probably in a number of (or perhaps even most) languages having the sound quality in question. For example, [s] in Swedish is evidently normally pronounced with the tip of the tongue resting against the lower incisors, and with the tongue blade against the alveolar ridge immediately behind the upper teeth (Fritzell 1973, p. 41; Söderpalm Talo 1976, p. 16). A recent investigation (Bladon and Nolan 1977) has shown the same type of gesture to be predominant in English.

The normal articulation for Swedish [ɕ] is probably the one shown in figs. 1 and 2. I have observed this articulation in two X-ray filmed subjects.

Design and Material

Figs. 1 and 2 illustrate the tje-pronunciation of two Southern Swedish male subjects, called hereafter A and B. They were X-ray filmed in profile while uttering the phrase

EBBE SCHISE TJASAR I SODASJO
V_1 V_2 V_3 V_3

and five other versions of that utterance, where the marked vowels were systematically varied while keeping the other sounds constant. Each of the vowels [i:,ɑ:,u:,ɛ:,y:,ɯ:] appeared once in the respective V_1, V_2 and V_3 positions in the six phrases. Thus [ɕ] appears in the following context:

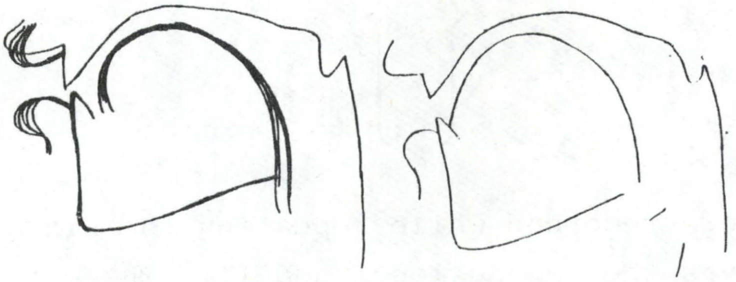


Figure 1

Profile X-ray tracings, speaker A. Left: [ɕ] before [i: ɑ: u: ɛ: y: ɯ:]. Right: the vowel [i:].

Figure 2

Profile X-ray tracings, speaker B. Left: [ɕ] before [i: ɑ: u: ɛ: y: ɯ:]. Right: the vowel [i:].

$$\text{ə}\ \ \text{ɕ}\ \left\{\begin{matrix} \text{ɩ:} \\ \text{ɑ:} \\ \text{u:} \\ \text{ɛ:} \\ \text{y:} \\ \text{ɯ:} \end{matrix}\right\}$$

Both speakers were recorded while repeating the six phrases twice, which gives the two series I and II. The diagrams in this article are based on series I. Series II showed essentially the same tendencies as series I.

The X-ray films were taken at the X-ray department of Lund hospital.[1] The film speed was 75 frames/sec. Tape recordings were made at the same time. Representatives of every case of [ɕ] have been taken from the frame where the tje-gesture shows maximum contact. This point usually occurs immediately after the midpoint of the fricative sound. On the basis of these frames the most important contour lines of the articulatory organs were drawn in profile diagrams. Figs. 1 and 2 show the contour lines for [ɕ] preceding the six different vowels in series I. As a contrast is shown each subject's articulatory position for the pronunciation of [i:].

Results

The following features of [ɕ] are strikingly constant for both subjects despite the varying vowel context and, besides, alike for the two speakers:

I. Angle between the lower and upper jaw, which was very small.

II. Position of the body of the tongue, which was raised and advanced as for [i:].

---

[1] For helping with the X-ray photography I take this opportunity to thank Sidney Wood and Gunilla Holje, Rolf Schoener and Gudmund Swahn from the X-ray department of Lund Hospital.

III. Position of the front of the tongue, which differs from that of the [i:] articulation. Figs. 1 and 2 show how the blade of the tongue for both subjects is withdrawn from the lower front teeth for [ɕ] (about 10 mm) but not for [i:]; subject B had a rather sharp edge. The surface of the tongue below this edge points steeply downwards. In subject A the bending is somewhat more gradual. Thus a cavity is formed between the surface of the tongue and the lower front teeth. This cavity is about 15 mm deep. Observe that this front tongue position means that the cross-section of the vocal tract is widened abruptly in front of the displacement for [ɕ].

These features seem to be essential for the production of the normal Swedish tje-sound. To this can be added another two features which, as far as can be seen, are also essential to the production of [ɕ] but which cannot be illustrated with the help of profile diagrams:

IV. The blade of the tongue forms, just by the marked forward edge, a maximum constriction against the alveolar ridge. The cross-section area of this constriction is of the same order of magnitude as that of [s], and it is, as in [s], groove-shaped rather than slit-shaped. At the constriction the air-stream generates turbulence, so here is a sound source.

V. The concentrated air stream which is formed at the constriction is directed against the edge of the lower front teeth, so that turbulence is also generated here. The sound associated with [ɕ] is, then, produced both at the narrow constriction and at the teeth.

I have obtained information concerning IV by direct observation of subjects A and B, and also a number of other speakers. For those who are not satisfied with this direct method, there is the possibility of proving the hypothesis instrumentally, using X-ray sections or palatography. I plan to experiment along these lines.

The assumption of V, that the teeth's effect upon the air stream is an important source of sound, is reinforced by two different factors: Firstly, that [ɕ] is changed radically even by a slight lowering of the jaw while the tongue is held constant relative to the upper jaw. Secondly, subject B who has protruding front teeth pushes his lower jaw forward about 4 mm for his pronunciation of [ɕ]. This causes the teeth to get into the air stream. With other sounds, however, his lower jaw moves little in a horizontal direction from its position of rest. See fig. 3.

In contrast to the features already discussed, the lip articulation for [ɕ] varies considerably in the different vowel contexts, especially with subject A. See figs. 1 and 2. All of these lip gestures are, however, variants of a non-rounded gesture. This can be observed when, for example, the sound is pronounced isolated. The basic lip articulation for [ɕ] is, then, the same as for [i:] and [j].

The contrast [ɕ] - [i:, j]

The tongue and lip articulations for [ɕ] and [i:, j] are, as we have seen, very similar. But they are not completely so. See figs. 1 and 2. What basically differentiates them is, partly, that the lip articulation for [ɕ] seems to be more likely to vary, and partly that the front of the tongue makes a specific articulation in [ɕ] which obviously implies a complication when contrasted with [i:, j].

That [ɕ] is a "difficult" sound is shown by the facts that a child learns the sound late in life, and that adult foreigners often fail to master its idiomatic pronunciation. [ɕ] is also a very unusual sound in the languages of the world. The difficulty of the sound seems to lie both in its production and perception. In contrast to this, [iː] and [j] are learnt at an early stage by the child, and are very common among the languages of the world. [iː] and [j] can be said to be both easy to produce and to perceive.

The contrast [ɕ] - [ç]

The typical German [ç] should normally be produced with the same tongue body and lip position as [iː], [j] and [ɕ] but, unlike [ɕ], with the tip of the tongue against the lower teeth or the bottom of the mouth as in normal [i] and [j] (see fig. 4). The maximum constriction in [ç] is formed further back than in [ɕ], with the back of the tongue against the hard palate. The cross-section of the point of constriction seems to be slit-shaped, while it is groove-shaped for [ɕ]. Unlike [ɕ], the cross-section of the vocal tract area increases gradually anterior to the point of constriction in [ç]. Variation in the angle between the lower and upper jaw in [ç], or the tongue tip's position (between the bottom of the mouth and the lower teeth), only produces a slight acoustic-perceptual effect, also unlike [ɕ]. Furthermore, [ç] sounds more or less the same, whether it is pronounced with inhalatory or exhalatory air stream, while [ɕ] is radically changed by inhalation. Two factors of production are eliminated in [ɕ] by inhalation: Firstly, there is no longer an air stream directed against the teeth and, secondly, there is no steep increase of the cross-section area at the air stream's outlet from the constriction. There are, however, no such production factors for [ç] producing a difference in inhalation and exhalation.

In the same way as [ɕ], [s], [ʂ] and [ʃ] also change their character when pronounced with an inhalatory air stream. The sibilant character of the sounds disappear; they are no longer sharp and strident. It seems obvious that [ɕ], just like these other sounds, is a sibilant. The most important articulatory condition for sibilance seems to be the perturbation of a concentrated stream of air at the teeth. It seems also that the

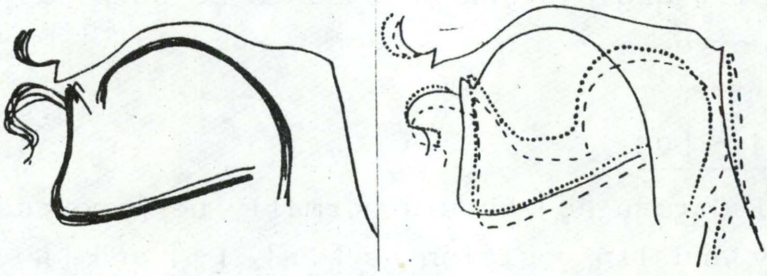


Figure 3

Profile X-ray tracings, speaker B. Left: [ɕ] before [i: ɑ: u: ɛ: y: ɯ:]. Right: the vowels [i](unbroken line), [ɑ:] (broken line) and [u:] (dotted line).

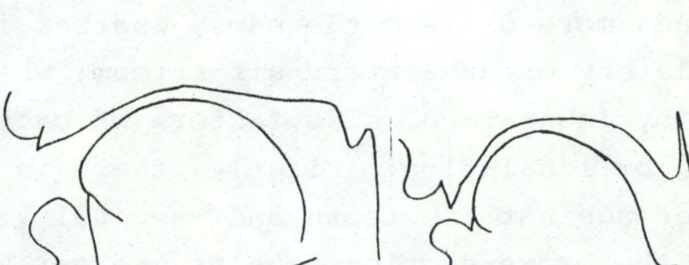


Figure 4

Profile X-ray tracings of [ç]. Left: speaker B; [ç] was spoken in isolation, a following [ɛ:] being intended. Right: [ç] as reproduced in Wängler (1958).

steep increase in cross-section area anterior to the constriction contributes to the efficiency of the sound source as far as sibilance is concerned.

Thus, [ɕ] is a sibilant, whereas [ç] is not. However, with regard to auditory brightness, the sounds resemble each other and both contrast with the brighter [s] and the darker [ʂ]. Fig. 5 shows the typical acoustic differences and similarities between [ɕ] and [ç], as pronounced by speaker B, who does not, however, use [ç] as a natural sound. Their spectra have approximately the same energy distribution, weak energy at the lower frequencies but from about 3 kHz and upwards a strong, wide energy band without very noticeable formant peaks. This acoustic similarity probably explains the similarity of their auditory colour.

The typical difference between the spectra of the sounds as pronounced by speaker B is the degree of steepness of the energy decline at lower frequencies. The lower limit at about 3 kHz is steep for [ɕ] and more drawn out for [ç]. This difference may be a general acoustic cue to the auditory difference between [ɕ] and [ç], a hypothesis which I intend to investigate further.

[s], [ɕ] and [s]

The contrast between [ç] and [ɕ] in fig. 5 probably illustrates the most important acoustic correlate of sibilance. This correlate should be comprised of a strong, broad energy band without very marked formant peaks which goes abruptly down at the lower frequency limit. This observation is strengthened by the spectra of [s], [ɕ] and [ʂ]. See fig. 6 which shows typical spectra of these sounds. Observe that their energy patterns are, in their main features, similar to what was observed for [ɕ]. The frequency positions for steep/abrupt lower limit of the patterns vary, however, systematically between the sounds. [ɕ] takes up the position between [s] and [ʂ].

The relation between the frequency position for the lower limit (and thereby the sound's lowest effective resonance), the

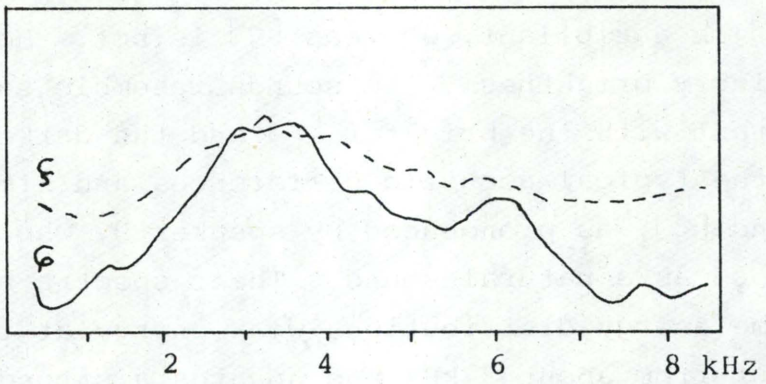


Figure 5

Typical [ç] and [ɕ] spectra, spoken in isolation by speaker B.

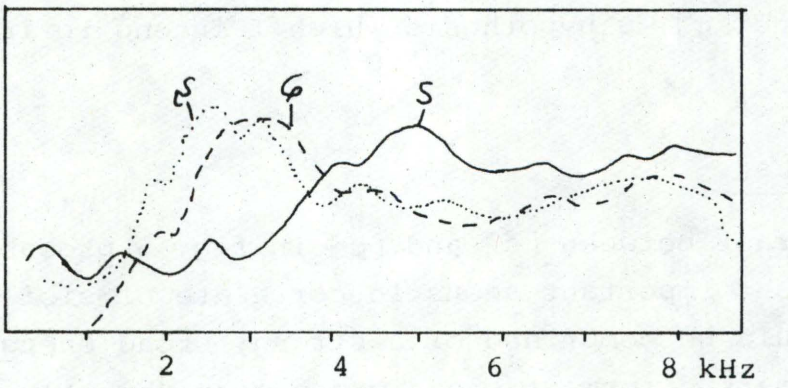


Figure 6

Typical [s], [ɕ], and [ʂ] spectra, spoken in isolation by speaker B.

relative size of the front cavity, and the auditory impression of brightness for these sounds is noticeable. The smaller the front cavity, the higher the frequency, and the brighter the quality. [ɕ] takes here a position between [s] and [ʂ]. The contrasts are not very great.

In several Swedish dialects that have all of these three sounds, there is to be found a crowded sibilant dimension. The [ɕ]-articulation is, as mentioned above, difficult to accomplish per se. The degree of difficulty of this articulation is further increased because of the small differences between [ɕ] and the two other sibilants.

Conclusion

The front cavity of the sibilant sounds should be given great attention. The position of the body and front of the tongue in [s], [ɕ], [ʂ] and [ʃ] should be studied with special reference to their influence on the dimensions of the cavity between the tongue and the lower teeth. Because this cavity (the most important resonator) is small, even slight changes in its size (just as for the vestibule between the teeth and lips) give rise to great acoustic-perceptual effects in these sounds.

In Linell, Svensson and Öhman (1971, p. 97), the Swedish tje-sound is classified as [-coronal] and [-strident]. From the above it should be clear that this is incorrect, and that the normal tje-pronunciation should, on the contrary, be [+strident] and [+coronal].

References

Bladon and Nolan 1977:	"A video-fluographic investigation of tip and blade alveolars in English", J.Ph.
Elert, C.-C. 1966:	Allmän och svensk fonetik
Fritzell, B. 1973:	Foniatri för medicinare
Linell, Svensson and Öhman 1971:	Ljudstruktur
Malmberg, B. 1968:	Svensk fonetik

Principles of the International Phonetic Association

Söderpalm Talo, E. 1976: Röst-, tal- och språkrubbningar

Wängler, H.-H. 1958: Atlas deutscher Sprachlaute

PERCEPTION OF STATIONARY FORMANT PATTERNS[1]

Hartmut Traunmüller[2]

The object of my research study was the role of intrinsic factors in perceiving stationary speech sounds. Features that are acoustically present within the segment in question are here called intrinsic factors, as distinguished from extrinsic (contextual) factors. As stimuli simple synthetic vowels with one or two stationary formants were used.

The point of departure of my discussion is not the acoustic signal but the representation of sounds along the basilar membrane, where the Bark scale of tonality z(F) is applicable. It has been presupposed that formants and pitch are the decisive features of the phonetic identity of sounds. The question to be answered was to what extent the positions of - and the distances between - the relevant parts of the sound representation are used in phonetic perception.

If distances are most essential, there seems to be no need to suppose a normalizing process in order to explain perceptual invariance for isolated vowels produced by children, women, and men, despite overlap in formant-frequency data of different phonemes, since these age- and sex-conditioned differences on the whole are reflected as a uniform displacement of

---

[1] Summary of my paper entitled "Einige Aspekte der Wahrnehmung quasistationärer Vokale" in PILUS 32, p. 8-13, Stockholm Univ., oct. 1976.

[2] Department of Speech Communication, Royal Institute of Technology (KTH), S-100 44 Stockholm 70, Sweden.

the sound representation along the basilar membrane. But there is a less than average displacement found in F_1 of non-open vowels produced by men.

One-formant vowels covering almost the whole range where phonetic interpreting is possible, were generated in steps of at most one Bark for F_o and the formant. For two-formant vowels F_1 was maintained somewhat above F_o, while F_2 on the one side, and F_o, F_1 on the other side were generated as above mentioned. F_o of the stimuli was non-stationary.

The vowels were identified by two Austrian and two Swedish subjects. In the dialect of the Austrian subjects there occur five distinctive degrees of opening among vowels, which results in good resolution among one-formant vowels, while in the dialect of the Swedish subjects there occur four maximally closed vowels, which results in good resolution among the 'closed' two-formant vowels.

The results showed:

1) that there is a lack of agreement between the vowel identifications of the different subjects, particularly with regard to those stimuli that were not perceived as natural vowels because of too high $z(F_o)$ or $z(F_1)$.

2) that the distance between F_1 and F_o $(z(F_1)-z(F_o))$ is decisive for the identity of one-formant vowels as long as the distance is less than approximately 6 Bark, otherwise the identity is determined mainly by $z(F_1)$.

3) that the phoneme boundaries between [i] - [y] - [ʉ] in two-formant vowels are determined mainly by $z(F_2)$. This cannot explain the phonetic similarity of /y/ produced by different speakers.

4) that one-formant vowels are identified not only as back vowels with corresponding F_1. Instead of the expected sequence [u o ɔ ɑ a] the following sequence [u o ɔ œ (ɶ) æ] was obtained from three of four subjects.

Result 1) points to the possibility that different persons may have different perceptual strategies. Results 2) and 3) are in accordance with the hypothesis that the phonetic identity primarily is determined by the distances between formants with a small mutual distance, including the voice fundamental (e.g. F_1, F_o or F_4, F_3, F_2 in front vowels). In addition, formant positions function as substitutes if adjacent formants are not perceived. No satisfactory explanation for result 4) has been found. It is, however, evident that this result cannot be brought to agreement with the idea that vowels are perceived by means of simple matching of their total spectra against memorized patterns.

(ARIPUC 12, 1978)

FORM FACTORS FOR POWER SPECTRA OF VOWEL NUCLEI

Ulf J. Stålhammar[1]

Abstract: The spectrum of the vowel nucleus is, within the present framework, regarded in terms of 'Form Factor Elements'. One such element is reflected in the structure of formant clusters and may be expressed in terms of bandwidth. It is shown that in two-formant synthesis a vowel can change its identity from /i/ to /e/ (or /y/ to /ø/ with a decreased pole span) simply by an increase of the higher pole bandwidth. Prior to this discovery, and indicating its potential, a four-formant synthesis experiment was conducted. In this part of the investigation the possibility of constructing /e/ vowel spectrum envelopes containing either higher, equal or lower spectral centres of gravity, as compared to envelopes generating the auditory impression of /i/, is demonstrated. N.B. in the process of generating the power spectra for /i/ and /e/, the frequency parameters F_4, F_1 and F_0 were held constant. The spectrum balance was achieved either by means of only frequency adjustments of F2 and F3 or, with these frequencies "frozen", by amplitude modifications of F2, F3 and F4. The two experiments emanate from an empirically found paradoxical relationship for one female voice between such parameters as Tongue Height plus Fronting versus Second Formant Prime, (F'_2) versus Centre of Gravity for the Spectral Components above F_1, (denoted here as G'_2).
The psychoacoustic evidence obtained focuses the attention towards the development of an "excitation area" theory of perception based on form factors.

---

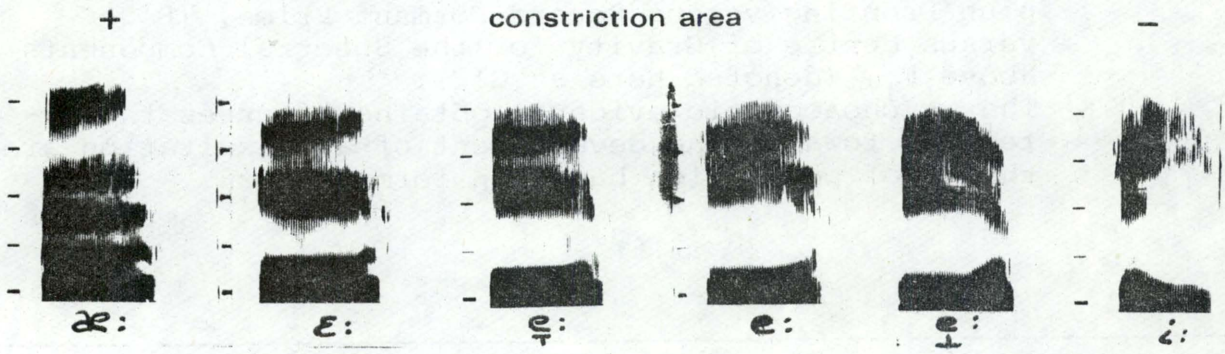
[1] Department of Speech Communication, Royal Institute of Technology (KTH), S 100 44 Stockholm 70, Sweden.

Introduction

The work to be described forms part of a research program dealing with vowel nuclei in production, synthesis and perception. In the present paper the latter two issues are concerned. The paper provides some background philosophy to and a summary presentation of the talk given at the "Svensk-Dansk Fonetiker Seminar 1977" under the title: "'Formfaktorn' och dess betydelse för vokalperception" - "The 'Form Factor' and its Significance in Vowel Perception". The research is carried on in support of perception and automatic recognition studies. The aim is to investigate how modified spectra can nevertheless be perceptually identified as one and the same vowel, and ultimately to express each vowel in terms of one parameter only, if possible, and to use the same parameter for male and female voices.

Background

Space geometric properties of the human vowel tract allow the second formant, F_2, to be positively correlated with tongue height/fronting until the extreme prepalatal region is reached. Within this latter region the two parameters will be reversely correlated as demonstrated in the spectrogram sequence of figure 1. This fact focuses on problems referring to automatic speech recognition (ASR), decision logic and human perception.



1.1 1.2 1.3 1.4 1.5 1.6

Figure 1

Effect of increasing tongue height/fronting upon formant patterns. - Speaker US.

Figure 2 contains a vowel plot in terms of the F_1/F_2 spaces of adult males, adult females and five to eight-year-old children of both sexes (Stålhammar, 1971). Each data point represents the mean over ten subjects. Attention should be given to the relation between the vowels /i/ and /e/. The figure shows the relationship $F_{2i} < F_{2e}$ for all three speaker categories. Figure 3 demonstrates that the relation is manifested, again as a mean, also in different contexts. Context 1 represents vowels in isolation, context 2 vowels embedded in C-C environment and context 4 vowels in fluent speech (Stålhammar, Karlsson, Fant, 1973). Obviously, it is clear that the i/e relationship cannot be adequately described in terms of the F2 feature alone. This is a drawback for ASR systems having as vowel identifier the distance F_2-F_1. From a perceptual point of view, however, /i/ is perceived as [+high] in relation to /e/. An examination of figure 1.6 might provide the answer, the effect of the weak F2 amplitude due to its large

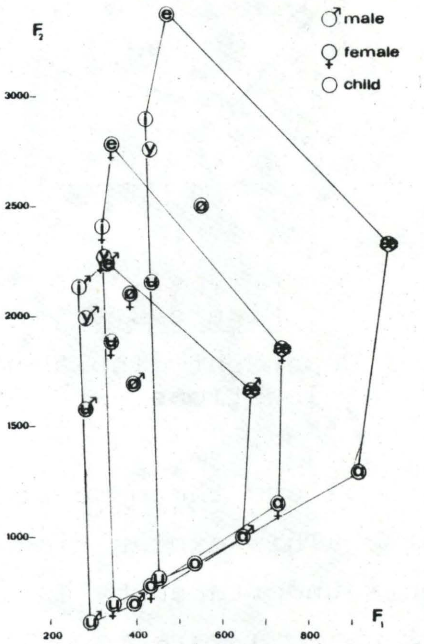


Figure 2

Sex dependent displacement of the F_1/F_2 locations.

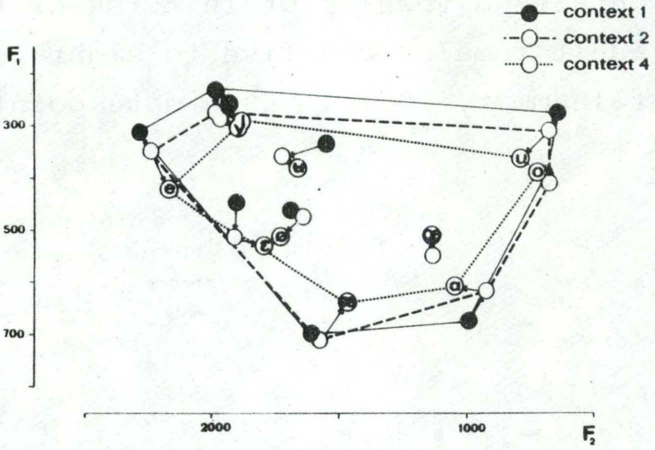


Figure 3

Context dependent displacement of the F_1/F_2 locations.

distance from F_3, which allows the perceptual focus to be oriented towards higher formants. This explanation is also supported by the matching experiments undertaken by Carlson, Fant, Granström (1975), in which test subjects were reported to match closer to F_3 than to F_2 for a synthetic /i/ vowel.

Consequently, the use of a new parameter (G'_2-F_1) in place of (F_2-F_1) should bring about an improvement from an ASR point of view; G'_2 represents the centre of gravity of spectral components above F_1. Under the described conditions which usually prevail in reality, a positive correlation between the parameters Tongue Height/Fronting and G'_2 is obtained.

While the newly defined parameter (G'_2-F_1) represents a definite improvement over the previously used (F_2-F_1) parameter, a contradictory case of a female voice has nevertheless been found.

Female case

Two vowel spectra originating from a female test subject, AKS, are shown in figure 4 in which the power spectra pertaining to the instants identified by the sampling arrows are presented. It is noticed that F_{1i}=285 Hz, F_{2i}=2600 Hz, F_{3i}=3670 Hz, F_{4i}= 5190 Hz; similarly, F_{1e}=305 Hz, F_{2e}=3030 Hz, F_{3e}=3700 Hz, F_{4e}=

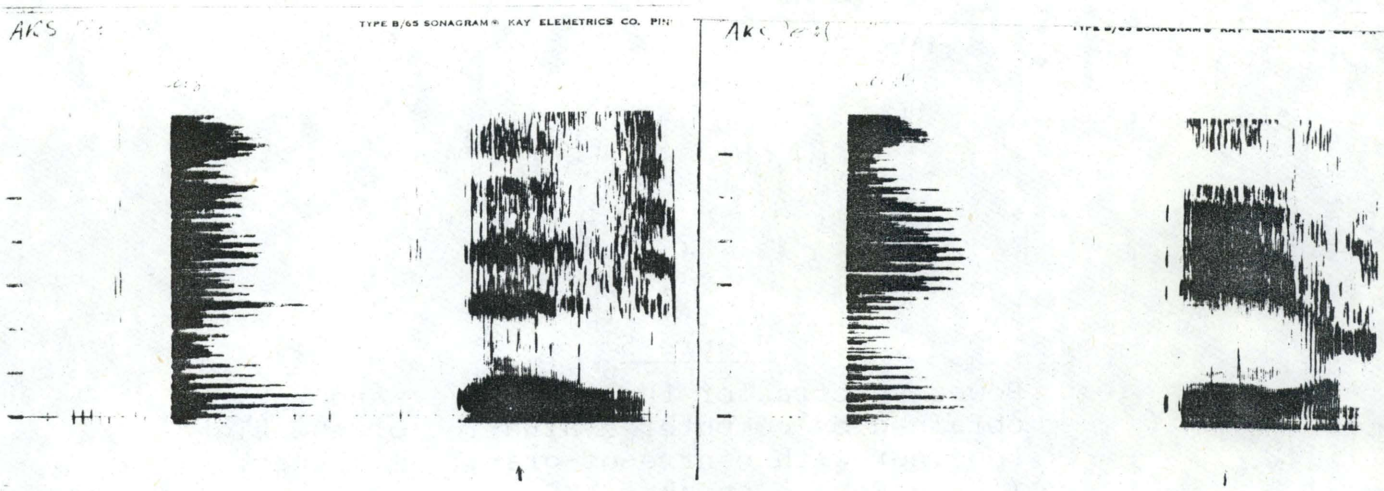


Figure 4

Visible speech and power spectrum cross-section patterns for vowels /i/ and /e/. - Female speaker AKS.

4795 Hz. Again it is observed that $F_{2i} < F_{2e}$ at the sampling locations, i.e. Δ = 430 Hz. More importantly, an unusually high amplitude of F_{2i} is observed. Based on visual examination it might be hypothesized that the G'_{2i} could be even lower than the G'_{2e} within the steady state portions of the two sounds. In order to scrutinize this hypothesis quantitatively, the frequency, $f_{G'}$, of the spectral centre of gravity was computed:

$$f_G = \frac{\sum f \cdot A_f}{\sum A_f} \; ; \qquad (1)$$

where A_f is the linear amplitude of the component of frequency f.

The cross-section patterns presented in figure 5 for the two vowels /i/ and /e/ were obtained by means of a computer operating at a sampling rate of 100 Hz.

Two separate evaluations of G'_2 were carried out using the frequency ranges 0.8-4.8 kHz (test 1) and 0.8-6.8 kHz (test 2), respectively for the evaluation of the spectral centre of gravity. Test 1 was undertaken mainly to focus on the contribution of F_2 and F_3 to the spectral centre of gravity, and test 2, with the high frequency limit increased to 6.8 kHz, to include the influence

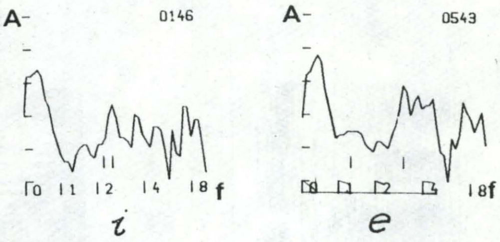


Figure 5

Power spectra for the vowels /i/ and /e/, obtained by computer evaluation of the signal together with centre-of-gravity markings, ('). - Female speaker AKS.

of higher spectral components upon the spectral centre of gravity. The following numerical values of G'_2 were obtained:

Test 1: $G'_{2i} = 2137$ Hz

$G'_{2e} = 2687$ Hz

$\Delta = G'_{2i} - G'_{2e} \approx -500$ Hz

Test 2: $G'_{2i} = 2375$ Hz

$G'_{2e} = 2825$ Hz

$\Delta = G'_{2i} - G'_{2e} \approx -500$ Hz

Although for test 2 the upper cut-off frequency is increased by 2 kHz the increase in G'_2 is about ten times smaller, i.e. only approximately 200 Hz; Δ remains essentially constant. Although usually /i/ is associated with a higher spectral centre of gravity, $G'_{2i} < G'_{2e}$ in this case.

Comparison with previous equivalent formant methods

An F'_2 calculation was carried out using the 1975 formula due to Fant (Carlson, Fant, Granström 1975):

$$F'_2 = \frac{F_2 + c(F_3F_4)^{1/2}}{1 + c} \qquad (2)$$

$$c = \left(\frac{F_1}{500}\right)^2 \cdot \left(\frac{F_2 - F_1}{F_4 - F_3}\right)^4 \cdot \left(\frac{F_3 - F_2}{F_3 - F_1}\right)^2$$

The following numerical values were obtained:

$$F'_{2i} = 2860 \text{ Hz}$$

$$F'_{2e} = 3450 \text{ Hz}$$

$$\Delta = F'_{2i} - F'_{2e} \approx -0.6 \text{ kHz}$$

The formula was later revised as follows (Bladon, Fant 1978):

$$F'_2 = \frac{F_2 + c^2(F_3F_4)^{1/2}}{1 + c^2} \qquad (3)$$

$$c = K(f)\frac{A_{34}}{A_2}$$

where A_{34} is the vocal tract transfer function in the valley between F3 and F4 at the frequency $F_{34} = (F_3F_4)^{1/2}$ and A_2 is the transfer function at the second formant peak, F2. The factor K(f) in the weighting function is intended to include the additional preemphasis originating from source, radiation and higher pole corrections and in addition a correction for differences in equal loudness levels.

With the new version the following numerical values were obtained:

$$F'_{2i} = 2725 \text{ Hz}$$

$$F'_{2e} = 3460 \text{ Hz}$$

$$\Delta = F'_{2i} - F'_{2e} \approx -0.7 \text{ kHz}$$

N.B. again with the /i/ as the lower counterpart.

Although the previous formulæ are adequate to represent most vowels, the contradictory example indicates that some additional modifications will be required to cover all vowels.

It should be observed here that for speaker AKS, given the F_n data presented above, F'_{2i} = 2860 Hz and F'_{2e} = 3450 Hz. In order to transform the F'_{2i} value to coincide with the F'_{2e} value by means of shifting only one formant at a time, the F_2 has to be shifted up to 3422 Hz, the F_3 up to 4031 Hz, or the F_4 down to 3283 Hz. It is understood that such a transformation of F_4 is not permissible if the resulting transformed value is smaller than F_3. F'_{2e} cannot in this case be reached by a shift in F_1 only. Similarly, in a CVC environment, for the observed female formant set (Stålhammar et al. 1973) for the vowel /I/, (F_1 350 Hz, F_2 2600 Hz, F_3 3075 Hz, F_4 4000 Hz) and the male set, (F_1 325 Hz, F_2 2315 Hz, F_3 2915 Hz, F_4 3400 Hz) it is noticed that each individual female formant occupies a higher frequency position, yet the female F'_2 = 2910 Hz and the male F'_2 = 3035 Hz, i.e. $F'_{2fe} < F'_{2ma}$. In order to transform the F'_{2fe} value to coincide with the F'_{2ma}, F_1 has to be shifted up to 505 Hz, F_2 to 3030 Hz, F_3 to 3140 Hz, or F_4 has to be lowered to 3845 Hz. An increase of F_4 results in a downshift of F'_2.

Since it is now possible to state that neither the (F_2-F_1) parameter nor the (G'_2-F_1) parameter nor the (F'_2-F_1) parameter cover all possible cases, there must be another, as yet unknown, factor of importance. A preliminary attempt to determine the nature of such a factor will be done by means of (1) a four-formant synthesis experiment and (2) by means of a two-formant synthesis experiment. However, for the sake of completeness, the rôle of formant transitions should first be considered.

The Vowel - a Spectral Chameleon

The low F_1 vowels in Swedish, /i/, /y/, /ʉ/ and /u/, are mostly characterized by pronounced F_n transitions when uttered in isolation (see e.g. the /i/ and /ʉ/ vowels in figure 6 and Stålhammar, Karlsson 1972). These vowels tend to be diphthongized towards a target of extreme tongue-palate closure for /i/ and /y/ and a target of labial closure following /ʉ/ and /u/. In order

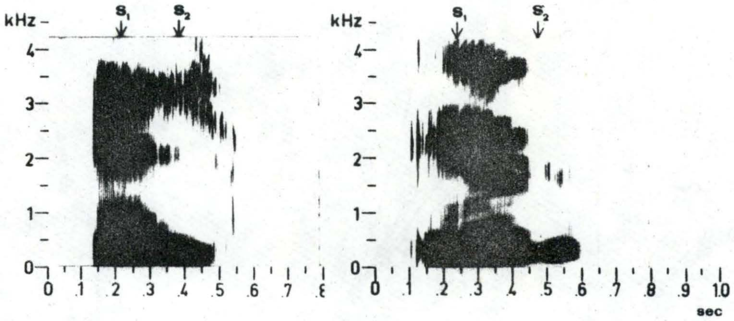


Figure 6

Spectra of an /i/ vowel, left, and an /ʉ/ vowel, right, together with indications of sample 1, (S_1), and sample 2, (S_2), locations. - Speaker US.

to obtain a more adequate representation of F_n for this category of vowels, the location of two samples, S_1 and S_2, are defined. S_1 is located at approximately 20% of the total duration from voice onset and the location of S_2 is derived from an articulatory target as follows: for /i/ and /y/ it is where the F_3 transition changes direction from positive to negative as in figure 6 or at other target criteria, and in /ʉ/ and /u/ where the F_2 reaches steady state. The values of F_1, F_2 and F_3 at the instant S_1 are redefined as being equal to zero; thus the endpoints which correspond to S_2 reflect the difference, ΔF_n, between the final target value (at S_2) and the initial value (at S_1) of the F_n's. In figure 7 these ΔF's are plotted for F_1, F_2 and F_3. The encircled digits indicate test subjects. A close examination of the plot reveals that the F-pattern shows great variability as a function of the two sampling locations. For the vowel /i/ the F_1 transition is negative and simultaneously associated with positive F_2 and F_3 transitions for subjects 3 and 7. Similarly, for subjects 0 and 4, F_1, F_2 and F_3 all show negative transitions. For subjects 5 and 8, F_1 is negative, F_2 constant and F_3 positive. For subject 2, F_1 is constant, F_2, F_3 negative. For subject 1, F_1 is also constant, F_2 negative; however, F_3 is positive. For subjects 6 and 9, F_1 and F_2 are negative, while F_3 is positive.

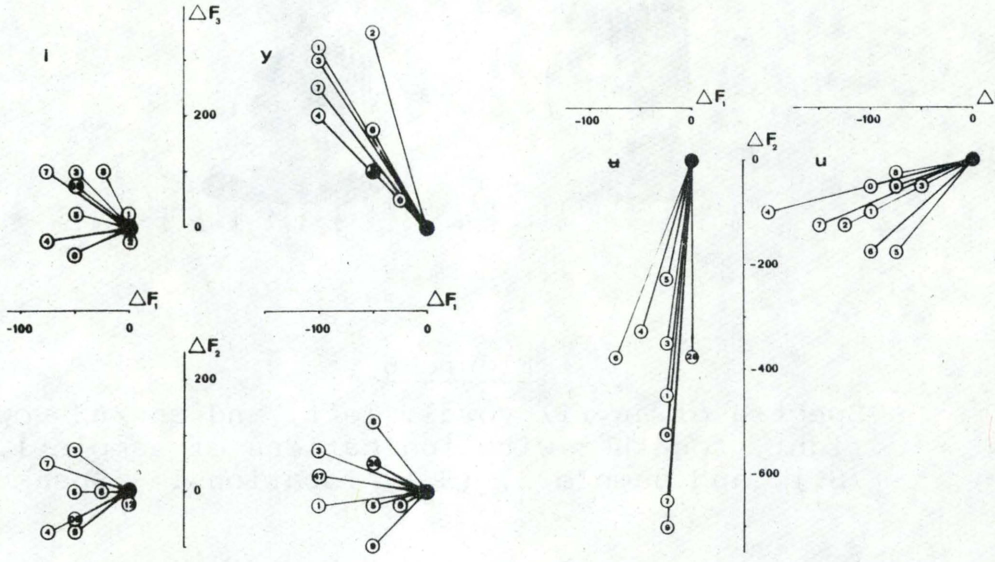


Figure 7

△F's for formant transitions between the instants S_1 and S_2 for various vowels. ● = normalized location for S_1. o = △F's, obtained for various subjects at S_2.

For the vowel /y/ all subjects display negative F_1 transitions combined with positive F_3 transitions, while F_2 shows a less consistent pattern. The vowels /ʉ/ and /u/ have clear negative transitions for both the lower formants. The negative $F_{2ʉ}$ transition for subject 9 is about 700 Hz.

In view of the inconsistent transition patterns for F_1, F_2 and F_3, it is imperative to search for more consistent parameters such as, e.g., the form factors described in this paper.

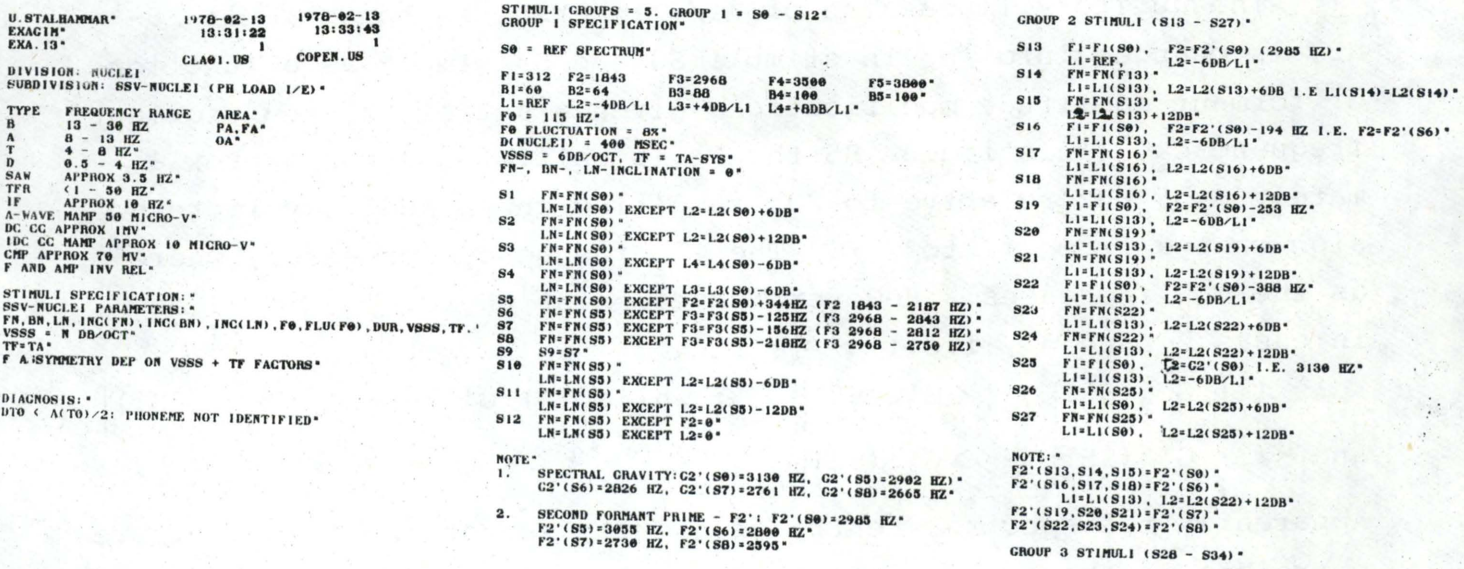
In order to eliminate the possibility of any spurious effects on the present study, the signal obtained from speaker AKS was examined in terms of the following questions:

1. Are the vowels produced by speaker AKS adequate and correct in all respects?
2. What about the higher formant amplitude decreases in the later part of the /i/ vowel and the negative higher formant transitions in the later part of the /e/ vowel?

With respect to question (1) a panel of 30 listeners judged the utterances of the speaker as being entirely proper and showing no spurious effects. With respect to question (2) the problem of amplitude decreases/transitions was avoided by means of retaining only the steady state portions of the vowels, i.e. approximately 50% of the total durations. After the truncations were made, the steady states were repeated to reestablish the original durations. According to the same panel of listeners the reconstituted signals retained a sufficient phonemic identity (/i/ and /e/, respectively).

Four-formant case

In order to retain control over formant amplitudes, the four-formant stimuli were produced by a computer simulation of parallel-formant synthesis. Input data forming the reference construct, SO, are presented in figure 8. The data are typical of a male vowel /i/ with a slight modification in F_1 so as to obtain an intermediate value between F_{1i} and F_{1e}. This weighted value was adopted



```
U.STALHAMMAR"              1978-02-13     1978-02-13
EXAGIM"                      13:31:22       13:33:43
EXA.13"                             1              1
                           CLA01.US       COPEN.US
DIVISION: NUCLEI"
SUBDIVISION: SSV-NUCLEI (PH LOAD I/E)"

TYPE    FREQUENCY RANGE     AREA"
B       13 - 30 HZ          PA,FA"
A       8 - 13 HZ           OA"
T       4 - 8 HZ"
D       0.5 - 4 HZ"
SAW     APPROX 3.5 HZ"
TFR     <1 - 50 HZ"
IF      APPROX 10 HZ"
A-WAVE MAMP 50 MICRO-V"
DC CC APPROX 1MV"
IDC CC MAMP APPROX 10 MICRO-V"
CMP APPROX 70 MV"
F AND AMP INV REL"

STIMULI SPECIFICATION:"
SSV-NUCLEI PARAMETERS:"
FN,BN,LN,INC(FN),INC(BN),INC(LN),F0,FLU(F0),DUR,VSSS,TF.'
VSSS = N DB/OCT"
TF=TA"
F A:SYMMETRY DEP ON VSSS + TF FACTORS"

"DIAGNOSIS:"
"DTO < A(TO)/2: PHONEME NOT IDENTIFIED"
```

```
STIMULI GROUPS = 5. GROUP 1 = S0 - S12"
GROUP 1 SPECIFICATION"

S0 = REF SPECTRUM"

F1=312  F2=1843     F3=2968     F4=3500     F5=3800"
B1=60   B2=64       B3=88       B4=100      B5=100"
L1=REF  L2=-4DB/L1  L3=+4DB/L1  L4=+8DB/L1"
F0 = 115 HZ"
F0 FLUCTUATION = 8%"
D(NUCLEI) = 400 MSEC"
VSSS = 6DB/OCT, TF = TA-SYS"
FN-, BN-, LN-INCLINATION = 0"

S1    FN=FN(S0)"
      LN=LN(S0) EXCEPT L2=L2(S0)+6DB"
S2    FN=FN(S0)"
      LN=LN(S0) EXCEPT L2=L2(S0)+12DB"
S3    FN=FN(S0)"
      LN=LN(S0) EXCEPT L4=L4(S0)-6DB"
S4    FN=FN(S0)"
      LN=LN(S0) EXCEPT L3=L3(S0)-6DB"
S5    FN=FN(S0) EXCEPT F2=F2(S0)+344HZ (F2 1843 - 2187 HZ)"
S6    FN=FN(S5) EXCEPT F3=F3(S5)-125HZ (F3 2968 - 2843 HZ)"
S7    FN=FN(S5) EXCEPT F3=F3(S5)-156HZ (F3 2968 - 2812 HZ)"
S8    FN=FN(S5) EXCEPT F3=F3(S5)-218HZ (F3 2968 - 2750 HZ)"
S9    S9=S7"
S10   FN=FN(S5)"
      LN=LN(S5) EXCEPT L2=L2(S5)-6DB"
S11   FN=FN(S5)"
      LN=LN(S5) EXCEPT L2=L2(S5)-12DB"
S12   FN=FN(S5) EXCEPT F2=0"
      LN=LN(S5) EXCEPT L2=0"

NOTE"
1.    SPECTRAL GRAVITY:G2'(S0)=3130 HZ, G2'(S5)=2902 HZ)"
      G2'(S6)=2826 HZ, G2'(S7)=2761 HZ, G2'(S8)=2665 HZ"

2.    SECOND FORMANT PRIME - F2': F2'(S0)=2985 HZ"
      F2'(S5)=3055 HZ, F2'(S6)=2800 HZ"
      F2'(S7)=2730 HZ, F2'(S8)=2595"
```

```
GROUP 2 STIMULI (S13 - S27)"

S13   F1=F1(S0),   F2=F2'(S0) (2985 HZ)"
      L1=REF,      L2=-6DB/L1"
S14   FN=FN(F13)"
      L1=L1(S13), L2=L2(S13)+6DB I.E L1(S14)=L2(S14)"
S15   FN=FN(S13)"
      L2=L2(S13)+12DB"
S16   F1=F1(S0),   F2=F2'(S0)-194 HZ I.E. F2=F2'(S6)"
      L1=L1(S13), L2=-6DB/L1"
S17   FN=FN(S16)"
      L1=L1(S16), L2=L2(S16)+6DB"
S18   FN=FN(S16)"
      L1=L1(S16), L2=L2(S16)+12DB"
S19   F1=F1(S0),   F2=F2'(S0)-253 HZ"
      L1=L1(S13), L2=-6DB/L1"
S20   FN=FN(S19)"
      L1=L1(S13), L2=L2(S19)+6DB"
S21   FN=FN(S19)"
      L1=L1(S13), L2=L2(S19)+12DB"
S22   F1=F1(S0),   F2=F2'(S0)-388 HZ"
      L1=L1(S1),   L2=-6DB/L1"
S23   FN=FN(S22)"
      L1=L1(S13), L2=L2(S22)+6DB"
S24   FN=FN(S22)"
      L1=L1(S13), L2=L2(S22)+12DB"
S25   F1=F1(S0),   F2=G2'(S0) I.E. 3130 HZ"
      L1=L1(S13), L2=-6DB/L1"
S26   FN=FN(S25)"
      L1=L1(S0),   L2=L2(S25)+6DB"
S27   FN=FN(S25)"
      L1=L1(S0),   L2=L2(S25)+12DB"

NOTE:"
F2'(S13,S14,S15)=F2'(S0)"
F2'(S16,S17,S18)=F2'(S6)"
      L1=L1(S13), L2=L2(S22)+12DB"
F2'(S19,S20,S21)=F2'(S7)"
F2'(S22,S23,S24)=F2'(S8)"

GROUP 3 STIMULI (S28 - S34)"
```

Figure 8

Specification of the 4-formant, SO-S12, and the 2-formant, S13-S27, stimuli.

so as to make the stimuli more sensitive to subsequent adjustments of the higher formant frequencies. Based on this reference, adjustments have been made according to figure 8; the resulting preemphasized (+6 dB/octave within 0.2-5 kHz) computer generated envelopes are displayed in figure 9. In addition, loudness envelopes of the same stimuli produced by a HP 8051 Loudness Analyzer are shown. The Analyzer performs a continuous 1/3 octave level analysis of the noise to be measured and computes, by the method indicated by Zwicker (ISO recommendation 532), the loudness S in sones$_G$. In figure 8 FN represents any formant, BN the formant bandwidth, and LN the formant amplitude. 30 subjects judged the quality of the stimuli.

Evaluation

Stimuli S0 and S1 showed a definite codability with /i/, S2 a high degree of codability with /i/, and S3 and S4 were decoded as /e/. Stimuli S5-S9 were classified as /e/. Stimulus S10 was ambiguous, and S11 and S12 showed a definite codability with /i/.

Comments

The auditory impression of /i/, inherent in the stimuli S0-S2, is changed into /e/ in stimuli S3 and S4 simply as a function of formant amplitude modifications given a set of "frozen" formant frequencies. In stimulus S5 the F_2 is shifted upwards approximately 0.35 kHz relative to S0, resulting in an auditory impression change from /i/ to /e/! The G'_2 frequency decreases, whereas the F'_2 increases. However, cases exist where $G'_{2e} > G'_{2i}$ as in cases 2. and 3., i.e.:

(1) $G'_2(S5)\big|_{/e/} < G'_2(S0)\big|_{/i/}$; while (2) $G'_2(S5)\big|_{/e/} > G'_2(S1)\big|_{/i/}$ and (3) $G'_2(S5)\big|_{/e/} > G'_2(S2)\big|_{/i/}$

Apparently the spectral centre of gravity is not the only decisive correlate in the i/e dimension. Since the upshift of F_2 in S5 results in a perceptual shift from /i/ to /e/ relative to S0, it is evident that the positive correlation between F_2 and the

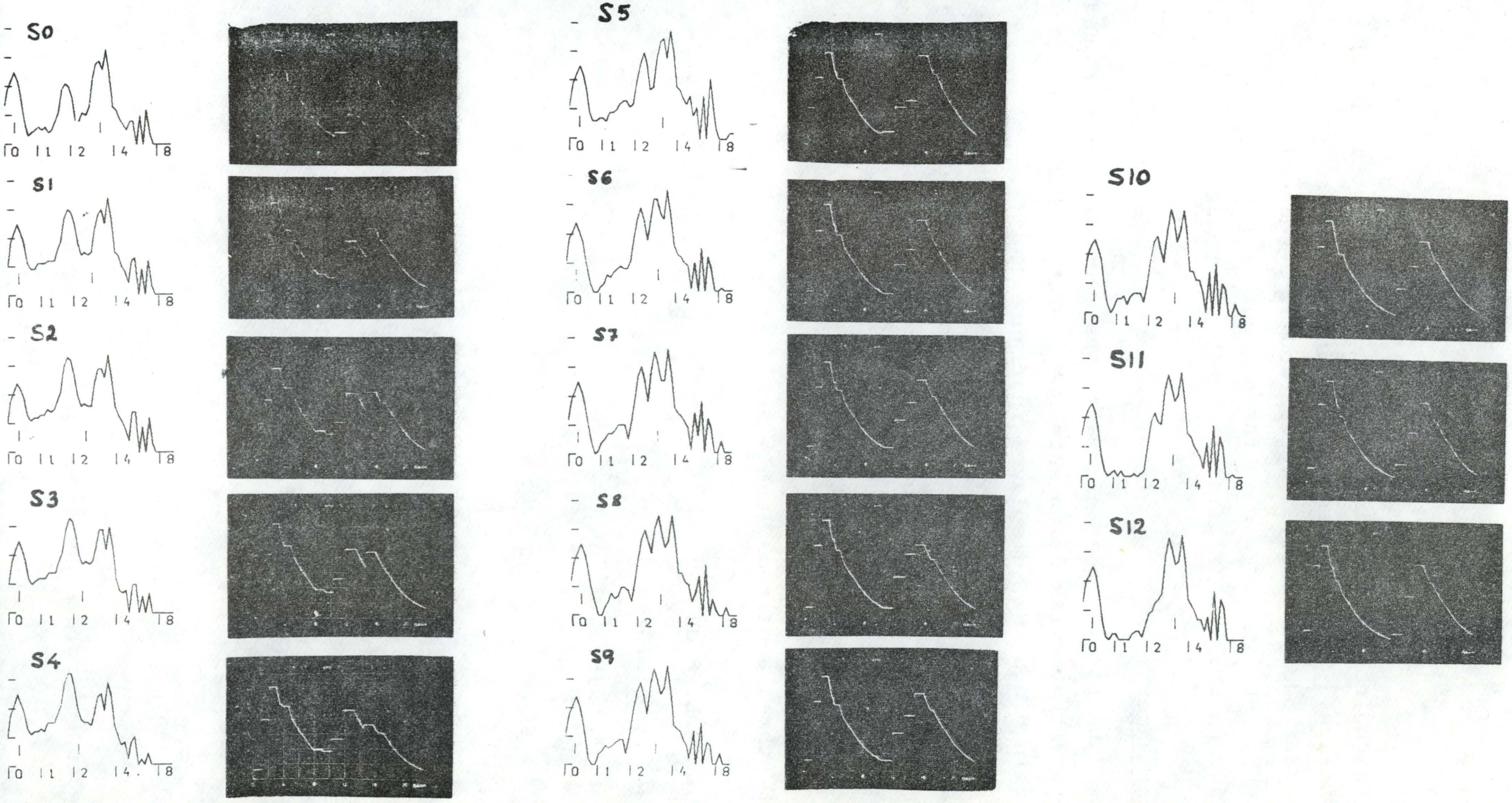


Figure 9

Left: Computer generated spectrum envelopes of the 4-formant synthetic stimuli, S0-S12.
Right: Loudness envelopes of the same stimuli produced by the HP 8051 Loudness Analyzer.

feature [+high] no longer exists when F_2 clusters with F_3 and higher formants. This is also evident in the samples S6-S8 where the /e/ vowel identity is preserved at a successive decrease of F_3. An L_2 decrease in the sequence S10 to S12 relative to S5 brings us back to /i/ (in S12 L_2 is set to zero).

Two-formant case

Input data forming the 2-formant constructs are to be found in figure 8 with the resulting envelopes, preemphasized +6 dB/ octave within 0.2-5 kHz, displayed in figure 10.

The 2-formant reference construct S13 is generated with the same F_1 as F_1 of the 4-formant reference construct S0, while the upper formants are reduced to one pole equal to F'_2(S0). Similarly, the higher pole frequency, F_2, of stimulus S16 is derived from F'_2 of the 4-formant construct S6; F_2(S19) = F'_2(S7); F_2(S22) = F'_2(S8). However, F_2(S25) = G'_2(S0). Furthermore, for each F_2 three amplitude levels are used, e.g.:

S13 $L_2 = L_1 - 6$ dB

S14 $L_2 = L_2(S13) + 6$ dB

S15 $L_2 = L_2(S13) + 12$ dB

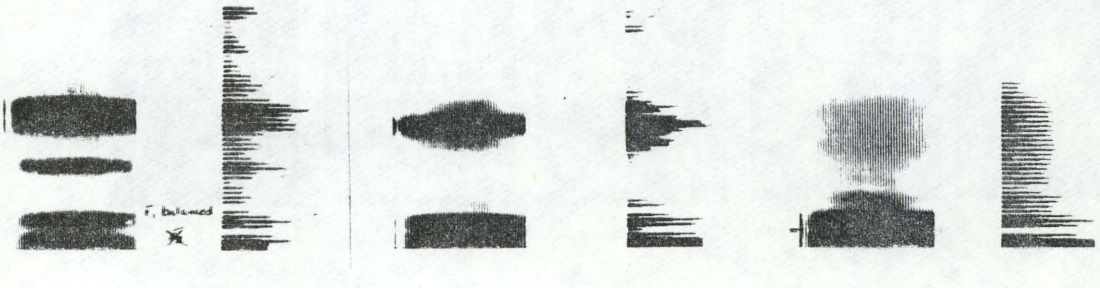


Figure 11

Broadband spectrograms and power spectra for stimuli S0, S13 and S30.

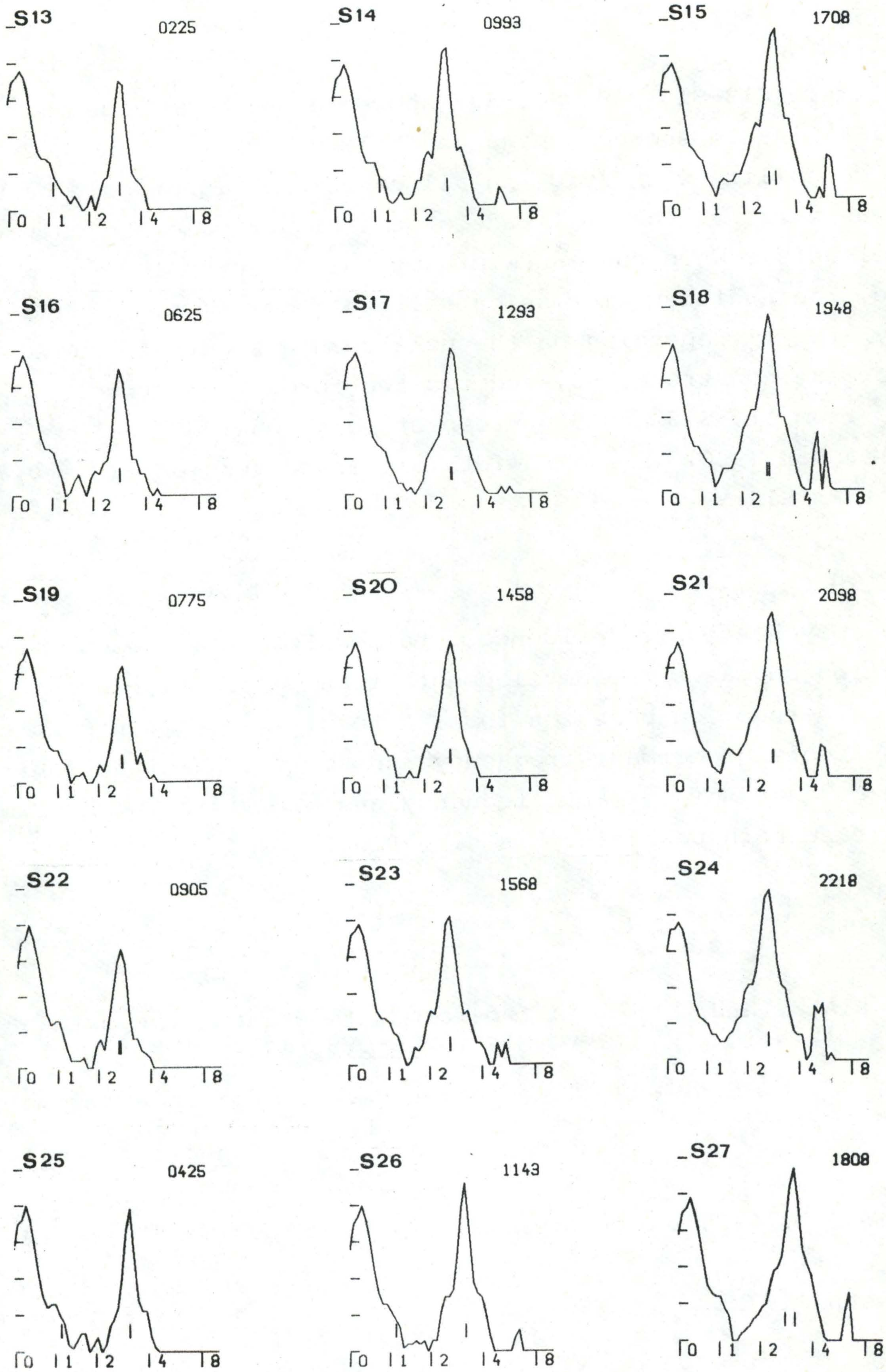


Figure 10

Computer generated spectrum envelopes of the 2-formant synthetic stimuli, S13-S27

Evaluation

Stimuli S13-S18 plus S25-S27 showed a definite codability with /i/, S19-S21 showed a high codability with /i/, and S22-S24 a high codability with /y/. In all cases the higher pole bandwidth was narrow. The amplitude of the higher pole had only minor effects. When the bandwidth of the higher pole was increased, the auditory impression shifted from /i/ to /e/ and from /y/ to /ø/ depending on the pole span. Figure 11 shows a four-formant construct, S0, and two two-formant constructs derived from it, S13 and S30, respectively, where $F_2(S13) = F'_2(S0)$ and $F_2(S30)$ is a bandwidth increased version, ($B = 0.75$ kHz), of $F_2(S13)$.

Conclusion

The psychoacoustic evidence obtained from the present investigation suggests the development of an "excitation area" theory of perception based on form factors. The combined integrated effects of formant frequency and amplitude in the four-formant case and of formant frequency and bandwidth in the two-formant case both point in this direction.

References

Bladon, R.A.W. and G. Fant 1978: "A two-formant model and the cardinal vowels", *STL-QPSR* 1, p. 1-8

Carlson, R., G. Fant and B. Granström 1975: "Two-formant models, pitch and vowel perception", pp. 55-82 in Fant, G. and Tatham, M.A.A. (eds.) *Auditory Analysis and Perception of Speech*, London

Stålhammar, U. 1971: Internal Report, STL-RIT, Stockholm

Stålhammar, U. and I. Karlsson 1972: "A Phonetic Approach to ASR", *1972 Conf. on Speech Communication and Processing* AFCRL Paper C10. IEEE cat.no. 72 CHO 596-7 AE, New York

Stålhammar, U., I. Karlsson and G. Fant 1973: "Contextual effects on vowel nuclei", *STL-QPSR* 4, p. 1-18

SYNTHESIS OF A FEMALE VOICE, A PRELIMINARY STUDY

Inger Karlsson[1]

Abstract: The ultimate aim of my work is to synthesize a female voice with the help of our OVEIII synthesizer. I have started this work by surveying the existing literature where only three articles describing experiments in this area could be found, none of which claimed to have been successful. We copied one of those experiments and also tried to copy one utterance made by a female speaker. None of these synthesized samples were acceptable as a female voice.

To gain more knowledge about the perceptual load of different parameters we ran a test where the fundamental frequency, the bandwidths, and the higher formants were varied. The listeners seemed to be aware of differences in average fundamental frequency that were greater than 10 Hz and were more sensitive to changes in the first formant than in the higher formants.

Experiments with female voice synthesis

The ultimate goal of my research is to produce a fully acceptable synthesis of a female voice. Little seems to have been done in this area. I have found only three relevant articles: U. Goldstein (1972), H. Sato (1974), T. Yasuhiro and K. Ozeki (1976). None of those claimed to have been wholly successful. Anyhow, we decided to try to copy the most promising of these studies, namely the one by Yasuhiro and Ozeki. They have used LPC technique, see B. Atal and S. Hanauer (1971), the output rate being set to 1.3 times the input rate, thus increasing the formant frequencies and bandwidths by 30%, and the fundamental

---

[1] Department of Speech communication, Royal Institute of Technology (KTH), S 100 44 Stockholm 70, Sweden

frequency being set to 2.1 times the original value. A sentence from a reading made by a male speaker was used as test material. The processed utterance was passed through a comb filter with zeros at 900, 2800, 4700 and 6600 Hz to get a more female-like result, and this way, according to the authors, "...an almost satisfactory female voice ..." was achieved. We tried to copy this experiment with the R. Carlson and B. Granström (1976) rule synthesis program, the formant frequencies and bandwidths being raised by the same amount as in the Yasuhiro and Ozeki experiment, and the resulting synthesis being filtered as described above. Examples of this synthesis, both filtered and unfiltered, were played at the seminar; neither of these examples had any great resemblance to a female voice. One possible explanation of this is that the relations between male and female formant frequencies and bandwidths are nonuniform, see G. Fant (1966) and O. Fujimura and J. Lindqvist (1971), whereas in this experiment they were assumed to be uniform. We therefore synthesized the same sentence with the formant values set to typical average values for female speakers, but this did not remarkably improve the "femaleness" of the synthesized voice. We have also tried to make a synthetic copy of an utterance made by a woman. The result can be seen in Fig. 1, where spectrograms of the original and the synthesized copy are shown. These samples were also played at the seminar. Though a similarity could be heard between the two, the similarity was not so much in voice quality as in prosody, and the synthesized voice could hardly be accepted as a female voice.

In all the examples mentioned above the parameters altered to get a more female voice were formant frequencies, bandwidths, and fundamental frequency. As we could ascertain by listening to the synthesized speech, these parameters do not carry all information that is needed to unambiguously define a certain sex. Further parameters to be taken into consideration are those of the voice source, as well as variations within and between sounds of the parameters that we used in the experiments described above.

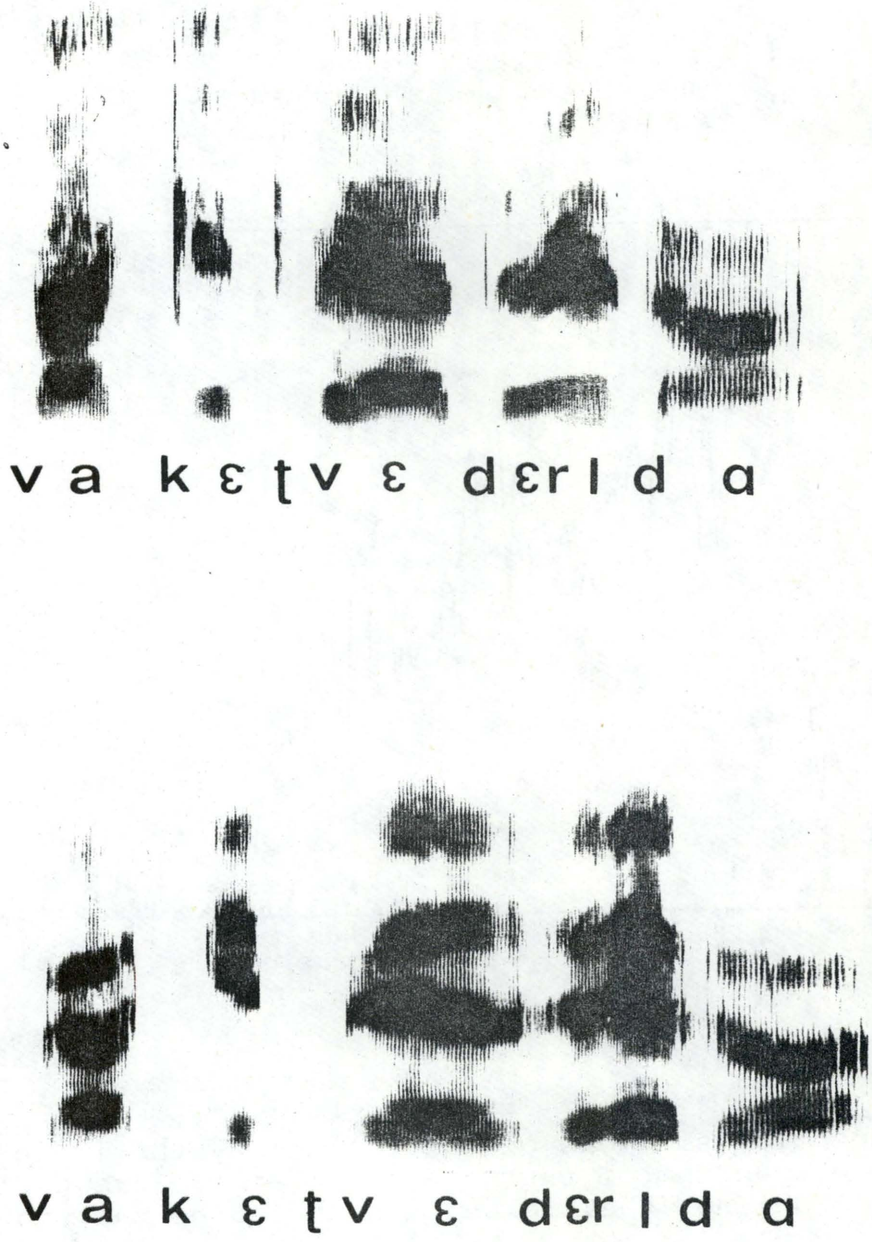


Figure 1

Upper part human voice, lower part synthetic copy of the utterance: "Vackert väder idag".

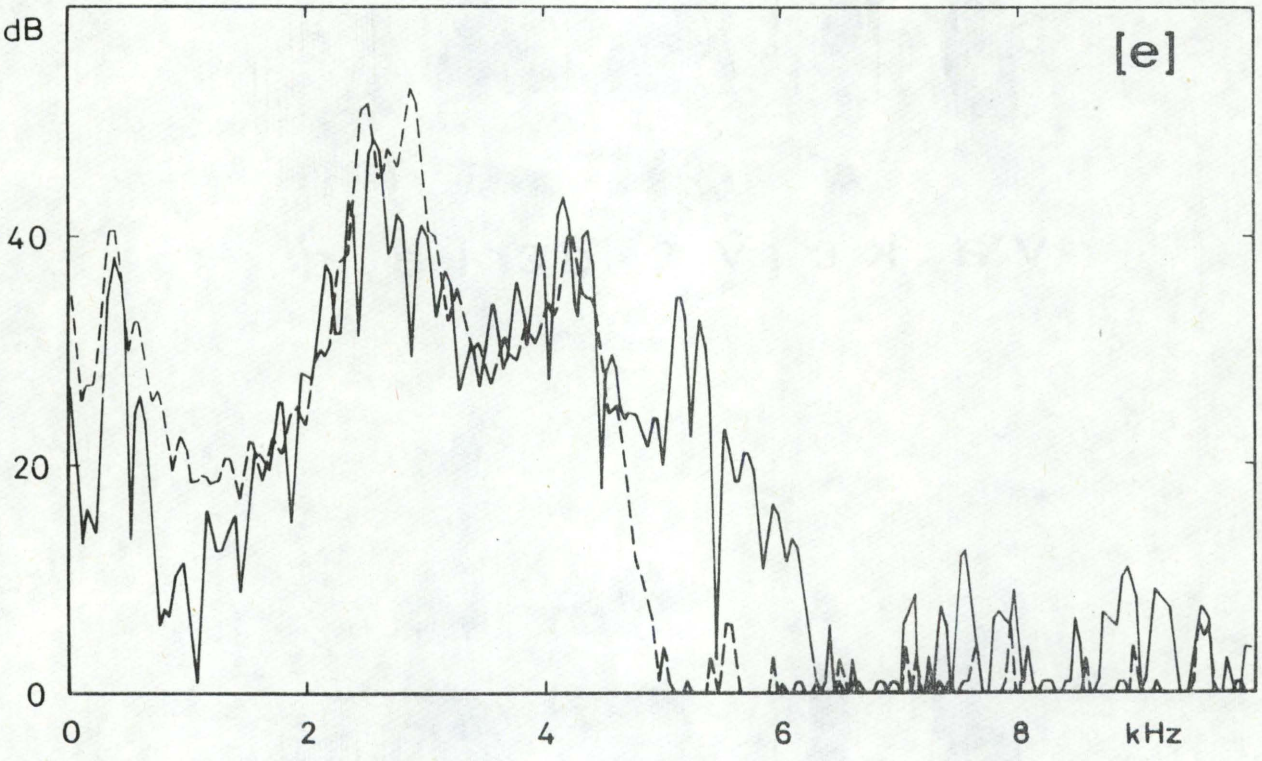


Figure 2

Spectrum sections of natural and synthetic speech. The spectrum of the natural speech is drawn with a solid line, the broken line indicates the synthetic speech.

Listening test

We have also tried to estimate the relative importance of different parameters for the perception of similarity of sounds. To do this we used a listening test where the test participants were asked to judge the similarity between two syllables using a scale ranging from 5 for full similarity to 1 for no similarity. The first syllable was the same throughout the whole test: a syllable pronounced by a woman. The second was a synthetic copy of the first where the frequency and amplitude of the fourth formant, the average value of the fundamental frequency, the bandwidths of the first and of the second + third formants were varied, one at a time. The other parameters were held at values that, as judged by spectrograms, were as similar to those of the first syllable as possible. Spectrum sections of the natural vowel and of the synthetic copy that was used as a reference for the different parameters are shown in Fig. 2.

The results of the test are given in Fig. 3. We can see that the fundamental frequency has to be altered at least 10 Hz before the similarity rating decreases. For the fourth formant an 8 dB decrease in amplitude does not influence the result, whereas a 500 Hz decrease in frequency shows only a weak tendency towards lower ratings. That a frequency increase gives such high rating values probably depends on the fact that the speaker we used in the test has a strong fifth formant. The bandwidth of the first formant seems to be an important parameter, whereas the listeners seemed to be unaware of changes in the second and third formant bandwidths. Due to the type of synthesizer we used (a terminal analog synthesizer) it is impossible to tell whether it is really the differences in bandwidths that are perceived; it may just as well be the amplitude differences. According to Kakusho et al. (1971), the ear seems to be fairly sensitive to amplitude differences, especially if these are located in the first formant. This suggests that voice quality depends much on

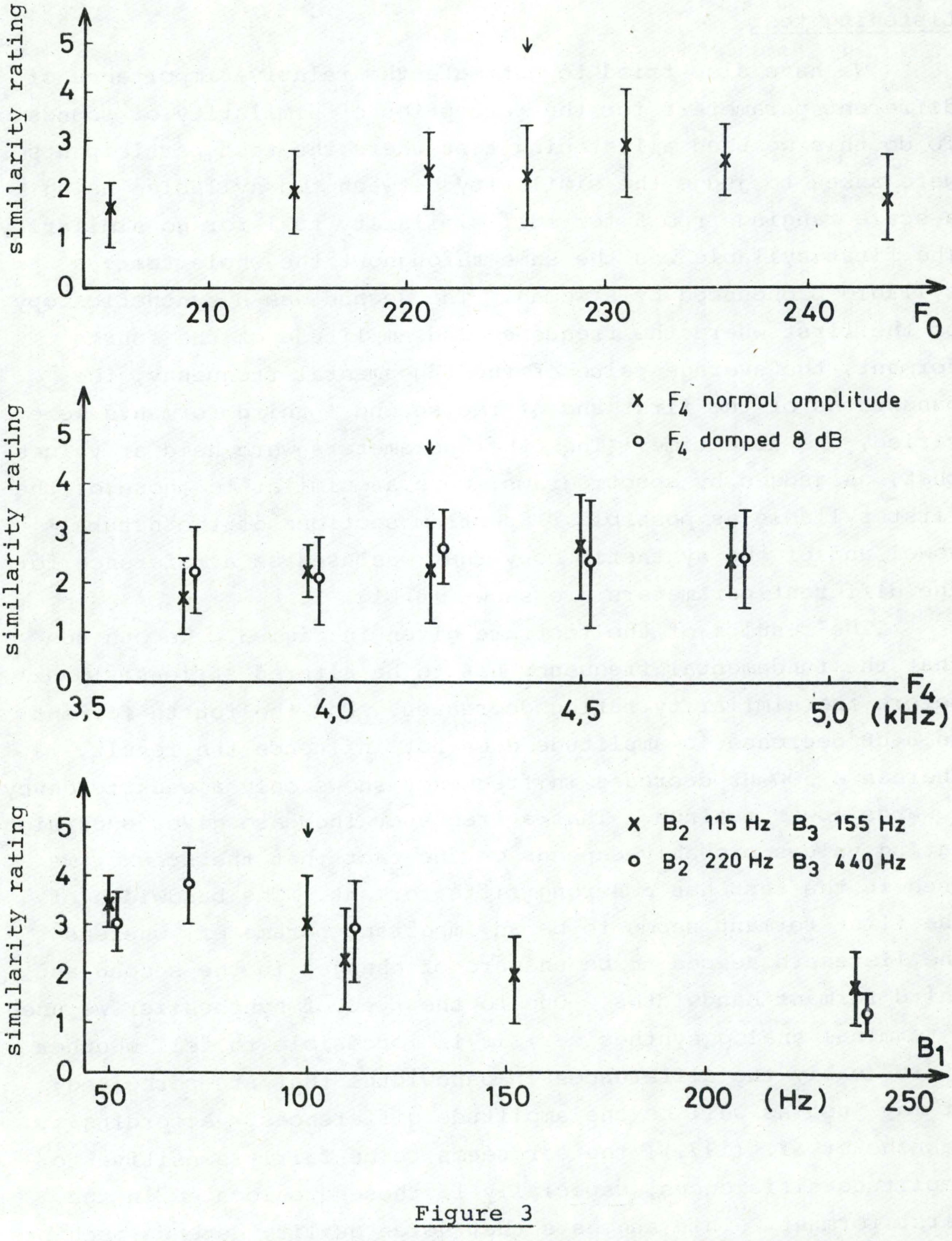


Figure 3

Results of the listening test. Mean values and standard deviations of the similarity ratings are shown. The reference sample that was judged the most similar to the natural speech sample from spectrograms is indicated with an arrow.

the lower parts of the spectrum, the fundamental frequency and the first formant region. It is interesting to compare with the K.Ågren and J.Sundberg (1976) study of tenor and alto voices: they found that the differences between those two groups can be ascribed, for the most part, to a stronger fundamental and a higher, +400 Hz, fourth formant for the alto voices. Combined with our results this would indicate that the next thing I ought to do to improve the female voice synthesis would be to raise the amplitude of the fundamental.

References:

Atal, B.S. and S.L. Hanauer 1971: "Speech analysis and synthesis by linear prediction of the speech wave", JASA 50, p. 637-655

Carlson, R. and B. Granström 1976: "A text-to-speech system based entirely on rules", Conf. Records, 1976 IEEE Int. Conf. on Acoustics, Speech and Signal Processing, Philadelphia USA, p. 686-689

Fant, G. 1966: "A note on vocal tract size factors and non-uniform F-pattern scalings", STL-QPSR 4, p. 22-30

Fujimura, O. and J. Lindqvist 1971: "Sweep-tone measurements of the vocal tract characteristics", JASA 49, p. 95-100

Goldstein, U. 1972: Comparisons of terminal-analog synthesis of male and female voices, Unpublished thesis, MIT

Kakusho, O., H. Hirato, K. Kato and T. Kobayashi 1971: "Some experiments of vowel perception by harmonic synthesizer", Acoustica 24, p. 179-190

Sato, H. 1974: "Acoustic cues of female voice quality", Electronics and Communication in Japan 57-A, p. 29-38

Yasuhiro, T. and K. Ozeki 1976: "An experiment on male to female voice conversion", J. of the Acoustical Society of Japan 32, p. 362-368

Ågren, K. and J. Sundberg 1976: "An acoustic comparison of alto and tenor voices", STL-QPSR 1, p. 12-16

My talk was followed by a discussion over the topic: Can you judge the sex from the voice? In this discussion I cited the following articles:

Ingeman, F. 1968: "Identification of the speaker's sex from voiceless fricatives", JASA 44, p. 1142-1144

Lass, N., K. Hughes, M. Bowyer L. Waters and V. Bourne 1976: "Speaker sex identification from voiced whispered and filtered isolated vowels", JASA 59, p. 675-678

Weinberg, B. and S. Bennett 1971: "Speaker sex recognition of 5- and 6-year-old children's voices", JASA 50, p. 1210-1213

MEASUREMENTS OF FORMANT BANDWIDTHS IN THE TIME DOMAIN

Lennart Nord[1]

Abstract: Two methods of estimating formant bandwidth in the time domain were tried. The vocal tract with the glottis closed was excited, either by pulses or by a gated sine wave signal, with its frequency adjusted to one of the resonances of the vocal tract. Bandwidth values were obtained from the decay curves. Measurements on Swedish vowels, articulated by two male and two female speakers were performed and the results compared to sweep-tone measurement data.

Introduction

Acoustic data such as bandwidth values are needed for a thorough modelling of the vocal tract, including losses. Sources of error when measuring bandwidths are glottal and nasal coupling which will broaden the formant peaks. A movement of the articulators, shifting the resonance frequency will also result in too large values.

Two techniques of measuring bandwidth in the time domain will be described. The experiments are part of a series of experiments, investigating the acoustical properties of the vocal tract. (For some aspects of the wall impedance of the vocal tract, see Fant, Nord and Branderud (1976).)

Theory

A resonance can be described in the frequency or in the time domain, see fig. 1. The bandwidth B in the frequency plane

---

1) Department of Speech Communication, Royal Institute of Technology (KTH), S-100 44 Stockholm 70, Sweden.

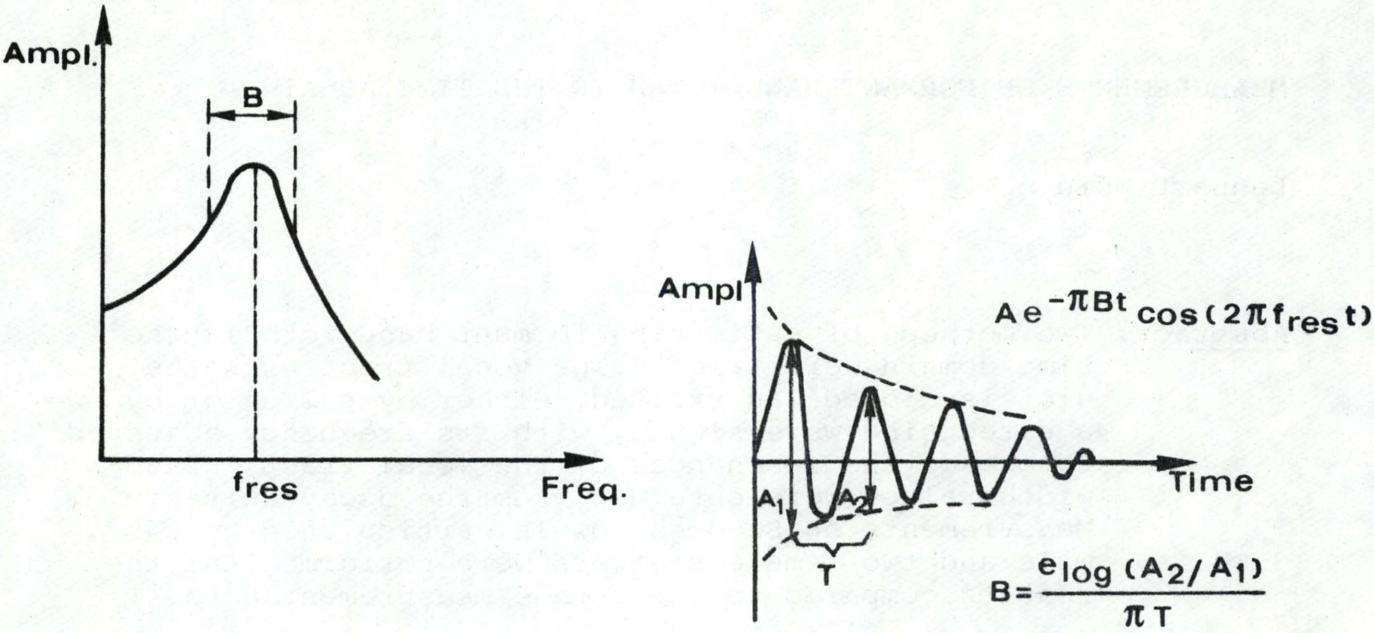


Figure 1

a) Frequency response, and b) impulse response for a resonator.

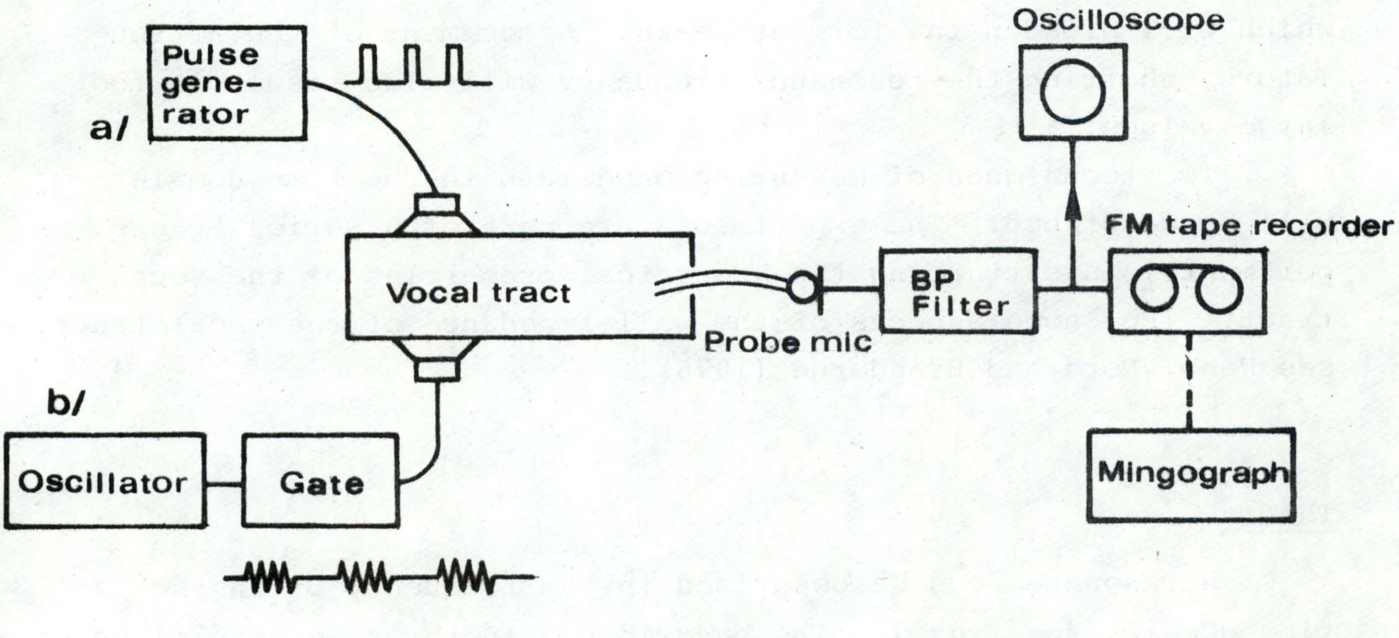


Figure 2

Two methods of bandwidth measurement.
a) Pulse excitation.
b) Gated sine wave excitation.

corresponds to the width of the resonance peak and can be measured, for example by means of a sweep-tone analysis technique, see Fujimura and Lindqvist (1971). If a resonance is excited by a pulse, the amplitude will oscillate and decay according to fig. 1b. In the time domain the damped oscillation can be described by the expression $A \cdot e^{-\pi Bt} \cdot \cos(2\pi f_{res} \cdot t)$, where the factor B thus determines the rate of decay.

Experiment

Method 1 See fig. 2. The vocal tract was excited by short pulses from an electromagnetic transducer (of a type normally used as a sound source for laryngectomized) driven by a pulse generator. The transducer had a contact area of approximately 7 cm^2 and was placed just above the larynx on one side of the throat. The placement was adjusted so that loud clicks were heard when the subject held his glottis and velum closed and his mouth open. The pulse width and amplitude were adjusted to give a strong first transient and negligible ringing. (The resonance of the transducer alone was well above the first formant but could in some cases interfere with the second formant.) The sound was registered by a small condenser probe microphone and displayed on an oscilloscope screen. The probe was a thin damped plastic tube inserted about 1 cm into the mouth. For a given articulation a number of pulse responses of the vocal tract were recorded on an FM tape recorder. The recorded signals were played back at a reduced speed (by a factor 16) and a mingograph tracing of the signal was obtained. Consecutive peak-to-peak amplitude values of the decaying signal were plotted on a logarithmic scale and a straight line was fitted visually to the points. The slope of the line gave the bandwidth value of the resonance. For vowels with two formants lying close together it was sometimes necessary to bandpass filter the microphone signal to get the bandwidth value of one of the resonances.

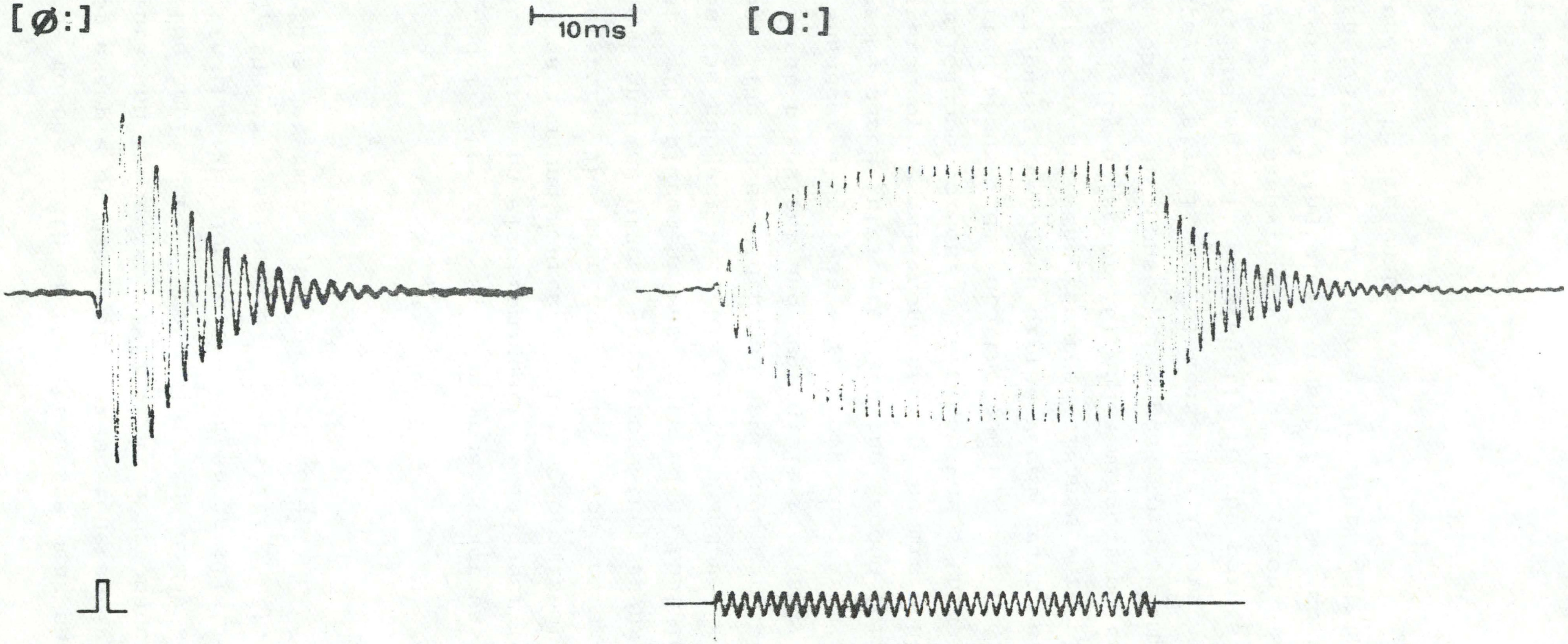


Figure 3

Mingograms of pulse excitation response and gated sine wave excitation response signals. The pulse and sine wave are indicated.

Method 2 The exciting pulse did not correspond to an ideal pulse, which meant that only little energy was transmitted through the vocal tract wall at higher frequencies. This made it difficult to measure anything more than the bandwidth of the first formant reliably.

Another way of exciting the tract was therefore tried. The transducer was fed with a sine wave signal of the resonance frequency of interest. By gating this signal, the rise and decay pattern of the recorded response signal could be treated in the same way as the pulse response. In fig. 3 both types of response signals are shown together with the indicated pulse and sine wave excitation signals. The latter method demanded less energy from the transducer as all the energy was used; however, the mechanical constraints of the system, even with this method, made it difficult to get bandwidth data from higher formants.

Measurements

Both methods were tried on a number of male and female subjects. Some practice was usually needed to make the subjects aware of the type of articulation needed with the glottis closed and no nasal coupling. Method two also required that the articulation was held fixed while the first resonance frequency of the tract was found with the oscillator. This could be avoided by some automated frequency scanning.

Results and discussion

Results for two male and two female speakers are plotted in fig. 4. No consistent difference was found between the measurement values for the two methods that were tried. Therefore, no distinction is made in fig. 4.

Included in fig. 4 are also average values from sweep-tone measurements made by Fujimura and Lindqvist (1971). The similari-

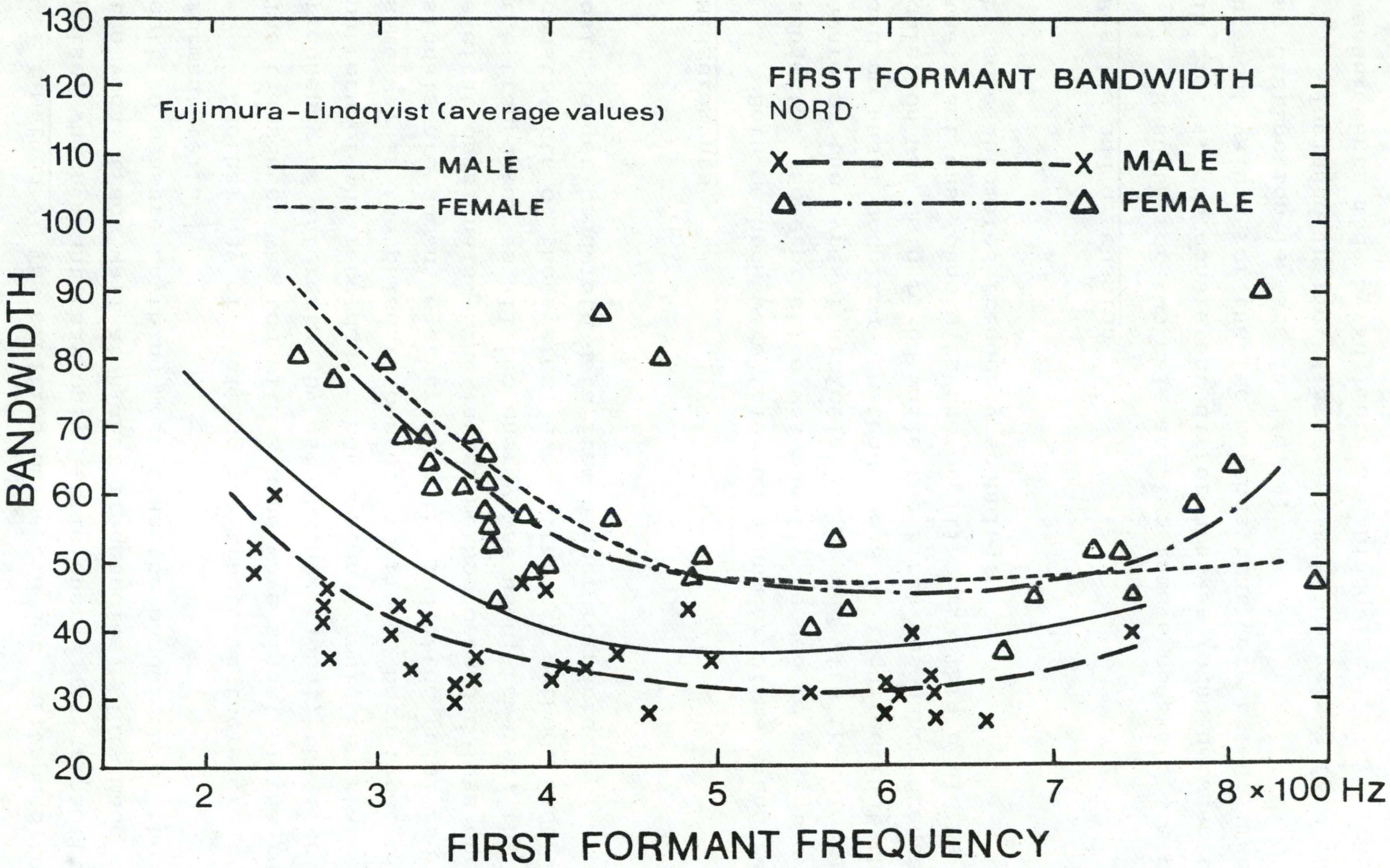


Figure 4

Bandwidth versus first formant frequency for two male and two female speakers. Average values from sweep-tone measurements by Fujimura and Lindqvist (1971) are included as a comparison.

ties are quite good with the same tendencies concerning male-female differences and high bandwidth values when F_1 is low, due to increased wall losses. For discussions on bandwidths, see articles by Fujimura and Lindqvist (1971) and by Fant (1972).

The time domain data in fig. 4 appear to have somewhat lower values than the sweep-tone measurements. This might be explained by the fact that measurements in the frequency domain take longer time and will give broader peaks if the resonance frequency shifts during the frequency sweep. Time domain analysis is generally faster and a changing of the articulation will not affect the bandwidth to the same degree.

If two formants are very close, such as F_1 and F_2 for high back vowels, a satisfactory filtering of F_1 cannot be made. The impulse response will not decay exponentially and B_1 cannot be evaluated in the manner described above. (Compare the problem of extracting bandwidths in the frequency plane when two formant peaks lie close together.)

An alternative method of pulse excitation is to use an electric spark as an acoustic source inside the mouth cavity. This technique has been described by House and Stevens (1958).

Summary

The two methods of estimating bandwidths in the time domain of the vocal tract, closed at the glottal end, gave promising results for the measurements of B_1. Values for some vowels articulated by two male and two female speakers were consistent with sweep-tone measurement data obtained by Fujimura and Lindqvist (1971).

References

Fant, G. (1972): "Vocal tract wall effects, losses and resonance bandwidths", STL-QPSR 2-3, p. 28-52

Fant, G., L. Nord and P. Branderud (1976): "A note on the vocal tract wall impedance", STL-QPSR 4, p. 13-20

Fujimura, O. and J. Lindqvist (1971): "Sweep-tone measurements of vocal tract characteristics", JASA 49, p. 541-558

House, A.S. and K.N. Stevens (1958): "Estimation of formant band widths from measurement of transient response of the vocal tract", JSHR 1/4, p. 309-315

FINAL LENGTHENING - A CONSEQUENCE OF ARTICULATORY AND PERCEPTUAL RESTRICTIONS?

Bertil Lyberg[1]

Introduction

The duration of speech segments varies according to the position in the word and in the phrase. In the literature some of these variations are known as initial and final lengthening.

Segmental variation has mainly been studied as an isolated phenomenon without any connection to parameters such as fundamental frequency and intensity variations.

In this study, attempts have been made to explain some of the temporal regularities of spoken Swedish such as, e.g., "final lengthening" by means of articulatory and perceptual restrictions on the speed of fundamental frequency change.

Segment durations

Observations

In an investigation reported previously (Lyberg 1977), vowel duration measurements were made for an invariant "probe" sequence [-ˈdɑːg-], where position at the word and phrase levels was systematically varied. The results from the study can be summarized by the following points:

a) The duration of the test word vowel is longer when the word is at the end of a phrase, at least when the test word vowel is followed by fewer than two unstressed syllables.

---

1) Department of Phonetics, Institute of Linguistics, Stockholm University, Fack S-10405 Stockholm 50, Sweden.

b) In phrase-final positions, the duration of the vowel segment decreases as a function of the number of preceding main stresses.

c) In all other positions, the duration of the vowel does not show any systematic variations depending on the position in the phrase.

Model based on a lengthening process

If the "shortest" inherent duration of the speech segments is used as a basic input parameter, it seems possible to describe the durational patterns of Swedish utterances by means of the following expression:

$$D = (u(2a_p + a_w - 2) + u(1 - 2a_p - a_w) \cdot k \cdot (1 + a_w)^{-\alpha} (1 + b_p)^{-\beta}) D_1$$

$$\text{where } u(n) = \begin{cases} 1 & \text{if } n \geqslant 0 \\ 0 & \text{if } n < 0 \end{cases}$$

D = the segment duration of a given main stress vowel in a word and in a phrase

D_1 = the inherent "shortest" duration of the vowel segment.

a_w = the number of syllables that follow at a given point.

a_p = the number of main stresses that follow at a given point.

b_p = the number of main stresses that precede at a given point.

α, β, k = constants.

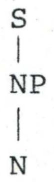
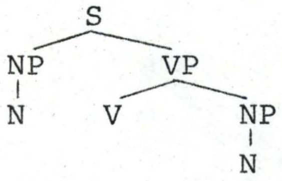
Contrary to the earlier models (Lindblom and Rapp 1973, Lyberg 1977) the constants used in the present model are totally independent of the metric structure of the test words.

Fundamental frequency change

Observations.

Some of the sentences of the speech material used in an earlier investigation (Lyberg 1977) were also used in this investigation in order to study the frequency contour for two of the speakers (JJ, PEN). The sentences used appear in Table 1.

Table 1

Sentence	syntactic structure
Dag Dagen Dagobert Idag	
Dag berömmer Dagobert Dagobert berömmer Dag	

The experimental results are presented for one of the two speakers in figs. 1 and 2. Every point in the diagrams represents a mean value of ten comparative samples.

The fundamental frequency contour at the end of the sentences consists of a maximum followed by a minimum value. This phrase-end signalling will always occur in the final word of the sentences and different cases will occur depending on the metric structure of the final word:

a) The final word consists of only one syllable. Both the maximum and the minimum have to be manifested in the same syllable.

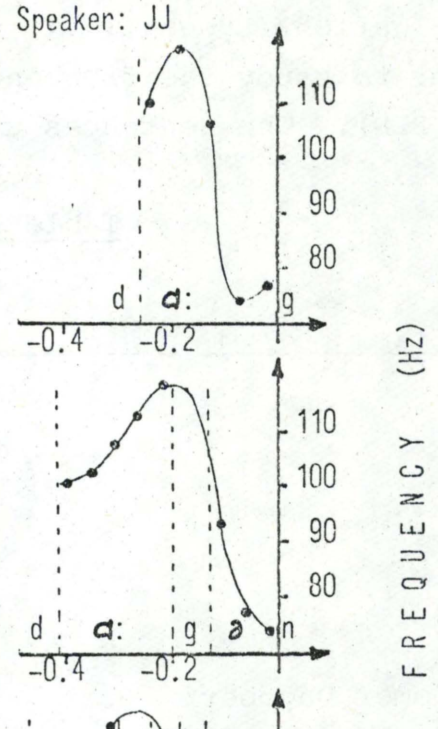


Figure 1

The fundamental frequency contours of the one-word sentences [dɑ:g], [dɑ:gən], [dɑ:gɔbæʈ], [ɪdɑ:].

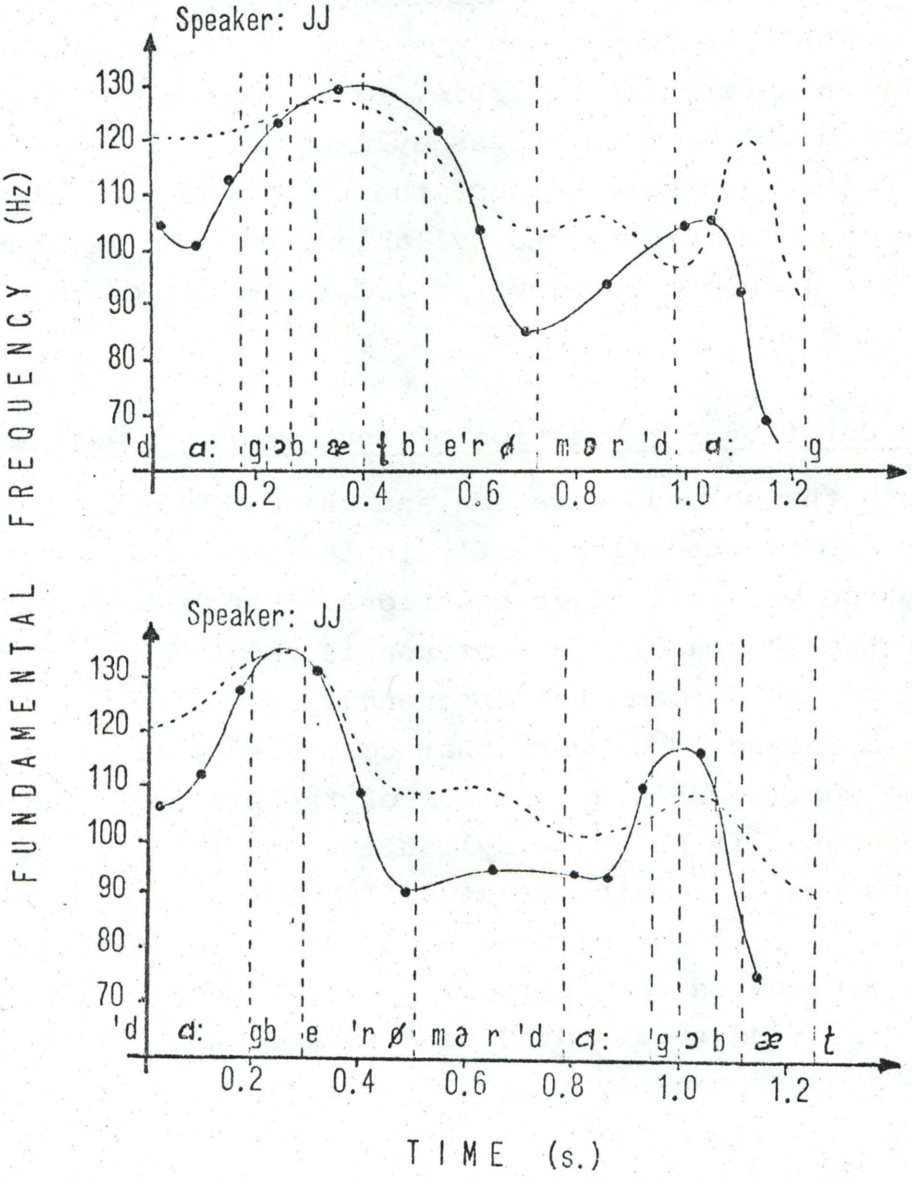


Figure 2

The fundamental frequency contours of the sentences [ˈdɑ:gɔbæʈbeˈrømərˈdɑ:g] and [ˈdɑ:gbeˈrømərˈdɑ:gɔbæʈ].

b) The final word consists of a main stressed syllable followed by an unstressed syllable. The minimum value is then located in the unstressed syllable but the maximum will still occur in the main stress syllable.

c) The final word consists of a main stress syllable followed by two unstressed syllables. Both the maximum and the minimum are located in the unstressed syllables.

d) The final word consists of a main stress syllable preceded by an unstressed syllable. The maximum and the minimum are located in the main stress syllable.

All the main stresses but the final one when it is followed by less than two unstressed syllables, are manifested by a minimum in the frequency contour (cf. Carlson and Granström 1973).

Segment duration - a function of fundamental frequency change?

Both the segment duration and the fundamental frequency contour are treated differently in the phrase final position as compared with all other positions. The question then arises whether the lengthening phenomenon is related to the phrase-end command of the fundamental frequency.

In a phrase final word that consists of only one syllable, both the maximum and the minimum of the phrase-end command have to be executed in the same syllable. But in a phrase final word that consists of a main stress syllable followed by an unstressed syllable, the maximum will be located in the main stress syllable and the minimum in the unstressed syllable.

If the time required to perform a certain frequency deviation is a monotonically rising function of the frequency deviation, it is obvious that the stressed vowel in a phrase final word like /dag/ is longer than the stressed vowel in e.g. /dagen/.

The segment duration is thus only prolonged when it is necessary in order to perform the required fundamental frequency deviation. The variation of the segment duration at different places in a sentence is then only a secondary effect of the signalling by means of the fundamental frequency.

References

Carlson, R. and B. Granström 1973: "Word accent, emphatic stress and syntax in a synthesis by rule scheme for Swedish", STL-QPSR 2-3/1973, p. 31-36

Lindblom, B. and K. Rapp 1973: "Some temporal regularities of spoken Swedish", PILUS (Papers from the Institute of Linguistics, University of Stockholm) No. 21

Lyberg, B. 1977: "Some observations on the timing of Swedish utterances", Journal of Phonetics 5, p. 49-59

(ARIPUC 12, 1978)

THE GLOTTAL GESTURE IN SOME DANISH CONSONANTS - PRELIMINARY OBSERVATIONS

Birgit Hutters[1]

1. Introduction

In the present paper some preliminary fiberoptic and glottographic observations of various Danish obstruents will be presented. Furthermore, some of our preliminary EMG[2] results will be mentioned. But it should be emphasized that the results are *very* preliminary, and thus the present account may deviate somewhat from the results as they will appear in the final report.

2. Procedure

Photo-electric glottograms of various obstruents in stressed, intervocalic position were made with the fiberscope positioned in the pharynx, serving as light source, and a photo-transducer placed on the frontal part of the neck. In addition to the light source function the fiberscope has a controlling function: during the glottographic recording a still-picture of the larynx is taken during each test sound, making it possible to control the stability of the light and the larynx. This stability is very important for the reliability of the glottographic recordings.

---

[1] Institute of Phonetics, University of Copenhagen.

[2] The EMG project at the Institute of Phonetics is part of a larger framework titled: "The glottal behaviour in Danish consonants, stress, and stød". Fiberoptic and glottographic investigations are also included in this main project.

Furthermore, the still furnishes immediate information about the state of the glottis and the surrounding structures at the moment of exposure. A synchronizing pulse is recorded in order to synchronize the still and the glottogram. Moreover, a microphone signal is recorded. Fig. 1 illustrates the curves used for segmentation and extraction of parameters, and the still of the glottis, taken in the exemplified test sound, is displayed as well. The obstruents mentioned in this paper are p t k b d g f s said in the frame sentence "De ville sige [- i(:) $\binom{l}{n}$ə]" (they would say ...), i.e. the segment preceding and following the test sound is [i] ("sige" is pronounced [si:]). All the test words are meaningful Danish words. So far, only one subject (Hu) has been closely investigated, but a cursory inspection of other subjects has been performed as well. The dialect of the subject treated here is Advanced Standard Copenhagen Danish.

3. Extraction of parameters

The acoustic and glottographic signals are segmented as seen in fig. 1:

The acoustic signal

V_1: onset of the preceding vowel

V_2: offset of the following vowel

C_1: onset of the consonant

C_2: offset of the consonant

E: explosion of the stop.

The glottographic signal

G_1: onset of the glottal gesture[1]

G_2: offset of the glottal gesture[1]

---

[1] It should be noted that both the definition of the parameter called the opening-closing gesture of the glottis, and the identification of this parameter in the glottograms cause problems, which will be treated in a later report.

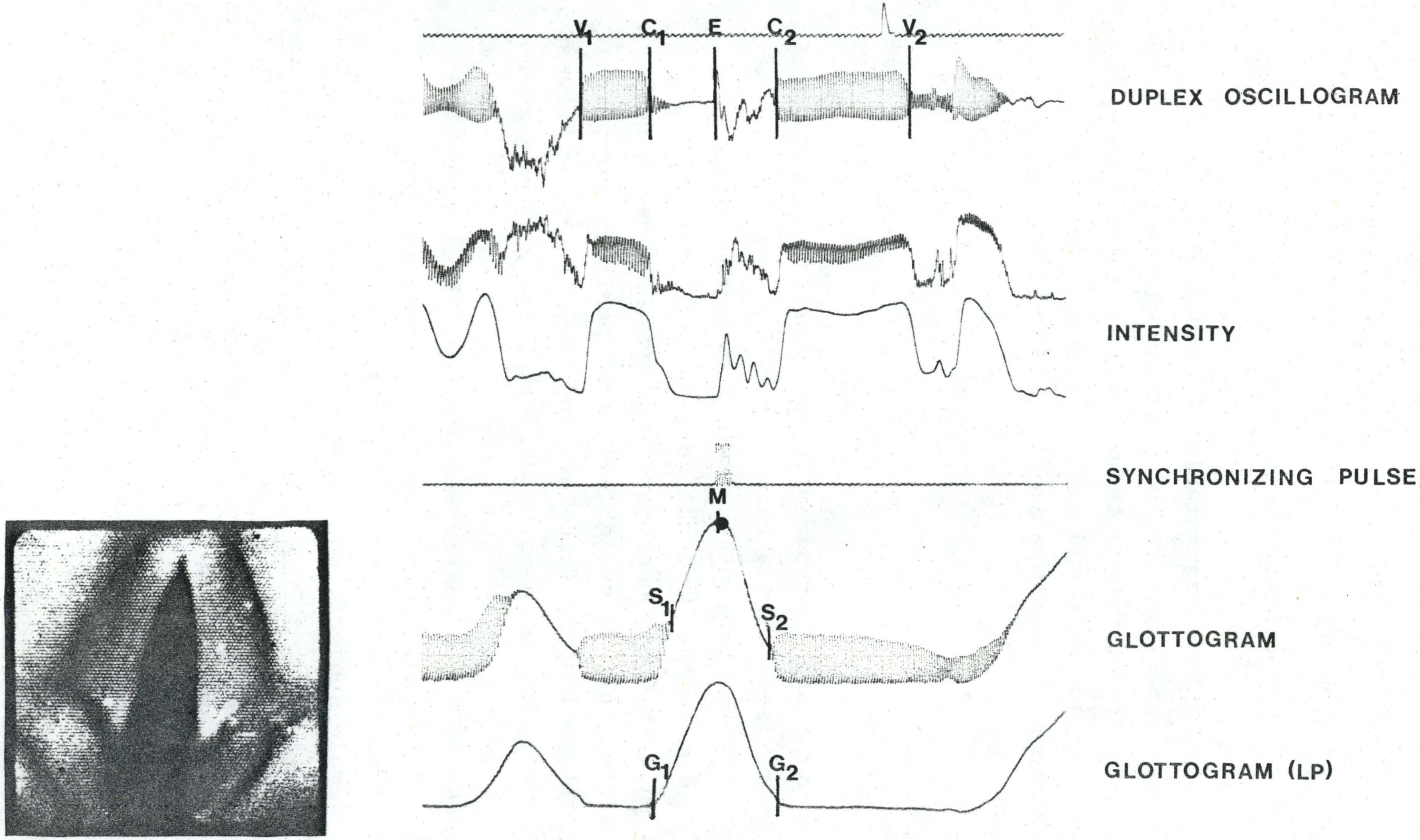


Figure 1

The curves used for delimitation of segments and extraction of parameters, and a still of the glottis, taken at the moment indicated by the filled circle on the glottogram.

S_1: offset of the vocal fold vibrations

S_2: onset of the vocal fold vibrations

M: moment of maximum glottis aperture.

On the basis of this segmentation a number of temporal parameters can be extracted, partly acoustic and glottal parameters, and partly parameters which may throw light on the temporal relationship between acoustic and glottal events. Besides the temporal parameters the maximum amplitude of the glottograms has been measured.[1]

4. Observations

The fiberoptic and glottographic observations will be presented as a comparison, partly between two different categories of obstruents (ptk vs. bdg; pt vs. fs), and partly between obstruents belonging to the same category (p vs. t vs. k; b vs. d vs. g; f vs. s). The stop consonants ptk and bdg are phonologically distinct only in syllable-initial position (principal rule). Both categories are voiceless and the difference between them is essentially one of aspiration, ptk being strongly aspirated.

Fig. 2 depicts the acoustic and glottographic parameters of the eight obstruents, shown with inclusion of the preceding vowel i. Each set of parameters is averaged over 10 measurements of one female subject, and presented in a stylized form.

---

1) It is well known that the relations between the amplitude of the glottogram and the glottis aperture are very complicated. These problems have been discussed in some detail in Hutters (1976) and will be treated again in a later report.

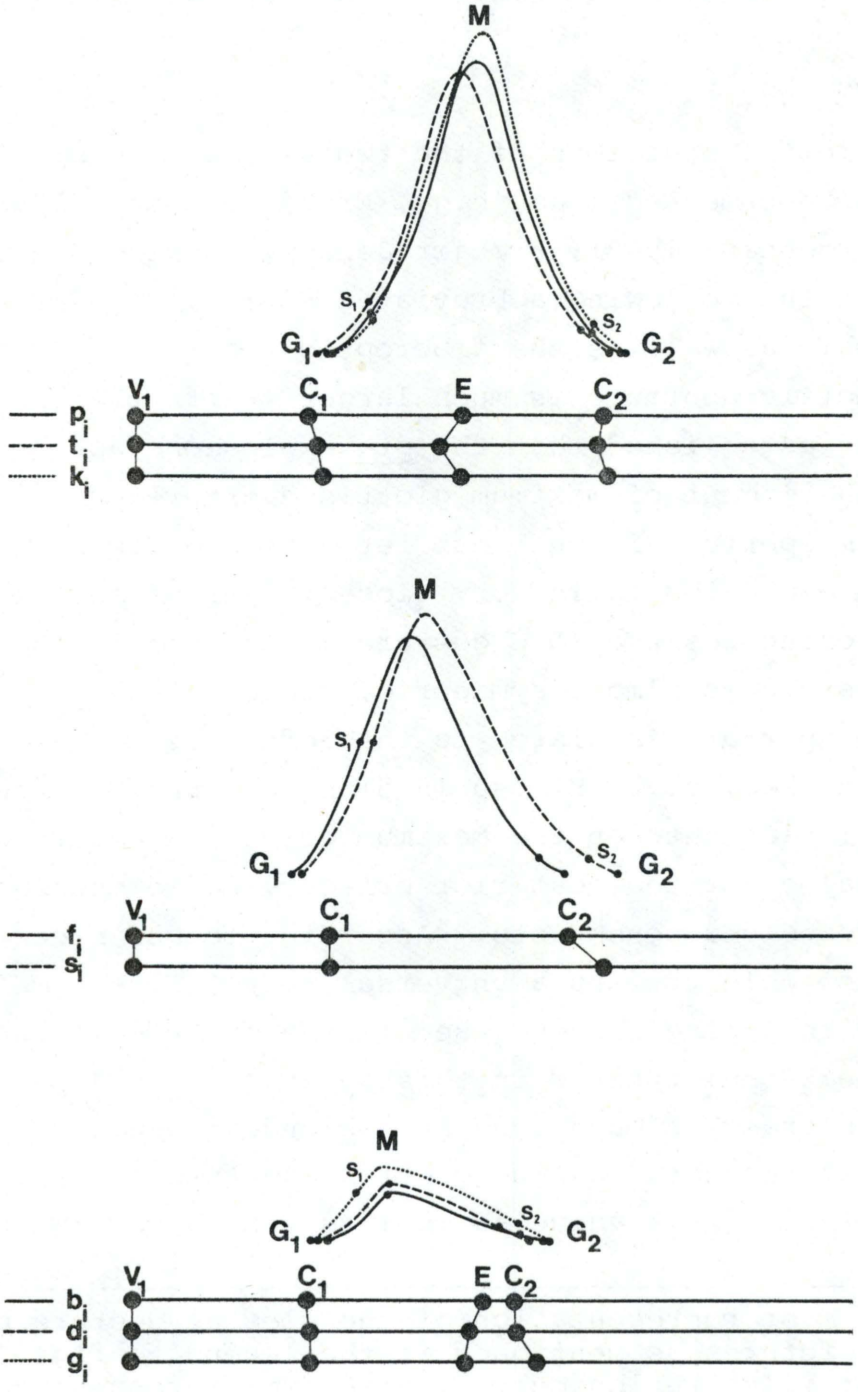


Figure 2

A stylized presentation of the acoustic and glottographic parameters of p t k b d g f s including the preceding vowel i, averaged over 10 measurements of one female subject. Line-up point is the onset of the preceding vowel.

Figs. 3 and 4 illustrate raw curves and stills of p and b (see the legend to figs. 3 and 4 for further explanation).[1]

4.1 ptk vs. bdg

The glottal behaviour of the two sets of stops is already rather well documented (see, e.g., Frøkjær-Jensen (1967, 1968), Fischer-Jørgensen (1968), Frøkjær-Jensen, Ludvigsen and Rischel (1971) - in the following abbreviated F-L-R). The present glottographic as well as the fiberoptic recordings confirm that maximum glottis aperture is much larger in ptk than in bdg. It has been established that the ptk explosion falls rather close to the moment of maximum glottis aperture, whereas in bdg the maximum aperture is much earlier than the moment of explosion. In bdg the explosion falls very close to the offset of the glottal opening-closing gesture (O-C gesture). As seen in the figures, the ptk gesture is almost symmetric, whereas in bdg the moment of maximum aperture is biased to the left.

Hirose (see, e.g., Hirose 1975) has shown that there is a high correlation between the maximum glottis aperture and the EMG peak value for the posterior cricoarytenoid muscle (PCA) in Japanese voiceless consonants. According to our recordings, this may not hold true as a universal rule. With some of our subjects - including the one used in the fiberoptic and glottographic recordings treated in this paper - the PCA values are of the same order of magnitude in ptk and in bdg, in spite of the big difference in maximum glottis aperture. The insufficiency of PCA activity (peak value as well as timing) as an indication

---

1) The form of representation of the glottal gesture based on still pictures as mentioned in the legend to figs. 3 and 4 is used for illustration only. This form of representation implies that the timing of the acoustic events is invariant across the tokens. No attempt has been made to meet this requirement, however. An investigation of the information given by representations based on still pictures and on moving pictures, respectively, and of the reliability of these two methods, is planned - and in part in progress - at the Institute.

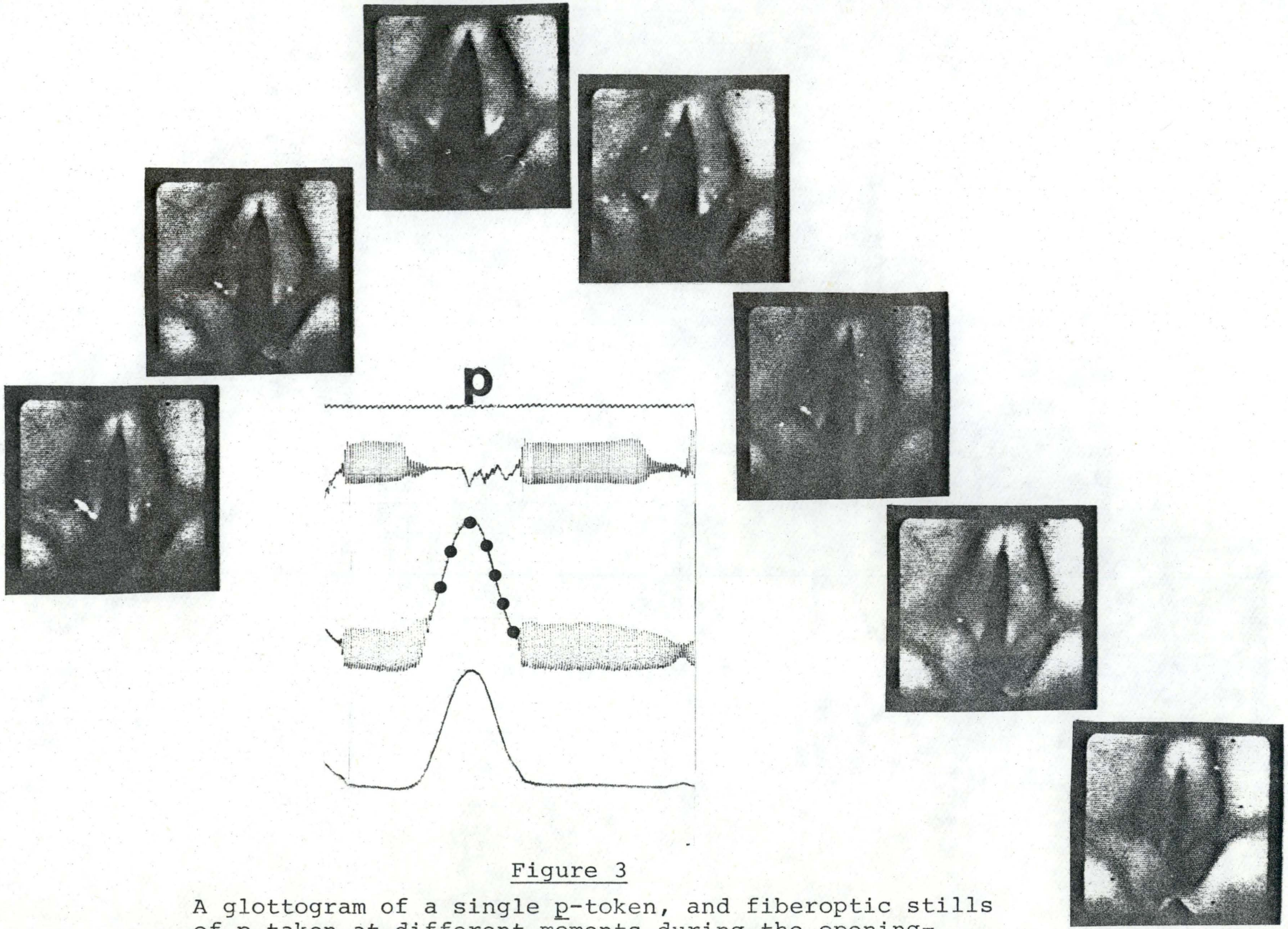


Figure 3

A glottogram of a single p-token, and fiberoptic stills of p taken at different moments during the opening-closing gesture of the glottis, as indicated by the filled circles (the stills originate from seven different renderings of the same sound).

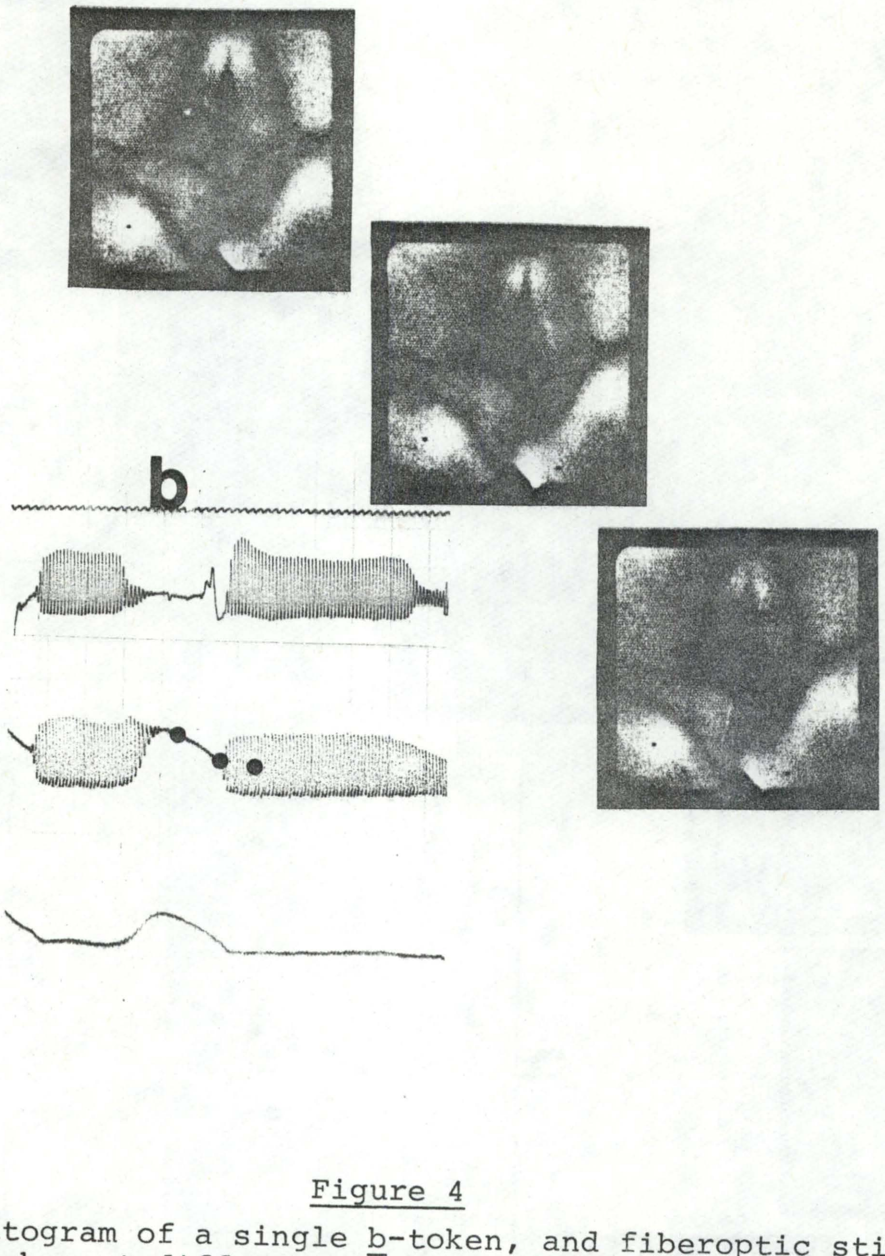


Figure 4

A glottogram of a single b-token, and fiberoptic stills of b taken at different moments during the opening-closing gesture of the glottis, as indicated by the filled circles (the stills originate from three different renderings of the same sound).

of the muscular activity underlying the actual articulatory realization is demonstrated very clearly by the EMG recordings of (voiced) v and l, which often show clear PCA activity, whereas the fiberoptic and glottographic recordings do not reveal a corresponding opening of the glottis.

4.2 pt vs. fs

Comparing the onset of the glottal O-C gesture to the acoustic onset of the consonant, it appears from my data that the onset of the O-C gesture lies earlier in the case of fricatives than in the case of stops. By measuring the distance from the onset of the preceding vowel to the onset of the O-C gesture I find this distance to be longer for stops than for fricatives, i.e. there seems to be a difference of timing of the glottal gesture. In most of our subjects a corresponding difference is seen in the EMG recordings.

So far the present recordings reveal that the maximum glottis aperture is smaller in fs than in pt, which is in agreement with F-L-R. The EMG recordings seem to show a corresponding difference between fs and pt if not just the PCA activity but also the activity of the interarytenoid muscle is taken into account.

As mentioned above, the glottal O-C gesture is almost symmetric in the aspirated stops; in the fricatives, however, the moment of maximum glottis aperture is biased slightly to the left.

Finally, I shall make some observations concerning the duration of voicing. In the fricatives the vibrations of the vocal folds continue rather far into the opening gesture, whereas in the stops the vibrations cease quickly after the onset of the glottal opening, due to the supraglottal closure. This difference in the duration of voicing is also observed, among others, by F-L-R. Moreover, they found the reverse relationship for the onset of the vibrations. In the fricatives the vibrations did not occur again until the adduction of the vocal folds was almost completed, whereas in the aspirated stops they started well in advance of the completion of the adduction. In the material presented here, however, this difference in onset time of the vocal fold vibrations between stops and fricatives does not appear. This discrepancy is probably due to a different quality of the following vowel. In the material presented in this paper the following vowel is i, whereas in the material of F-L-R the following vowel is a vowel of e or ε quality. In another part of my material the following vowel is a vowel of a quality, and here the vibrations of the vocal folds do indeed reoccur early in the adduction phase of the glottal O-C gesture for stops but not for fricatives, thus confirming the observation made by F-L-R. Such differences in the offset and onset of vocal fold vibrations are due to different aerodynamic conditions, and consequently, not only the consonant is decisive but also the surrounding sounds. Furthermore, the anatomy of the speech organs - i.e. the size of the larynx and the supraglottal cavities - may be of importance for the voicing conditions.

4.3 p vs. t vs. k

The aspirated stops ptk differ in more respects than just place of articulation. As early as 1954, Eli Fischer-Jørgensen showed that the duration of "aspiration" (affrication plus aspiration) is shortest in p, longer in k, and longest in t. This has been confirmed in a number of subsequent investigations and is confirmed also by my data, although the difference between p and k does not seem to be significant in this material (Hu). The difference in the duration of "aspiration" is correlated to affrication: the more the consonant is affricated, the longer is the duration of "aspiration". It is worth mentioning that the duration of the supraglottal closure - at least in the subject treated here - shows the opposite relationship, i.e. the p closure is longest, the t closure shortest. In fig. 2 it is seen that the p explosion lies closer to the maximum glottis aperture than does the t explosion, the k explosion lying between p and t. But the differences in distance between the explosion and the moment of maximum aperture of the glottis is not caused solely by differences in closure duration but also by differences in the timing of the moment of maximum glottis aperture.

In several investigations of the glottal behaviour in stops it has been observed that there seems to be a correlation between maximum glottis aperture and place of articulation. The more the place of articulation is retracted, the larger is the maximum glottis aperture (see, e.g., Fischer-Jørgensen 1968, Ondráčková 1970, Sawashima 1974, Pétursson 1976). This seems to be partly confirmed by my glottographic and fiberoptic recordings. The difference between p and k seems clear, whereas the maximum aperture in t varies considerably.

4.4 b vs. d vs. g

Some minor differences are also observed between b, d, and g. As pointed out by Eli Fischer-Jørgensen (1954, 1968), the duration from the moment of explosion to the onset of the following vowel is longer in g than in b and d, due to the affrication of g. My data shows, furthermore, that the duration of the open phase is longer in d than in b, due to the affrication observed in the younger generation's pronunciation of d.[1] The difference in the degree of affrication is correlated to a difference in the timing of the explosion in relation to the offset of the glottal O-C gesture.

Concerning the maximum glottis aperture, the relationships are similar to those found in ptk, in that we find the smallest aperture in b and the largest aperture in g.

4.5 f vs. s

It is obvious that the duration of f is shorter than that of s, and a corresponding difference is seen in the duration of the glottal O-C gesture.

Concerning maximum glottis aperture, it is seen that the aperture of s is larger than that of f.

---

1) It has to be recalled that the following vowel is i, which creates optimum conditions for affrication.

5. Final remarks concerning the control of the glottis

Although the observations of glottal behaviour (and its relation to supraglottal events) in various Danish consonants presented here are rather random and very preliminary, I might venture some more general remarks concerning the glottal behaviour and some assumptions about its underlying control mechanisms.

F-L-R put forward the interesting hypothesis that the opening-closing gesture in Danish ptk is due to neural commands, whereas the gesture in bdg may be a consequence of the aerodynamic conditions. The two gestures are called active and passive, respectively. This hypothesis was later rejected as a consequence of EMG recordings showing PCA activity in ptk as well as in bdg (see Fischer-Jørgensen, Hirose 1974). However, it is conceivable that the PCA activity observed in bdg has nothing whatsoever to do with the preplanned motor control of these stops,[1] but is rather caused by some peripheral reflex mechanism, elicited by the change in, say, the intraoral pressure, and having no special function in speech. The stills of the glottis in bdg show that the PCA activity influences only to a very limited degree the distance between the vocal processes and, furthermore, the maximum glottis aperture is not between the processes but in the middle of the ligamental part of the vocal folds as in the case of vocal fold vibrations.

Many years ago, it was realized that the offset of voicing in Danish bdg is solely a matter of the aerodynamic conditions (Fischer-Jørgensen 1963), and the same explanation may be advanced concerning the opening-closing movement in these consonants. If it is true that glottal behaviour in Danish bdg is primarily a consequence of the aerodynamic conditions and that the observed PCA activity is elicited by some reflex mechanism, then this glottal behaviour should appear also in the

---

[1] Any discussion in such terms must be taken with reservations since it is a fundamental problem whether the preplanned motor control involved in speech production is of a totally different nature than generally assumed.

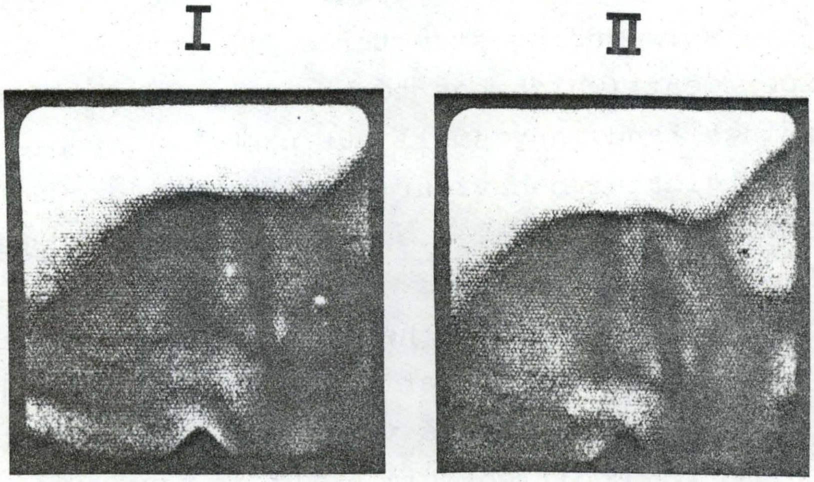


Figure 5

Fiberoptic stills of the larynx taken during [b̥] (I) and during an externally implemented closure (II). See the text for further explanation.

case of an externally implemented closure which can be obtained in the following way: the subject is phonating a sustained i vowel into an airtight mask with a small aperture. When the subject's phonation is momentarily interrupted, at unexpected points in time, by closing the external orifice of the mask, the aerodynamic conditions are changed in a fashion similar to that of Danish bdg. In fig. 5 are shown two stills of the glottis in these two conditions, and it is clear that the state of the glottis and the surrounding structures are very much alike. This primitive experiment (which has to be elaborated and supplemented by EMG recordings) and many small but probably stable differences observed in the glottal behaviour in consonant production may indicate, as far as I see it, a causal relation between the motor control of the laryngeal muscles and some sort of reflex mechanism, elicited by the conditions in other parts of the speech apparatus.

If the suggested interpretation of the PCA activity in Danish bdg is confirmed, the hypothesis ventured by F-L-R can be maintained in a slightly modified form. But in any case, we have to realize that motor behaviour of the speech organs is centrally as well as peripherally controlled with a complex interplay between the two control mechanisms, which ought to influence the interpretation of our experiments and our speech production models.

References

Fischer-Jørgensen, E. 1963: "Beobachtungen über den Zusammenhang zwischen Stimmhaftigkeit und interoralem Luftdruck", Zs.f.Ph. 16, p. 19-36

Fischer-Jørgensen, E. 1968: "Voicing, tenseness and aspiration in stop consonants", ARIPUC 3, p. 63-114

Fischer-Jørgensen, E. and H. Hirose 1974: "A preliminary electromyographic study of labial and laryngeal muscles in Danish consonant production", Haskins SR-39/40, p. 231-254

Frøkjær-Jensen, B. 1967: "A photo-electric glottograph", ARIPUC 2, p. 5-19

Frøkjær-Jensen, B. 1968: "Comparison between a Fabre glottograph and a photo-electric glottograph", ARIPUC 3, p. 9-16

Frøkjær-Jensen, B., C. Ludvigsen and J. Rischel 1971: "A glottographic study of some Danish consonants", F&S, Akademisk Forlag, p. 123-140 (ARIPUC 7, 1973, p. 269-295)

Hirose, H. 1975: "The posterior cricoarytenoid as a speech muscle", Ann. Bull. Univ. Tokyo, Logopedics-Phoniatrics 9, p. 47-66

Hutters, B. 1976: "Problems in the use of the photoelectric glottograph", ARIPUC 10, p. 275-312

Ondráčková, J. 1972: "Vocal-chord activity", Folia Ph. 24, p. 405-419

Pétursson, M. 1976: "Aspiration et activité glottale", Phonetica 33, p. 169-198

Sawashima, M. and S. Niimi 1974: "Laryngeal condition in articulation of Japanese voiceless consonants", Ann. Bull. Univ. Tokyo, Logopedics-Phoniatrics 8, p. 13-17

(ARIPUC 12, 1978)

MODEL EXPERIMENTS ON LIP, TONGUE AND LARYNX POSITIONS FOR PALATAL VOWELS

Sidney Wood[1]

Abstract: Languages contrasting [y] and [i] prefer prepalatal constrictions. The larynx is low for [y] and high for [i], the lips are less rounded for [y] than for [u] and the tongue blade tends to be raised for [y]. The consequences of these manoeuvres were studied in a series of model experiments. Their advantages are twofold: they provide the best plain-flat [i]-[y] spectral contrast and they ensure stable resonance conditions in the vocal tract.

Introduction

X-ray tracings collected from the literature (Wood 1975) revealed language-specific preferences for either prepalatal or midpalatal tongue body positions for palatal vowels (Wood 1977). Languages contrasting [y] or [ɨ] qualities with [i] prefer the prepalatal position. There should be little difference of F_2 between the two positions since this formant is least sensitive to constriction location perturbations throughout the prepalatal and midpalatal region (Stevens 1972). Three-parameter model nomograms (Stevens and House 1955, Fant 1960) show that when the mouth-opening is narrowed, this zone is shifted anteriorly away from the glottis. Rounding the lips for [y], unaccompanied by other articulatory modifications would yield a vowel with F_2 more sensitive to coarticulatory displacements of the palatal constriction. There is a common belief that the tongue is retracted for

---

1) Department of Linguistics, Phonetics Laboratory, Lund University, Sweden.

[y], compared with [i]. This is probably a misinterpretation of the lower F_2. Retraction for [y] would take the constriction even further from the zone where F_2 is least sensitive to location perturbations, and make the vowel even more sensitive to coarticulation. Published X-ray tracings do not support retraction for [y] (Wood 1975). On the contrary, the vocal tract is often narrowed anteriorly to the prepalatal constriction (especially Danish, private communication from Eli Fischer-Jørgensen and Nina Thorsen). This is achieved by raising the tongue blade. This narrowing should occur in the zone (now advanced by lip-rounding for [y]) where F_2 is least sensitive to location perturbations. The effect of this should be to make the vowel less sensitive to lingual coarticulation. The effect of tongue body position on vowel spectra was studied in a series of model experiments by systematically altering a prepalatal configuration to a velar configuration in steps. The consequences of tongue blade raising will not be described here but will be deferred to a forthcoming report.

The larynx is lower for rounded vowels and higher for spread-lip vowels. X-ray tracings indicate an average overall range of about 15 mm for vertical larynx movement. Trained singers who have learned to control their larynxes may utilize a larger range (Sundberg and Nordström 1976). There are suggestions that larynx lowering compensates for labial undershooting (Riordan 1977), whereas X-ray films indicate that the larynx is lower the more the lips are rounded (Wood 1975, 1977), and that the lower lip alone very successfully compensates for disturbances to lip-rounding (Wood, forthcoming). I offer the following hypothesis: increased lip-rounding shifts further from the glottis the zone where F_2 is least sensitive to location perturbations; simultaneous larynx-lowering should restore that region to the vicinity of the hard palate so that the spectral consequences of tongue body manoeuvres and coarticulatory location perturbations will remain similar for all palatal vowels irrespective of rounding. To investigate this, the model experiments were repeated with both high and low larynx position.

Whenever differences of lip-rounding have been reported for [y] and [u], they have always indicated less close rounding for [y] (Lyttkens and Wulff 1885, Hadding et al. 1976, McAllister et al. 1974, Bannert et al. 1976 for Swedish; Benguerel and Cohen 1974, Riordan 1976 for French). The same can be seen on X-ray tracings of Danish vowels (private communication, as above). The closer the lips are rounded for [y] the greater is the shift away from the glottis of the zone where F_2 is least sensitive to location perturbations. I hypothesize that beyond a certain degree of rounding for [y] this zone is so far advanced that the palatal tongue gesture no longer produces its desired effect. The model experiments were repeated with four different lip conditions - spread, neutral, moderately rounded and closely rounded.

To test the generality of the results, key parts of the investigation were repeated with both longer and shorter vocal tracts. This will not be reviewed here. The check did confirm that the results would apply to all vocal tract sizes.

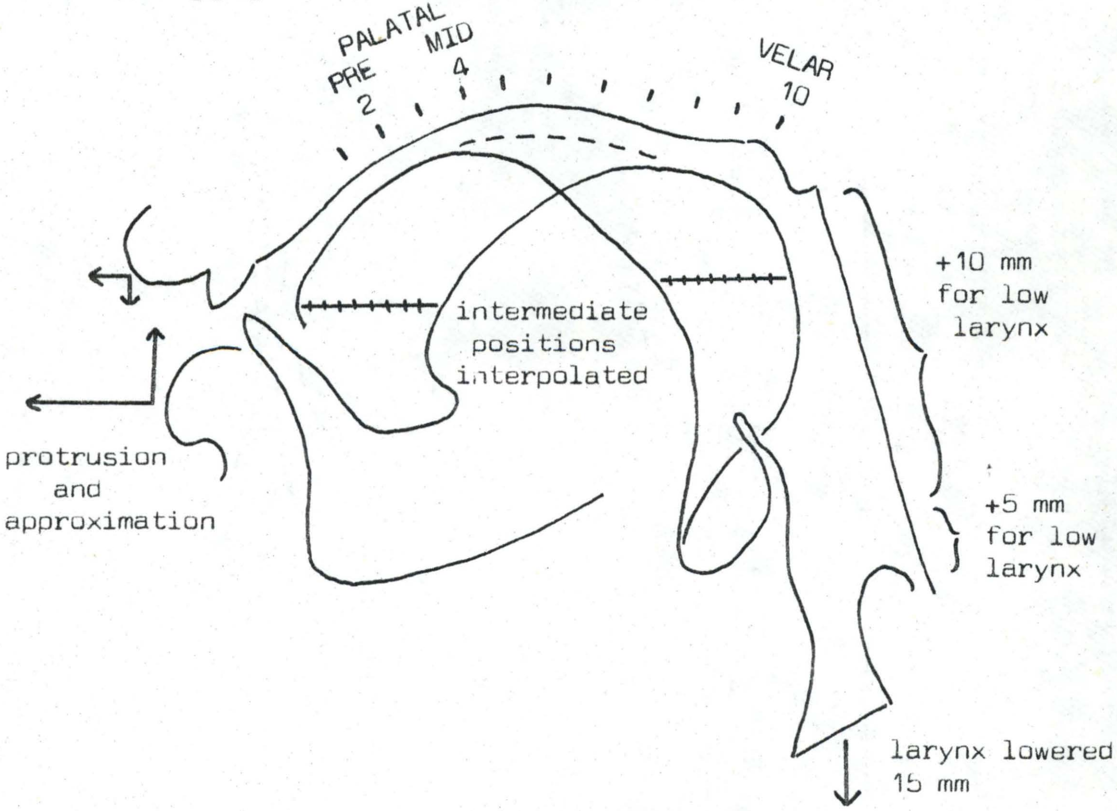


Figure 1

The modifications made to the model vocal tract: 10 tongue body positions from pre-palatal to velar, 4 lip positions from spread to closely rounded, high and low larynx positions.

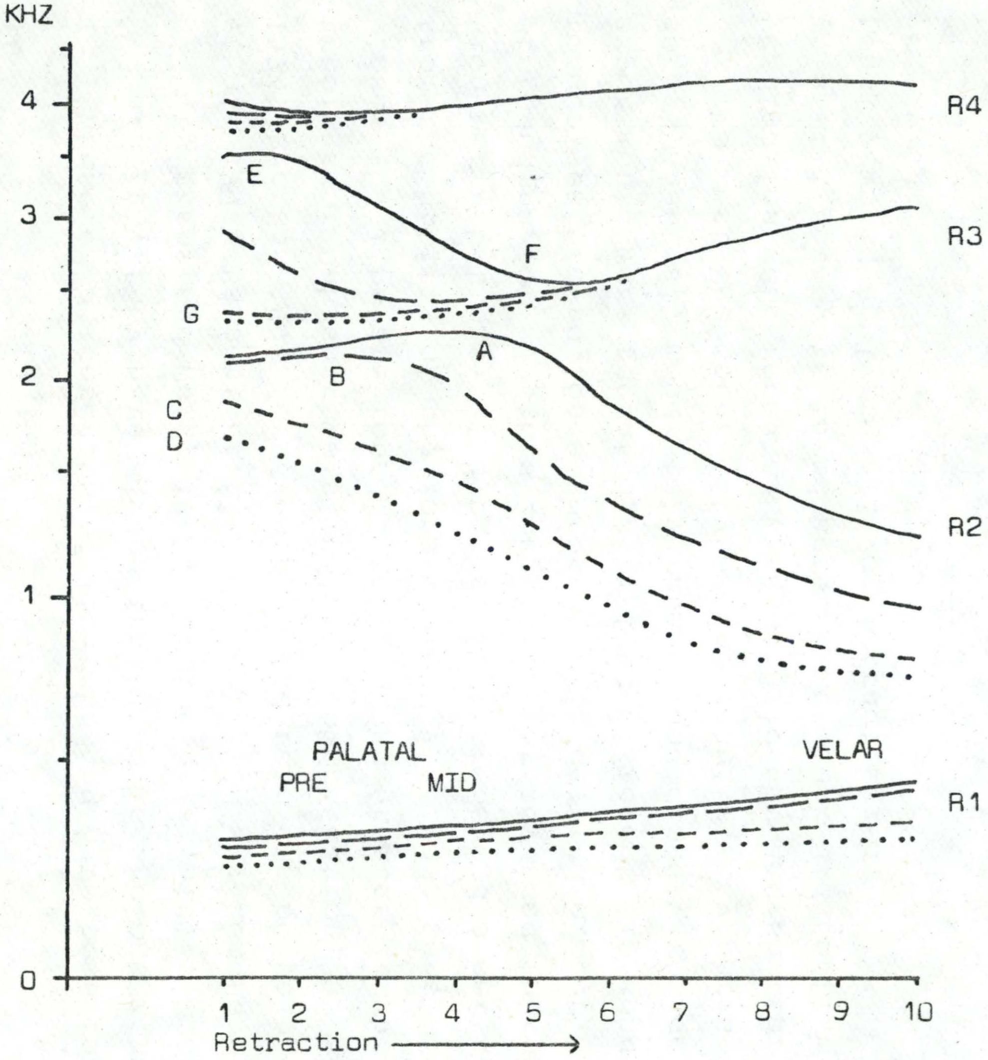


Figure 2

Consequences of four different lip positions at 10 tongue positions (high larynx).

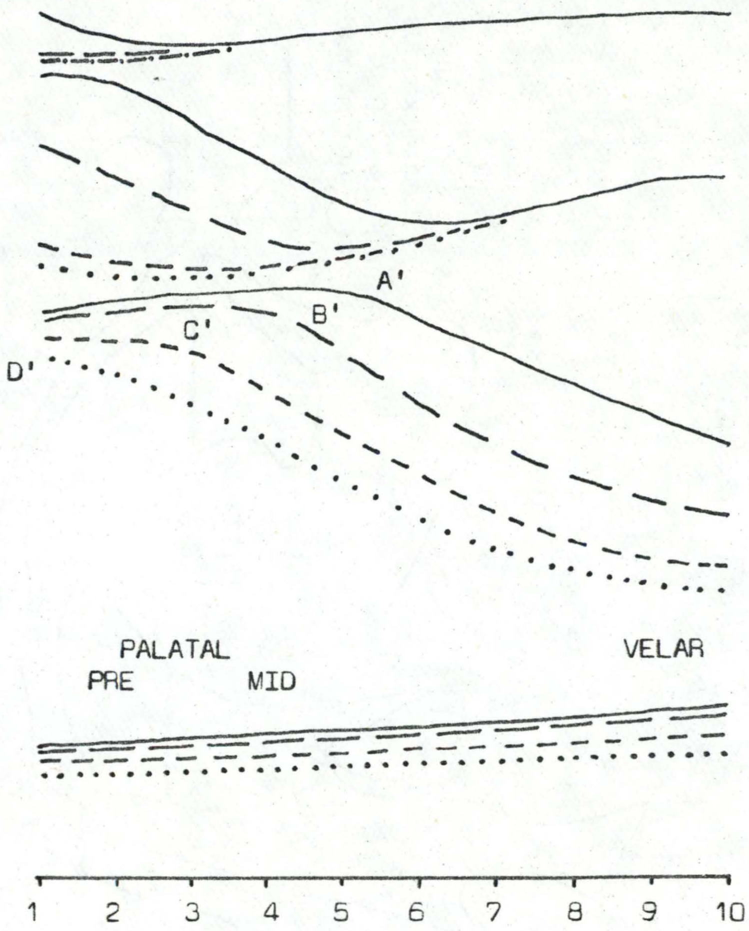


Figure 3

Consequences of four different lip positions at 10 tongue positions (low larynx).

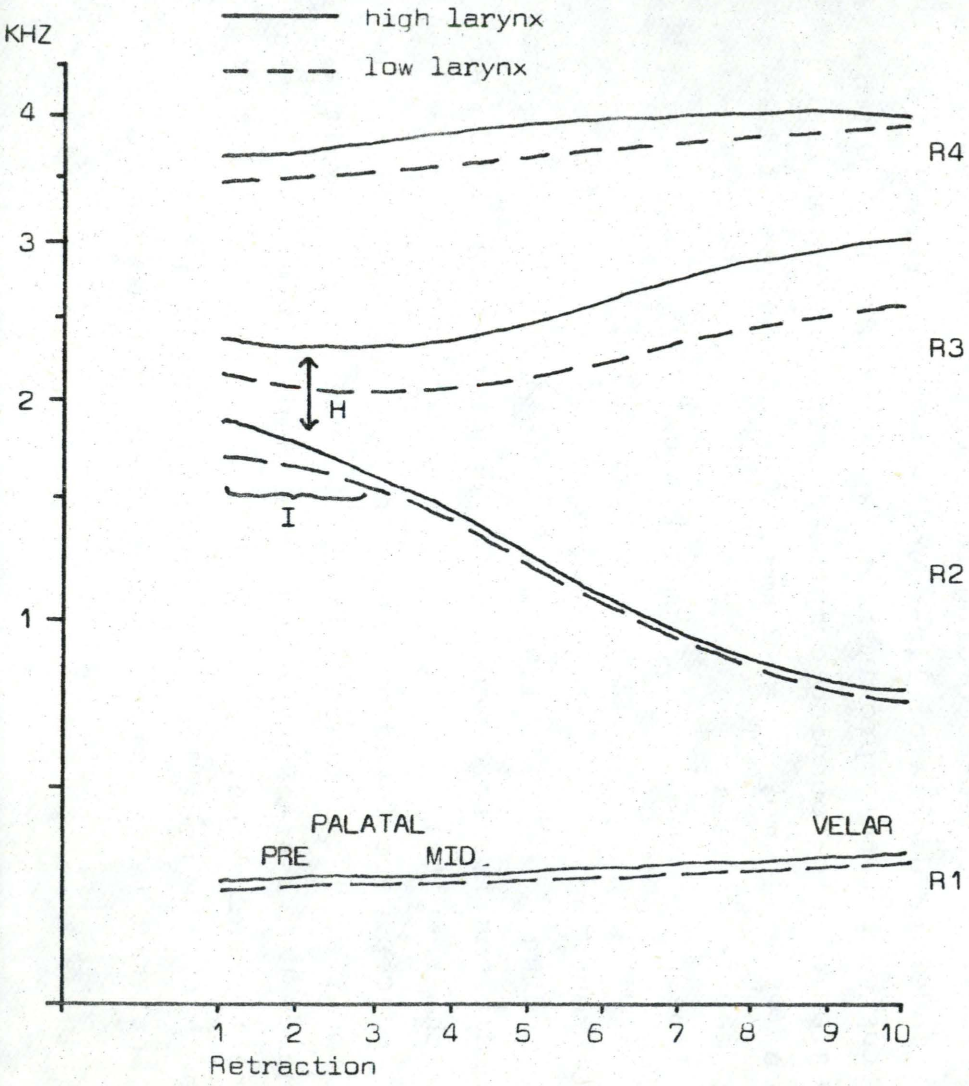


Figure 4

Consequences of larynx-lowering at 10 tongue positions (moderate rounding).

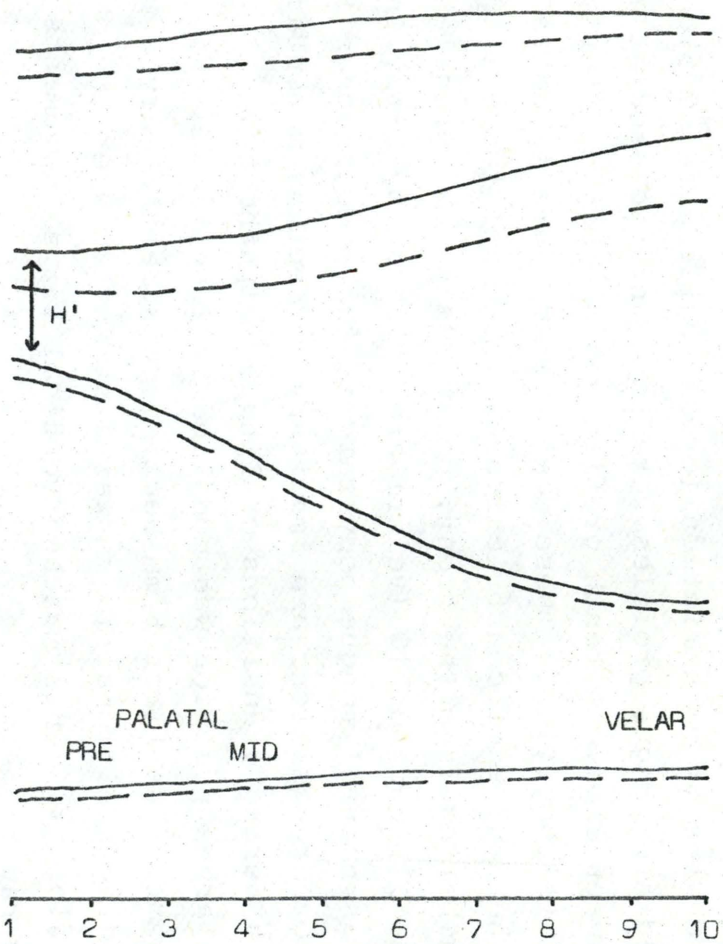


Figure 5

Consequences of larynx-lowering at 10 tongue positions (close rounding).

Procedure

Two X-rayed configurations (prepalatal [i] and velar [u]) were adjusted so that they both had the same high larynx position and the same degree of constriction (cross-section area 0.5 cm^2). Seven intermediate tongue profiles were then interpolated at even distances between them. A tenth profile was projected anteriorly to the prepalatal position. These modifications are illustrated in fig. 1. The degree of constriction was constant for all 10 profiles. Numbering from the front, positions no. 2 (prepalatal), no. 4 (midpalatal) and no. 10 (velar) are attested by X-ray investigations as natural tongue body targets for monophthongs in speech. The other positions are imaginary, except as momentary passing points during transitions or lingual diphthongs.

The low larynx position was modelled by lengthening the pharynx by 15 mm. The first 5 mm section outside the larynx was duplicated and an additional 10 mm was distributed over the pharynx above the epiglottis. This matched vertical larynx movements observed on X-ray films.

The lips were modelled according to Lindblom and Sundberg (1971). The moderate degree of lip-rounding was arbitrarily set at full protrusion but no approximation (w_m = -10 mm, h_m = 0 mm). The close degree of lip-rounding comprised both protrusion and approximation (w_m = -10 mm, h_m = -2 mm). The dimensions of the modelled lip sections were:

spread lips	0.5 cm x 2.5 cm^2
neutral lips	0.5 cm x 0.95 cm^2
moderate rounding	1.5 cm x 0.66 cm^2
close rounding	1.5 cm x 0.33 cm^2

Resonance conditions for intermediate lip conditions can be interpolated from the results.

The resonances were found by computing the pressure distribution in the modelled vocal tract for increasingly higher excitation frequencies until standing waves were found. The only losses included were for radiation.

The length of the modelled vocal tract from the glottis to the central incisors (high larynx) was 14.5 cm. Three X-rayed male subjects had this length for [i]. Two others had 15 cm. The corresponding measure for Fant's Russian subject was 16 cm.

Results

Fig. 2 shows how the frequencies of the first four resonances changed when the tongue body position was retracted in 10 steps from pre-palatal to velar, for four different lip positions and high larynx. Fig. 3 shows the same for the low larynx. Fig. 4 shows the consequences of larynx-lowering at the 10 tongue body positions with moderate lip-rounding. Fig. 5 shows the same for close lip-rounding.

Discussion

For spread lips and high larynx, F3 was high with the prepalatal tongue position (E on fig. 2) and low with the midpalatal position (F on fig. 2). This is why the prepalatal [i] sounds sharper than the midpalatal [i]. It is well known that for the plain-flat spectral contrast all the formants are lowered for [y] relative to [i] and the spectrum is "flattened" by bringing F3 close to F2. Fig. 2 (or fig. 3) shows that lip-rounding lowered F1 and F2 generally, F3 only for anterior tongue body positions and F4 hardly at all. Fig. 4 (or fig. 5) shows that laryngeal depression lowered F4 and F3 generally, F2 mainly for anterior tongue body positions and F1 slightly. These consequences of lip-rounding and larynx-lowering are predictable from acoustic theory (Fant 1960, pp. 63 and 64). For lowering the individual formant frequencies of palatal vowels, lip-rounding and laryngeal depression are complementary rather than interchangeable, except with the prepalatal configuration where the two manoeuvres combine to lower F2 and F3 (although the 200-300 Hz laryngeal contribution to F3 was small compared with the labial contribution). With the prepalatal configuration, lip-rounding lowered F3 much more than F2 (about 1000 Hz and 300 Hz respectively). It is this very large lip contribution to F3 of the prepalatal [y] that brings F3 close to F2 to flatten the spectrum. With the midpalatal configuration the situation was reversed: lip-rounding lowered F2 more than F3 and consequently the difference between F3 and F2 actually increased. Lip-rounding does not flatten the spectrum with a midpalatal configuration. Only the prepalatal position is favourable for spectral flattening.

The effect of degree of lip-rounding, tongue body position and larynx-lowering on the plain-flat spectral contrast can be studied in figs. 4 and 5. With moderate lip-rounding (fig. 4) and high larynx, retraction from the prepalatal position caused F2 and F3 to diverge (H) with consequent loss of spectral flattening. With the larynx lowered, the F2 and F3 curves were virtually parallel in the prepalatal region (I) so that coarticulatory location pertubations would have no appreciable effect on spectral flattening there. With close lip-rounding (fig. 5) retraction caused F2 and F3 to diverge (H') whether the larynx was high or low. Thus, moderate lip-rounding, with the prepalatal configuration and low larynx, is the more favourable combination for preserving the flattened spectrum of [y] (and hence the plain-flat contrast) against coarticulatory location perturbations.

The results illustrate Stevens's finding that F2 of [i] is hardly sensitive to location perturbations. For spread lips and high larynx, F2 varied about 5 Hz per mm of constriction displacement within the prepalatal and midpalatal region (A on fig. 2). At tongue body positions further back than mid-palatal, F2 was very sensitive to location perturbations and fell about 50 Hz per mm of retraction. F_2 of [i]-like vowels is very similar at both the prepalatal and midpalatal positions. With neutral lips and high larynx the least sensitive zone of F2 is more anterior (B on fig. 2) as anticipated. For the two rounded conditions the least sensitive zone is considerably advanced (C and D on fig. 2).

With the larynx lowered 15 mm the least sensitive zone of F2 was less advanced and was located in the palatal region (A', B', C' on fig. 3), as hypothesized. For moderate lip-rounding, the least sensitive region of F2 coincided with the prepalatal constriction (C' on fig. 3), while F2 was still very sensitive to location perturbations at the midpalatal position. In this respect, the prepalatal tongue body position, with moderate lip-rounding and lowered larynx, is more favourable for [y]. The midpalatal position is less favourable, even with the lowered larynx.

With close lip-rounding, the least sensitive zone for F2 is very much advanced even with the larynx lowered (D' on fig. 3). In this respect, close lip-rounding is less favourable for [y] at either tongue body position and with any larynx height. Other experiments (report forthcoming) show that raising the tongue blade will in this case make F2 less sensitive to location perturbations.

Conclusions

The prepalatal position is more favourable for the plain-flat contrast. F3 is high in [i] and low in [y], providing the greatest contrast. F_3 is lowered more than F_2 by lip-rounding so that both formants come closer together in [y], ensuring spectral flattening. The midpalatal position is less favourable since F3 is already low for [i], providing a smaller [i]-[y] contrast. Further, lip-rounding lowers F_3 less than F_2 so that the [y] spectrum would not be flattened.

With more than moderate lip-rounding, the spectral flattening of [y] becomes increasingly sensitive to lingual coarticulation, F2 and F3 responding differently to location perturbations. With moderate lip-rounding and the larynx lowered, F2 and F3 respond similarly to location perturbations throughout the prepalatal region so that the spectral flattening of [y] is safeguarded against coarticulatory location perturbations.

The laryngeal contribution to formant frequency lowering for prepalatal [y] is small compared with the labial contribution. I conclude therefore that the main function of larynx-lowering for [y] is to stabilize resonance conditions with the consequence that lingual gestures have a similar effect on the formants of both spread-lip and rounded palatal vowels. This requires the larynx to be lower as lip-rounding increases, a correlation that can be observed on X-ray films. This stabilizing effect can be backed up by raising the tongue blade.

References

Bannert, R., E. Gårding and S. Wood 1976: "Vokaler och vokalsystem", in: Gårding, E. (ed.) *Kontrastiv fonetik och syntax med svenska i centrum*, p. 27-60

Benguerel, A.-P. and H.A. Cohen 1974: "Coarticulation of upper lip protrusion in French", *Phonetica* 30, p. 41-55

Fant, G. 1960: *The acoustic theory of speech production*. The Hague

Hadding, K., H. Hirose and K.S. Harris 1976: "Facial muscle activity in the production of Swedish vowels: an electromyographic study", *J.Ph*. 4, p. 233-245

Lindblom, B. and J. Sundberg 1971: "Acoustical consequences of lip, tongue, jaw and larynx movements", *JASA* 50, p. 1166-1179

Lyttkens, I. and F. Wulff 1885: *Svenska språkets ljudlära, beteckningslära och aksent*. Lund

McAllister, R., J. Lubker and J. Carlson 1974: "An EMG study of some characteristics of the Swedish rounded vowels", *J.Ph*. 2, p. 267-278

Riordan, C.J. 1976: "Electromyographic correlates of the phonological /y/-/u/ distinction in French", *J.Ph*. 4, p. 1-16

Riordan, C.J. 1977: "Control of vocal tract length in speech", *JASA* 62, p. 998-1002

Stevens, K.N. 1972: "The quantal nature of speech: evidence from articulatory-acoustic data", in: David, E.E. and P.B. Denes (eds.) *Human communication, a unified view*, p. 51-66

Stevens, K.N. and A.S. House 1955: "Development of a quantitative description of vowel articulation", *JASA* 27, p. 484-495

Sundberg, J. and P.-E. Nordström 1976 "Raised and lowered larynx - the effect on vowel formant frequencies", *STL-QPSR* 2-3/1976, p. 35-39

Wood, S. 1975: "The weaknesses of the tongue-arching model of vowel articulation", *Working Papers* 11, p. 55-108, Department of Linguistics, Lund University

Wood, S. 1977: "A radiographic analysis of constriction locations for vowels", *Working Papers* 15, p. 101-131, Department of Linguistics, Lund University

A SIMPLIFIED EXPLANATION, IN PHYSICAL TERMS, OF THE ACOUSTICAL CONSEQUENCES OF TONGUE AND LIP MOVEMENT IN VOWEL PRODUCTION[1]

Nina Thorsen

Abstract: This paper presents an excerpt from a larger exposition, attempting to give non-technicians a basic insight in the relationship between articulation and acoustics of vowels and consonants.

By means of Newton's second law (force equals the product of mass and acceleration) and Boyle-Mariotte's law (at constant temperature the product of pressure and volume for a given quantity of air is constant) one can explain the fact that "When a part of a pipe is constricted its resonance frequency becomes low or high according as the constricted part is near the maximum point of the volume current ... or of the excess pressure ..." (Chiba & Kajiyama, 1958, p. 151). This is achieved mainly by considering the relative forces that act on a thin slice of air, oscillating back and forth through the open end of a quarter-wavelength resonator at its first resonance frequency: a decrease of the volume of the pipe near its closed end increases the forces that keep the slice in motion and thus raises its frequency, and vice versa. Inversely, diminishing the opening of the resonator decreases the forces that keep the slice in motion, and thus lowers its frequency, and vice versa.

1. Introduction

This paper does not pretend to be scientific and original in the ordinary sense of the words. I just try to explain to phoneticians without any special training in physics and mathematics, in a simpler fashion than do most of the articles and books on the subject, the often cited fact that "When a part of a pipe is constricted its resonance frequency becomes low or high according as the constricted part is near the maximum point of the volume current ... or of the excess pressure ..." (Chiba & Kajiyama, 1958, p. 151).

---

[1] Translation of a contribution for the 9èmes Journées d'Etude sur la Parole, Lannion 31 mai - 2 juin 1978.

2. Initial simplifications

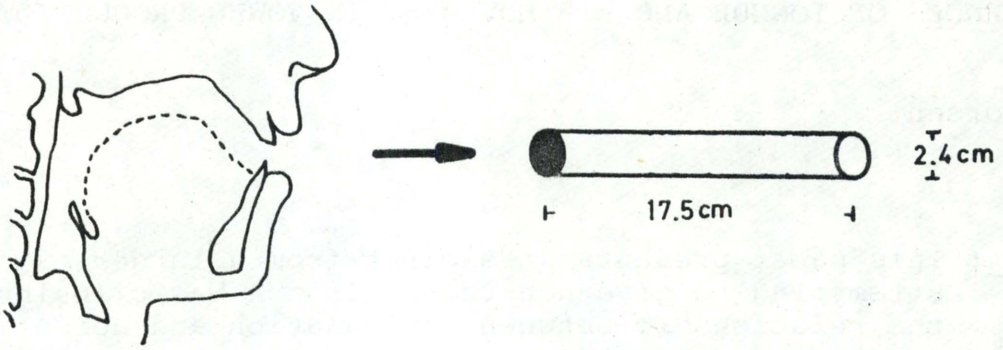


Figure 1
The vocal tract and a simplified model of the vocal tract.

Suppose the vocal tract is a cylindrical pipe, 17.5cm long and with a diameter of 2.4cm, closed at one end, open at the other, see fig. 1. Let us say, further, that the walls of the pipe are perfectly hard and non-yielding (i.e. they do not absorb acoustic energy) and that there is no radiation of energy from the open end of the pipe to the exterior (i.e. there is no diffraction of sound from the lips. This is, of course, a monstrous absurdity, and in practice it would mean that we could not hear each other speak, but it is a convenient simplification and one which is not a serious obstacle to the qualitative considerations that follow.) Thus we are dealing with an ideal uniform quarter-wavelength resonator, with resonances at approximately 500, 1500, 2500, ... Hz (cf. p. 9).

3. The vibratory pattern in the uniform quarter-wavelength pipe at its first resonance frequency

As point of departure, let us consider the uniform pipe of fig. 1 (i.e. the neutral vowel). Let us look at the column of air, after it has been made to oscillate at its first resonance, and let us suppose that this oscillation continues with undiminished amplitude so long as we are interested in studying it. (In practice this is impossible without a constant supply of energy; this is of no consequence for the present treatment.)

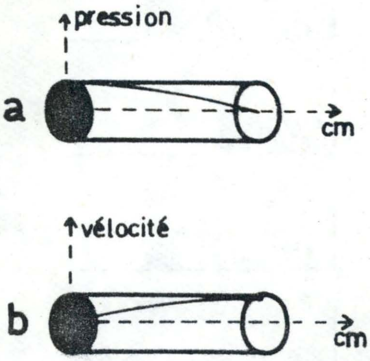


Figure 2 The distribution of pressure variation (a) and volume velocity (b) in a uniform quarter-wavelength pipe at its first resonance.

We know that at the open end of the pipe (the lips) the variation of volume velocity[1] is maximum, i.e. the air particles perform oscillations in and out through the opening with maximum elongation. At the closed end (glottis) the pressure variation[2] is maximum. We also know that there is pressure variation and volume velocity, respectively, all along the pipe but that the pressure variation decreases from closed to open end (where it is zero) and that the velocity decreases from open to closed end (where it is zero), see fig. 2. (These facts can also be illucidated in an intuitively comprehensible fashion, but not without exceeding the limits of this paper.)

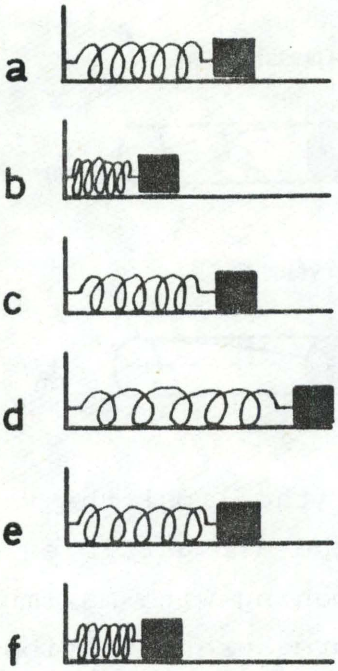
However, as long as we are dealing with only the first resonance, we can conceive of the column of air as if its movement, i.e. the volume velocity, were concentrated at the open end and as if the pressure variation were concentrated at the closed end of the pipe. (Thus we are dealing with a system of concentrated constants, with one degree of liberty, i.e. it can oscillate at one, and only one, frequency.) In this case the acoustic system can be likened to a mechanical system, composed of a mass and a spring, attached to a hard wall, sliding on a perfectly smooth surface, which means that no friction occurs between the surface and the mass, when it oscillates, see fig. 3.

All movement presupposes a force: if the mass is displaced to the left (3b) the compression of the spring exerts a force to the right, and when we let go of the mass this force will set the

---

1) "variation of volume velocity" is occasionally abbreviated "volume velocity", or just "velocity" in the following.

2) "pressure variation" is occasionnally abbreviated "pressure" in the following.



Vibratory pattern of a spring and mass, attached to a hard wall, sliding on a smooth surface.

Figure 3

mass in motion, towards its rest position. The mass will pass the rest position (3c), because every body that has a mass has inertia as well, which means that a movement will continue some time after the force which initiated it has ceased to operate. Thus, the spring becomes more and more stretched and exerts a growing force to the left which will eventually stop the motion of the mass (3d) and a movement to the left begins, towards the rest position. Because of its inertia, the mass will once more pass its position of equilibrium (3e), the spring is compressed anew and exerts a growing force to the right until the mass is stopped (3f) and a movement to the right commences, and so on and so forth. If no energy is lost anywhere, the mass will oscillate eternally with undiminished elongation. Its frequency depends on the tension of the spring and the size of the mass: the greater the tension, and the smaller the mass, the higher its frequency of oscillation, and vice versa. The elongation of the mass depends only on the initial displacement which sets the system in motion.

In the same fashion we can consider the behaviour of a thin slice of air, S, at the open end of the pipe, see fig. 4 (the movements and thickness of this slice are greatly exaggerated in the figures). What keeps this slice of air in motion is the combined action of (1) the pressure variations that arise in the

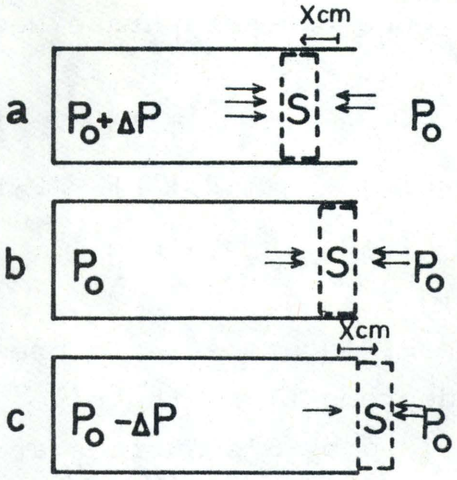


Figure 4 Vibratory pattern of a thin slice of air at the open end of the uniform quarter-wavelength pipe.

pipe due to the motion of S and (2) S's inertia. When S commences a movement to the right (4a) it is because the pressure in the pipe is greater than the atmospheric pressure outside the pipe, and when S passes its rest position (4b) it is because S has a certain (however small) mass and therefore inertia. Thereby the pressure within the pipe decreases, and the atmospheric pressure constitutes a (relative to the pressure in the pipe) growing force to the left, which eventually stops S and sets it in motion back towards the position of equilibrium (4c). This oscillation, too, will continue eternally, with undiminished amplitude, if no energy is lost anywhere from the system.

There is a close tie between the forces that act on S, S's mass, and S's motion, which is given by Newton's second law:

[1.1] $F = m \cdot G$ (force equals the product of mass and acceleration)

a) the force, in our case, is the product of pressure and the area of the surface to which the pressure applies:

[1.2] $F = P \cdot A$

This area is constant (see fig. 4). Only the pressure in the pipe varies.

b) S is simultaneously influenced by two antagonistic forces, one due to the pressure in the pipe and one due to the atmospheric pressure outside. The resultant force is due to the difference between these two pressures.

c) S's mass is constant.

Thus, one can paraphrase Newton's second law:

[1.3] $P \cdot A = m \cdot G$ i.e.

[1.4] $((P_0+\Delta P) - P_0) \cdot A = m \cdot G$ thus $\Delta P = \frac{m \cdot G}{A}$ or

[1.5] $\Delta P = k \cdot G$

where P_0 is atmospheric pressure, ΔP is the pressure increment (or decrement) in the pipe, k is a constant, equal to the mass of S divided by S's surface area, and G is S's acceleration, which can be taken as an indication of S's mean velocity. Thus S's velocity varies according to the difference in pressure within and outside the pipe. This difference is positive and negative, intermittently, and S thus moves from left to right and back again through the opening of the pipe.

4. Non-uniform pipes with constant opening

It can be shown that

a) S's elongation depends only on the magnitude of the initial force.

b) S's elongation and frequency are independent of each other.

c) S's frequency depends only on the relative changes of pressure that are induced in the pipe due to S's motion, which, in their turn, are determined by the total volume of the uniform pipe.

All this is a consequence of Newton's second law and of another law which states that, at constant temperature, the product of pressure and volume for a given quantity of air is constant (Boyle-Mariotte's law):

[2.1] $P \cdot V = k$

In our case it means that when the volume of the column of air is increased by S's movement out of the pipe, the pressure in the pipe decreases, and vice versa. The most important fact to note is that as long as the volume increments and decrements are small compared to the total volume of the pipe, the pressure and volume variations are proportional to one another:

Let P_0 and V_0 be pressure and volume, respectively, in the rest condition:

$$[2.2] \qquad P_0 \cdot V_0 = k$$

We decrease the volume by ΔV and get a pressure increase of ΔP_1:

$$[2.3] \qquad (P_0+\Delta P_1)(V_0-\Delta V) = k \quad \text{i.e.}$$

$$[2.4] \qquad \Delta P_1(V_0-\Delta V) = k - P_0(V_0-\Delta V) = P_0 \cdot V_0 - P_0(V_0-\Delta V) = P_0 \cdot \Delta V$$

We decrease the volume by $2\Delta V$ and get a pressure increase of ΔP_2:

$$[2.5] \qquad (P_0+\Delta P_2)(V_0-2\Delta V) = k \quad \text{i.e.}$$

$$[2.6] \qquad \Delta P_2(V_0-2\Delta V) = k - P_0(V_0-2\Delta V) = P_0 \cdot V_0 - P_0(V_0-2\Delta V) = 2 \cdot P_0 \cdot \Delta V$$

$$\text{thus } [2.7] \qquad \Delta P_2 = \frac{2 \cdot P_0 \cdot \Delta V}{(V_0-2\Delta V)} = \frac{2 \cdot \Delta P_1(V_0-\Delta V)}{(V_0-2\Delta V)} \simeq 2 \cdot \Delta P_1 \quad \text{if } \Delta V <<< V_0$$

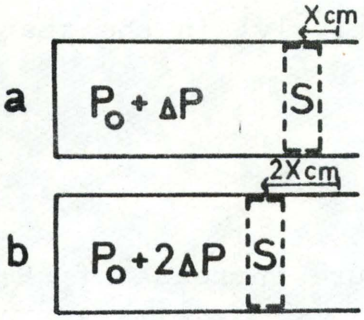
It follows that if a volume decrement of $\Delta V cm^3$ causes a pressure increment of $\Delta P \mu bar$, a volume decrement of $2\Delta V cm^3$ will render a pressure increment of $2\Delta P \mu bar$. (In practice the volume changes are very small indeed, since the elongation of the air particles is of an order of magnitude of a few millionths of a millimeter.)

Re (a) and (b) (p. 6)

Let us say that the initial force which sets S going is a displacement to the left by Xcm (see fig. 5a). This produces a pressure of $(P_0+\Delta P)$bar in the pipe. When we let go of S it starts moving to the right, and we know of its acceleration (and thus its mean velocity) that it is porportional to the difference between the pressures within and outside of the pipe.

$$[3.1] \qquad (P_0+\Delta P) - P_0 = k \cdot G_1 \qquad \text{i.e.} \quad G_1 = \frac{\Delta P}{k}$$

If, instead, we commence by giving S a displacement of 2Xcm to the left (see fig. 5b), the pressure within the pipe will be $(P_0+2\Delta P)$bar. We get an acceleration as follows:



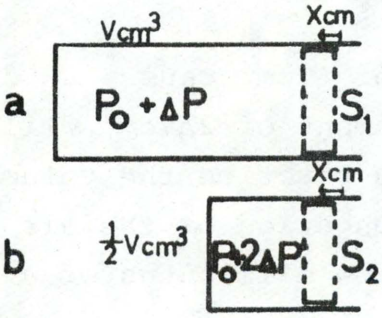
The relationship between initial elongation of S and pressure increment in the uniform pipe.

Figure 5

[3.2] $$G_2 = \frac{(P_0+2\Delta P) - P_0}{k} = \frac{2\Delta P}{k}$$

thus [3.3] $$G_2 = 2G_1$$

The mean velocity of S in the second case is twice that of S in the first case, but the elongation is also twice that of the first case, and thus the frequency of oscillation is identical in the two cases, and is independent of the elongation.



The relationship between pressure increment and elongation of S in two uniform pipes with volumes Vcm^3 (a) and $\frac{1}{2}Vcm^3$ (b).

Figure 6

Re (b) and (c)

If one displaces S_1 and S_2 of fig. 6 by Xcm to the left, in pipes having volumes of Vcm^3 and $\frac{1}{2}Vcm^3$, the relative volume decrement in the lower pipe is twice that of the upper pipe, and thus the pressure increment in the lower pipe is twice that of the upper pipe. The force which acts on S_2 is thus twice as large as the one that acts on S_1. The mean velocity of S_2 is therefore twice that of S_1, and since their elongations are identical, the frequency of oscillation of S_2 must be twice as high as that of S_1. (This is in complete accord with what we obtain from the formula for resonance frequencies in uniform quarter-wavelength pipes:

$$f_n = \frac{c}{4 \cdot L}(2n-1)$$

where c is the speed of sound in air, L is the length of the pipe, n is the number of the resonance. If c=35000cm/sec, L_1=17.5cm (a), and L_2=8.75cm (b) we get:

$$\text{(a)}: f_1 = \frac{35000}{70} = 500\text{Hz} \qquad \text{(b)}: f_1 = \frac{35000}{35} = 1000\text{Hz} \text{)}$$

The model of oscillation described above is extremely simplified, because velocity and pressure are not concentrated at the open and closed ends, respectively, of the pipe, see fig. 2. In practice this means that a volume change will have the greatest influence on the first resonance if it is located near the closed end of the pipe, where pressure variation is at its maximum.

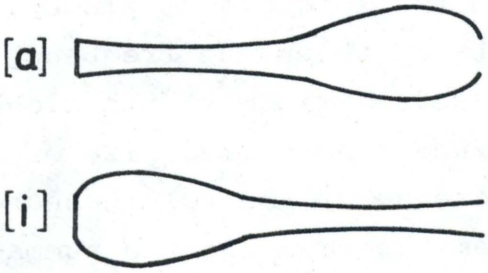
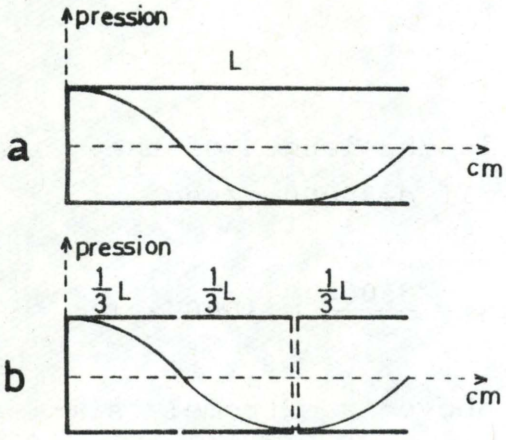


Figure 7 Models of two vowels [ɑ] and [i].

We may conclude that the frequency of the first resonance of the pipe in fig. 7 above, which is a model of the vowel [ɑ], must be higher than that of the uniform pipe ([ə]) and that, inversely, the first resonance of the pipe in fig. 7 below, which is a model of the vowel [i], is lower than that of [ə], which is confirmed be empirical facts. (See also the summary.)

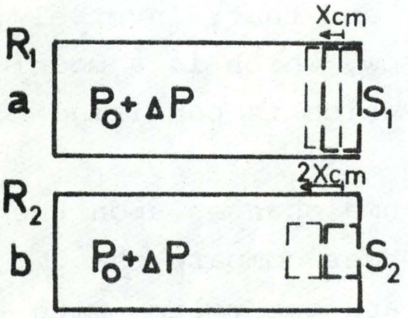
If we wish to consider the effect of volume changes upon the second, third, etc., resonances, we can no longer compare the system with a single slice of air (one mass) and one volume with pressure variation (one spring). The pressure distribution along the pipe at its second resonance frequency is depicted in fig. 8. If the column of air oscillates only at its second resonance the system behaves exactly as if it were combined of three pipes, each 1/3 Lcm long. The two imaginary pipes to the left in fig.



The pressure distribution in a uniform pipe at its second resonance frequency. Figure 8

8b are joined with the open ends against each other, and the pipe to the right is joined, closed ends together, to the one in the middle. Now we can reason about the system in the same way as for the first resonance, only there are two places where a change of volume will have an appreciable effect on the (second) resonance, namely at the closed end and at a distance of 2/3 Lcm from the closed end. At the third resonance there will be three places, at the fourth four places, etc., where volume changes will affect the resonance frequency appreciably. Each resonance has a pressure maximum at the closed end of the pipe, and thus all resonance frequencies rise or lower as a consequence of a decrease or increase of the volume near the closed end (but not always to the same degree, see the summary).

5. Uniform pipes with varying degrees of opening



The relationship between the elongation of two slices of air with areas of Acm^2 (a) and ½Acm^2 (b) and pressure increment in the uniform pipe. Figure 9

Let us now consider what happens if we decrease or increase the volume at the open end of the pipe. The volume change in itself is of no consequence, since there is no pressure variation at the open end of the pipe, but the vertical closening or opening of the pipe is essential, and it is that only which is depicted in fig. 9. We compare two pipes of identical diameters and lengths. The upper pipe, R_1, is fully open, i.e. the area of opening is $A\text{cm}^2$. The lower pipe, R_2, has a circular opening of $\frac{1}{2}A\text{cm}^2$. The two slices, S_1 and S_2, are equally thick, Bcm. We employ the laws of Newton and of Boyle-Mariotte: $F = m \cdot G$ and $P \cdot V = k$.

The masses of S_1 and S_2 are known if we know the volume of the slices and the density of the air, ρ:

[4.1] $m_1 = A \cdot B \cdot \rho$ [4.2] $m_2 = \frac{1}{2}A \cdot B \cdot \rho$

For a given pressure increment, ΔP, in R_1 and R_2, we get forces, F_1 and F_2, that act on S_1 and S_2 as follows:

[5.1] $F_1 = \Delta P \cdot A$ [5.2] $F_2 = \Delta P \cdot \frac{1}{2}A$

But force also equals the product of mass and acceleration, thus:

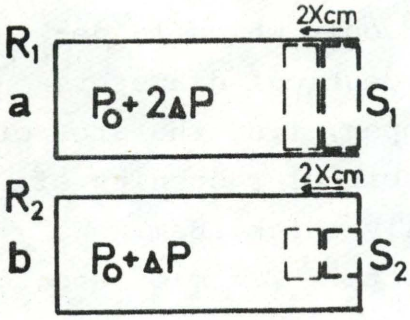
[6.1] $F_1 = \Delta P \cdot A = m_1 \cdot G_1$ i.e. $G_1 = \dfrac{\Delta P \cdot A}{m_1} = \dfrac{\Delta P \cdot A}{A \cdot B \cdot \rho} = \dfrac{\Delta P}{B \cdot \rho}$

[6.2] $F_2 = \Delta P \cdot \frac{1}{2}A = m_2 \cdot G_2$ i.e. $G_2 = \dfrac{\Delta P \cdot \frac{1}{2}A}{m_2} = \dfrac{\Delta P \cdot \frac{1}{2}A}{\frac{1}{2}A \cdot B \cdot \rho} = \dfrac{\Delta P}{B \cdot \rho}$

thus [6.3] $G_1 = G_2$

The two slices of air will have the same acceleration (mean velocity). BUT they do not have the same elongation. In order to induce in R_1 and R_2 the same volume decrement, and thus the same pressure increment, S_2 will have to be displaced twice as far into the pipe as S_1, because the surface area and volume of S_2 are only half those of S_1. If the two slices have the same mean velocity, but the distance covered by S_2 is twice that covered by S_1, S_2's period will be twice that of S_1, and consequently the frequency of oscillation of S_2 will be half that of S_1.

If, instead, we commence by giving S_1 and S_2 the same displacement, as in fig. 10, we know that the volume decrement in R_1 is twice that of R_2. The pressure increment in R_1 is thus twice that of R_2, e.g. $2\Delta P$ as against ΔP. These values are sub-



The relationship between initial elongation of S_1 and S_2 and pressure increment in the uniform pipe. Figure 10

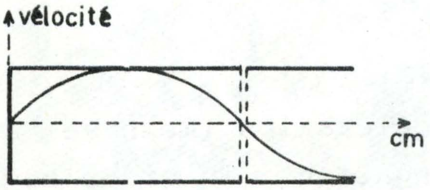
stituted in the expression for acceleration:

[7.1] $G_1 = \frac{2\Delta P}{B \cdot \rho}$ [7.2] $G_2 = \frac{\Delta P}{B \cdot \rho}$

i.e. [7.3] $G_1 = 2G_2$

The mean velocity of S_1 will be twice as large as that of S_2, and since the elongations are identical, the frequency of oscillation of S_1 will be twice as high as that of S_2.

Since the volume velocity is not concentrated at the open end, but is distributed all along the pipe (see fig. 2) we may conclude that an occlusion will have a greater effect on the first resonance frequency near the open end of the pipe, where velocity is at its maximum.



The velocity distribution in a uniform pipe at its second resonance. Figure 11

If we consider the second, third, etc., resonances we must again look at the velocity distribution all along the pipe. At the second resonance (see fig. 11) there are two places where the variation of volume velocity is maximum, namely 1/3 Lcm from the closed end, and at the open end. At the third resonance there will be three places, at the fourth four places, etc., where an occlusion will have an appreciable effect on the resonance. Each resonance has a velocity maximum at the open end and thus all resonances are lowered by an occlusion at the open end (but not always to the same degree, se the summary below).

6. Summary

A volume increase in the pipe produces a lowering of a resonance (and vice versa), the more so the nearer it is located to a pressure maximum for that resonance, and an occlusion produces a lowering of a resonance (and vice versa), the more so the nearer it is located to a velocity maximum for that resonance, and, all things being equal, the greater the change in volume or opening/closing, the greater the change in frequency. However, in practice we cannot separate these two types of changes within the vocal tract. Because of the limitations imposed by the articulatory organs, variations in the cross-sectional area within the vocal tract are simultaneously volume changes and occlusions/openings. Therefore the general formulation "When a part of a pipe is constricted its resonance frequency becomes low or high according as the constricted part is near the maximum point of the volume current or of the excess pressure."

It follows that if the constriction or expansion is situated exactly between a pressure and a velocity maximum, it will have no effect. Further: the vocal tract is an integrated system whose configuration is determined by the position of the tongue and lips. The tongue cannot simultaneously perform an extended constriction in the pharynx and at the hard palate, on the contrary, a pharyngeal constriction produces a relative expansion near the hard palate, and vice versa, see fig. 7. The cumulative effect is a "double" raising ([ɑ]) or lowering ([i]) of the first resonance frequency.

We have considered the acoustic consequences of changes in the cross-section of the vocal tract for each resonance separately. In practice a vowel is, of course, always composed of several resonances, that conjointly form one complex oscillation. This is of no consequence for our considerations - one can treat this complex oscillation as a superposition of sinusoidal oscillations and consider the effect of a change in the vocal tract for each component separately.

What is more important is the fact that as soon as one does not as point of departure take the uniform pipe, but a pipe already deformed (like the one in fig. 7 below, [i]) one cannot quantify the changes in frequency in as simple a fashion as for the uniform pipe. This is due to the fact that the distribution of pressure and velocity is no longer sinusoidal (as it is in figs. 2, 8, 11). One example will suffice: for the model of [i] the velocity at the second resonance is very nearly zero in the front third of the pipe near the opening, and an occlusion at the opening (rounding of the lips) will thus have very little effect on the second resonance. But since the velocity at the third resonance is maximum (greater, in fact, than for the uniform pipe) it will decrease radically due to an occlusion (rounding) at the opening. (For diagrams of the distribution of pressure and velocity for several vowels and resonance frequencies, see the works cited in the references.) This is why one can say, not wholly unjustified, that certain resonances are, in certain cases, more dependent on changes in one part of the vocal tract than in another, and this is true especially of the narrow vowels.

7. Conclusion

In real life, i.e. speech, the situation is far more complicated than this demonstration would lead one to believe. The walls of the vocal tract are not hard, and there is a considerable radiation of energy to the exteriour. Apart from the loss of energy, this radiation causes a tuning of the resonances, which is not of the same magnitude for high and low frequencies, and it is, further, heavily dependent on the degree of opening at the lips. The voice source, i.e. the pulse train from the glottis, constitutes another complicating element, among other things by the coupling it allows between the sub- and supraglottal cavities. Apart from all that, the mathematics and physics employed above do not suffice: in order to quantify the consequences of tongue and lip movement, one must solve higher order differential equations.

References

Chiba, T. & M. Kajiyama 1958:	The Vowel, its Nature and Structure, Tokyo.
Fant, G. 1970:	Acoustic Theory of Speech Production, 2nd ed., Paris.
Ungeheuer, G. 1962:	Elemente einer akustischen Theorie der Vokalartikulation, Berlin.

(ARIPUC 12, 1978)

ON THE IDENTIFICATION OF SELECTED DANISH INTONATION CONTOURS

Nina G. Thorsen

Abstract: 10 Copenhagen and 4 non-Copenhagen speakers identified 15 human utterances, differing only in their fundamental frequency course, as being either declarative, non-final, or interrogative (forced choice).
Responses are very clearly correlated with Fo: the most steeply falling intonation contours are identified as being declarative, the least falling (i.e. "flat") ones as being interrogative, and contours in the middle of the continuum as being non-final. Copenhagen speakers clearly perceive three categories, whereas non-Copenhagen speakers seemingly operate with only two. Several, mutually interdependent, parameters in the Fo course may account for the results, the two most powerful ones, however, being the levels of the last stressed and the succeeding unstressed syllable in the utterance. In a subsequent experiment, 7 Copenhagen speakers identified the same utterances as being either declarative or non-declarative. The majority of the (formerly) non-final sentences were now labelled non-declarative, rather than being split into partly declarative, partly non-declarative categories.
When a subset of the same utterances were mutilated, identification criteria changed, and identification deteriorated almost progressively with the number of syllables being cut away from the end of the utterance (but not seriously - until only the first stress group remained), whereas syllables cut away from the beginning hardly affected identification at all.

1. Introduction

In a previous volume of ARIPUC the results of a preliminary analysis of intonation in Advanced Standard Copenhagen Danish (ASC) were presented. They were summarized in a model for fundamental frequency in short sentences, cf. fig. 1. For a detailed account of the procedure that led to the formulation of this model, the reader is referred to Thorsen (1976). However, those features that are relevant to the present experiment will be

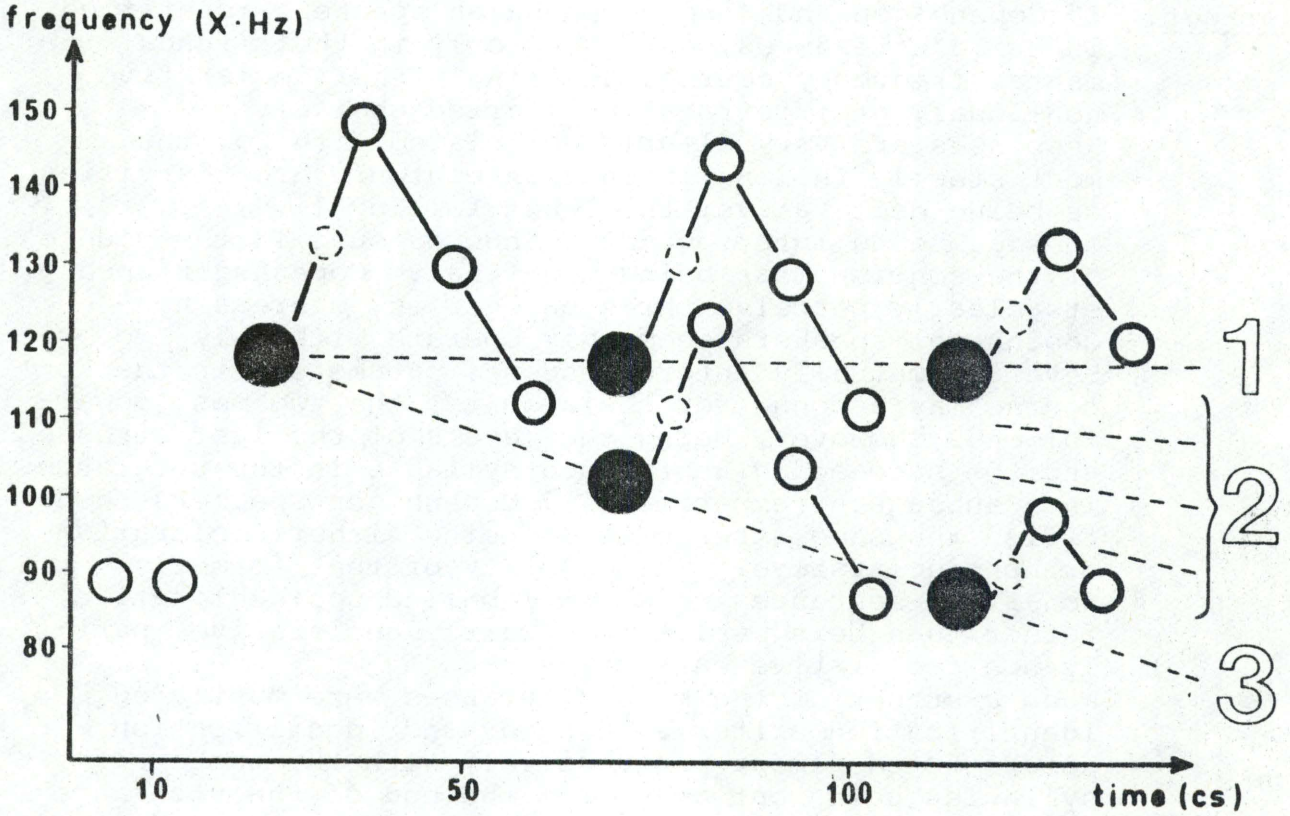


Figure 1

A model for fundamental frequency in short sentences in Advanced Standard Copenhagen Danish. 1: statement questions, 2: interrogative sentences with word order inversion and/or with interrogative particle, and non-final periods (variable), 3: declarative sentences.
The heavy circles represent stressed syllables, the empty circles unstressed syllables, and the broken circles represent syllables with assimilated or elided /ə/.
The full lines represent the fundamental frequency pattern associated with stress groups and the broken lines denote the intonation contour.

briefly outlined here: An underlying assumption is that the complex course of Fo in an utterance is the outcome of a superposition of several components.
(1) A sentence component which supplies the INTONATION CONTOUR.
(2) The contour is overlaid by a stress group component which furnishes the STRESS GROUP PATTERNS (both exemplified in fig. 1).
(3) To the resultant of those two components is added a stød component, rendering STØD MOVEMENTS. (However, as stød words had been excluded from the material, the model does not include this particular feature.) These first three components are language specific and thus "speaker controlled". (4) Finally, intrinsic fundamental frequencies of segments and coarticulatory variations at segment boundaries supply a MICROPROSODIC COMPONENT, which is not consciously controlled by the speaker, but due to inherent properties of the speech production apparatus. - The same point of view about "layers" in intonation has been expressed previously by several authors, see e.g. Bolinger (1970), Bruce (1977), Cohen and t' Hart (1967), Collier and t' Hart (1975), t'Hart (1966), t'Hart and Cohen (1973), and Lehiste and Peterson (1961), and for a more thorough account, see Thorsen (forthcoming).

As may be seen from the full lines of fig. 1, the stress group pattern can be described as a (relatively) low stressed syllable followed by a high-falling tail of unstressed syllables. This pattern is a predictable and recurrent entity, allowing, however, for contextual variations in the magnitude of the rise from stressed to unstressed syllable, which decreases from the beginning to the end of the utterance. In fact, it was this observation which led to the definition of the stress group in Danish as a stressed syllable plus all succeeding unstressed syllables (within the same non-compound sentence), irrespective of intervening word- or morpheme boundaries, and it also led to the definition of the intonation contour as the course described by the stressed syllables alone, cf. the dotted lines of fig. 1. The same concept of the intonation contour can be found in Bolinger's (1958, 1970) treatment of American English, and it is similar to the "declination" line from which the "hat" patterns set off in Dutch, cf. e.g. Collier and t'Hart (1975) and t'Hart and Cohen (1973).

Fig. 1 may be accounted for in a different manner, namely in terms of top-lines and base-lines which are tangents to the maxima and minima, respectively, in the Fo course. This would be in accordance with the way Bruce and Gårding (1978) describe Swedish, and, likewise, with the description by Breckenridge and Liberman (1978) of American English, except that in Swedish and American the top-line is tangent to the stressed syllables, and the base-line to the unstressed syllables. A feature common to all three languages is the fact that both lines decline (more or less steeply), but the top-line declines faster than the base-line, so that top- and base-line together create a wedge-shape. (Assuming that the stressed syllables determine the intonation contour, this would mean that the top-line carries the perceptual cue in Swedish and American, where it is the base-line in ASC Danish.) One objection that could be raised towards applying this description to Danish is that the stressed syllables do not always constitute the minima in the Fo course, - sometimes the "base-line" (i.e. the intonation contour) is transgressed by the tail of unstressed syllables, see e.g. contour "1" of fig. 1.

The intonation contour tends to vary systematically with sentence type, as suggested by fig. 1: declarative sentences having the most steeply falling contours (about 25%/sec), at one extreme, and statement questions (i.e. questions where only the intonation contour signals their interrogative function) having "flat" contours, at the other extreme. In between these two are found other types of questions as well as non-final periods. Further, there seems to be a certain trade-off between syntax[1] and intonation contour: the more syntactic information is contained in the sentence about its non-final or interrogative function, the more declarative-like, i.e. the more steeply falling, is its intonation contour, cf. Thorsen (1976, pp. 134-135). Such a trade-off has also been observed for other languages, see e.g. Bolinger (1964), Cohen and t'Hart (1967), Daneš (1960), von Essen (1956), Hadding-Koch (1961), and Mikoš (1976). - It was this observation that led to the formulation of the following three questions, which the experiments to be reported below were designed to answer:

---

[1] I do not wish, by this use of the terminology, to exclude the possibility of treating intonation as an integral part of the syntax of the language, and 'syntax' (and 'syntactic') should therefore just be regarded as a convenient abbreviation for "other signals, such as word order inversion, interrogative particles, and the like".

(1) will listeners identify three types of contours, in correlation with the actual course of Fo?
(2) if the answer to the first question is affirmative: are these categories linguistic, rather than purely phonetic?
(3) how early, or how late, in the sentence are the various contours perceptually differentiated?

2. Test 1

Test 1 attempts to answer the first question above, whether listeners will identify three categories of contours. The search for two or more (linguistic and/or attitudinal) categories of sentence intonation is anything but new, see e.g. Daneš (1960), Delattre et al. (1965), von Essen (1956), Gårding and Abramson (1965), Hadding-Koch and Studdert-Kennedy (1963, 1964, 1965, 1974), Isačenko (1965), Isačenko and Schädlich (1963), Johansson (1978), Mikoš (1976), Studdert-Kennedy and Hadding (1972, 1973), and Uldall (1960, 1961). The primary motivation for conducting yet another experiment on the identification of intonation contours is the fact that it has not been done with Danish material before. Furthermore, the material and procedure deviate somewhat from that of previous investigations.

2.1 Test material - test 1

A small part of the material recorded for the original analysis was selected for identification. I.e. natural speech was employed. The reason for this choice lies first and foremost in a curiosity to see how the contours, as actually produced by a speaker, would be identified, and whether the obviously and rather systematically different contours can serve a perceptual and linguistic purpose. Secondly, experiments with real speech would be a natural prerequisite to ones with synthetic speech, i.e. the relevant parameters for systematic variation in synthetic speech would appear in this fashion. Thirdly, the objection that could, rather maliciously and not wholly reasonably perhaps, be raised against the many investigations conducted with synthetic speech (or "semi-synthetic" speech, i.e. vocoder reproduced segments with externally controlled Fo) that listeners are exposed to stimuli that could never occur in real speech, is muted from the outset. (Note that all of the perceptual experiments cited in the bibliography, except for parts of the material of Gårding and Abramson (1965) and parts of Johansson's (1978) stimuli, have been performed with synthetically produced stimuli.)

A further advantage is the fact that the stimuli presented to subjects here are fairly long, comprising three stress groups (7 syllables in all), whereas the studies by Hadding-Koch and Studdert-Kennedy (1963, 1964, 1965, 1974) and Studdert-Kennedy and Hadding (1972, 1973), which most closely resemble, in outline and purpose, the present study were conducted with utterances containing only one stressed syllable ('For Jane' and 'November', respectively).

2.1.1 Stimuli - test 1

A subset of the material for analysis of intonation contours contained a statement, 5 different types of questions, and 3 different types of non-final periods (i.e. 9 sentences in all, which had been recorded 5 times by four subjects), all variations on the same theme:

"..mange busser fra Tiflis .."

('.. many buses out of Tiflis ..'), cf. Thorsen (1976, pp. 91-92). The statement, one type of question, and one type of non-final period were selected from the recordings of one of the male subjects (SH), according to the following criteria: the averages over the five recordings of each type should be well spaced on the frequency scale, whereas the five recordings of each sentence type must show a certain dispersion, so as to create an at least quasi-continuous series of stimuli, from the most steeply falling to flat contours. (But this is, of course, where natural speech is at a disadvantage, compared to synthetic speech: it was not possible to procure a series of stimuli spaced completely equidistantly on the "slope-continuum".) These demands are best fulfilled by SH's non-terminal main clause:

> "Der går mange busser fra Tiflis, så vi kan godt lade bilen stå."

('There are many buses out of Tiflis, so we may well leave the car.') and by the question with word order inversion:

> "Går der mange busser fra Tiflis?"

('Are there many buses out of Tiflis?'). The statement was:

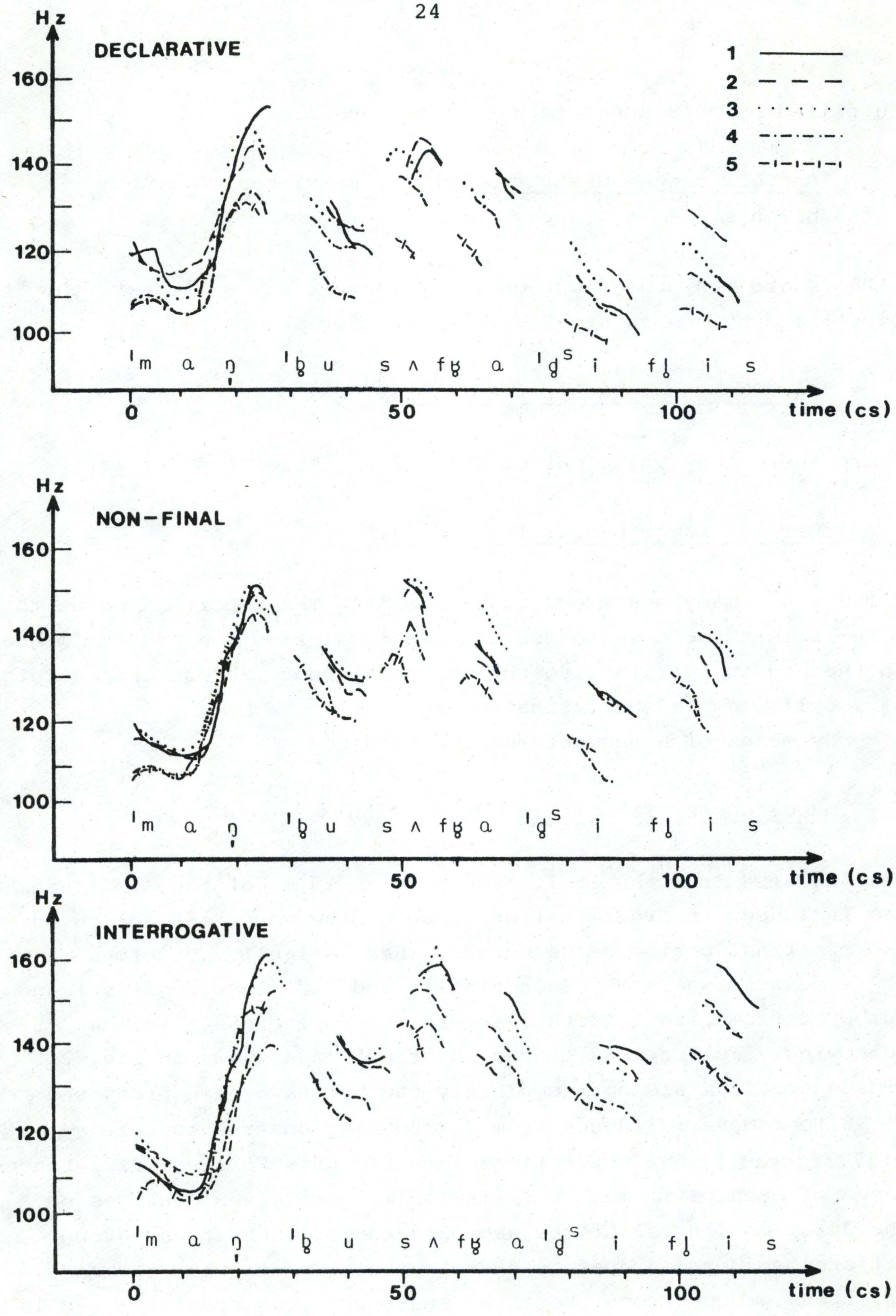
> "Der går mange busser fra Tiflis."

('There are many buses out of Tiflis.'). The important point to note is that these sentences are identical from the [m] in "mange" to the [s] in "Tiflis", so the only difference across this stretch of 7 syllables is the intonation contour.

By means of a segmentator the sequence

> mange busser fra Tiflis ([ˈmɑŋ̩ ˈb̥usʌ fʁ̥ɑ ˈd̥ˢifl̥is])

was isolated from the 15 items. Only 50-70 ms of the final [s] was included, in order that no trace of the word "så" following the non-final period be detected. These stimuli are termed "D" (declarative), "NF" (non-final), and "I" (interrogative) and numbered from 1 to 5 (first, second, etc. recording) in the following. Fo tracings of the 15 stimuli are shown in fig. 2. (These tracings are not completely raw but have been processed so as to remove influence from surrounding obstruents, cf. Thorsen (1977,forthc.), but corrections for intrinsic Fo level differences have not been performed.) - In table 1 are given measures of the duration (in cs) of the segments and of each stress group. Differences between average durations of a particular segment or stress group across the three sentence types are small and in no case statistically significant.



<u>Figure 2</u>

Fundamental frequency tracings
of the 15 stimuli of test 1.

Table 1

Durations (in cs) of segments and
stress groups in the 15 stimuli.

	m	ɑ	ŋ(ə)	ˈmɑŋ(ə)	b̥	u	s	ʌ	fʁ̥	ɑ	ˈb̥usʌfʁ̥ɑ	d̥ˢ	i	fl̥	i	s	ˈd̥ˢifl̥is	Total
D-1	6	12	10	28	8	11	5	6	8	6	44	12	9	11	8	6	46	118
D-2	6			27	7	10	7	6	8	5	43	14	7	10	9	6	46	116
D-3	5	9	11	25	7	11	6	5	5	6	40	13	11	9	10	6	49	114
D-4	5	10	10	25	6	11	6	6	8	6	43	11	11	8	10	5	45	113
D-5	5	8	11	24	8	9	7	5	6	5	40	13	9	11	9	6	48	113
Average	5.4			25.8	7.2	10.4	6.2	5.6	7.0	5.6	42.0	12.6	9.4	9.8	9.2	5.8	46.8	114.8
NF-1	5			24	9	10	6	5	8	5	43	15	8	11	8	7	49	116
NF-2	5			26	9	8	6	5	8	5	41	15	7	13	5	7	47	114
NF-3	5	8	12	25	8	10	6	6	8	5	43	14	7	14	6	5	46	114
NF-4	5			25	8	9	5	6	9	4	41	13	8	10	9	6	46	112
NF-5	6	8	11	25	6	9	6	6	8	7	42	12	6	13	7	7	45	112
Average	5.2			25.0	8.0	9.2	5.8	5.6	8.2	5.2	42.0	13.8	7.2	12.2	7.0	6.4	46.6	113.6
I-1	6			27	10	10	6	5	10	5	46	13	8	11	10	6	48	121
I-2	5			27	9	10	4	6	8	5	42	13	9	10	10	5	47	116
I-3	5	11	12	28	8	10	5	8	9	5	45	10	10	9	8	6	43	116
I-4	5	9	13	27	7	8	8	7	8	4	42	12	8	11	9	5	45	114
I-5	5	8	11	24	7	10	5	7	9	4	42	12	10	11	9	5	47	113
Average	5.2			26.6	8.2	9.6	5.6	6.6	8.8	4.6	43.4	12.0	9.0	10.4	9.2	5.4	46.0	116.0

2.2 Test procedure - test 1

A test tape containing 5 different randomizations of the 15 stimuli, led by 5 dummies, i.e. 80 stimuli in all, was prepared. Stimuli were introduced by numbers (read by the author) with a 3 sec interval for responding. The total duration of one run of the test tape was 11 minutes.

Subjects were instructed that they would hear sequences cut out from a larger context (they were given several examples of possible contexts for each sentence type), and their task was to decide whether the cut originated from one of three sentence types: "declarative", "non-final", or "interrogative" (forced choice). - Three categories to choose from were, of course, motivated by the fact that the 15 stimuli represented 3 different types of sentences. Subjects were given forced choice with no possibility of responding "do not know" or of placing stimuli on a multi-valued scale, - mainly to facilitate the subsequent interpretation of results. - The test was presented twice, with at least one day's interval, giving a total of 10 responses to each stimulus by each subject. 10 subjects listened over head-phones, 4 subjects over loudspeakers in a class-room.

2.3 Subjects - test 1

14 subjects took the first test. 10 were trained phoneticians (among them the speaker), 3 were language students, with no phonetic training, and one was a trained singer with no background in phonetics. Two of the phoneticians and two of the language students were born and went to school outside the Copenhagen area, - the rest were genuine speakers of ASC. It turned out that Copenhagen and non-Copenhagen speakers behaved significantly different, and the two groups are treated separately in the following account of the results.

2.4 Results - test 1

First of all, there is no "order effect" to be detected with any of the subjects. That is, neither "correct" nor "incorrect" responses can be seen to be due to a particular stimulus/response immediately preceding. (An order effect would have been surprising, since stimuli occurred in 5 different randomizations on the tape, and since, further, a stimulus number was announced, by a different voice, and always with a declarative intonation, before each stimulus.) Secondly, the 10 Copenhagen speakers did not differ significantly among themselves, and they are, accordingly, pooled in figures and tables to follow.

Figs. 3 and 4 and tables 2 and 3 present the results, for Copenhagen and non-Copenhagen speakers, respectively. In the figures, stimuli have been reorganized according to the number of responses they received, from maximally interrogative, through maximally non-final, to maximally declarative.

It is immediately obvious from the shape of fig. 3 that there are stimuli in each of the three categories that have been well identified by the Copenhagen speakers: two declarative sentences (D-4 and D-5), two non-final periods (NF-1 and NF-2) and three interrogative sentences (I-1, I-2, and I-3) received 85% or more responses for the respective categories. It is equally obvious that non-Copenhagen speakers do not fare so well on the non-final category, which can hardly be said to have any peak in the identification function at all. Further, non-Copenhagen responses are, on the whole, far more overlapping: 9 stimuli received responses of all three kinds, as opposed to the Copenhagen speakers, who only gave all three kinds of responses to two stimuli (NF-3 and NF-2).

Stimuli that are grouped within parentheses in the two figures are stimuli whose distributions of responses in χ^2 tests turned out not to be significantly different. In both figures, 6 such groups appear, but the members of each group are not exactly identical. Neither is, by the way, the rank order of the stimuli, i.e. NF-3 is shifted one to the left and NF-4 is shifted two to the left in fig. 4 as compared to fig. 3.

The non-Copenhagen speakers thus seem to have been in rather great doubt about the second category, the non-final one, and one might cautiously conclude that these speakers simply do not operate with more than two categories. (However, 4 speakers, who were not even representatives of the same non-Copenhagen dialect,

Table 2

10 Copenhagen speakers' response
to the 15 stimuli in test 1.

response stimulus	DECLARATIVE	NON-FINAL	INTERROGATIVE	Total
D-1	84	16		100
D-2	75	25		100
D-3	79	21		100
D-4	97	3		100
D-5	100			100
NF-1		88	12	100
NF-2	4	90	6	100
NF-3	2	79	19	100
NF-4	81	19		100
NF-5	34	66		100
I-1		5	95	100
I-2		11	89	100
I-3		15	85	100
I-4		61	39	100
I-5		57	43	100
Total	556	556	388	1500

Table 3

4 non-Copenhagenen speakers' response to the 15 stimuli in test 1.

response stimulus	DECLARATIVE	NON-FINAL	INTERROGATIVE	Total
D-1	23	16	1	40
D-2	18	20	2	40
D-3	21	17	2	40
D-4	31	8	1	40
D-5	39	1		40
NF-1		27	13	40
NF-2	7	23	10	40
NF-3	2	15	23	40
NF-4	16	23	1	40
NF-5	12	25	3	40
I-1			40	40
I-2			40	40
I-3			40	40
I-4		20	20	40
I-5	2	15	23	40
Total	171	210	219	600

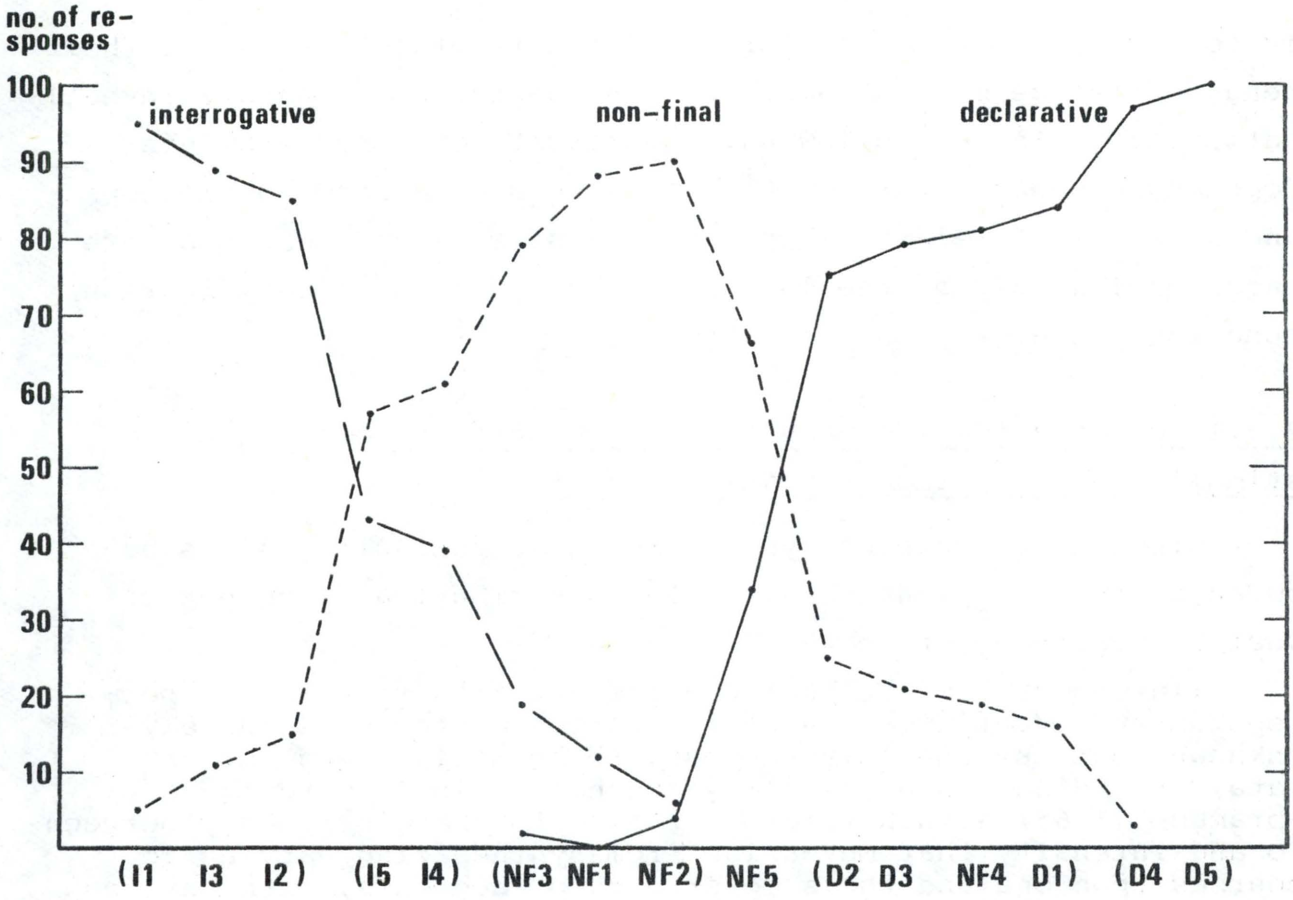


Figure 3

Identification functions for the 15 stimuli in test 1, as identified by 10 Copenhagen speakers.

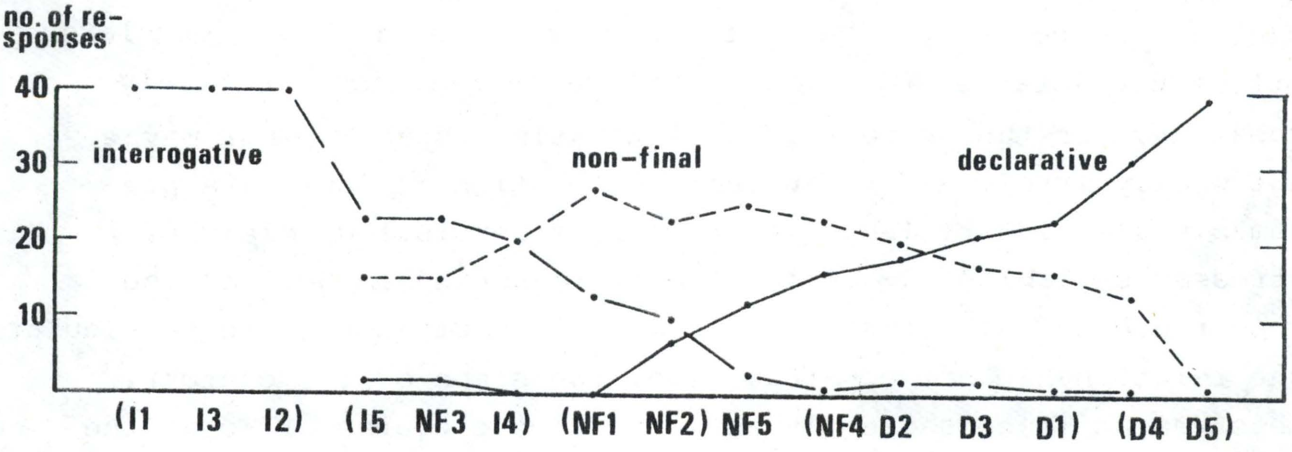


Figure 4

Identification functions for the 15 stimuli in test 1, as identified by 4 non-Copenhagen speakers.

is too small a population for any definite statements about their behaviour to be made - and this point warrants a separate investigation.) - In the following, the results obtained from non-Copenhagen speakers are omitted from consideration: conclusions and analyses of relationships between stimuli and responses are based exclusively on the 10 Copenhagen speakers' identification functions, i.e. fig. 3 and table 2.

2.4.1 Correlations between physical properties of the stimuli and their identification - test 1

Since there were no systematic durational differences between stimuli, explanations for the distribution of responses must be searched for in the Fo course.

(Intensity has not been considered. Its role in the perception and identification of intonation contours is largely unknown, but is generally supposed to be minimal and, at any rate, subordinate and ancillary to that of Fo. - Gårding and Abramson (1965) attach some importance to the discrepancy between Fo and intensity that may arise from synthesizing varying Fo courses upon one and the same stretch of segments, with one and the same distribution of intensity, and Breckenridge and Liberman (1978) found that even small amplitude adjustments have rather large effects upon judgments of the pitch of the second stressed vowel in a "maMAmamaMAma" sequence. Both studies, however, employed synthetic speech and there is as yet no reason to doubt that in natural speech fundamental frequency is the dominant factor in the perception of intonation.)

Since stress group patterns are recurrent entities, cf. fig. 1, the course of the unstressed syllables is to a very large extent predictable and might therefore be redundant, strictly speaking, for the perception of intonation contours (but maybe not wholly irrelevant). Further, intonation contours are presumably also identifiable in utterances consisting solely of stressed syllables. - This line of reasoning emphasizes the importance of the stressed syllables in an utterance, in particular the relationship between them, i.e. the slope of the contour. Less weight is consequently assigned to the course of Fo at the very end. This is in explicit contradistinction to the viewpoint expressed by Daneš (1960), Isačenko (1965), and Isačenko and Schädlich (1963) who considered only the terminal Fo course in an utterance to have any differentiating function. And although a number of writers have conceded that earlier parts of the course of Fo may play a role, they have all proceeded to test the significance of the "terminal contour", see e.g. von Essen (1956), Gårding and Abramson (1965), Hadding-Koch and Studdert-Kennedy (1963, 1964, 1965, 1974), Studdert-Kennedy and Hadding (1972, 1973), and Uldall (1960, 1961). Let us see how far the working hypothesis, underlying the present experiments, will get us:

In fig. 5 are depicted Fo tracings of those three groups of stimuli, members of which could not be shown to differ significantly among themselves as far as the distribution of responses was concerned, and which were well identified, cf. fig. 3. They are, not unexpectedly, seen to be fairly well separated, and this separation is most evident in the final stressed vowel ("T$\underline{i}$flis"), less evident in the second stressed vowel ("b$\underline{u}$sser") but still without any overlap between the three groups. The unstressed syllables of the second and third stress groups tend to distribute themselves in the same manner as the stressed syllables, but some overlap is to be found. It is hard to find any order at all in the first stress group. This is all as it should be, granted that different contours, departing from the same point, are more clearly separated, the further we progress in time, cf. fig. 1.

In fig. 6 are shown Fo tracings of stimuli that have received ambiguous responses (i.e. stimuli at or near the cross-over points of the identification functions of fig. 3: I-4, I-5 and NF-5, D-2) as well as one well identified stimulus from each category (I-1, NF-1, D-5). I-4 and I-5 are well removed from I-1, and much closer to NF-1 (especially in the third stress group). Likewise, NF-5 and D-2 are well removed from D-5 (but not particularly close to NF-1).

2.4.1.1 A closer look at stimuli

Figs. 5 and 6 may thus account for the gross trends in the results, but they cannot account for the finer distinctions we observe in the identification functions. In order to find one

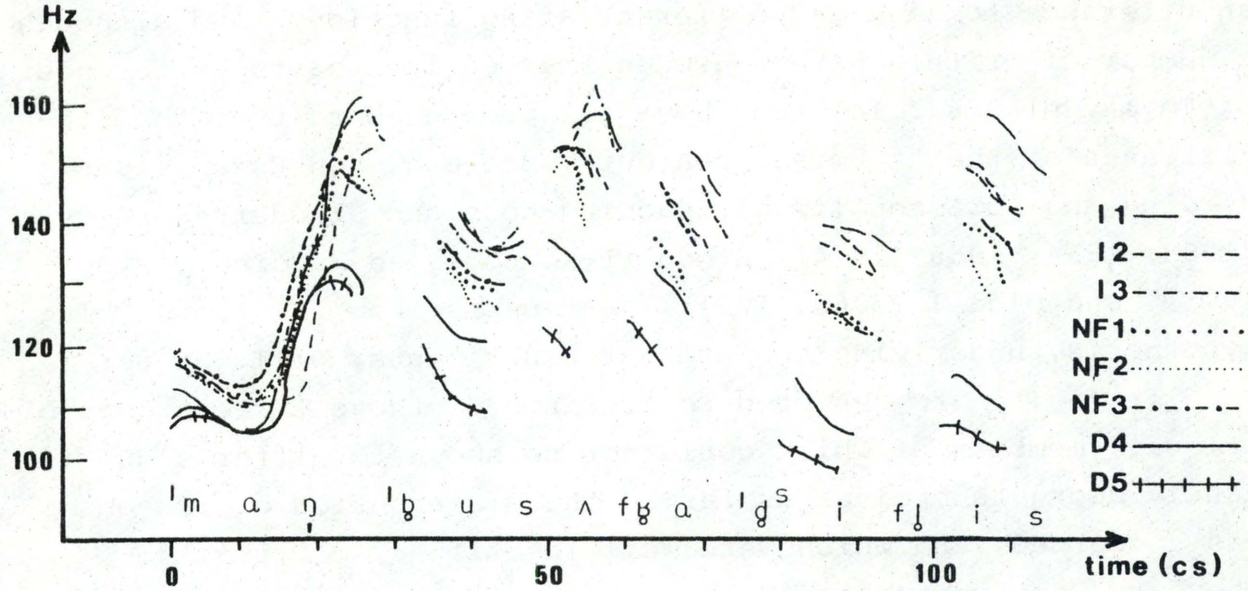


Figure 5

Fundamental frequency tracings of groups of stimuli that have been well identified by 10 Copenhagen speakers, cf. fig. 3.

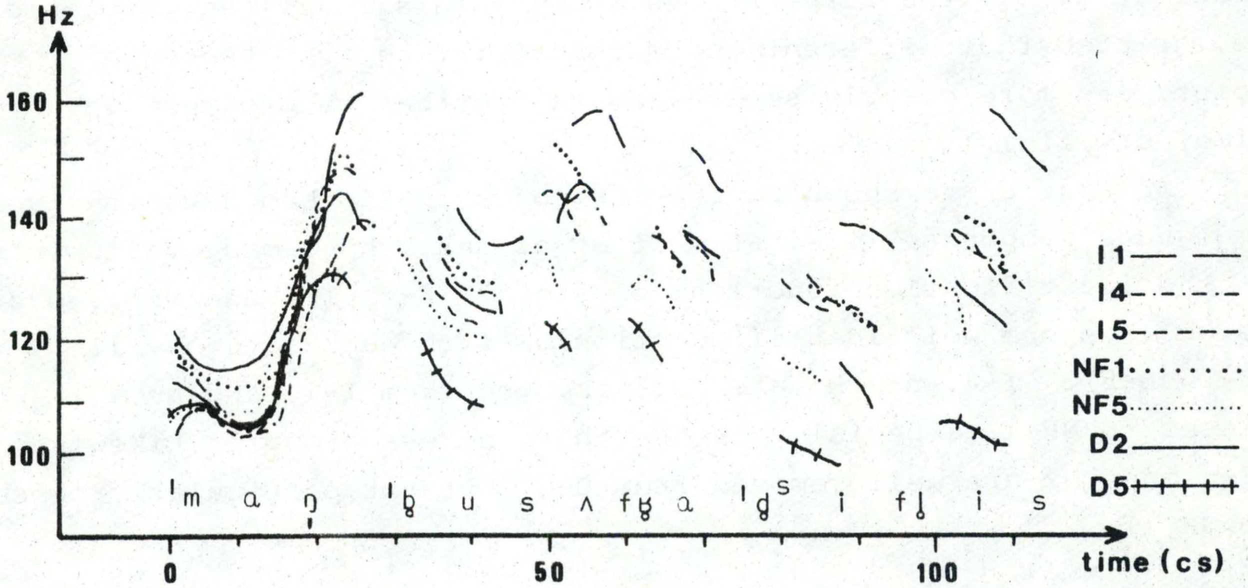


Figure 6

Fundamental frequency tracings of stimuli at or near the cross-over points of the identification functions of fig. 3 (I-4, I-5 and NF-5, D-2) as well as stimuli that have been well identified (I-1, NF-1, D-5) by 10 Copenhagen speakers.

or more relevant parameters, i.e. parameters which will yield a high correlation with the distribution of responses, table 4 was prepared. It contains a number of measures of various points in the Fo tracings, as well as certain relations which have been calculated on the basis of these measures.

Measuring Fo is not a wholly uncontroversial business. As a rule, in all the vowels, except [ˈɑ], that point has been measured which lies at 2/3 of the distance in time from the beginning of the vowel, in accordance with Rossi's (1971, 1978) results on the perception of Fo movements that are too short and/or too slight to be perceived as anything but level pitches. However, delimiting the beginning and end of a vowel (which determine the "2/3 point") is not a fool-proof procedure, and moving the point of measurement just 1 cs left or right may introduce quite considerable differences in the Fo measurement. Thus, the measures in table 4 should be taken cum grano salis.

For the first stressed vowel, [ˈɑ], two measures are given: its minimum value (which is turned into account in the calculation of the total rise of Fo in that, first stress group), and the value at 2/3 of the distance from the beginning of its rise (which is turned into account in fig. 7). From [ˈu] and [ˈi], and their time interval, a "base-line" slope, S_b, has been calculated, and, likewise, from [˳ʌ] and [˳i], a "top-line" slope, S_t, is determined. The distance between each stressed and the immediately succeeding unstressed vowel (or sonorant consonant) is also given.

Note that stimuli have been grouped in table 4, as in fig. 3, according to the rank order of their responses, from maximally interrogative (rank 1), through maximally non-final, to maximally declarative (rank 15), as indicated in small script next to the stimulus-designations. The 6 groups of fig. 3 are maintained, and for each group averages have been calculated. Every measure has been assigned a rank, from highest (1) to lowest (15), as indicated in small raised script to the right of each value. Further, the averages have been assigned ranks in the same fashion, from highest (1) to lowest (6), as indicated in small lowered script.

If we let the "natural order" be the ranks assigned to stimuli according to responses, i.e. the order in which stimuli have been tabulated from top to bottom in table 4, then it is

Table 4 (continued on the next page)

Fundamental frequency measurements of segments and relations between certain segments in the 15 stimuli of test 1. Stimuli have been re-grouped according to identification by 10 Copenhagen speakers, cf. fig. 3. Numbers in small script indicate ranks from highest (^{1}) to lowest (^{15}) on a 15-point rank order scale. For an account of the measuring procedure, see the text.

	ˈɑ_{min} (1)	$\text{ˈɑ}_{\frac{2}{3}}$ (2)	ŋ_{max} (3)	ˈu	$_{\circ}\text{ʌ}$	$_{\circ}\text{ɑ}$	ˈi
I-1 ^{1}	107 $^{10\frac{1}{2}}$	116 $^{6\frac{1}{2}}$	162 ^{1}	136 $^{1\frac{1}{2}}$	158 ^{1}	147 ^{1}	138 ^{1}
I-3 ^{2}	112 $^{4\frac{1}{2}}$	119 $^{1\frac{1}{2}}$	160 ^{2}	136 $^{1\frac{1}{2}}$	156 ^{2}	141 ^{2}	134 ^{3}
I-2 ^{3}	111 ^{7}	112 ^{11}	153 ^{4}	135 ^{3}	150 ^{4}	139 ^{4}	137 ^{2}
Av. $_{1}$	110.0 $_{3}$	115.7 $_{3}$	158.3 $_{1}$	135.7 $_{1}$	154.7 $_{1}$	142.3 $_{1}$	136.3 $_{1}$
I-5 ^{4}	104 ^{15}	108 $^{13\frac{1}{2}}$	148 $^{8\frac{1}{2}}$	125 ^{10}	140 $^{10\frac{1}{2}}$	134 $^{7\frac{1}{2}}$	126 ^{5}
I-4 ^{5}	105 $^{13\frac{1}{2}}$	108 $^{13\frac{1}{2}}$	140 ^{13}	128 $^{6\frac{1}{2}}$	142 $^{8\frac{1}{2}}$	134 $^{7\frac{1}{2}}$	127 ^{4}
Av. $_{2}$	104.5 $_{6}$	108.0 $_{6}$	144.0 $_{4\frac{1}{2}}$	126.5 $_{3}$	141.0 $_{3}$	134.0 $_{4}$	126.5 $_{2}$
NF-3 ^{6}	114 ^{2}	119 $^{1\frac{1}{2}}$	152 $^{5\frac{1}{2}}$	132 ^{4}	150 ^{4}	140 ^{3}	124 $^{7\frac{1}{2}}$
NF-1 ^{7}	112 $^{4\frac{1}{2}}$	116 $^{6\frac{1}{2}}$	152 $^{5\frac{1}{2}}$	130 ^{5}	150 ^{4}	133 $^{10\frac{1}{2}}$	125 ^{6}
NF-2 ^{8}	112 $^{4\frac{1}{2}}$	118 $^{3\frac{1}{2}}$	150 ^{7}	128 $^{6\frac{1}{2}}$	146 ^{6}	130 ^{12}	124 $^{7\frac{1}{2}}$
Av. $_{3}$	112.7 $_{1}$	117.7 $_{1}$	151.7 $_{2}$	130.0 $_{2}$	148.7 $_{2}$	134.3 $_{2\frac{1}{2}}$	124.3 $_{3}$
NF-5 $^{9}_{4}$	108 $^{9}_{4}$	110 $^{12}_{4\frac{1}{2}}$	144 $^{11}_{4\frac{1}{2}}$	125 $^{10}_{4}$	133 $^{14}_{5}$	128 $^{14}_{5}$	115 $^{9}_{4}$
D-2 ^{10}	116 ^{1}	118 $^{3\frac{1}{2}}$	144 ^{11}	126 ^{8}	143 ^{7}	136 ^{5}	113 ^{10}
D-3 ^{11}	110 ^{8}	113 ^{10}	148 $^{8\frac{1}{2}}$	125 ^{10}	140 $^{10\frac{1}{2}}$	133 $^{10\frac{1}{2}}$	112 ^{11}
NF-4 ^{12}	107 $^{10\frac{1}{2}}$	115 $^{8\frac{1}{2}}$	144 ^{11}	122 $^{12\frac{1}{2}}$	137 ^{12}	134 $^{7\frac{1}{2}}$	108 $^{12\frac{1}{2}}$
D-1 ^{13}	112 $^{4\frac{1}{2}}$	117 ^{5}	154 ^{3}	119 ^{14}	142 $^{8\frac{1}{2}}$	134 $^{7\frac{1}{2}}$	105 ^{14}
Av. $_{5}$	111.3 $_{2}$	115.8 $_{2}$	147.5 $_{3}$	123.0 $_{5}$	140.5 $_{4}$	134.3 $_{2\frac{1}{2}}$	109.5 $_{5}$
D-4 ^{14}	106 ^{12}	115 $^{8\frac{1}{2}}$	134 ^{14}	122 $^{12\frac{1}{2}}$	134 ^{13}	129 ^{13}	108 $^{12\frac{1}{2}}$
D-5 ^{15}	105 $^{13\frac{1}{2}}$	105 ^{15}	131 ^{15}	112 ^{15}	121 ^{15}	119 ^{15}	96 ^{15}
Av. $_{6}$	105.5 $_{5}$	110.0 $_{4\frac{1}{2}}$	132.5 $_{6}$	117.0 $_{6}$	127.5 $_{6}$	124.0 $_{6}$	102.0 $_{6}$

---

1) ˈɑ_{min} is measured at the turning point in the fall-rise of [ˈɑ].

2) $\text{ˈɑ}_{\frac{2}{3}}$ is measured at a point in time $\frac{2}{3}$ from the beginning of the rise of [ˈɑ].

3) ŋ_{max} is the maximum value of [ŋ].

(continued from the preceding page) Table 4

ₒi		S_b(4)		S_t(5)		ɑ$_{min}$- ŋ$_{max}$		ɑ$^2_3$ - ŋ$_{max}$		ˈu-ₒʌ		ˈi-ₒi			
153	1	-4.2	2	9.3	1	55	1	46	1	22	2	15	1	I-1	1
144	2½	4.3	5	23.5	6	48	2	41	2½	20	3½	15	6	I-3	2
144	2½	-4.4	1	11.5	3½	42	4½	41	2½	15	9½	7	11½	I-2	3
147.0	1	-1.4	1	14.8	3	48.3	1	42.7	1	19.0	1	10.7	2	Av.	1
134	6	-2.1	3	11.3	2	44	3	40	4	15	9½	8	9	I-5	4
134	6	2.0	4	15.4	5	35	12	32	10½	14	12	7	11½	I-4	5
134.0	2	0.0	2	13.4	2	39.5	2	36.0	2	14.5	4	7.5	5	Av.	2
137	4	17.0	8	24.0	7	38	8	33	9	18	5½	13	2	NF-3	6
134	6	10.4	7	29.1	8	40	6	36	6	20	3½	9	7	NF-1	7
130	8	8.5	6	30.8	9	38	8	32	10½	18	5½	6	13	NF-2	8
133.7	3	12.0	3	28.0	4	38.7	3	33.7	4	18.7	2	9.3	3	Av.	3
127	9 4	20.8	9 4	11.5	3½ 1	36	11 5	34	8 3	8	15 6	12	3½ 1	NF-5	9 4
124	10	26.5	10	36.5	11	28	13½	26	13½	17	7	11	5	D-2	10
115	12	27.7	11	45.5	13	38	8	35	7	15	9½	3	15	D-3	11
120	11	29.8	13	33.3	10	37	10	29	12	15	9½	12	3½	NF-4	12
113	13½	30.4	14	54.7	14	42	4½	37	5	23	1	8	9	D-1	13
118.0	5	28.6	5	42.5	5	36.3	4	31.8	5	17.5	3	8.5	4	Av.	5
113	13½	29.2	12	45.3	12	28	13½	19	15	12	13	5	14	D-4	14
104	15	34.0	15	70.4	15	26	15	26	13½	9	14	5	9	D-5	15
108.5	6	31.6	6	57.9	6	27.0	6	22.5	6	10.5	5	6.5	6	Av.	6

---

4) S_b is the slope of the intonation contour ("base-line" slope), in cs/sec, based on the measurements of [ˈu] and [ˈi], and their time interval.

5) S_t is the slope of the "top-line", in cs/sec, based on the measurements of [ₒʌ] and [ₒi], and their time interval.

immediately obvious that certain parameters deviate less from this natural order than do others. And if we just look at the ranks of the averages, S_b, [ˈi] and [ₒi] are even seen to follow the natural order exactly, and the rank order for the averages of [ˈu] is reversed in only one place. These parameters, therefore, look like good candidates and worth inspecting more closely, in the search for correlations between the physical properties of stimuli and the responses they yielded. - Fig. 7 is a stylized graph of the intonation contours and the final rise from stressed to unstressed [i], based on the measures of [ˈu], [ˈi] and [ₒi]. To complete the picture of the contour, the value of ˈɑ$^{2}_{3}$ has been included (plus 15% to compensate for intrinsic Fo level differences between [ˈu] and [ˈi], on the one hand, and [ˈɑ] on the other, for this particular speaker, cf. Reinholt Petersen, 1976).

Fig. 7 clarifies figs. 5 and 6 and "explains" some things that were less transparent in those figures: for instance, I-4 and I-5 were seen in fig. 6 to be very nearly concurrent with NF-1, and one was left to wonder why, then, had these two interrogative sentences not been labelled "non-final" more than about 50% of the time by listeners? - It is clear from fig. 7 that the slopes of I-4,5 are different from those of the well identified non-finals, i.e. they are very nearly flat, whereas the slopes of NF-1,2,3 are falling. - In short, the six groups of fig. 3 and table 4 emerge rather clearly.

Since the members from each group do not differ significantly among themselves, as far as distribution of responses goes, one might leave it at that and be satisfied that slope, final stressed and final unstressed vowel all correlate exactly with responses when averages are considered. - However, this is slightly unsatisfactory, partly because stimuli *are* different, and *did* receive different responses, partly because it is not possible to say which of the three parameters is the most crucial for identification. - Accordingly, the data in table 4 were turned into account for the calculation of correlations between responses, on one hand, and each of the parameters of table 4, on the other.

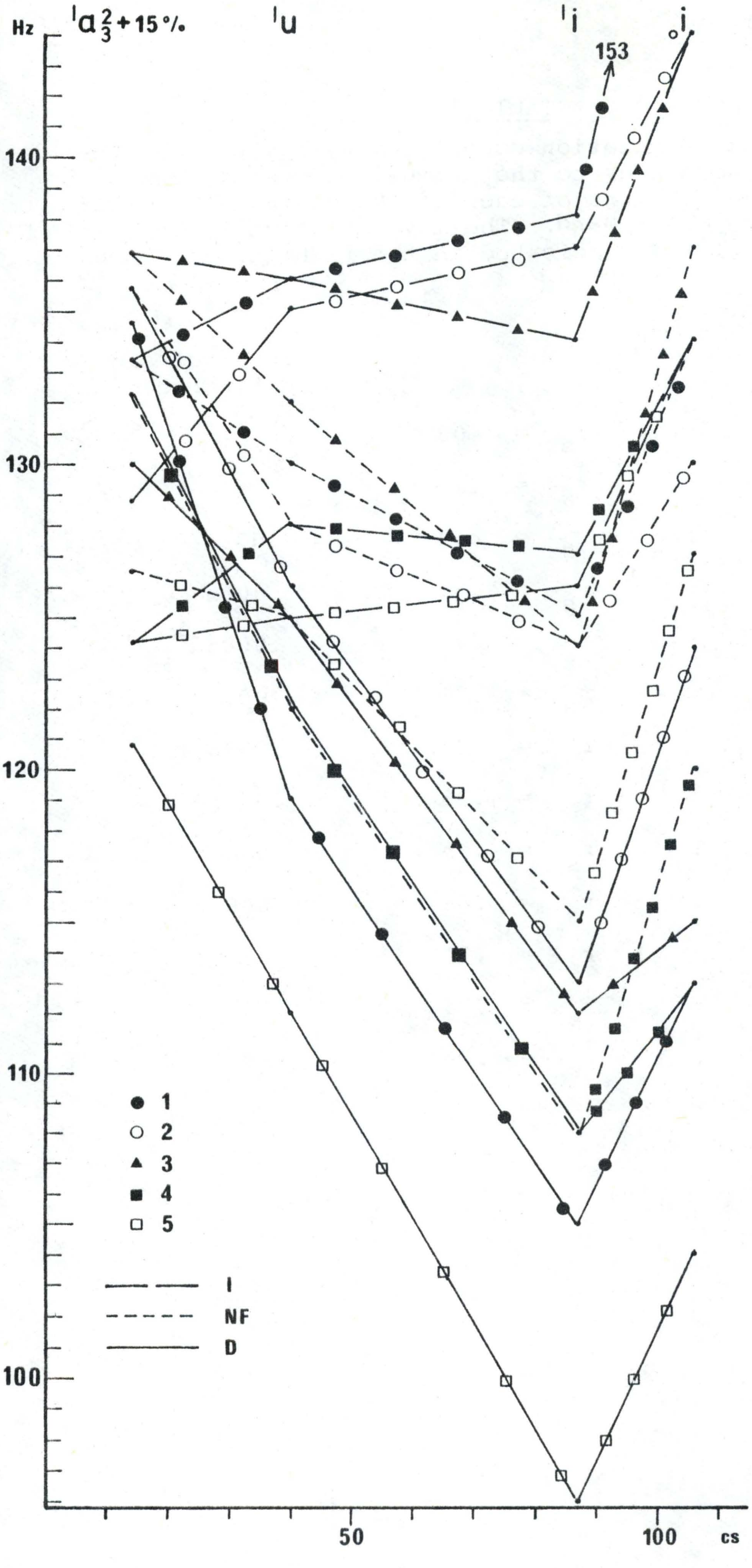


Figure 7

Stylized intonation contours plus the final rise from stressed to unstressed [i] of the 15 stimuli in test 1. See further the text.

Table 5

Spearman rank correlation coefficients, r_s, between the rankings of responses to the stimuli of test 1, on one hand, and the rankings of each of the parameters in table 4, on the other hand. The level of significance (if better than .10) is indicated in the right hand column.

Correlation between response rankings and rankings of			level of significance
ˈɑ$_{min}$	:	r_s = .08	-
ˈɑ$^{2}_{3}$	:	r_s = .18	-
ŋ$_{max}$	:	r_s = .64	.005
ˈu	:	r_s = .89	.0005
$_{\circ}$ʌ	:	r_s = .77	.0005
$_{\circ}$ɑ	:	r_s = .71	.005
ˈi	:	r_s = .98	.0005
$_{\circ}$i	:	r_s = .98	.0005
S_b	:	r_s = .95	.0005
S_t	:	r_s = .88	.0005
ɑ$_{min}$-ŋ$_{max}$	:	r_s = .71	.005
ɑ$^{2}_{3}$-ŋ$_{max}$	:	r_s = .74	.005
ˈu-$_{\circ}$ʌ	:	r_s = .38	.10
ˈi-$_{\circ}$i	:	r_s = .30	-

Table 5 gives Spearman rank correlation coefficients (calculated, when necessary, with corrections for tied ranks) between the rankings of responses to the stimuli, and the rankings of the measures of various points in the fundamental frequency trancings, as presented in table 4.

It is obvious from table 5 that S_b, ˈi, and ˳i, i.e. the slope of the intonation contour (or "base-line" slope), the levels of the final stressed and the final unstressed vowel, respectively, all correlate very highly with responses.

2.4.1.2 Terminal rises

Before we proceed, it is interesting to note that the magnitude of the final rise, i.e. ˈi-˳i, shows only a slight and non-significant correlation with responses (and no correlation at all (r_s = .05) when this rise is expressed in percentage of the level of the stressed vowel). This must be taken to mean that the magnitude of the final rise was not, in this experiment, a perceptual cue to the identification of intonation contours, but its placement on the frequency scale is (i.e. the levels of the final stressed and succeeding unstressed vowels).

However, Danish listeners are not insensitive to terminal rises: In the course of some informal experiments, conducted at the Institute of Linguistics, Uppsala University, with the ILS-system for analysis and synthesis of speech, the declarative sentence "Der går mange busser fra Tiflis." was recorded (by the author). It was then re-synthesized with the final unstressed syllable, "-lis", being raised 60Hz, in steps of 10Hz (everything else being kept constant). These 6 utterances, together with the original, were later played back to a group of phoneticians in Copenhagen. The first two steps up of "-lis" made no apparent difference. At +30Hz and +40Hz listeners reported that they heard a statement, but a rather obstinate one. At +50Hz the utterance began to sound interrogative (to some listeners), and at 60Hz the utterance was classified as a question by the majority, but a rather surprised or disbelieving echo-question. Other listeners would not accept the utterance as a question at all, but dismissed it as being wholly unnatural.

This illustrates very nicely a point where real and synthetic speech part company: it is possible to make listeners respond in certain (and often predictable) ways to cues, or rather combinations of cues (cf. below), which are not found in real speech, but analyses of and perceptual experiments on real speech may then serve to dissolve any ambiguity as to which of the cues, found to be operative in synthetic speech, is the dominant one in speech perception. The interpretation presented here for Danish could probably also be applied to the results obtained for Swedish (and American): Hadding-Koch (1961) found in her analysis of Swedish that statements tend to end in a terminal fall, questions in a terminal rise (but there were differences to observe in earlier parts of the utterance, also). In experiments with synthetic stimuli Studdert-Kennedy and Hadding (1972, 1973) found that if a low peak (in the stressed vowel of 'November') is heard, listeners tend to interpret the utterance as a statement unless it is followed by a large terminal rise (in the final, unstressed syllable). If a high peak is heard, listeners tend to interpret the utterance as a question, unless they also hear a large terminal fall. In other words: everything else being equal, the higher the terminal rise, the more questions are heard, and vice versa, or, again, everything else being equal, the higher the terminal peak, the more questions are heard, and vice versa. They conclude that peak and perceived terminal glide are the two factors that seem to govern linguistic judgments of intonation contours (although they are not just and simply additive in their effects). However, in consequence of the line of argument for Danish above, I would be inclined to give top priority to their peak level.

2.4.2 Correlations and concordances between various points in the course of fundamental frequency in the stimuli

Although table 5 is informative, it may also be slightly misleading. The high correlations we observe between responses and S_b, [ˈi], and [ₒi] (and [ˈu] and S_t), respectively, are of course partly due to the fact that these parameters are intercorrelated themselves. If [ˈu] is high, [ˈi] is likely to be high also, etc.; in fact, this is precisely the kind of prediction that can be read off from the model in fig. 1.

The Spearman rank correlation coefficients (with corrections for tied observations, when necessary) between a number of pairs of measurements in the tracings, as well as the Kendall coefficient of concordance between several measures, have been calculated and are presented in table 6. The level of significance, when better than .10, is indicated in the right hand column.

Table 6 focuses upon the maxima (first unstressed syllable in each stress group) and minima (the stressed syllables) and their correlations to each other. A couple of aspects seem worth commenting upon. - If we set the limit between high and low correlations at r_s = .70, we note, first of all, that all correlations involving the first stressed vowel, $ˈɑ^{2}_{3}$ or $ˈɑ_{min}$ 1), are low and so are, with one exception ($ŋ_{max}$ ~ ˳ʌ) all correlations involving $ŋ_{max}$. The course of fundamental frequency during the first stress group, in short, does not correlate very highly with later parts of the utterance.

Correlations between the stressed vowel and the rise to the succeeding unstressed vowel, in the second and third stress groups, are also low. Apart from that, every point of measurement in the second and third stress groups correlate highly with every other point, and "top-line" and "base-line" slopes, S_t and S_b, are highly correlated. (Further, the stressed vowels correlate highly with S_b and the unstressed vowels with S_t - but one should bear in mind that these slopes have been calculated from [ˈu] and [ˈi], and [˳ʌ] and [˳i], respectively, and are thus anything but independent of these measures.)

The degree of concordance between the three maxima (unstressed syllables) is considerable and is higher than that between the

---

1) $ˈɑ_{min}$ renders lower coefficients than does $ˈɑ^{2}_{3}$, and since $ˈɑ^{2}_{3}$ thus seems to be the more "consistent" of the two measurements, correlation coefficients between $ˈɑ_{min}$ and <u>all</u> other points have not been calculated.

Table 6 (continued on the next page)

Spearman rank correlation coefficients, r_s, between the rankings of a number of pairs of parameters in table 4, and Kendall coefficients of concordance, W, between the rankings of several such parameters. The level of significance (if better than .10) is indicated in the right hand column.

			level of significance
ˈi ~ S_b	:	r_s = .98	.0005
ˈi ~ ˈu	:	r_s = .92	.0005
ˈi ~ ɑ$_{\mathrm{min}}$	:	r_s = .01	-
ˈi ~ ˈɑ$^2_3$	:	r_s = .11	-
ˈi ~ ˳i	:	r_s = .96	.0005
ˈi ~ ˳ʌ	:	r_s = .75	.005
ˈi ~ ŋ$_{\mathrm{max}}$	:	r_s = .57	.025
ˈi ~ ˈi - ˳i	:	r_s = .19	-
ˈu ~ S_b	:	r_s = .82	.0005
ˈu ~ ˈɑ$_{\mathrm{min}}$	:	r_s = .38	.10
ˈu ~ ˈɑ$^2_3$	:	r_s = .45	.05
ˈu ~ ˳ʌ	:	r_s = .90	.0005
ˈu ~ ˳i	:	r_s = .94	.0005
ˈu ~ ŋ$_{\mathrm{max}}$	:	r_s = .67	.005
ˈu ~ ˈu - ˳ʌ	:	r_s = .48	.05

Table 6 (continued from preceding page)

					level of significance
ˈɑ$^{2}_{3}$~S_b	:	r_s	=	.02	-
ˈɑ$^{2}_{3}$~ŋ$_{max}$	:	r_s	=	.58	.025
ˈɑ$^{2}_{3}$~ₒʌ	:	r_s	=	.66	.005
ˈɑ$^{2}_{3}$~ₒi	:	r_s	=	.28	-
ˈɑ$^{2}_{3}$~ˈɑ$^{2}_{3}$-ŋ$_{max}$	:	r_s	=	.12	-
ˈɑ$_{min}$~ŋ$_{max}$	:	r_s	=	.49	.05
ˈɑ$_{min}$~ˈɑ$_{min}$-ŋ$_{max}$	:	r_s	=	.34	-
ŋ$_{max}$~ₒʌ	:	r_s	=	.84	.0005
ŋ$_{max}$~ₒi	:	r_s	=	.66	.005
ₒʌ~ₒi	:	r_s	=	.82	.0005
S_b~S_t	:	r_s	=	.89	.0005
ŋ$_{max}$~ₒʌ~ₒi	:	W	=	.84	.005
ˈɑ$^{2}_{3}$~ˈu~ ˈi	:	W	=	.66	.01
ˈɑ$_{min}$~ˈu~ˈi	:	W	=	.61	.025
ˈɑ$^{2}_{3}$~ŋ$_{max}$~ˈu~ₒʌ~ˈi~ₒi	:	W	=	.73	.0005
ˈɑ$_{min}$~ŋ$_{max}$~ˈu~ₒʌ~ˈi~ₒi	:	W	=	.70	.0005

three minima (stressed vowels), which is somewhat surprising, - but it might possibly reflect a greater uncertainty about measuring Fo in [ˈɑ], i.e. finding the relevant point to measure, than in other vowels. Finally, the concordance between all six points, maxima and minima, is high as well. (Again, when ˈɑ$^{2}_{3}$ is used, the concordance is slightly higher than when ˈɑ$_{min}$ is used.)

The fact that the two stressed vowels, [ˈu] and [ˈi], correlate highly with [˳ʌ] and [˳i], respectively, but not with the magnitude of the rise from stressed to unstressed vowel, together with the fact that responses show low correlations with these rises, cf. table 5, may be interpreted thus: the distance between stressed and first unstressed vowel in a stress group, which decreases from beginning to end of the utterance, does not, also, decrease as a function of the kind of intonation contour upon which the stress group patterns are imposed. In other words, top-lines and base-lines decline (more or less, depending on the sentence) and top-lines decline more rapidly than do base-lines. But top-lines do not decline relatively more rapidly, the steeper the base-line is:

The differences between S_t and S_b were calculated and ranked. The correlation coefficient between this ranking and that of S_b is .18 only, - whereas it is .89 between the rankings of S_b and S_t. - Thus, base-line and top-line are highly correlated, and top-lines are steeper than base-lines, cf. table 4, but the difference in slope between top- and base-lines is, within certain limits, more or less random, averaging about 15Hz/sec. - However, only a subset of the recordings of one subject is involved, and nothing definite can be said about this point yet.

Since all points in the course of fundamental frequency in the second and third stress groups correlate rather highly with each other, the high correlations listed in table 5 between responses and [ˈi], [˳i], S_b, [ˈu], and S_t cannot be said to be mutually independent. On the contrary, there is good reason to believe that identification has been based on the whole Fo course during the second and third stress groups, rather than on one single parameter, but, as mentioned in the abstract, the level of the final stressed and unstressed vowels can almost, alone, account for the way responses were distributed over the 15 stimuli.

One example will suffice to illustrate the point made above: [ˈi] and [˳i] gave the highest correlation with responses, and they correlate highly with each other as well. The Kendall rank correlation coefficient, τ, between response and [ˈi] is .93, between response and [˳i] it is .94, and between [ˈi] and [˳i] it is .91.

The Kendall partial rank correlation coefficient between response and [ˈi], with the effect of [ₒi] partialled out, is .57, and the partial rank correlation between response and [ₒi], with the effect of [ˈi] partialled out, is .60. Both coefficients are a good deal lower than the non-partial correlations, and we may conclude that the relation between response and [ˈi] is *not* independent of [ₒi], and, vice versa, the relation between response and [ₒi] is *not* independent of [ˈi]. Since the non-partial correlation coefficients decrease by approximately the same amount we cannot determine which of the two, [ˈi] or [ₒi], is more dependent upon the other.

2.5 Conclusion - test 1

In conclusion we may say about the results of test 1, that they confirm the implications of the model in fig. 1 and the working hypothesis outlined in section 2.4.1: intonation contours are more widely separated in later than in early parts of the utterance, and the stressed syllables in this later part are more decisive for identification than are the unstressed syllables, cf. table 5. Since the correlation/concordance between points in the Fo course in the second and third stress groups is high, cf. table 6, it is not possible to establish a single parameter which, independently of all other parameters, will account for listeners' identification. Rather, identification may be determined by all of the later part of the Fo course, probably, however, with the *level* of the final stress group having slightly more weight than that of the second stress group. The terminal *rise* from stressed to unstressed syllable was of no consequence whatsoever.

Copenhagen speakers identified three clear categories, non-Copenhagen speakers did not. - Whether these three categories are purely phonetic, i.e. a matter of dividing a physical (near-) continuum into X number of classes, or whether this categorization reflects some real linguistic categories is a matter for test 2 to resolve. - However, I did think that there was some support for a hypothesis of linguistic categories already in the fact that speakers do indeed produce more than two kinds of intonation contours, which do tend to arrange themselves along three slopes, with a good deal of overlap, however, between individual items.

3. Test 2

If the categorization performed by subjects in test 1 were purely phonetic, or non-linguistic, i.e. the continuum had been divided into three equally large bits, without regard to any linguistic analysis, then one would expect listeners to divide the same continuum into two equally large bits, when presented with only two categories to choose from. Accordingly, the test tape from test 1 was presented to 7 of the 10 Copenhagen speakers (and two of the four non-Copenhagen speakers), about 6 months later. The procedure resembled exactly that of test 1, except that stimuli now were to be labelled either "declarative" or "non-declarative". (Subjects did not know that the two tapes were identical.)

3.1 Results - test 2

Table 7 and fig. 8 present the 7 Copenhagen speakers' distribution of responses in test 1 to the 15 stimuli. Comparing fig. 8 to fig. 3, we see that the response distributions are almost identical: the same 6 groups turn up in fig. 8 as in fig. 3, but the ordering of stimuli within some groups is slightly different. Thus, I-3 and I-1, I-4 and I-5, NF-1 and NF-3, and NF-4 and D-2 have changed places. However, the Spearman rank correlation coefficient between response rankings by 10 and by 7 listeners, respectively, is .99, and we may safely conclude that the 7 listeners employed in test 2 are a representative subset of the 10 listeners employed in test 1.

Table 8 and fig. 9 present the results of test 2. As in test 1, there was no "order-effect" to be found with any subject, nor did subjects differ significantly among themselves.

It is evident that stimuli have not, in test 2, been divided into two equally large classes. There are differences in the rank ordering of stimuli in test 2 and test 1 (for the 7 listeners) but the Spearman rank correlation coefficient between these rankings is .96 and the response distributions in the two tests are thus highly correlated. The only non-predictable difference between figs. 8 and 9 is that whereas the cross-over point between the non-final and declarative categories lies to the right of stimulus NF-5 in fig. 8, the corresponding cross-over point between the non-declarative and declarative categories lies to the left of NF-5 in fig. 9. But the cross-over point is still sharp and well-defined, and only three groups are now formed, cf. the parentheses around stimuli designations in fig. 9. (A group is formed by stimuli whose response distributions cannot, by a χ^2 test, be proved to be significantly different at the .10 level.)

About the change of NF-5 from one category to another in the two tests: in fact, its identification in test 2 as mainly declarative (67% of the time) is less curious than its identification in test 2 as mainly non-final (79% of the time by the 7 Copenhagen speakers). If we go back to (table 4 and) fig. 7, we see that NF-5, in slope and position on the frequency scale, runs very close to those stimuli that have been identified mainly as declarative (D-1,2,3,4,5 and NF-4). This would have warranted an identification in test 1 as declarative. However, it is the "top" contour in that group of "low" ones in fig. 7, and its identification as mainly non-final in test 1 might possibly reflect an attempt by listeners to balance responses (i.e. to try and give 33% responses in each category) - and if one of the "low" contours is to be affected by this attempt, it is precisely NF-5.

The two non-Copenhagen speakers who took test 2 did not, this time, differ remarkably from the Copenhagen speakers. The cross-over point between the two categories is in the same place as for Copenhagen speakers. But to the left of NF-5 everything is unanimously labelled "non-declarative", whereas the identification functions to the right of the cross-over point are slightly more slowly rising and falling than for Copenhagen speakers. Again, however, 4 and 2 (non-homogeneous) non-Copenhagen speakers, respectively, are not sufficient for sound conclusions to be drawn, nor for attempts at explaining their behaviour.

In conclusion, the stimuli that were labelled "non-final" in test 1 were not split up into two equally sized bits, each going to the declarative and non-declarative categories, respectively. With one exception, the formerly "non-final" stimuli were now categorized as "non-declarative" and this may be taken as an indication that linguistic categories were at play. (A further support

Table 7

7 Copenhagen speakers' response
to the 15 stimuli in test 1.

response stimulus	DECLARATIVE	NON-FINAL	INTERROGATIVE	Total
D-1	61	9		70
D-2	52	18		70
D-3	56	14		70
D-4	70			70
D-5	70			70
NF-1		59	11	70
NF-2	3	63	4	70
NF-3	1	58	11	70
NF-4	56	14		70
NF-5	20	50		70
I-1		3	67	70
I-2		4	66	70
I-3		1	69	70
I-4		32	38	70
I-5		34	36	70
Total	389	359	302	1050

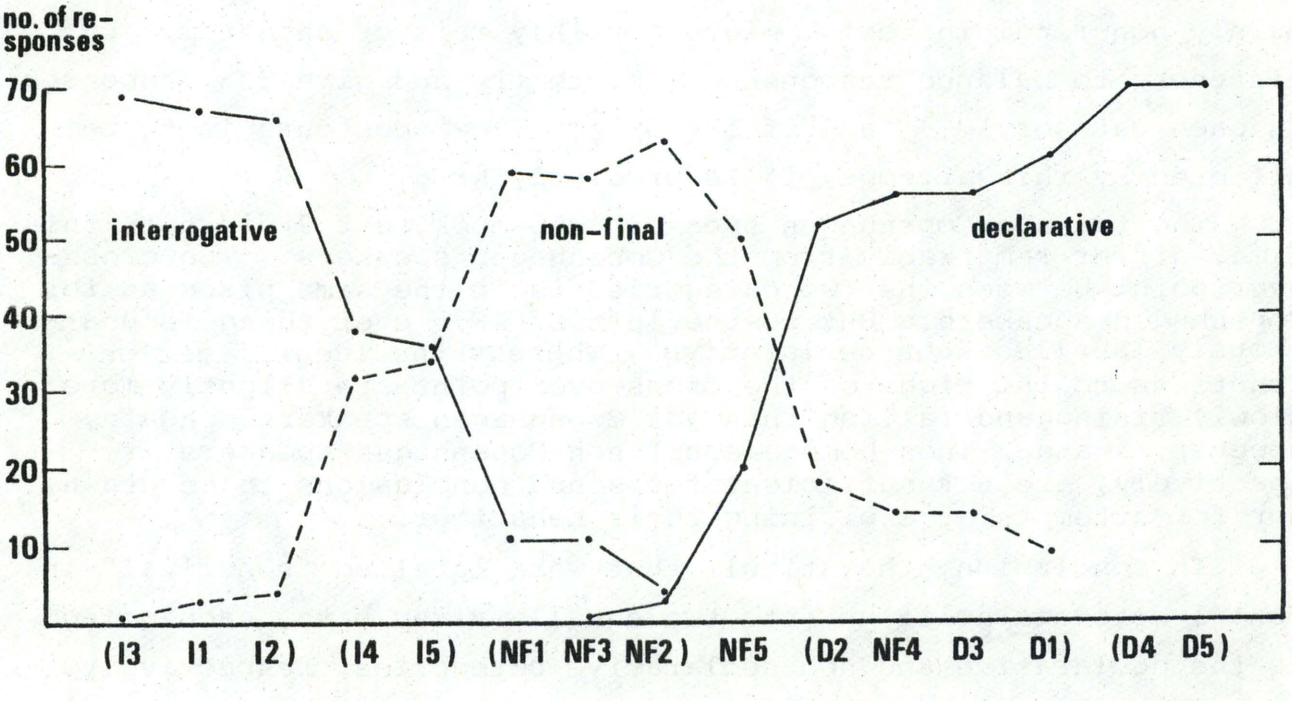


Figure 8

Identification functions for the 15 stimuli in
test 1, as identified by 7 Copenhagen speakers.

Table 8

7 Copenhagen speakers' response
to the 15 stimuli in test 2.

response stimulus	DECLARATIVE	NON-DECLARATIVE	Total
D-1	68	2	70
D-2	66	4	70
D-3	63	7	70
D-4	70		70
D-5	70		70
NF-1		70	70
NF-2	8	62	70
NF-3	3	67	70
NF-4	59	11	70
NF-5	47	23	70
I-1		70	70
I-2		70	70
I-3		70	70
I-4		70	70
I-5	1	69	70
Total	455	595	1050

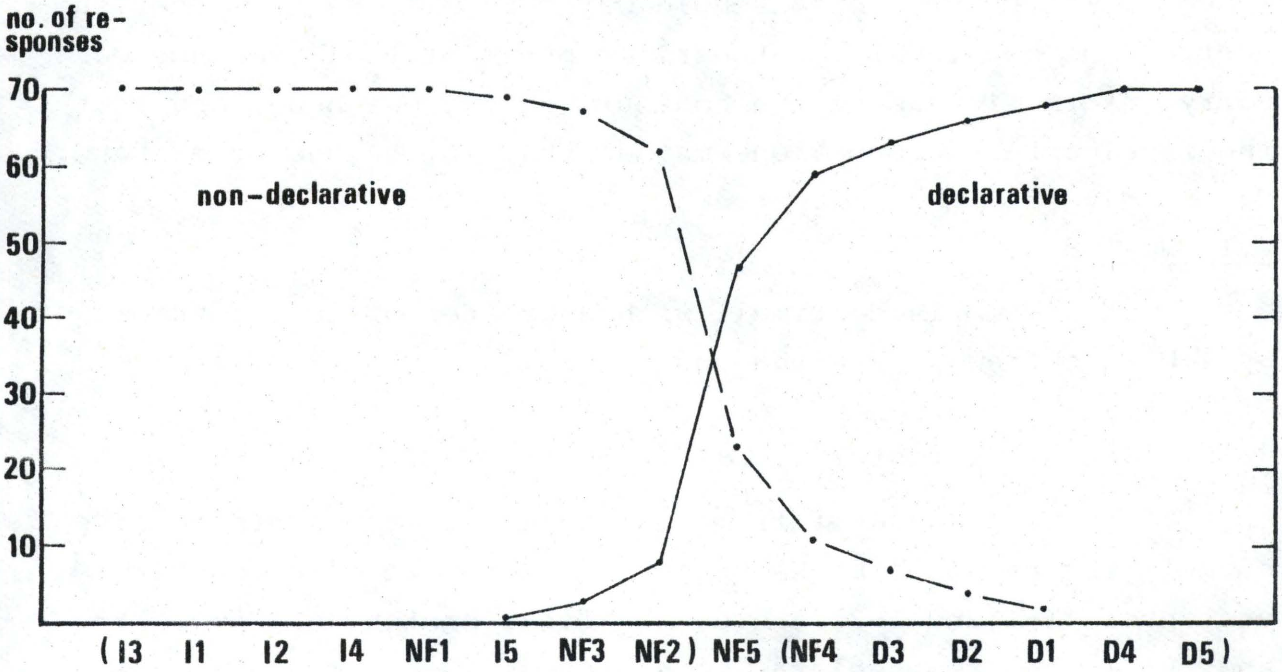


Figure 9

Identification functions for the 15 stimuli in test 2, as identified by 7 Copenhagen speakers.

of this hypothesis would have been if running the same test, but with response labels "interrogative" and "non-interrogative", had given a cross-over point in the identification functions in the neighbourhood of stimuli I-4/NF-3, cf. fig. 3. This ought, as a matter of fact, to be tried.)

4. Tests 3 and 4

Tests 3 and 4 were designed to answer the third question in the introduction: how early or how late in the utterances are the various contours perceptually differentiated? From the results of test 1 (table 5, more precisely), we might expect stimuli to be rather well identified even when the third stress group is missing, given that responses to test 1 correlated fairly highly with [ˈu] and [˳ʌ], - but identification should not be much better than chance, given only the first stress group to judge from. Likewise, we may predict that identification should not be affected much in the absence of the first and second stress groups. If the predictions hold water it will be a further support of the hypothesis that earlier parts of the utterance do carry information about the contour, on the one hand, and that the final stress group alone may do the job, on the other hand.

4.1 Test 3

Test 3 comprised stimuli with a greater or lesser number of syllables cut away from the end.

4.1.1 Stimuli - test 3

7 of the 15 stimuli from test 1 served as parents for the stimuli in test 3. The seven stimuli were two interrogative sentences (I-1 and I-2), three non-final sentences (NF-1, NF-2, and NF-3), and two declarative sentences (D-1 and D-4) - all stimuli that had been well identified by the Copenhagen speakers in test 1, cf. figs. 3 and 8. Three non-finals were included

because this was the category which, although well identified, subjects had been least sure of. D-5 was omitted because it was so very much different from the other declarative sentences, cf. figs. 1 and 7, and in a test with relatively few stimuli no one stimulus ought to stand out clearly from the rest. - The rationale behind choosing, for test 3, stimuli that had been well identified in test 1 was simple: stimuli that had been ambiguous, given the whole of three stress groups to judge from, would be even more so when they were mutilated. However, as Robert Porter has pointed out to me in a personal communication, this may be an erroneous assumption. When stimuli are mutilated, identification criteria must necessarily change, - and originally ambiguous stimuli might turn out to reveal some interesting features. (However, even with only 7 parent stimuli, the test was already rather long - but these 7 might, of course, still have included one or two of the originally ambiguous stimuli.)

By means of a segmentator, the 7 parent stimuli were being relieved step-wise of the last 5 syllables, thus:

cut no. 1: mange busser fra Tif

cut no. 2: mange busser fra

cut no. 3: mange busser

cut no. 4: mange bus

cut no. 5: mange

7 parents times 5 cuts give a total of 35 stimuli. Two test tapes, each containing 2x35 randomized stimuli, plus 5 introductory dummies, i.e. two tapes with 80, differently randomized, stimuli on each were prepared. Stimuli were introduced by numbers (read by the author) with a 3 sec interval for responding. The total duration of one run of each test tape was 8 minutes.

Subjects were instructed exactly as in test 1, cf. section 2.2.

One test tape was presented three times, the other twice, giving a total of 10 responses to each stimulus by each subject.

No subject did more than one run of a test tape per day. All (four) subjects listened over head-phones.

4.1.2 Subjects - test 3

4 of the 10 Copenhagen speakers (all phoneticians) from test 1 served as subjects in test 3. Four is, certainly, a small number, and the results presented below thus may not be conclusive, but they may at least serve as indications of certain trends.

4.1.3 Results - test 3

As in tests 1 and 2, there is no "order-effect" to be detected, and subjects do not differ significantly among themselves.

Fig. 10, top left, and table 9 present the 4 subjects' distribution of responses in test 1 to the 7 parent stimuli employed in test 3. Compared to fig. 3, NF-1 and NF-3 have changed places (still being members of the same group), but the Spearman rank correlation coefficient between the rankings of responses to the 7 stimuli used in test 3, by 10 subjects (fig. 3), and the ranking of responses by four subjects to these stimuli (fig. 10, top left) is .96 - and as for the 7 subjects in test 2, we may conclude that the 4 subjects in test 3 are a representative subset of the 10 subjects employed in test 1.

Fig. 10 and table 10 present the results. (In fig. 10, stimuli have been grouped according to the same criteria as in figs. 3, 4, 8 and 9, cf. sections 2.4 and 3.1.)

4.1.3.1 Cut no. 1: "mange busser fra Tif" - test 3

Fig. 10, top right, shows three clear peaks in the identification functions. Comparison with response distributions to the parent stimuli, fig. 10 top left (and tables 9 and 10), shows the same order of stimuli, but a raised proportion of "declarative"-responses, at the expense of both "interrogative" and "non-final" responses. And a χ^2 test on the distributions of declarative-responses in the parent and cut no. 1 cases shows

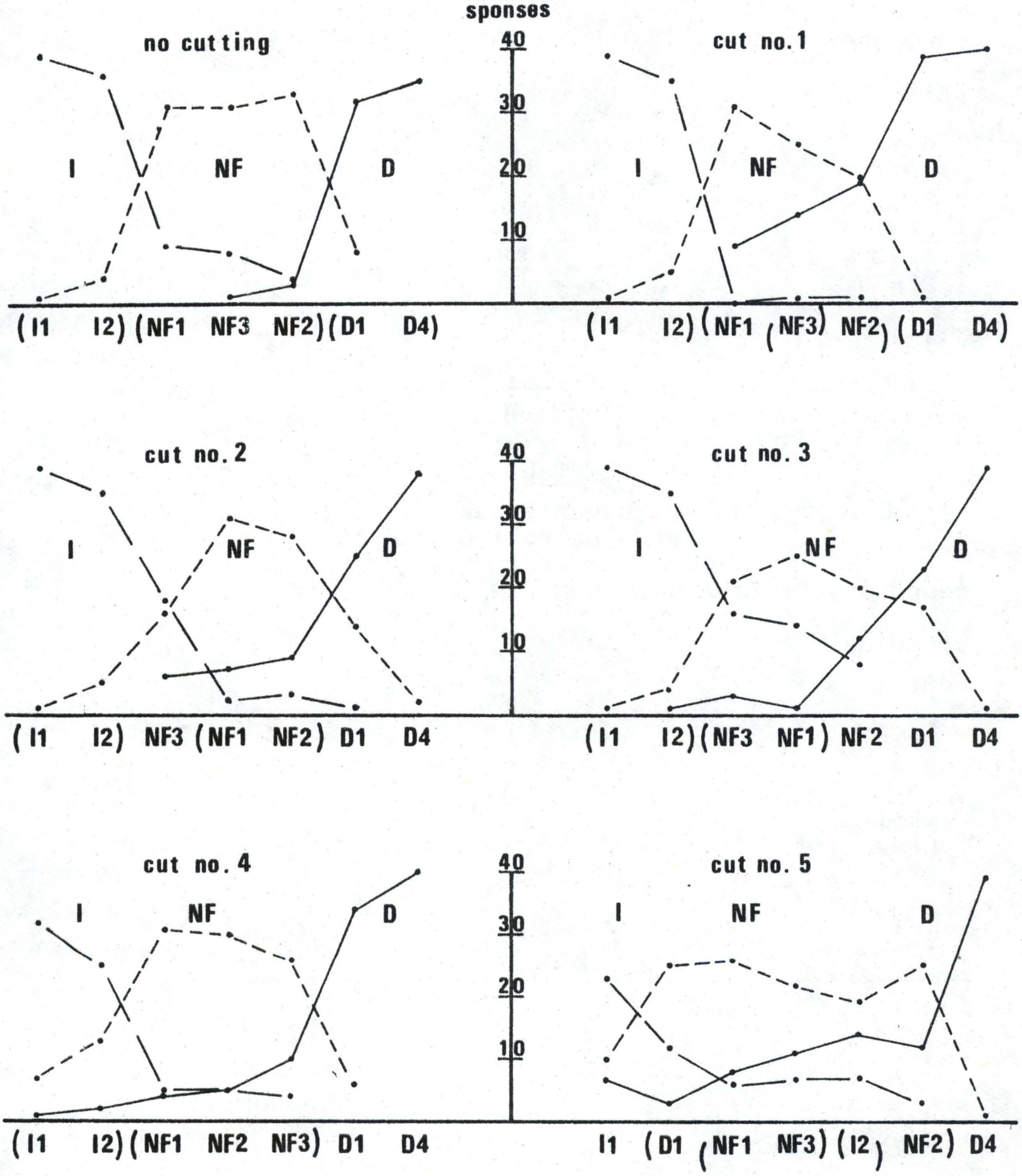


Figure 10

Identification functions for the 7 stimuli in test 3, as identified by 4 Copenhagen speakers.

Table 9

4 Copenhagen speakers' response in test 1 to the 7 stimuli of test 3.

response stimulus	DECLARATIVE	NON-FINAL	INTERROGATIVE	Total
D-1	32	8		40
D-4	40			40
NF-1		31	9	40
NF-2	3	33	4	40
NF-3	1	31	8	40
I-1		1	39	40
I-2		4	36	40
Total	76	108	96	280

Table 10

4 Copenhagen speakers' response to the 7 stimuli of test 3.

response stimulus	DECLARATIVE	NON-FINAL	INTERROGATIVE	Total
cut 1				
D-1	39	1		40
D-4	40			40
NF-1	9	31		40
NF-2	19	20	1	40
NF-3	14	25	1	40
I-1		1	39	40
I-2		5	35	40
Total	121	83	76	280
cut 2				
D-1	25	14	1	40
D-4	38	2		40
NF-1	7	31	2	40
NF-2	9	38	3	40
NF-3	6	16	18	40
I-1		1	39	40
I-2		5	35	40
Total	85	97	98	280

(continued on the next page)

Table 10 (continued
from the preceding page)

4 Copenhagen speakers' response
to the 7 stimuli of test 3.

response stimulus	DECLARATIVE	NON-FINAL	INTERROGATIVE	Total
cut 3				
D-1	23	17		40
D-4	39	1		40
NF-1	1	25	14	40
NF-2	12	20	8	40
NF-3	3	21	16	40
I-1		1	39	40
I-2	1	4	35	40
Total	79	89	112	280
cut 4				
D-1	34	6		40
D-4	40			40
NF-1	4	31	5	40
NF-2	5	30	5	40
NF-3	10	26	4	40
I-1	1	7	32	40
I-2	2	13	25	40
Total	96	113	71	280
cut 5				
D-1	3	25	12	40
D-4	39	1		40
NF-1	8	26	6	40
NF-2	12	25	3	40
NF-3	11	22	7	40
I-1	7	10	23	40
I-2	14	19	7	40
Total	94	128	58	280

that this difference is statistically significant beyond the .001 level.

Losing the last unstressed syllable thus pulls in the direction of declarative-identification. The easy explanation would be that the terminal movement is, after all, essential and, everything else being equal, the lack of a final rise increases the number of declarative-identifications. - However, this is too facile: It is important to note that as soon as one cuts away bits and pieces, the utterance is no longer "natural" and does not necessarily resemble and contain the same amount of information as a corresponding non-mutilated one, in this case one that never had a final unstressed syllable (like: "mange busser fra Brest", for instance). In fact, another part of the material for the original analysis (Thorsen, 1976) contained questions and non-final periods, all variations on: ".. mange timer i statistik.." (the final syllable in "statistik" is stressed). And although the level of this final stressed syllable, "-stik", varies according to sentence type in the same fashion as does "Tif-", there are also minor differences to be observed in the movement of Fo within the stressed [i]: the higher it is on the frequency scale, the less falling it also is (and in some instances, with some subjects, it is even slightly rising, - but this is the exception rather than the rule). This variation in Fo movement is not, generally, found in the stressed vowel of "Tiflis". Thus, if variation in Fo movement is a general feature of stress groups consisting only of one (stressed) syllable, it may be interpreted as a compensation, in the absence of the high rise that a succeeding unstressed syllable would have brought about.

A more appealing account of the discrepancy (which is, after all, not very great) between fig. 10 top right and top left would therefore be that the Fo movement of the vowel in "Tif-" in NF-3 and NF-2 (which are those two stimuli that contribute most to the difference between response distributions) are too steeply falling (cf. fig. 2) to perceptually compensate for the lack of a final unstressed syllable, wherefore NF-3 and NF-2 begin to resemble declarative utterances. (But this presupposes that subjects judged the stimuli in cut no. 1 as if it were a complete utterance.) As it turns out, the pattern almost repeats itself when we get to cuts no. 3 and 4 (mange busser vs. mange bus), and I think the explanation just offered is a fairly likely one. (See, however, a general reservation towards this comparison between parents and offspring at the end of section 4.2.5.)

In table 11 are given Spearman rank correlation coefficients between rankings of responses by the four subjects to the stimuli in test 3 and rankings of a number of the Fo measurements presented in table 4. Since cuts no. 2 and no. 3 had the same response rankings, they are treated together. Note that the correlation coefficients in table 11 are not directly comparable to those in table 5 which are based on rankings of 15 stimuli, whereas table 11 is based on rankings of 7 stimuli. But the coefficients in table 11 are comparable within and across cuts.

In cut no. 1 [ˈi] and [ˌʌ] are seen to correlate highly with responses, with [ˈu] as a close follow-up. [ˌi] is no longer there, and [ˌʌ] seems to have taken over its role, being now the last maximum in the utterance.

4.1.3.2 Cut no. 2: "mange busser fra" - test 3

It is apparent from fig. 10, mid left, that losing the whole of the third stress group is not detrimental to the identification of three categories, - and, further, we are back at a distribution which resembles that of the identification of parent stimuli rather closely, but the order of stimuli has been shifted in one place: NF-3 is moved one step to the left in cut no. 2. However, the identification functions are slightly less steep at the crossover points than in fig. 10, top left.

Table 11 says that [ˈu] and [ˌʌ] are most highly correlated with responses (and [ˈɑ] is closely following them), and they have evidently taken over the role played by [ˈi] and [ˌi] in the parent stimuli, being now the last stress group in the stimuli. (Note that this cut ends in a terminal fall (as does any stress group that contains more than one post-tonic syllable) cf. fig. 2, but this does not diminish the number of interrogative-responses.)

Table 11 (continued on the following page)

Spearman rank correlation coefficients, r_s, between the rankings of responses to the stimuli of test 3, on one hand, and the rankings of certain Fo measurements in these stimuli, cf. table 4. The level of significance (if better than .10) is indicated in the right hand column.

Correlation between response rankings and rankings of				level of significance
Cut no. 1	ˈɑ$_{min}$	:	$r_s = 0$	
	ˈɑ$^{2}_{3}$	:	$r_s = -.21$	-
	ŋ$_{max}$	:	$r_s = .63$	.10
	ˈu	:	$r_s = .93$	.005
	$_{\circ}$ʌ	:	$r_s = .96$	.0005
	$_{\circ}$ɑ	:	$r_s = .75$	.005
	ˈi	:	$r_s = .96$	.0005
	S$_{b}$	:	$r_s = .82$	.025
	ˈɑ$_{min}$−ŋ$_{max}$	:	$r_s = .71$	.025
	ˈɑ$^{2}_{3}$−ŋ$_{max}$	:	$r_s = .79$	.025
	ˈu−$_{\circ}$ʌ	:	$r_s = .24$	-
Cuts no. 2 and 3	ˈɑ$_{min}$	:	$r_s = .07$	-
	ˈɑ$^{2}_{3}$	:	$r_s = -.08$	-
	ŋ$_{max}$	:	$r_s = .63$	.10
	ˈu	:	$r_s = .96$	.0005
	$_{\circ}$ʌ	:	$r_s = .96$	.0005
	$_{\circ}$ɑ (cut 2, only)	:	$r_s = .86$	.01

(continued from the preceding page) <u>Table 11</u>

Correlation between response rankings and rankings of			level of significance
<u>Cuts no. 2 and 3</u> (continued)	ˈɑ$_{min}$-ŋ$_{max}$	: r_s = .65	.01
	ˈɑ$\frac{2}{3}$-ŋ$_{max}$	: r_s = .75	.005
	ˈu-$_{\circ}$ʌ	: r_s = .19	-
<u>Cut no. 4</u>	ˈɑ$_{min}$	: r_s = -.07	-
	ˈɑ$\frac{2}{3}$	: r_s = .75	.005
	ŋ$_{max}$	: r_s = .58	-
	ˈu	: r_s = .86	.01
	ˈɑ$_{min}$-ŋ$_{max}$	: r_s = .71	.05
	ˈɑ$\frac{2}{3}$-ŋ$_{max}$	: r_s = .75	.005
	ˈɑ$_{min}$-ˈu	: r_s = .85	.01
	ˈɑ$\frac{2}{3}$-ˈu	: r_s = .89	.005
<u>Cut no. 5</u>	ˈɑ$_{min}$	: r_s = .21	-
	ˈɑ$\frac{2}{3}$	: r_s = .03	-
	ˈɑ$_{min}$-ŋ$_{max}$	: r_s = .82	.025
	ˈɑ$\frac{2}{3}$-ŋ$_{max}$	: r_s = .79	.025
	ŋ$_{max}$	: r_s = .88	.005

4.1.3.3 Cut no. 3: "mange busser" - test 3

The top in the identification function for the non-final category is rather broad-banded, cf. fig. 10, mid right, but the shapes of the three curves still bear a good resemblance to fig. 10, top left. The order of stimuli is the same as in cut no. 2, i.e. NF-3 is shifted one to the left, compared to fig. 10, top left, and the proportion of interrogative-responses has risen at the expense of the non-final category, compared to the parent stimulus. The correlations between responses and Fo are the same as for cut no. 2.

4.1.3.4 Cut no. 4: "mange bus" - test 3

Even in this case, fig. 10 lower left, three peaks in the identification functions appear, and the order of stimuli is again only disturbed in one place, as compared to fig. 10, top left, i.e. NF-3 has been shifted one to the right. Compared to fig. 10 top left, the proportion of declarative-responses has risen, mainly at the expense of the interrogative-category. Compared to fig. 10, mid right (cut no. 3), both the declarative and non-final categories grow, whereas the interrogative one shrinks. This last pattern was also observed between parent stimuli and cut no. 1, and serves to confirm the explanation offered for this phenomenon in section 4.1.3.1.

In table 11, "cut no. 4", have been included the correlations between responses and the measured rise from [ˈɑ] to [ˈu], which are seen to be as high as (or higher than) that between responses and [ˈu]. (And, as in test 1, $ˈɑ_3^2$ still renders a slightly better correlation than $ˈɑ_{min}$.) This rise is indicative of the slope between the two stressed vowels, and seems to play the role here that S_b did in test 1 (and cut no. 1).

As long as two stressed vowels remain (which together suffice to determine a slope), identification is not very seriously affected - and is not very different from identification of the parent stimuli.

4.1.3.5 Cut no. 5: "mange" - test 3

Now the picture changes completely. The order of stimuli in fig. 10, lower right, is a complete mess, the groups no longer resemble those of fig. 10, top left, and only one stimulus, D-4, received unambiguous responses. The non-final responses dominate. The strongest correlation between responses and fundamental frequency is found in the *level* of $ŋ_{max}$, cf. table 11, but since this point does not correlate too well with the rest of the contour, cf. table 6, there is no real cue in it as to the kind of utterance it occurred in.

4.1.4 Conclusion - test 3

It seems that the predictions made in the introduction to tests 3 and 4 (section 4.) are substantiated by the results. Earlier parts of the utterance *do* carry information about the intonation contour and identification deteriorates but slightly when stimuli are mutilated - until nothing is left but the first stress group, which contains no relevant information. Subjects use all the cues they are given, still, however, with a slight predominance of the level of the last stressed (and unstressed) syllable in the utterance. - And now for the second part of the prediction, that identification should not be affected much by the removal of the first and second stress groups:

4.2 Test 4

Test 4 comprised stimuli with the first and second stress groups cut away from the beginning.

4.2.1 Stimuli - test 4

9 of the 15 stimuli from test 1 served as parents for test 4. They were: I-1, I-2, I-3 - NF-1, NF-2, NF-3 - D-1, D-3, D-4, i.e. the three best identified stimuli from each category (disregarding D-5). For a discussion of this choice, see section 4.1.1.

By means of a segmentator, the 9 parents were deprived of the first and second stress groups, thus:

cut no. 1: busser fra Tiflis

cut no. 2: Tiflis

9 parents times 2 cuts give a total of 18 stimuli. A test tape containing 5x18 randomized stimuli, plus 6 introductory dummies, i.e. 96 stimuli in all, was prepared. Stimuli were again introduced by numbers (read by the author) with a 3 sec interval for responding. The total duration of one run of the test tape was 11 minutes.

Subjects were instructed exactly as in tests 1 and 3, cf. section 2.2.

The test tape was presented twice, with at least one day's interval, giving a total of 10 responses to each stimulus by each subject.

4.2.2 Subjects - test 4

The same subjects were employed as served in test 3. (In fact, they took test 3 and test 4 intermittently.)

4.2.3 Results - test 4

The results of test 4 may be got over with quickly: Fig. 11, top, and table 12 present the 4 subjects' distribution of responses in test 1 to the 9 parent stimuli of test 4. Compared to fig. 3, NF-1 and NF-3 have changed places, but the Spearman rank correlation coefficient between response rankings by 10 subjects in test 1 to the 9 stimuli of test 4 and response rankings by 4 subjects to these stimuli is .98 - and the 4 subjects are still, of course, good representatives of the original 10.

Fig. 11 and table 13 present the results. (In fig. 11, stimuli have been grouped as in the earlier figures, cf. sections 2.4 and 3.1.)

4.2.4 Cut no. 1: "busser fra Tiflis" - test 4

Fig. 11, mid, shows three very clear peaks in the identification functions. In fact, the categories are more clear-cut than those for the parent stimuli. The task must have been easy, probably due to the fact that there were fewer different stimuli than in test 1.

In table 14 are given Spearman rank correlations coefficients between rankings of responses, on the one hand, and rankings of a number of the Fo measurements presented in table 4.[1] [˳i], [ˈi], [ˈu], S_t, [˳ʌ], and S_b all correlate highly with responses, [˳i] slightly more so than the others, and the same conclusion as about test 1, section 2.4.2 may be made: that these parameters are all intercorrelated and no one single parameter can be made much more responsible for identification than the others.

4.2.5 Cut no. 2: "Tiflis" - test 4

Here again, fig. 11, bottom, subjects evidently found no difficulty in recognizing three categories, and in fact, the distribution of responses almost exactly is 1/3 to each category, cf. table 13, "cut 2". Compared to the parent stimuli, cut no. 2 received a higher proportion of declarative-responses and a smaller proportion of interrogative responses, i.e. a certain number of responses were moved one down on the interrogative-declarative scale (leaving the non-final category intact). The same sort of argument that was used in section 4.1.3.1 about the difference between parent stimuli and "mange busser fra Tif" may serve again: The original material for analysis also contained one-word statements and questions, among them "Tiflis." and "Tiflis?". Not only did the level of these words vary, but so did the Fo movement in the stressed syllable, which was falling

---

1) Note, again, that the coefficients in table 14 are not directly comparable to those in table 5, being based on rankings of only 9 items.

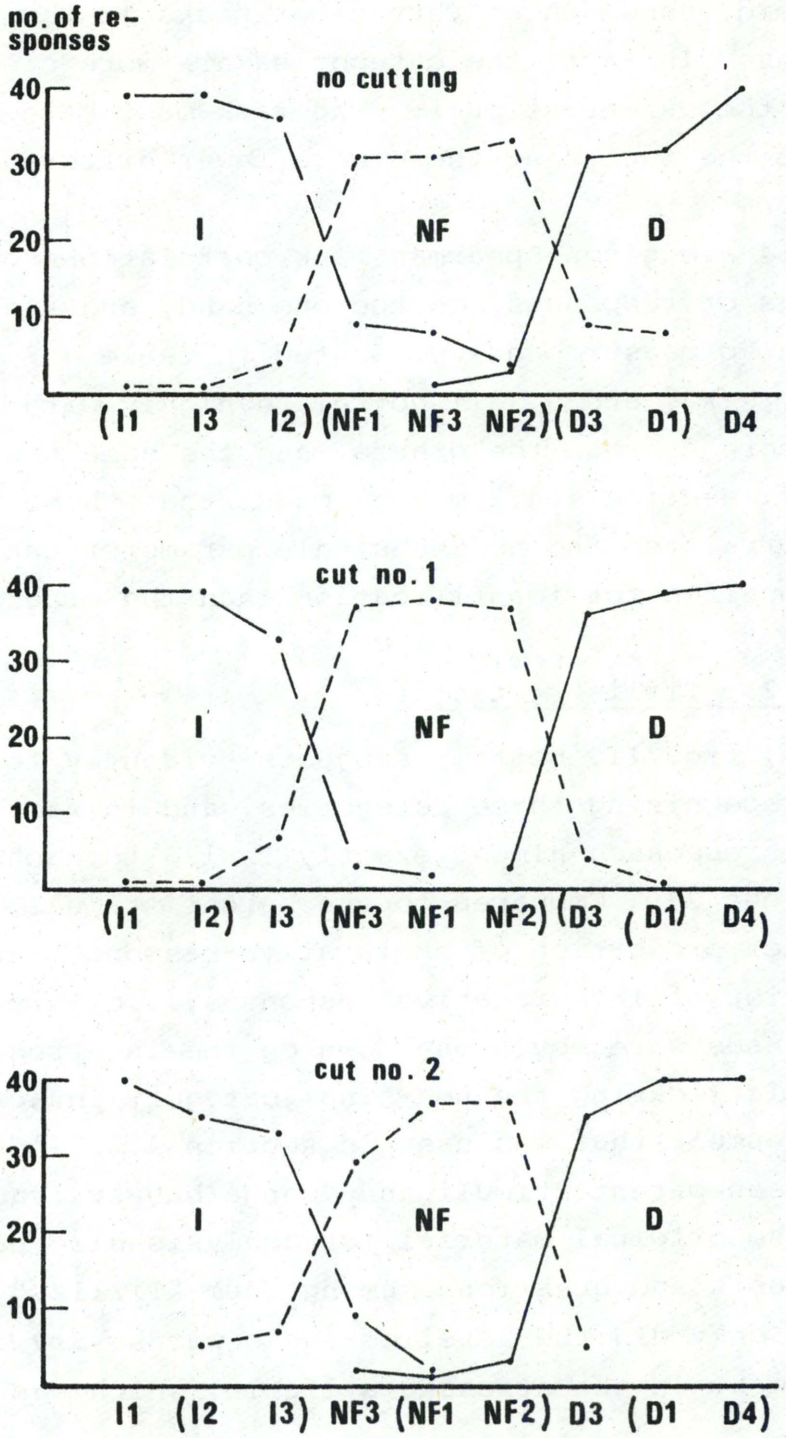


Figure 11

Identification functions for the 9 stimuli in test 4, as identified by 4 Copenhagen speakers.

Table 12

4 Copenhagen speakers' response in test 1 to the 9 stimuli of test 4.

response stimulus	DECLARATIVE	NON-FINAL	INTERROGATIVE	Total
D-1	32	8		40
D-3	31	9		40
D-4	40			40
NF-1		31	9	40
NF-2	3	33	4	40
NF-3	1	31	8	40
I-1		1	39	40
I-2		4	36	40
I-3		1	39	40
Total	107	118	135	360

Table 13

4 Copenhagen speakers' response to the 9 stimuli of test 4.

response stimulus	DECLARATIVE	NON-FINAL	INTERROGATIVE	Total
cut 1				
D-1	39	1		40
D-3	36	4		40
D-4	40			40
NF-1		38	2	40
NF-2	3	37		40
NF-3		37	3	40
I-1		1	39	40
I-2		1	39	40
I-3		7	33	40
Total	118	126	116	360
cut 2				
D-1	40			40
D-3	35	5		40
D-4	40			40
NF-1	1	37	2	40
NF-2	3	37		40
NF-3	2	29	9	40
I-1			40	40
I-2		5	35	40
I-3		7	33	40
Total	121	120	119	360

Table 14

Spearman rank correlation coefficients, r_s, between the rankings of responses to the stimuli of test 4, on one hand, and the rankings of certain Fo measurements in these stimuli, cf. table 4. The level of significance (if better than .10) is indicated in the right hand column.

Correlation between response rankings and rankings of			level of significance
Cut. no. 1	ˈu	: r_s = .95	.0005
	ₒʌ	: r_s = .91	.0005
	ₒɑ	: r_s = .80	.0005
	ˈi	: r_s = .95	.0005
	ₒi	: r_s = .98	.0005
	S_b	: r_s = .91	.0005
	S_t	: r_s = .95	.0005
	ˈu-ₒʌ	: r_s = .33	-
	ˈi-ₒi	: r_s = .66	.05
Cut no. 2	ˈi	: r_s = .97	.0005
	ₒi	: r_s = .99	.0005
	ˈi-ₒi	: r_s = .68	.025

in the statement and rising in the question. This variation was not found in the longer sentences. Apparently, in one word utterances Fo movement in the stressed vowel takes over and compensates for the lack of a contour proper. And this lack of compensation in the mutilated stimuli of test 4's cut no. 2 may account for the discrepancy between fig. 11 top and bottom.

As a matter of fact, it might be argued that since cut no. 2 received an almost perfect distribution of responses (very nearly 1/3 in each category), the lack of a compensatory Fo movement in the stressed vowel of "Tiflis" had no effect at all, and should therefore be deemed unimportant for the identification of intonation contours. (In this connection it is only fair to say that we do not know what the identification functions for the parent stimuli of tests 3 and 4 had looked like, had they been the only stimuli in test 1 - there might have been very little discrepancy to account for between identification functions in test 3, cut 1, and test 4, cut 2, respectively, and the parents.)

The correlation between responses and [ₒi] and [ˈi] are very high indeed, in fact, it almost reaches unity with [ₒi].

4.2.6 Conclusion - test 4

Evidently, and not surprisingly, the end of the utterance alone may suffice for identification of the intonation contour it rides upon. Thus, the predictions made in the introduction to tests 3 and 4 (section 4.) are borne out, and the model (fig. 1) as well as the results of test 1 are confirmed.

5. General conclusion and discussion

In a previous analysis of Advanced Standard Copenhagen Danish intonation, the intonation contour (as defined by the course of the stressed syllables in an utterance) was found to vary according to sentence type, being most steeply falling in declarative sentences, flat in statement-questions, and varying in between these two extremes in other types of questions and in non-final periods.

A certain trade-off between intonation contours and syntax[1] was found: the more syntactic information about the non-declarative function is contained in a sentence, the more declarative-like is its intonation contour. Thus, we may acoustically recognize roughly three types of intonation contours, but each with a certain dispersion and with a certain overlap as well. As it turned out, Copenhagen listeners can identify such utterances solely on the basis of their fundamental frequency course (i.e. given no contextual and no syntactical information): the less falling the intonation contour, the more interrogative utterances are heard, the more steeply falling the contour, the more declarative utterances are heard, and contours in the middle of the continuum are identified as being non-final. (Note that identification is not always in agreement with the speaker's intention, however.) Judging from the identification functions, perception seems to be categorical: three clear peaks are found, one for each category of interrogative, non-final, and declarative, respectively. (Non-Copenhagen speakers, on the other hand, did not, in these 15 stimuli by a Copenhagen speaker, identify three categories, but any conclusions as to the reason why must await an acoustical analysis of intonation contours in various dialects.)

Close inspection of the course of fundamental frequency in the stimuli revealed rather high correlations between response distributions, on one hand, and the course of Fo, on the other, with a slight predominance of the <u>level</u> of the final stress group in the utterance (whereas the final rise from stressed to unstressed syllable gave no such high correlation to responses). However, as all points of measurements during the second and third stress groups were highly correlated/concordant (ie., if one point is high, so are all the others), and since partial rank correlations could not determine one single parameter as being more independent of the others than any other parameter, we may conclude that identification is determined not just by the level of the final part of Fo, but by earlier parts as well. That this is actually the case was demonstrated by the identification

---

1) See the footnote on page 2o.

of stimuli where a greater or lesser number of syllables had been cut away from the end. Only when just the first stress group remains is identification seriously affected. This is not surprising, since intonation contours are least different in this first part of the utterance and since, further, the course of Fo during the first stress group showed a rather low correlation with later parts. - On the other hand, utterances are identified very well indeed when only the last stress group is presented, which is also to be expected, since this is the stretch where different contours are most widely differentiated acoustically. - All in all, we may conclude with Bolinger (1951, p. 206) that "... the basic entity of intonation is a pattern, not a pattern in the relatively abstract sense of grammatical recurrences, but in the fundamental, down-to-earth sense of a continuous line that can be traced on a piece of paper."

As for the number and kinds of categories established: Three perceptual categories of intonation contours seem to be the rule, rather than the exception: Gårding and Abramson (1965) established three in American English, "neutral statement", "yes or no question", and "counting in a series" (plus two that I would be inclined to call attitudinal: "anger", and "delighted surprise"). Hadding-Koch and Studdert-Kennedy (1963, 1965, 1974) found for Swedish and American subjects three categories, "statement", "question", and "talking-to-self" (the linguistic status of this last category is not transparent and was one whose meaning they had some difficulty in explaining to their subjects - I suspect that labelling it "non-final" or "continuation" might have made the task easier for the subjects and rendered the same results). Isačenko and Schädlich (1963) found three categories in German, "nicht-Frage", "progredient" or "weiterweisend", and "Frage". Johansson (1978) established three categories in Swedish and labelled them "declarative", "continuation", and "interrogative". Uldall (1961) found two categories in English, "question" and "statement" - but some stimuli were highly ambiguous and one is left to wonder whether the introduction of a third category might not have resolved this ambiguity.

The acoustical manifestation of the three categories varies from one language to another (although the similarities may be greater than the differences) and so do the perceptual cues to their identification (but even the perceptual processing of intonation contours may be less divergent than different authors on the subject would have us believe). Thus, any variation aside, there is little doubt that in a number of languages (or dialects) subjects are able to identify three types of intonation contours. Whether they all represent a linguistic reality is another question: it is characteristic that most of the perceptual experiments have had subjects work in a forced choice situation (Uldall, 1961, is an exception). Sven Öhman has objected (in a personal communication) that this is a doubtful procedure, and that the suspect third category, "non-final" or "continuation", may be a waste-paper basket into which everything is thrown that does not fit either the declarative or interrogative categories. - Had all the investigations been conducted with synthetic stimuli (and without prior acoustic analyses), one would have had to give the non-final category the benefit of a doubt. But the present investigation, and partly also those of Gårding and Abramson (1965) and Johansson (1978), has been carried out with real speech stimuli, chosen from a set of statements, questions,and non-final periods whose intonation contours varied accordingly, and subjects were thus not confronted with stimuli that are unlikely to occur in real speech, on the contrary: what they heard and identified was something that speakers actually produce, i.e. a linguistic reality. Therefore, I think, there are good grounds for accepting declarative, non-final, and interrogative intonation contours on an equal footing. (We may even have to continue the search, cf. Mikoš (1976) who found three types of interrogative contours, "yes-no questions", "wh-questions", and "alternative questions", plus a "declarative".)

Postscriptum

Although I do not, of course, wish to throw doubt upon the results reported above to a degree that would make them inconclusive and invalid, I do think it necessary to add a word of reservation: The four tests were based on only 15 utterances by a single speaker. This speaker, SH, is probably neither more, nor less, typical of ASC Danish than any of the other three speakers employed in the previous analysis, and the 15 stimuli were chosen among SH's recordings mainly for "technical" reasons, cf. section 2.1.1. We cannot know for sure (only make qualified guesses about) how the listeners would have reacted to other stimuli by SH or to stimuli by other ASC speakers. (And we do not know whether the non-ASC subjects might not have identified three categories of intonation contours in utterances by speakers of their various dialects.) Likewise, we do not know how great, or how small, a relative weight sentence intonation contours carry in speech perception in real-life communication where context and syntactic information are not suppressed the way they were in the present experiments.

References

Bolinger, D.L. 1951:	"Intonation: levels versus configurations", *Word* 7, p. 199-210
Bolinger, D.L. 1958:	"A theory of pitch accent in English", *Word* 14, p. 109-149
Bolinger, D.L. 1964:	"Intonation as a universal", *Proc.* Ling. IX, 833-48
Bolinger, D.L. 1970:	"Relative height" in: P. Léon, G. Faure, and A. Rigault (eds.) *Prosodic Feature Analysis*, Paris, p. 109-125
Breckenridge, J. and M.Y. Liberman 1978:	"The declination effect in perception", Bell Laboratories (manus.)
Bruce, G. 1977:	*Swedish word accents in sentence perspective*, Lund
Bruce, G. and E. Gårding 1978:	"A prosodic typology for Swedish dialects", paper read at the Symposium on the prosody of Nordic languages, Lund 14-17 June (will appear in *Travaux de l'Institut de Linguistique de Lund* 13)

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

Collier, R. and J. t'Hart 1975: "The role of intonation in speech perception" in: Cohen and Nooteboom (eds.) *Structure and process in speech perception*, Springer Verlag

Daneš, F. 1960: "Sentence intonation from a functional point of view", *Word* 16, 34-54

Delattre, P., E. Poenack and C. Olsen 1965: "Some characteristics of German intonation for the expression of continuation and finality", *Phonetica* 13, 134-161

von Essen, O. 1956: "Hochdeutsche Satzmelodie", *Zs.f.Ph.* IX, 1, 75-85

Gårding, E. and A.S. Abramson 1965: "A study of the perception of some American English intonation contours", *Studia Linguistica* 19, 61-79

Hadding-Koch, K. 1961: *Acoustic-phonetic studies in the intonation of southern Swedish*, Gleerup, Lund

Hadding-Koch, K. and M. Studdert-Kennedy 1963: "A study of semantic and psychophysical test responses to controlled variations in fundamental frequency", *Studia Linguistica* 17, 65-76

Hadding-Koch, K. and M. Studdert-Kennedy 1964: "An experimental study of some intonation contours", *Phonetica* 11, 175-185

Hadding-Koch, K. and M. Studdert-Kennedy 1965: "Intonation contours evaluated by American and Swedish test subjects", *Proc.Phon.* V, 326-331

Hadding-Koch, K. and M. Studdert-Kennedy 1974: "Are you asking me, telling me, or talking to yourself?", *J.Ph.* 2, 1, 7-14

t'Hart, J. 1966: "Perceptual analysis of Dutch intonation features", *IPO Annual Progress Report* 1, 47-51

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

Isačenko, A.V. 1965: "Deutsche Satzintonation als Konfigurationen zweier diskreter Tonebenen (mit Demonstration von Tonbändern)", *Proc.Phon.* V, 371-373

Isačenko, A.V. and H.-J. Schädlich 1963: "Untersuchungen über die deutsche Satzintonation", *Studia Grammatica* 7, 7-67

Johansson, I. 1978: *Funktionella aspekter på satsintonationen i svenska* (Functional aspects of sentence intonation in Swedish), Acta Universitas Umensis, Umeå Studies in the Humanities 16

Lehiste, I. and G.E. Peterson 1961: "Some basic considerations in the analysis of intonation", *JASA* 33, p. 419-425

Mikoš, M.J. 1976: "Intonation of questions in Polish", *J.Ph*. 4, 247-253

Reinholt Petersen, N. 1976: "Intrinsic fundamental frequency of Danish vowels", *ARIPUC* 10, p. 1-27

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

Rossi, M. 1978: "La perception des glissandos descendants dans les contours prosodiques", *Phonetica* 35, p. 11-40

Studdert-Kennedy, M. and K. Hadding 1972: "Further experimental studies of fundamental frequency contours", *Proc.Phon.* 7, 1024-1031

Studdert-Kennedy, M. and K. Hadding 1973: "Auditory and linguistic processes in the perception of intonation contours", *L&S* 16, 4, 293-313

Thorsen, N. 1976: "An acoustical investigation of Danish intonation: preliminary results", *ARIPUC* 10, p. 85-147 (will also appear, in slightly revised form, in *J.Ph*. 6, 1978)

Thorsen, N. 1977: "On the interpretation of raw fundamental frequency tracings", *ARIPUC* 11, p. 1-11

Thorsen, N. forthcoming: "Interpreting raw fundamental frequency tracings of Danish" (an enlarged version of Thorsen, 1977, to appear in *Phonetica*)

Uldall, E. 1960: "Attitudinal meanings conveyed by intonation contours", *L&S* 3, 223-234

Uldall, E. 1961: "Ambiguity: question or statement? or 'Are you asking me or telling me?'", *Proc.Phon.* 4, 779-783

(ARIPUC 12, 1978)

ON THE UNIVERSAL CHARACTER OF PHONETIC SYMBOLISM WITH SPECIAL REFERENCE TO VOWELS[1,2]

Eli Fischer-Jørgensen

1. Introduction

In various articles (1964, 1967, 1970, 1971) Bertil Malmberg has mentioned the fact that "l'arbitraire du signe" is never absolute in a language; there will always be layers of the vocabulary, representing a more primitive stage of the language, in which the relation between sound and meaning is partly motivated, i.e. onomatopoetic and expressive words, particularly often found in child language and in poetry. Partly inspired by Roman Jakobson, Malmberg calls attention to the need for a systematic investigation of this vocabulary in various languages, supplemented by psycho-linguistic tests in order to find out what is universal in the expressive function of these partly motivated signs. The present article is meant to give a contribution (though a very modest one) to the solution of this problem.

As far as purely onomatopoetic words are concerned, it is evident that they have a universal basis, although their form is also determined by the phonological system of the language in question. It is more problematic whether expressive words, which do not imitate sounds, have a universal basis.

---

[1] Published in <u>Sign and Sound, Studies presented to Bertil Malmberg on the occasion of his sixty-fifth birthday</u> (eds. B. Sigurd and J. Svartvik), <u>Studia Linguistica</u> 32/1-2, 1978.

[2] I am grateful to Niels Davidsen-Nielsen for improvements of my English style.

2. Westermann's description of phonetic symbolism in West African languages

In 1937 D. Westermann published an interesting paper on sound and meaning in some West African languages. In these languages (Ewe, Twi, Gã, Guang, Nupe, and Temne, which are all genetically related), expressive words (which Westermann calls "Lautbilder") play a much more important part than in European languages. The words may be old and common to related languages, but new ones can be formed any moment and are immediately understood by the listeners because they are created according to definite principles. Almost all phonetic features may have an expressive value: reduplication, vowel length, vowel quality, intensity and manner of the articulation of consonants, and pitch.

According to their expressive value the vowels of these languages can be divided into three categories: (1) e ε a (2) u o ɔ (3) i. In Westermann and Ward 1933 the a of West African languages is said to be most often close to Italian a, but in the cardinal vowel diagram of Ewe (the main source of Westermann's examples) it is placed very close to Cardinal vowel No. 4. This means that a in these languages may be a front or a central vowel. In Westermann's article on sound and meaning it is said that in the cases where a language has two a-sounds, the back a will belong to the u o ɔ category. The two categories e ε a and u o ɔ are thus generally front (and central) unrounded vs. back rounded. They are used to express opposite meanings, which are characterized very generally as "intensive" versus "extensive", but which can be divided into the following four groups (Westermann 1937, p. 166):

	Runder dunkler Vokal	Flacher heller Vokal
a)	massig, dick, plump, gedunsen, bauchig, hohl, rund, tief	dünn, lang gestreckt, gerade, flach, ausgebreitet, offen
b)	schlammig, schleimig, lose, weich, schwach	hart, fest, steif, stark

c)	schwer, langsam, unbeholfen, lässig	leicht, rasch, behende, gewandt, lebhaft
d)	dunkel, trübe, dumpf	hell, glänzend, heller Klang.

Generally, in expressive words, e ɛ a are combined with high tone and u o ɔ with low tone, but it is not always the case, and besides reinforcing the expressive value of vowel quality tone differences have other expressive values; in particular low tone is used to indicate big and lax things, and high tone to indicate small and tense things, or a high degree of something. And the tone may be shifted to give shades of meaning so that, for instance, a word which indicates things that are normally big will normally have a low tone, but if in certain situations the thing indicated is small, the tone is changed to high.

It is a characteristic feature of these languages that the vowel i does not come within the category of front unrounded vowels, but has its own specific expressive values, indicating something narrow, tight, squeezed or very dark.

Westermann gives ample documentation from several languages of the expressive functions described.

3. Testing Westermann's vowel categories on Danish subjects

In order to test whether these expressive values of the vowels in West African languages are found at least as latent possibilities in quite unrelated languages as well, I undertook a very simple experiment. Two groups of Danish students (43 in 1951, and 56 in 1963), who had just started a course in elementary phonetics, were asked to match the two vowel categories e ɛ a and u o ɔ with the members of a selected set of adjective pairs taken from Westermann's description. The pairs of adjectives were for (a) thin/thick (Danish tynd/tyk), and flat/round (flad/rund); for (b) hard/soft (hård/blød), tight/loose (fast/løs), and weak/strong (svag/stærk); for (c) light/heavy (let/tung), quick/slow (hurtig/langsom), and agile/clumsy (behændig/klodset);

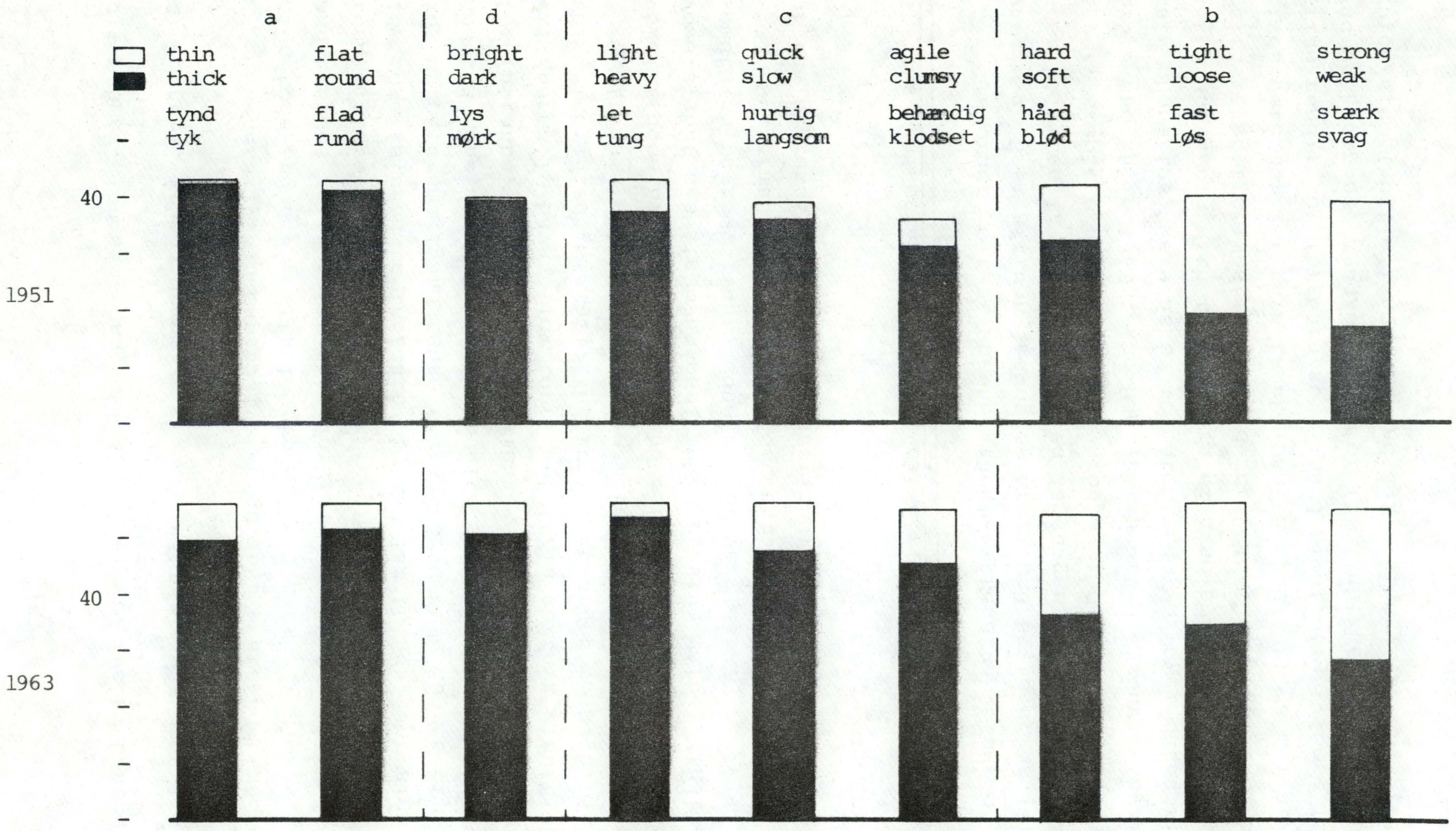


Figure 1

Two vowel categories e ɛ æ ☐ and u o ɔ ■ matched with pairs of adjectives by two groups of Danish students.

and for (d) bright/dark (lys/mørk).

The two vowel categories were written on the blackboard in phonetic transcription, and the adjective pairs were also listed on the blackboard with the two members in random order, i.e. in some pairs the member that corresponded to West African e ɛ a was first, in some last. The students were asked to match adjectives and vowels so that the members of each adjective pair were distributed on the two vowel categories. (It should be noted that Danish a will be pronounced as a front vowel [æ] in isolation, whereas it is a relatively back vowel before and after r and, when it is short, also before velars and (for most speakers) before labials.)

This was a rather crude test since the vowels were not tested separately, and there was only a forced choice answer, so that if, e.g., "thick" was felt to match with u o ɔ, then "thin" had to be matched with e ɛ a. The answers can therefore only prove that there is a relative difference between e ɛ a and u o ɔ in this respect.

It is striking, however, how close the agreement is with the way the vowels are used in the West African languages, and there is also a very close agreement between the two groups of students. For the six pairs belonging to Westermann's groups (a) (c) and (d), 79 to 98% of the Danish students were in agreement with the West Africans. (In a few cases the students did not make any decision (the maximum is 7, for agile/clumsy in the 1951-experiment). In the calculation of percentages these few cases have been distributed evenly on the two categories.) All these answers are significantly different from chance at the 0.01% level. For group (b), however, (comprising hard/soft, tight/loose and weak/strong) only hard/soft shows a positive agreement with the use in the African languages (76 and 67%, significant at the 0.1 and 1% level), whereas the two other pairs did not give any significant difference. The answers are displayed graphically in figure 1 , where they are ordered from left to right according to the degree of agreement (a,d,c,b).

It might be assumed that the vowels of the Danish adjectives have influenced the answers. In the examples flad/rund, let/tung and behændig/klodset the vowels support the answers given, but not in the other cases which have just as high a percentage, so this cannot be the reason. Moreover, the results are supported by other experiments on Danish vowels (E. Fischer-Jørgensen 1967). That Danish *e ɛ æ* are felt to be brighter than *u o ɔ* has been proved by various experiments with a large number of subjects, who have been asked to group vowels as bright or dark, or to evaluate their brightness on a seven point scale, or to match vowels with a grey scale or with colour charts. As for the other oppositions, experiments with seven point scales have shown *e ɛ æ* to be felt as thicker than *u o ɔ*, and experiments with a smaller number of vowels using the same method have shown *ɛ* to be harder, lighter, and more flat than *ɔ* and *u*. But for tight/loose the order did not correspond to Westermann's two categories (but depended on the degree of openness of the vowel), and just as in the first mentioned experiment, there was no clear distribution for strong/weak (but Westermann also had very few examples for this latter opposition). These other experiments thus support the results of the *e ɛ a* /*u o ɔ* test in every respect. The only two pairs which have not been tested by other methods are agile/clumsy and quick/slow. (It should be noted that in these latter tests the point of view has been shifted. Westermann stated that e.g. the vowel *u* was used to express something dark, but in these tests the question was whether the vowel *u* could be called a dark sound. However, this difference is not crucial to the issue, since a dark sound must be able to symbolize something dark.)

As for the vowel *i*, it was not included in the test undertaken to investigate the agreement with the West African phonetic symbolism, but it was included in all the other tests, and in all cases it behaved like the other front vowels, and *i* even had an extreme position, being the most thin, flat, bright, light, and hard of all the vowels. As for the specific expressive values mentioned by Westermann, there is agreement in so far as *i* was considered to be more small, tight, tense, and compact than the other vowels (and of these qualities only small/big showed a correlation with front/back), but *i* was never considered to be dark, on the contrary it was the brightest of all vowels.

4. Evidence from other languages

The almost complete agreement between the value of vowels in expressive West African words and the results of tests with Danish subjects clearly shows that these values are not dependent on specific languages or cultures. Moreover, experiments with speakers of various other languages, reported in the literature, support the hypothesis of almost universal values.

The connection of bright-dark with front and back vowels, respectively, has been mentioned very often. For instance, subjects who have associations between vowels and colours generally tend to combine front vowels with bright colours and back vowels with dark colours, and front vowels are also often combined with specifically bright hues (particularly yellow), and back vowels with specifically dark hues (particularly blue). Cases of audition colorée for single persons showing this tendency are mentioned by Roman Jakobson 1941, p. 66-67 (German and Czech), G. Reichard, R. Jakobson and E. Werth 1949 (Serbian/Hungarian), D.U. Masson 1951 (English). Collections of more extensive material comprising many persons and showing the same tendency are reported by Flournoy 1893 and A. Argelander 1927 (German, French, and English). Normally the vowel *i* goes with the front vowels, but there are also examples of *i* being felt as black. This is e.g. the case for some French subjects and a great number of English subjects.

Newman (1933) used comparisons of nonsense words and also found that front vowels were considered to indicate brightness and back vowels darkness. Similarly, 97% of Fónagy's subjects (1963) considered *i* to be brighter than *u*, cf. also Peterfalvi 1965.

As for the other oppositions it is well known that Sapir (1929) undertook some experiments on sound symbolism and found that nonsense words with relatively open vowels were considered to indicate big things as compared with words with relatively close vowels (cf. also Tarte 1974). Miron (cit. Chastaing 1964) asked American and Japanese subjects to place CVC-syllables on 7-point scales and found that syllables with front vowels were considered smaller, thinner, lighter, and quicker than syllables with back vowels. The subjects used by Fónagy (1963) found i to be smaller, quicker, lighter, thinner, and (partly) harder than u, whereas the answers to "stronger?" were less clear. In most cases front unrounded vowels have been compared with back rounded vowels, but it has generally been assumed that the front-back difference was the important one. Westermann, however, believed that the unrounded-rounded difference was the decisive one for the symbolism. The above mentioned experiments with Danish vowels, comprising rounded front vowels, showed that both rounding and backness contribute to make vowels seem darker, thicker, heavier, softer, and bigger.

The results of these different investigations are in very good agreement with each other. Taylor's assumption (1963) that people's responses in tests should be due to accidental frequencies of some sounds in semantically related words of their mother tongue can thus be refuted.

5. Expressive use of sounds in real languages

The finding of universal symbolic functions by means of tests does not mean that these potential symbolic values are utilized to any great extent in the normal vocabulary of all languages. The West African languages certainly constitute an extreme case. But studies of other languages have also revealed a certain number of expressive words. Since it is difficult to judge about expressiveness in one's mother tongue, comparative or statistical methods are preferable.

Otto Jespersen (1922) found that the vowel i was used in words for "little", and in words for small things, children and young animals, in a large number of languages much more frequently than should be expected (cf. also Thorndike 1946). Roman Jakobson (1960) interpreted the findings of Murdock concerning phonetic similarities between the words for mother and father in many unrelated languages. Chastaing (1964) studied dictionaries and frequency lists and found that in French 90% of the words for "small" contain front vowels (particularly i and y), whereas 80% of the words for "big" contain back vowels. Similarly, most of the French words for "dark" have back vowels, and most of the English words for "bright" have front vowels (but not vice versa). Thus, the symbolic values of vowels are used to a certain extent also in European languages.

It has also been attempted to test the universal character of expressive sounds in normal words by translating pairs of words with opposite meanings into a foreign language and ask subjects not knowing this language to guess what means what. Ertel (1969) gives a survey of a large number of experiments of this type, carried out by various psychologists, most of them with a positive result. He has himself undertaken such experiments with German subjects using 25 different foreign languages. In contradistinction to most other investigators he does not use single German word pairs but 8 different adjective pairs, covering the same semantic area, for comparison with each foreign word pair. He gets mostly positive results, but they are only significant for the semantic area "activity" ("Erregung"). Even with the precautions taken by Ertel this method seems rather dubious, because so much depends on the words chosen. In one case (Maltzman et al. 1956), English subjects were also asked to compare word pairs in two unknown languages (Japanese and Serbo-Croatian). This gave a negative result, which was not astonishing. Since they did not know the meaning, they could not base their guesses on relations between sound and meaning. And since the words were at best partly expressive, they could not know which sounds to compare.

Grammont, whose "Traité de phonétique" 1933 contains a long chapter on "phonétique impressive", takes his starting point in the onomatopoetic function of sounds, arguing that when e.g. i is used to imitate sounds that can be described as light and fine, it can also be used to express other light and fine objects. But this does not permit the conclusion that i is really used in this way.

6. Explanations of phonetic symbolism

Once the universal character of the expressive values of sounds has been documented, the next task will be to explain these values. This problem has been discussed in various books and papers, by Flournoy (1893), Mahling (1926), Argelander (1927), Wellek (1931), Fónagy et al. (1963), Peterfalvi (1965)[1], Ertel (1969), Karwoski and others. It will only be treated briefly here.

Some authors prefer articulatory explanations of phonetic symbolism, e.g. Westermann and Fónagy. In some cases it is quite convincing, e.g. when the consonant m is used in words for "press , "squeeze", "closed", etc. in West African languages. The same is true of the symbolic value of i in these languages ("squeeze", "press", "narrow"), and it is probable that it is this articulatory starting point that leads to the symbolic value "dark". As mentioned above, a number of English subjects also characterized the vowel i as black. This was a written inquiry, and it has been suggested that they thought of the name of the letter [ai] (but [ai] is not dark in Newman's experiment 1933), or that they have seen i as a black line, and o (which was called black) as a white spot with a black ring around it. This is not completely improbable, but in any case the symbolic value of i

---

[1] The later and more extensive work by Peterfalvi (1970) has not been accessible to me until the moment when this paper went into press.

in the African languages cannot be explained in this way, so perhaps the English responses have also been based on articulation, cp. that Danish subjects considered i, y and u to be the most tight and compact of all vowels (E. Fischer-Jørgensen 1967). And perhaps Newman (1933) is right in assuming that the expressive value "large" of open vowels has something to do with their big mouth opening (but why, then, is u bigger than i?).

However, in many cases it is more probable that it is the auditory impression which is at the base of the symbolic value, e.g. bright-dark. Fónagy's articulatory explanation that for i the tongue is raised towards the bright world outside, whereas for u it is moved back towards the dark pharynx can hardly be taken completely seriously. The explanation of front vowels as bright and back vowels as dark cannot be separated from the explanation of the fact that in general high tones, or sounds dominated by high frequencies, are considered to be bright, and low tones, or sounds dominated by low frequencies, are considered to be dark as generally known and also shown in many experiments.

Now, this is only one step in the explanation, for why is "high" combined with "bright", and why are frequencies called "high" and "low"? Fónagy thinks it is because we stretch our neck (and raise our larynx) when singing high tones. There may be something to it, but it can hardly be the whole explanation. Fónagy has, however, a good argument for his articulatory point of view: Deaf children who have learned to articulate give the same answers as normal children: i is smaller, thinner, quicker, brighter, more cheerful and kinder than u (cf. also Ertel 1969, p. 73ff). But the same is true of Fónagy's blind subjects. How can they consider i bright and u dark?

Part of the explanation is probably that there are interrelations between the qualities designated by the adjectives. The qualities "high, bright, thin, light, small" as against "low, dark, thick, heavy, big" generally go together. If deaf persons feel i as small and thin, it will also be bright. This means that phonetic symbolism cannot be explained independently of the general phenomenon of synesthesia.

Various explanations have been given of this phenomenon. Karwoski, Odbert and Osgood 1942 emphasize the importance of a general cultural, linguistic tradition with conventionalized synesthetic relations, e.g. the normal application of adjectives to different senses (a warm room, a warm colour, warm greetings, etc.), and the metaphoric use in poetry. This is certainly important, and may influence a subject's response in a concrete test situation. And it would be of great interest to make a comparative study of the use of adjectives in different languages from this point of view (Ertel mentions such a study by S.E. Asch, which has not been accessible to me). - But what, then, is the origin of such a tradition, and why are the similarities between synesthesia in quite unrelated cultures so striking? And why do small children have more synesthesia than grown-up people? (cf. Werner 1948).

Some authors have pointed to general experiences of the outer world: Big animals have lower voices than small ones, big things are heavier than small things, the sky is high and bright, whereas it is dark down in the basement, etc. Therefore small, light, high, and bright are related and opposed to big, heavy, low, and dark (e.g. Argelander 1927, Fónagy 1963). This is certainly of importance. But it can hardly explain everything.

Others have assumed that there are some deeper interrelations between our senses, on a more or less psychological or physiological level. Flournoy talks about similarities of the emotional connotations for different senses, Karwoski et al. (1942) find more abstract parallelisms in polarities and graduation (cf. also Argelander 1927 and Peterfalvi 1965). This would explain why it is possible to equate e.g. a given tone with a given shade of grey and a given smell (cf. the experiments of Hornbostel 1931) within a given reference frame, once the orientation of the dimensions has been equated (e.g. low = dark); but this latter point is important. Ertel (1969) attempts to reduce phonetic symbolism to the three fundamental dimensions set up by Osgood for the study of connotational meaning: Activity, potency, and valuation. But his attempt to apply these dimensions to the symbolic use of vowels is not very successful. There also seem to be some more concrete physiological interrelations, cp. that the impression of colours can be influenced by the simultaneous perception of sound and vice versa (e.g. Mahling 1926), and that the calming and exciting effects of colours have been demonstrated also in animal experiments (Argelander 1927). There is still much to be done in this field of study.

References

Argelander, A. 1927: *Das Farbenhören und der synästhetische Faktor der Wahrnehmung*, Jena

Chastaing, M. 1964: "Nouvelles recherches sur le symbolisme des voyelles", *J. de psych*. p. 75-88

Ertel, S. 1969: *Psychophonetik*. *Untersuchungen über Lautsymbolik und Motivation*, Göttingen

Fischer-Jørgensen, Eli 1967: "Perceptual dimensions of vowels", *To Honor Roman Jakobson*, p. 657-71

Flournoy, Th. 1893: *Les phénomènes de synopsie*, Paris-Genève

Fónagy, I. 1963: *Die Methaphern in der Phonetik*, The Hague: Mouton

Grammont, M. 1933: *Traité de phonétique*, Paris

Hornbostel, 1931: "Ueber Geruchshelligkeit", *Pflügers Archiv für die gesamte Physiologie*, 227, p. 517-38

Jespersen, O. 1922: "Symbolic value of the vowel I", *Philologica* I. Reprinted in *Linguistica, selected papers in English, French and German*, 1933, p. 283-333

Jakobson, R. 1941: *Kindersprache, Aphasie und allgemeine Lautgesetze*, Uppsala: Almqvist & Wiksell. Reprinted in: *Selected Writings*, 1962, p. 328-401

Jakobson, R. 1960: "Why "Mama" and "Papa"?", *Perspectives in Psychological Theory, dedicated to Heinz Werner*, New York. Reprinted in: *Selected Writings*, 1962, p. 538-545

Karwoski, T.F., H.S. Odbert and C.E. Osgood 1942: "Studies in synesthetic thinking II. The role of form in visual response to music", *The J. of General Psych*. 26, p. 199-222

Mahling, F. 1926: "Das Problem der "Audition colorée". Eine historisch-kritische Untersuchung", *Archiv f. die gesamte Psychologie* 57, p. 165-301

Malmberg, B. 1964: "Couches primitives de structure phonologique", *Phonetica* 11, p. 221-227

Malmberg, B. 1967: "Neue Perspektiven der phonetischen Forschung", *Studia Linguistica* 22, p. 75-90

Malmberg, B. 1970: "De Ferdinand de Saussure à Roman Jakobson. L'arbitraire du signe et la substance phonique du langage", *Proceedings of the Sixth Int. Cong. of Phon. Sc.*, Prague 1967, p. 599-603

Malmberg, B. 1971: "Notes sur le signe arbitraire", *Form and Substance* (ed. Hammerich, Jakobson, Zwirner), Copenhagen: Akademisk forlag, p. 31-38

Masson, D.I. 1951: "Synesthesia and sound spectra", *Word* 8, p. 39-41

Newman, St. S. 1933: "Further experiments in phonetic symbolism", *Am. J. of Psych*. 45, p. 53-75

Odbert, H.S., T.F. Karwoski and A.B. Eckerson 1942: "Studies in synesthetic thinking I. Musical and verbal associations of color and mood", *The J. of General Psych*. 26, p. 153-173

Peterfalvi, J.-M. 1965: "Les recherches expérimentales sur le symbolisme phonétique", *Am. J. of Psych*. 65, p. 439-473

Peterfalvi, J.-M. 1970: "Recherches expérimentales sur le symbolisme phonétique", *Monographies françaises de psychologie* 19, Paris (Not utilized)

Reichard, Gladys A., R. Jakobson and Elisabeth Werth 1949: "Language and synesthesia", *Word* 5, p. 224-233

Sapir, E. 1929: "A study in phonetic symbolism", *J. of exp. Psych*. 12, p. 225-239

Tarte, R.D. 1974: "Phonetic symbolism in adult native speakers of Czech", *LS* 17, p. 87-94

Taylor, I.K. 1963: "Phonetic symbolism re-examined", *Psych. Bull*. 60, p. 200-209

Thorndike, E.L. 1946: "On Orr's hypothesis concerning the front and back vowels", *British J. of Psych*. 36.10.14

Wellek, A. 1931: "Zur Geschichte und Kritik der Synästhesie-Forschung", *Archiv f. die gesamte Psych*. 79, p. 325-384

Werner, H. 1948: *Comparative psychology of mental development*

Westermann, D. 1937: "Laut und Sinn in einigen westafrikanischen Sprachen", *Archiv f. vergl. Phonetik* I, p. 154-172 and 193-212

Westermann, D. and Ida C. Ward 1933: *Practical phonetics for the students of African languages*, London: Humphrey Milford

(ARIPUC 12, 1978)

THE INFLUENCE OF ASPIRATION ON THE FUNDAMENTAL FREQUENCY OF THE FOLLOWING VOWEL IN DANISH: SOME PRELIMINARY OBSERVATIONS

Niels Reinholt Petersen

Abstract: Some preliminary observations are reported which seem to indicate that the fundamental frequency of a stressed vowel is higher after an aspirated than after an unaspirated stop consonant, and that the difference is not restricted to the beginning of the vowel but is also found at the middle and frequently all through the vowel.

1. Introduction

It is well known that the fundamental frequency of a vowel is higher after a voiceless than after a voiced stop consonant, other things being equal (see e.g. Di Cristo and Chafkouloff 1977, Fischer-Jørgensen 1968, House and Fairbanks 1953, Johansson 1976, Löfqvist 1975, Lehiste and Peterson 1961, Lea 1973, Mohr 1971). This tendency is most marked at the onset of the vowel, but the difference has been found to persist even at the end of the vowel (e.g. Löfqvist 1975, Hombert 1976).

The effect of aspiration on the fundamental frequency of the following vowel seems to be less well established. For Danish, where the only distinction between ptk and bdg is one of aspiration (both series being voiceless), Fischer-Jørgensen (1968) found no difference in Fo at the onset of the following vowel. Jeel (1975), on the other hand, found a difference at that point of measurement, which was statistically significant. Jeel also measured Fo at the point of minimum Fo in the vowel and at the end of the vowel (it is not explicitly stated where the minimum point is located, but it seems to be in the middle third of long vowels (cp. section 2.3 below) and somewhat earlier in short vowels). At these points of measurement there was a tendency for Fo to be higher after aspirated than after unaspirated stop, but the difference was significant in neither case. Nor do data from other languages having an aspirated-unaspirated opposition such as Thai (Gandour 1974), Korean (Han and Weitzman 1970, Kagaya 1974) or Hindi (Kagaya and Hirose 1975) show any consistent trend for the effect of aspiration on the Fo of the following vowel.

Hombert and Ladefoged (1977) have employed a cross-language approach in an attempt to clarify the matter. On the basis of their comparison of Fo contours after French and English *p*, *t*, and *k*, they conclude (p. 34) that "it is clear from these data that there is not a direct correlation between the duration of the aspiration after a voiceless consonant and the onset fundamental frequency of the following vowel since English and French voiceless consonants have very similar perturbatory effects on the Fo of the following vowel". Things become less clear, however, if one compares their data on English *ptk* and *bdg*, the latter being referred to (p. 34) as "voiceless unaspirated or "voiced"". It turns out, then, that Fo at the onset of the vowel is considerably higher after *ptk* than after *bdg* (about 20 Hz and 60 Hz for the male and female subject, respectively), and that the difference, although decreasing, still persists after 12 voice periods.

The aim of the present paper is to report in a very preliminary form some observations concerning the influence of aspiration on the fundamental frequency of the following vowel in Danish. The questions to be considered below are the following: 1) Do aspirated and unaspirated stops influence Fo of the following vowel differently? 2) How far into the vowel does the difference, if any, persist? 3) Do stressed and unstressed vowels behave differently with respect to the influence of aspiration on the Fo of the following vowel?

2. Method

2.1 Material

The material to be considered here consisted of the vowels [i], [ɑ], [u] occurring in nonsense words of the type /pV'pV:pV/ ([b̥ʰVˈb̥ʰV:b̥ʰV]) and /bV'bV:bV/ ([b̥Vˈb̥V:b̥V]), the vowels being identical in the three syllables. The test words were embedded in the carrier sentence "stavelserne i .... forkortes" ('The syllables of .... are shortened'). Each test word occurred five times in a randomized list.

2.2 Recordings and speakers

The recordings took place in an anechoic chamber by means of professional recording equipment. Two female subjects (BH and JG) and three male subjects (JB, SH, and NR (the author)) acted as speakers. They were all phoneticians, and all speakers of Advanced Standard Copenhagen Danish (see Basbøll 1968). Each subject read the list twice, so that ten tokens of each test word were obtained.

2.3 Data processing

The recordings of three of the subjects (JG, JB, and NR) were processed by the Multiple Channel Processing system which is implemented on the Institute's PDP/8 minicomputer. The MCP system (which is described in some detail in Holtse and Stellinger 1976) samples and averages slowly varying analog signals (in this case the output from hardware fundamental frequency and intensity meters) at a rate of 200 samples per second within selected time windows of 1250 ms. Up to seven channels may be sampled simultaneously. The line-up points used in the averaging process were the onset of the vowels in the CV'CV:CV test words. These points were determined from intensity tracings. Since only one line-up point can be specified in each channel for the averaging, the signal from the Fo meter was sampled in parallel in three channels, the first channel being averaged around the onset of the first (pretonic) vowel, the second channel around the onset of the second (long stressed) vowel, and the third channel around the onset of the third (posttonic) vowel. The output from the intensity meter was sampled in the fourth channel. The averaged curves were displayed on a graphic terminal screen. Figs. 1 to 18 show photographs of the screen. The average curves in the figures represent 10 tokens of each test word, i.e. both readings of the word list.

The MCP tracings were supplemented by a set of measurements of the fundamental frequency at the point of minimum Fo in the long stressed vowel. This set included data from all five subjects.

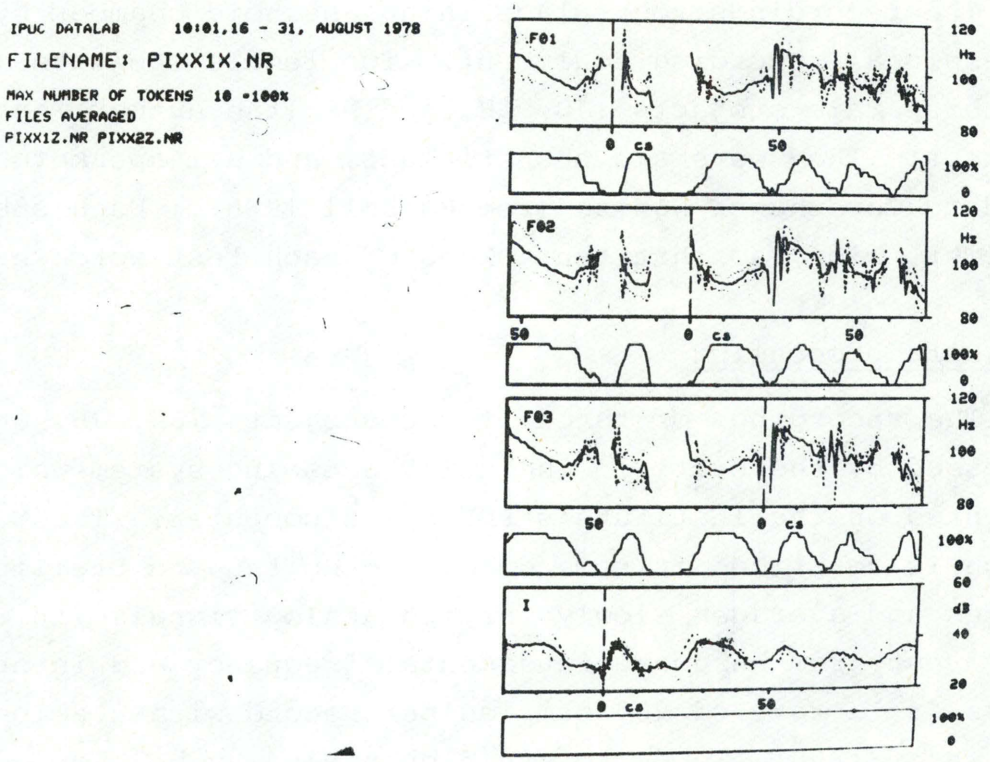
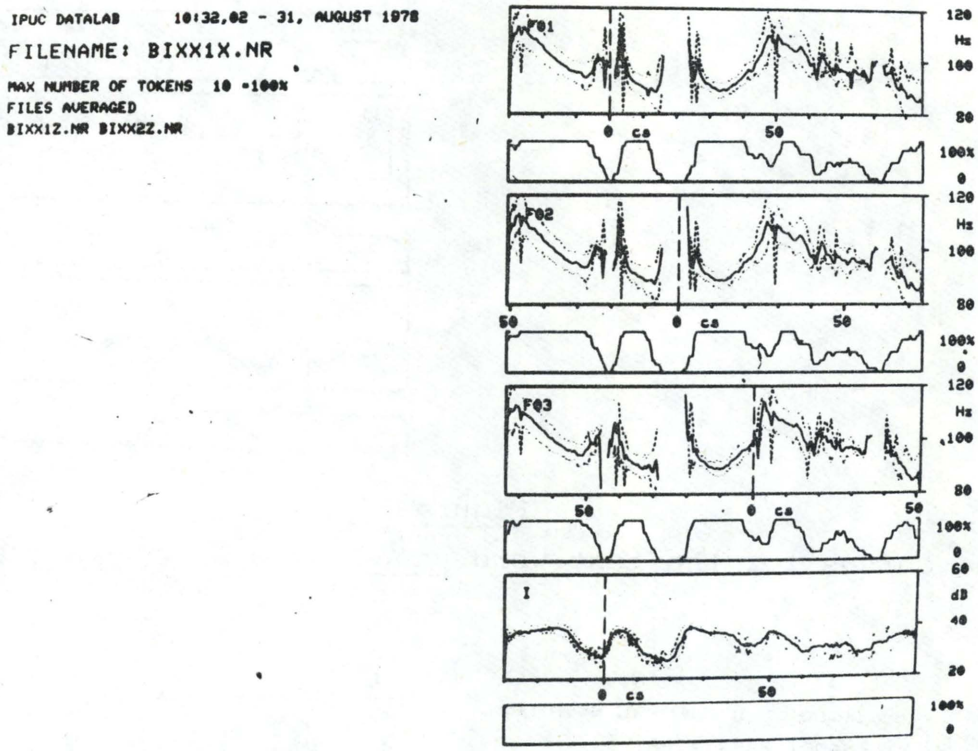


Figure 1

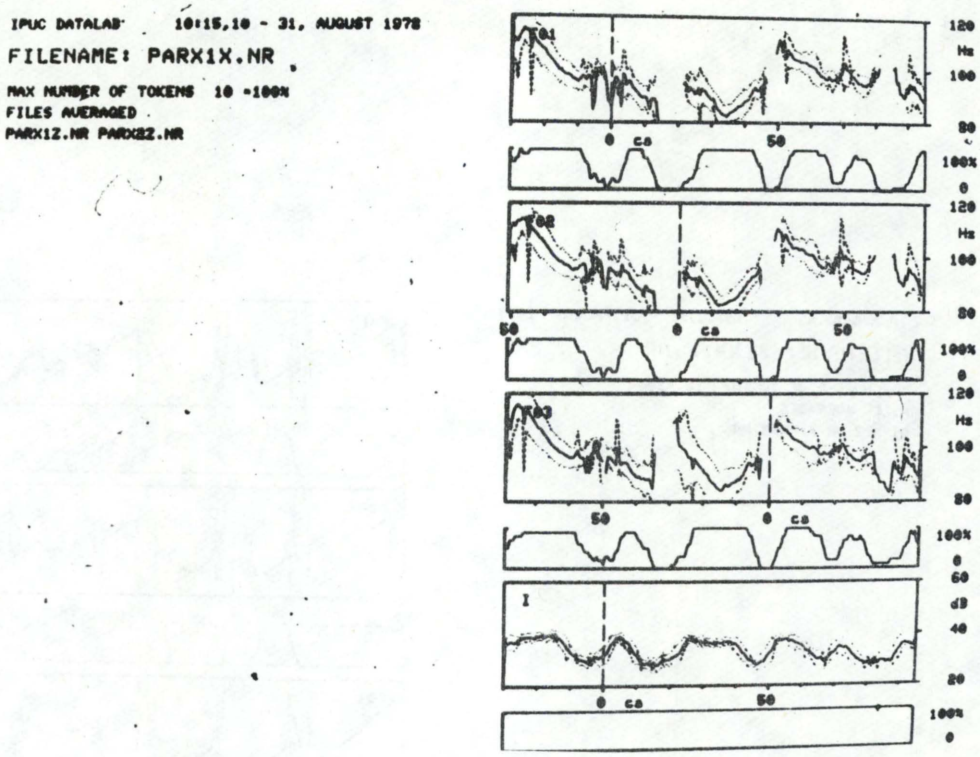
Average curves of the testword pi'pi:pi spoken by subject NR. The upper three channels (F01, F02, and F03) contain F_o tracings averaged with the onset of the fist, second, and third vowel of the testword, respectively, as line-up points (see section 2.3). The fourth channel contains the intensity averaged around the onset of the first vowel. The unbroken curves indicate the average, and the dotted lines on either side indicate the average plus and minus one standard deviation. The vertical broken lines at o cs on the time axis indicate the line-up points. Immediately below each channel the number of measurements (expressed as the percentage of the maximum number) included in the averaging of that channel is shown as a function of time. (For F_o channels any measurement lower than 60 Hz is excluded from the averaging.)



```
IPUC DATALAB       10:32,02 - 31, AUGUST 1978
FILENAME: BIXX1X.NR
MAX NUMBER OF TOKENS  10 =100%
FILES AVERAGED
BIXX1Z.NR BIXX2Z.NR
```

Figure 2

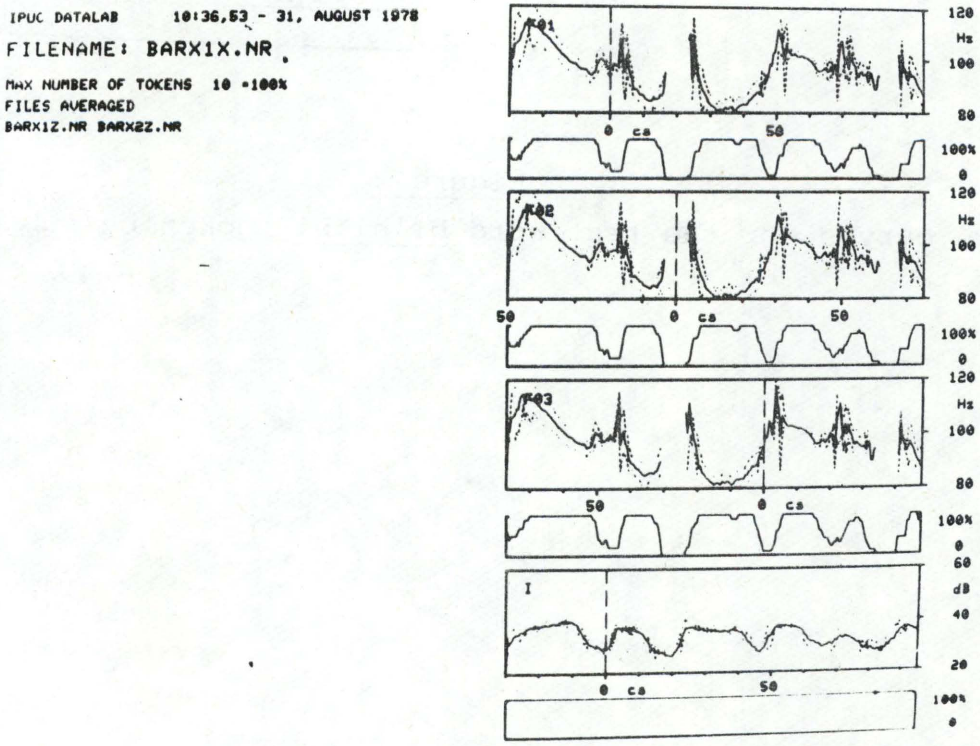
Average curves for the test word bi'bi:bi spoken by subject NR.



```
IPUC DATALAB     10:15,10 - 31, AUGUST 1978
FILENAME: PARX1X.NR
MAX NUMBER OF TOKENS  10 =100%
FILES AVERAGED
PARX1Z.NR PARX2Z.NR
```

Figure 3

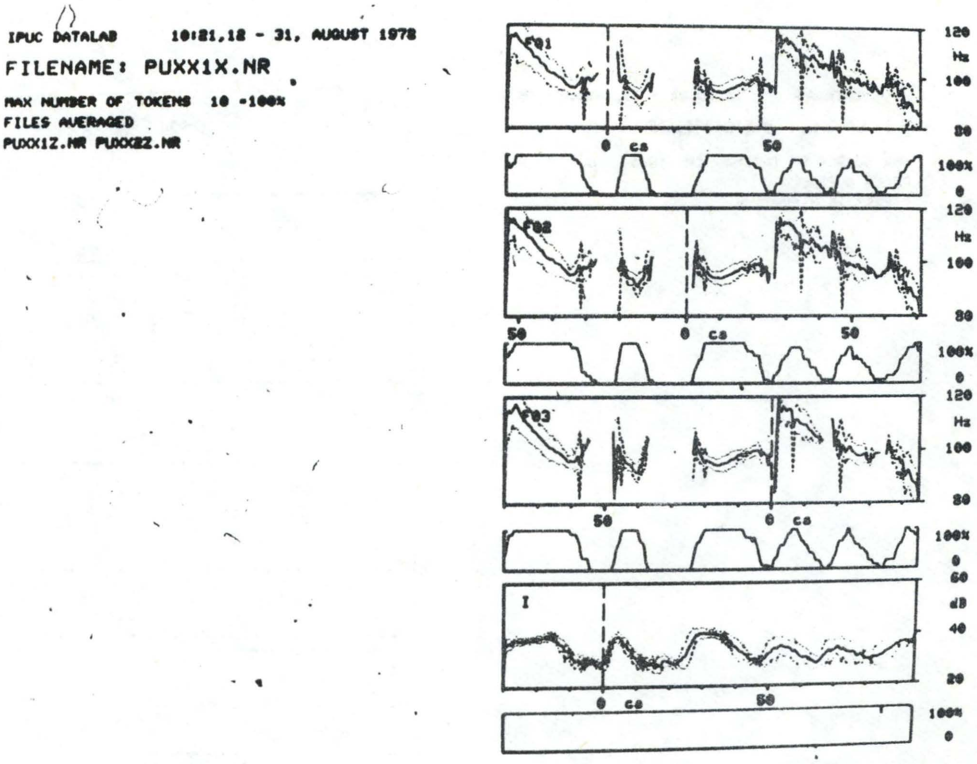
Average curves for the test word pa'pa:pa spoken by subject NR.



```
IPUC DATALAB     10:36,53 - 31, AUGUST 1978
FILENAME: BARX1X.NR
MAX NUMBER OF TOKENS  10 =100%
FILES AVERAGED
BARX1Z.NR BARX2Z.NR
```

Figure 4

Average curves for the test word ba'ba:ba spoken by subject NR.



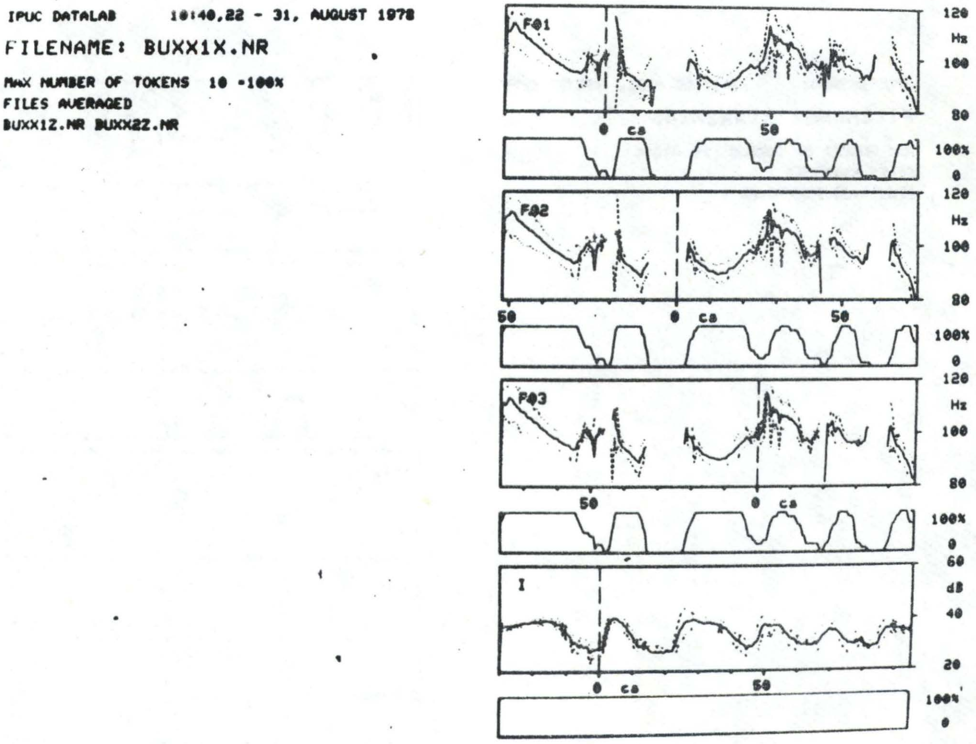
IPUC DATALAB 10:21,18 - 31, AUGUST 1978

FILENAME: PUXX1X.NR

MAX NUMBER OF TOKENS 10 -100%
FILES AVERAGED
PUXX1Z.NR PUXX2Z.NR

Figure 5

Average curves for the test word pu'pu:pu spoken by subject NR.



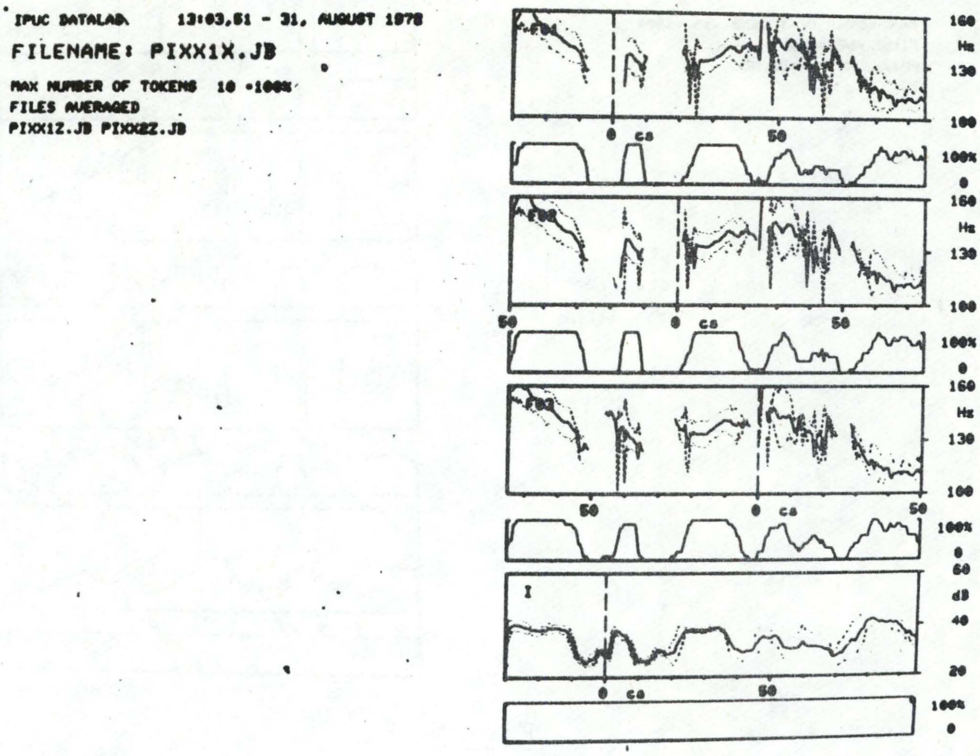
IPUC DATALAB 10:40,22 - 31, AUGUST 1978

FILENAME: BUXX1X.NR

MAX NUMBER OF TOKENS 10 -100%
FILES AVERAGED
BUXX1Z.NR BUXX2Z.NR

Figure 6

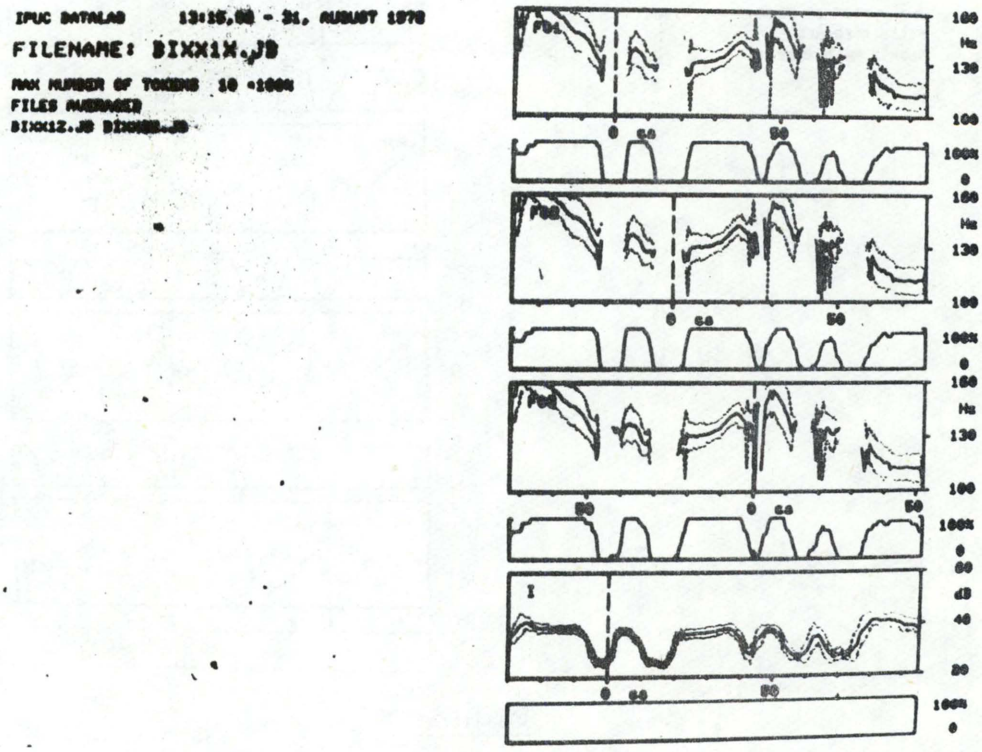
Average curves for the test word bu'bu:bu spoken by subject NR.



IPUC DATALAB 13:03,61 - 31, AUGUST 1978
FILENAME: PIXX1X.JB
MAX NUMBER OF TOKENS 10 =100%
FILES AVERAGED
PIXX12.JB PIXX22.JB

Figure 7

Average curves for the test word <u>pi'pi:pi</u> spoken by subject JB.



IPUC DATALAB 13:15,[illegible] - 31, AUGUST 1978
FILENAME: BIXX1X.JB
MAX NUMBER OF TOKENS 10 =100%
FILES AVERAGED
BIXX12.JB BIXX22.JB

Figure 8

Average curves for the test word <u>bi'bi:bi</u> spoken by subject JB.

```
IPUC DATALAB        13:07,06 - 31, AUGUST 1978
FILENAME: PARX1X.JB
MAX NUMBER OF TOKENS  10 =100%
FILES AVERAGED
PARX1Z.JB PARX2Z.JB
```

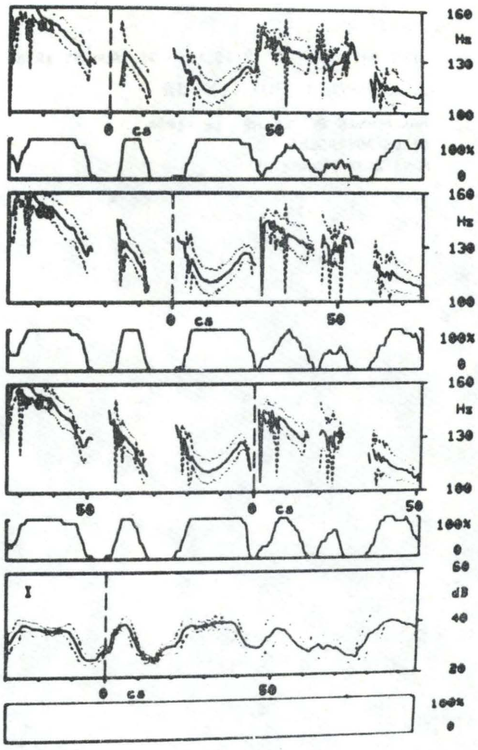


Figure 9

Average curves for the test word pɑ'pɑːpɑ spoken by subject JB.

```
IPUC DATALAB        13:21,27 - 31, AUGUST 1978
FILENAME: BARX1X.JB
MAX NUMBER OF TOKENS  10 =100%
FILES AVERAGED
BARX1Z.JB BARX2Z.JB
```

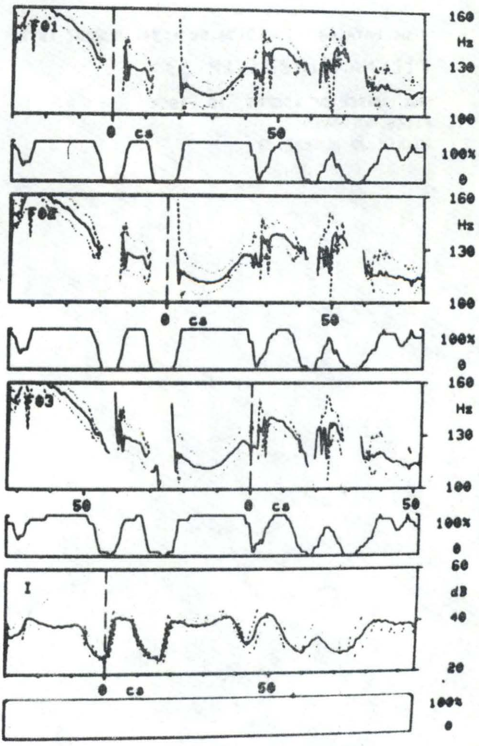
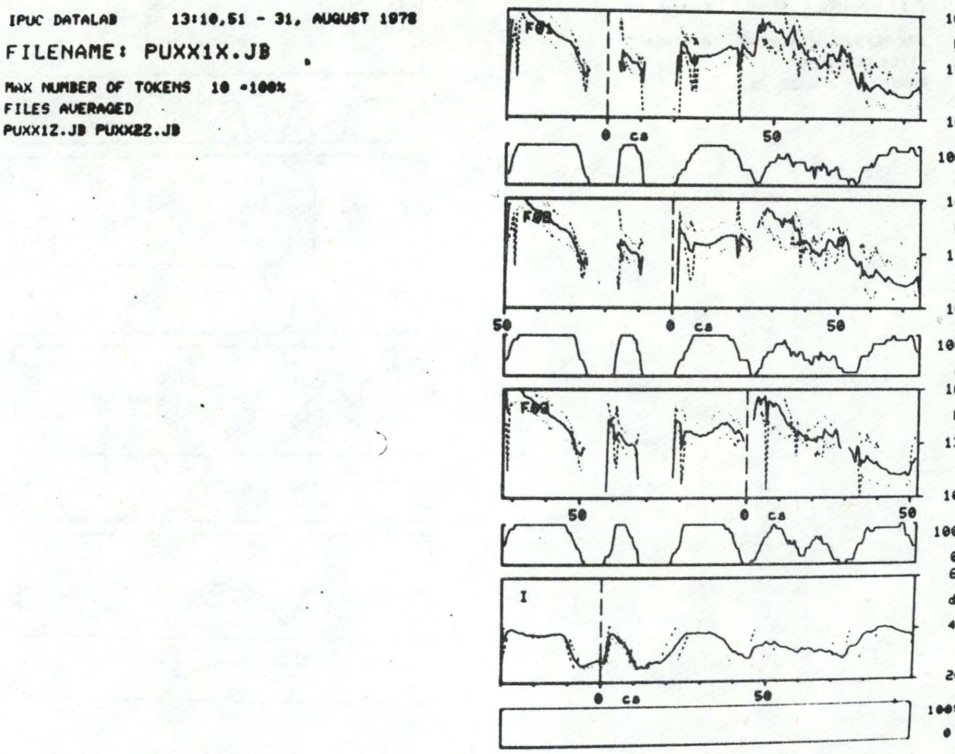


Figure 10

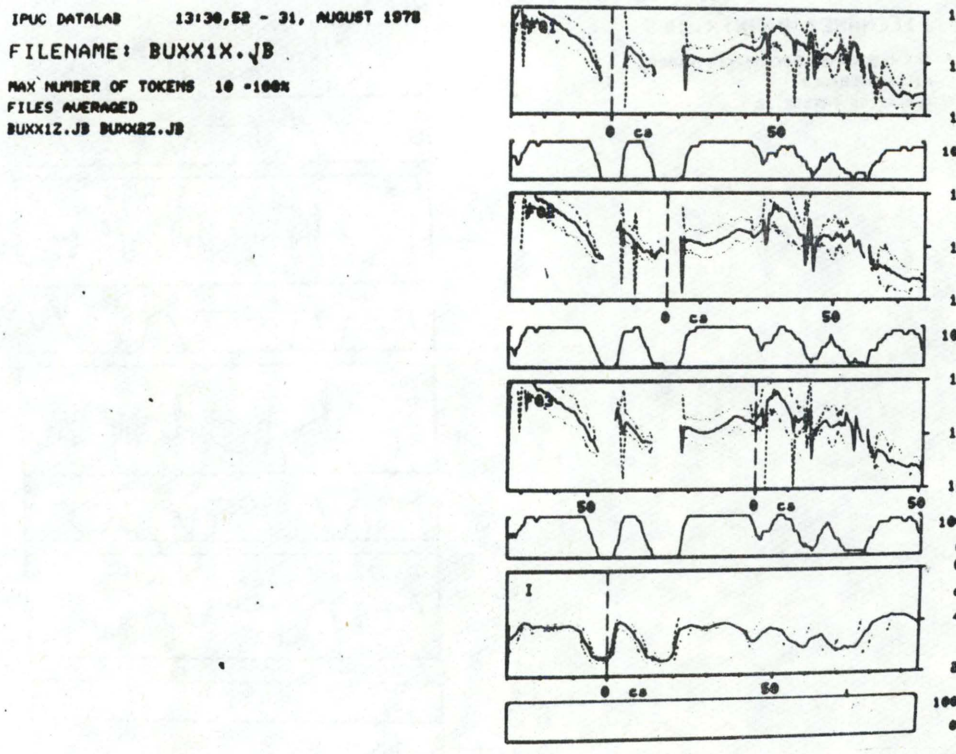
Average curves for the test word bɑ'bɑːbɑ spoken by subject JB.



```
IPUC DATALAB        13:10,51 - 31, AUGUST 1978
FILENAME: PUXX1X.JB
MAX NUMBER OF TOKENS  10 =100%
FILES AVERAGED
PUXX1Z.JB PUXX2Z.JB
```

Figure 11

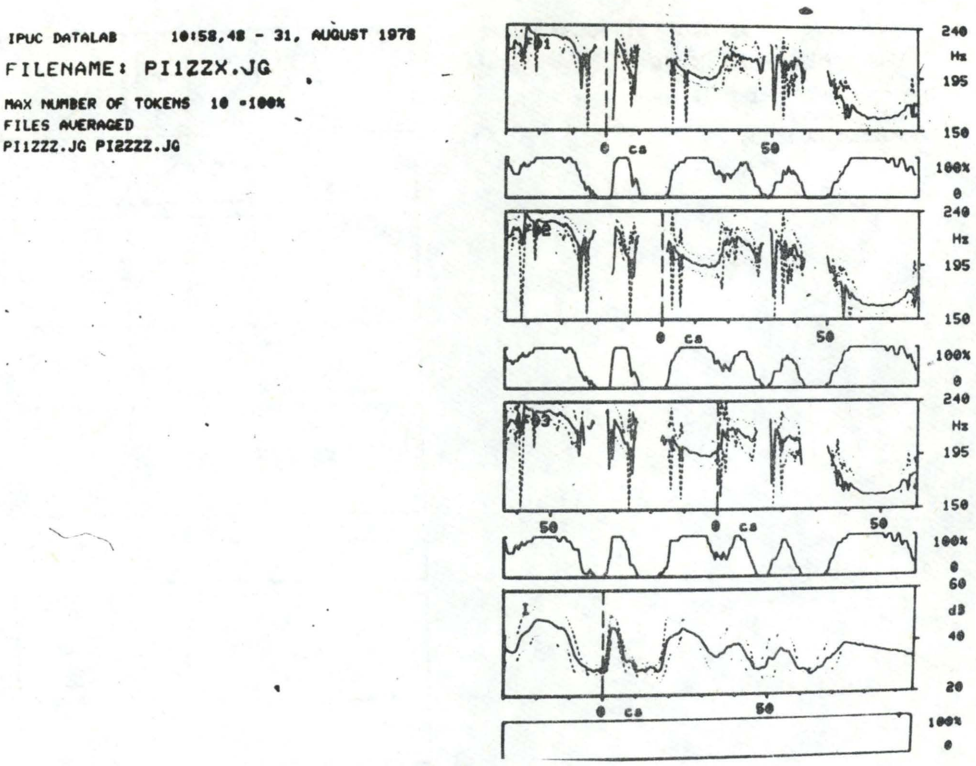
Average curves for the test word pu'pu:pu spoken by subject JB.



```
IPUC DATALAB        13:30,52 - 31, AUGUST 1978
FILENAME: BUXX1X.JB
MAX NUMBER OF TOKENS  10 =100%
FILES AVERAGED
BUXX1Z.JB BUXX2Z.JB
```

Figure 12

Average curves for the test word bu'bu:bu spoken by subject JB.



IPUC DATALAB 10:58,48 - 31, AUGUST 1978

FILENAME: PI1ZZX.JG

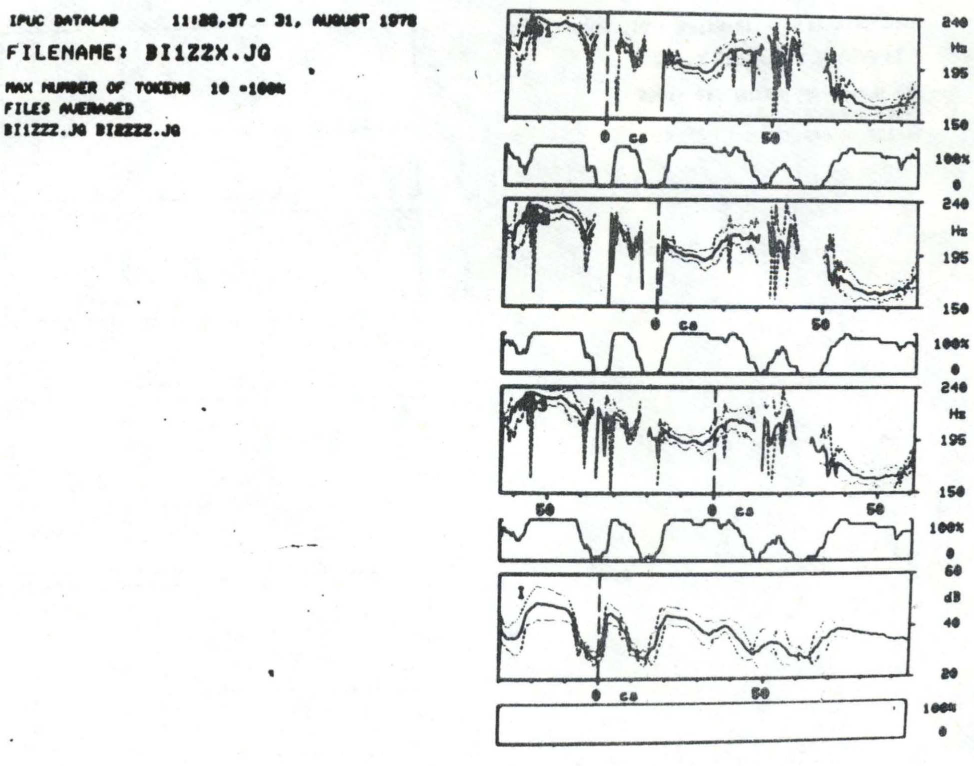
MAX NUMBER OF TOKENS 10 =100%

FILES AVERAGED

PI1ZZZ.JG PI2ZZZ.JG

Figure 13

Average curves for the test word pi'pi:pi spoken by subject JG.



IPUC DATALAB 11:28,37 - 31, AUGUST 1978

FILENAME: BI1ZZX.JG

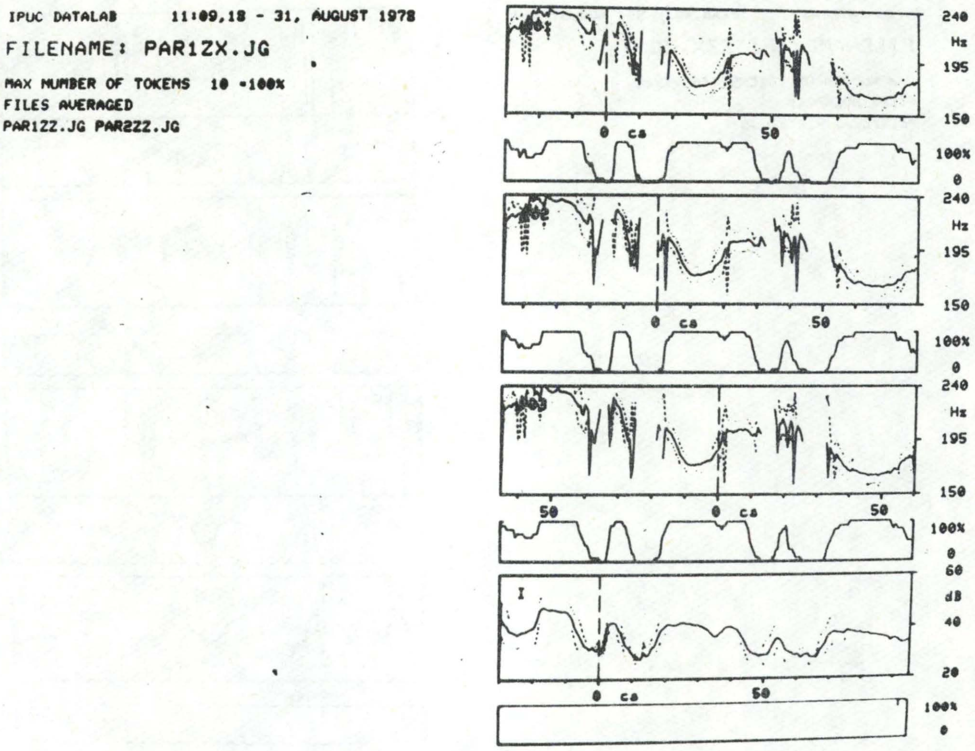
MAX NUMBER OF TOKENS 10 =100%

FILES AVERAGED

BI1ZZZ.JG BI2ZZZ.JG

Figure 14

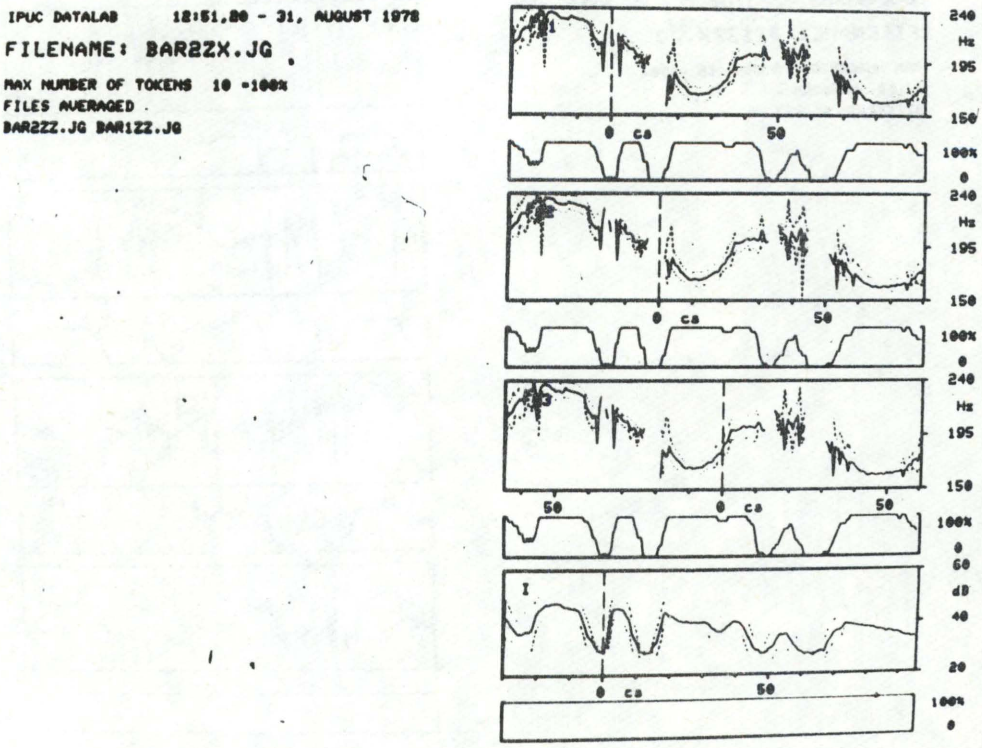
Average curves for the test word bi'bi:bi spoken by subject JG.



```
IPUC DATALAB        11:09,18 - 31, AUGUST 1978
FILENAME: PAR1ZX.JG
MAX NUMBER OF TOKENS  10 =100%
FILES AVERAGED
PAR1ZZ.JG PAR2ZZ.JG
```

Figure 15

Average curves for the test word pa'pa:pa spoken by subject JG.



```
IPUC DATALAB        12:51,20 - 31, AUGUST 1978
FILENAME: BAR2ZX.JG
MAX NUMBER OF TOKENS  10 =100%
FILES AVERAGED
BAR2ZZ.JG BAR1ZZ.JG
```

Figure 16

Average curves for the test word ba'ba:ba spoken by subject JG.

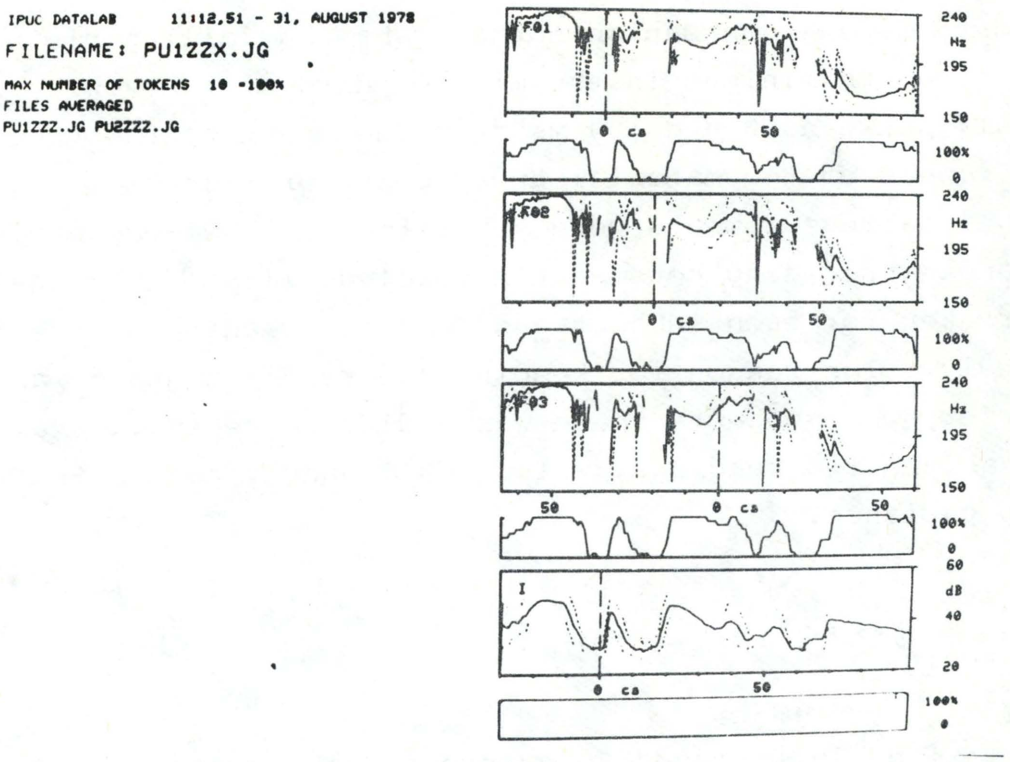


Figure 17

Average curves for the test word pu'pu:pu spoken by subject JG.

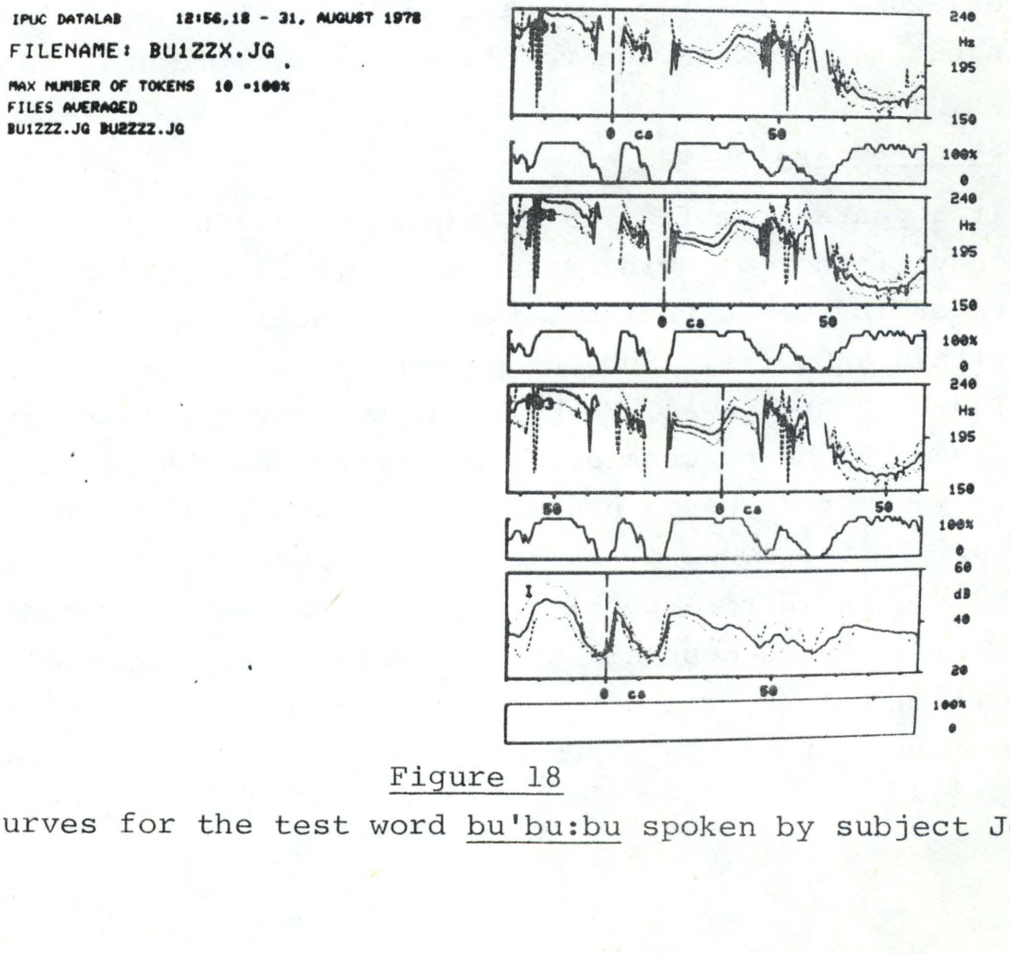


Figure 18

Average curves for the test word bu'bu:bu spoken by subject JG.

As mentioned in section 1 above, the stressed vowels of Advanced Standard Copenhagen Danish normally have a falling-rising Fo movement with its minimum in the middle third of the vowel, i.e. roughly between 50 and 150 ms from the onset of the vowel in the long vowels of the material under consideration here.

These measurements were submitted to a two-way analysis of variance (preceding consonant x vowel quality). The analysis was undertaken for each subject and for each reading of the list separately. The reason for keeping the readings apart was the fact that for some subjects there was a difference in general Fo level between the two readings, which would unduly add to the within group variance.

3. Results

In fig. 19 averaged Fo curves are shown for *p*- and *b*-words superimposed upon one another in order to facilitate comparison. The curves are drawings made on the basis of photographs of the graphic terminal screen (similar to those displayed in figs. 1 to 18), enlarged to exactly the same scale. (The time axis is correct within but not between the vowels of a test word.)

3.1 Stressed syllables

It appears from fig. 19 that *p* and *b* influence the fundamental frequency of the following stressed vowel differently. But the pattern of influence varies between subjects, and to some extent also within subjects. Subject NR has the clearest difference in the middle of the vowel, Fo being higher after *p* than after *b*. There seems to be no consistent difference at the beginning of the vowel. Subject JB has a higher Fo all through the vowels *i* and *u* after *p*. In the vowel *ɑ*, on the other hand, the difference is limited to the initial portion of the vowel. For subject JG, Fo is higher throughout all three vowels *i*, *ɑ*, and *u* following *p*, although in *i* and *ɑ* the difference grows smaller toward the end of the vowel. In *u* the difference remains constant during the vowel.

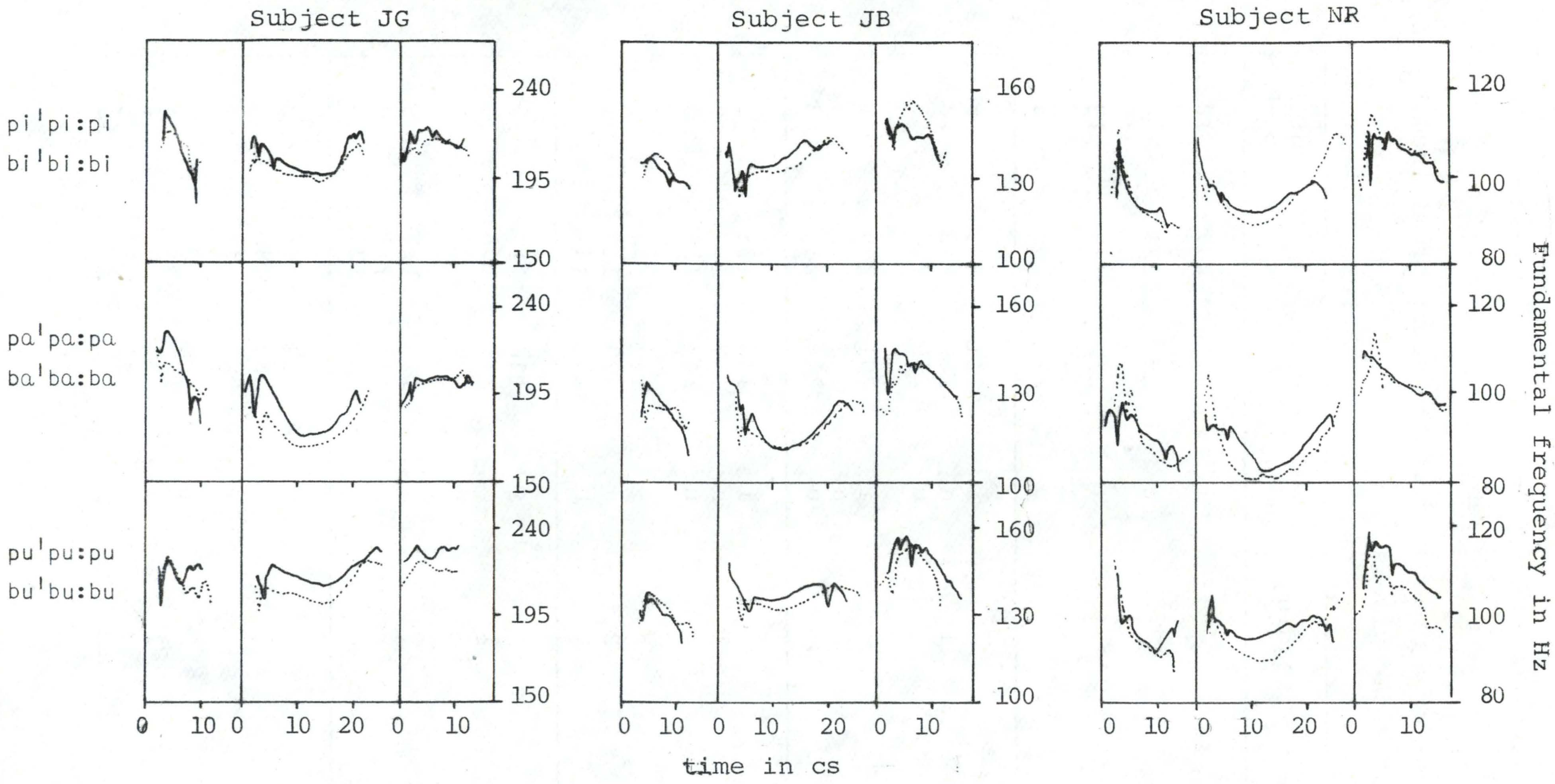


Figure 19

Average curves for p-words (unbroken lines) and b-words (dotted lines) superimposed upon one another (for further explanation, see text).

Table 1

Averages of F_o measured at the point of minimum F_o in long stressed vowels.

subject	reading	vowel qual.	p_	b_	diff. p_-b_
BH	1	i	219	209	10
		ɑ	188	185	3
		u	225	221	4
	2	i	212	199	13
		ɑ	175	172	3
		u	213	209	4
JG	1	i	211	205	6
		ɑ	186	178	8
		u	230	217	13
	2	i	201	200	1
		ɑ	177	172	5
		u	214	211	3
JB	1	i	140	138	2
		ɑ	117	117	0
		u	144	139	5
	2	i	143	140	3
		ɑ	120	119	1
		u	145	141	4
SH	1	i	126	119	7
		ɑ	111	108	3
		u	126	121	5
	2	i	126	121	5
		ɑ	111	111	0
		u	127	122	5
NR	1	i	93	91	2
		ɑ	83	82	1
		u	95	92	3
	2	i	95	92	3
		ɑ	87	83	4
		u	97	92	5

The results of the measurements of fundamental frequency at the minimum points of the stressed vowels are summarized in tables 1 and 2. Table 1 contains the means for the vowels i, ɑ, and u after p and b. It appears that the differences are quite small, varying from 0 to 13 Hz with an average of 4.4 Hz, but as can be seen from table 2, the effect of the preceding consonant was in all cases significant at the 5 per cent level or better. (There was also a highly significant effect of vowel quality.) The interaction between preceding consonant and vowel quality was in no cases significant ($p > 0.5$).

Table 2

Levels of significance obtained for the effect of preceding consonant on the Fo measured at the point of minimum Fo in long stressed vowels.

	BH	JG	JB	SH	NR
reading 1	$p<0.01$	$p<0.01$	$p<0.05$	$p<0.01$	$p<0.01$
reading 2	$p<0.05$	$p<0.05$	$p<0.05$	$p<0.01$	$p<0.01$

3.2 Unstressed syllables

From fig. 19 it appears that no consistent pattern of influence of the preceding consonant can be found in the fundamental frequency of the unstressed vowels, whether pretonic or posttonic. There are indeed cases in which Fo is higher after p than after b, but cases of no difference are equally frequent, and there are a few examples of the opposite relation. The pattern of influence of the preceding consonant seems not to be affected by the position of the unstressed syllable, i.e. whether it is pre- or posttonic.

From the curves in figs. 1 to 18, displaying the number of measurements as a function of time, it appears as if b in posttonic syllables may sometimes be voiced (or rather that the energy during the closure has been high enough to trigger the Fo meter). For the subjects NR and JB, however, this apparent voicing may be due to differences between tokens in the position in time of the line-up point in relation to offset and onset of voicing in the surrounding vowels. Taken token by token the posttonic stops were very rarely voiced with these subjects. For subject JG, on the other hand, it is evident that her posttonic stops, both *b* and *p*, are voiced in the majority of cases. This can also be seen from the Fo curves in the figures, which show an almost continuous movement during the entire tonic and posttonic part of the words in *b*-words as well as in *p*-words. One explanation for the very frequent voicing of JG's stops is probably her high speaking rate; her distance in time between the onset of the first vowel and the onset of the third vowel in the test words is about 35 cs on an average. The corresponding distances for NR and JB are 50 and 45 cs, respectively. An inspection of intensity tracings and spectrograms of the test words uttered by JB revealed a tendency for the voicing to be slightly weaker in *p* than in *b*. Before the vowel *i* *p* was followed by a phase of voiced aspiration, before *u* and *ɑ* such aspiration could be seen in a few cases only.

Neither *b* nor *p* were voiced in pretonic syllables for any of the subjects.

4. Discussion

Although the material under consideration is rather limited the main results seem to be quite clear, namely that the fundamental frequency of a vowel in a stressed syllable is higher after a voiceless aspirated than after a voiceless unaspirated stop, and that the difference is not restricted to the initial portion of the vowel, but is found also in the middle and in a great number of cases all through the vowel.

These results seem to be somewhat in disagreement with what would be predicted from current hypotheses dealing with the effect of aspiration on the fundamental frequency of the following vowel.

It has been suggested that the high Fo after an aspirated stop could be explained by the high rate of airflow upon release of the stop (e.g. Ohala 1973). It is true that the glottis aperture is far larger in Danish aspirated stops than in unaspirated stops (Frøkjær-Jensen, Ludvigsen,and Rischel 1971, Hutters 1978), and also that the airflow is higher (Fischer-Jørgensen 1968). This could explain a higher Fo at the onset of the vowel following an aspirated stop, but not the persistence of the difference between the effects of the aspirated-unaspirated distinction as far into the vowel as is found in the present data.

Hombert, Ohala, and Ewan (1976) and Ohala (1978) suggest that the voiced-voiceless opposition affects the vertical tension of the vocal cords both within the consonant and in the following vowel. Under the assumption that the vertical tension of the vocal cords is reflected by the height of the larynx they employ larynx height data measured by means of the "thyroumbrometer" (Ewan and Krones 1974) in support of the hypothesis. Ewan and Krones investigated the vertical movement of the larynx in vowel-stop-vowel sequences in English, French, Thai and Hindi. They found the larynx to be significantly higher in unvoiced stops than in voiced stops, and - what is interesting from the point of view of the present investigation - they found that in Thai and Hindi the distinction between voiceless aspirated and voiceless unaspirated was not accompanied by significant differences in larynx height, neither in the stop nor in the vowel following it. According to the vertical tension hypothesis, then, no difference should be expected between aspirated and unaspirated voiceless stops with regard to their influence on the Fo of the following vowel. This is in agreement with the Thai data of Gandour (1974) and the Hindi data of Kagaya and Hirose (1975), but not with the data on Danish reported above. Unfortunately Ewan and Krones (1974) do not accompany their larynx height data with simultaneous Fo tracings.

Another "tension hypothesis" has been advanced by Halle and Stevens (1971), who suggest that the vocal cords should be stiffer, i.e. have a greater longitudinal tension, in aspirated than in unaspirated stops. This would predict the higher Fo after aspirated stops actually found in the present material. On the other hand, EMG data on the behaviour of laryngeal muscles, among them the vocalis, in Danish stops reported by Fischer-Jørgensen and Hirose (1974) do not indicate that the vocal cords should be any stiffer in <u>ptk</u> than in <u>bdg</u> in Danish.

The results for the unstressed syllables are much less consistent than the results for the stressed ones, and the conclusion to be tentatively drawn is that the fundamental frequency of unstressed vowels is not affected by a difference in aspiration of the preceding stop consonant. This is not very surprising, since the aspiration of stops in unstressed syllables is considerably shorter than in stressed syllables. It is also in line with fiberscopic observations reported by Birgit Hutters (personal communication) that the difference in glottal gesture between p and b in unstressed syllables is far smaller than in stressed syllables, the gesture of p being more similar to that of b.

References

Basbøll, H. 1968:	"The phoneme system of Advanced Standard Copenhagen", ARIPUC 3, p. 33-54
Di Cristo, A. and M. Chafkouloff 1977:	"Les faits microprosodiques du français: voyelles, consonnes, coarticulation", 8ièmes journées d'études sur la parole, 25-27 mai 1, p. 147-158
Ewan, W.G. and R. Krones 1974:	"Measuring larynx movement using the thyroumbrometer", J.Ph. 2, p.327-335
Fischer-Jørgensen, E. 1968:	"Voicing, tenseness, and aspiration in stop consonants, with special reference to French and Danish", ARIPUC 3, p. 63-114
Fischer-Jørgensen, E. and H. Hirose 1974:	"A preliminary electromyographic study of labial and laryngeal muscles in Danish stop consonant production", Haskins SR 39/40, p. 231-254
Frøkjær-Jensen, B., C. Ludvigsen and J. Rischel 1971:	"A glottographic study of some Danish consonants", F&S, p. 123-140
Gandour, J. 1974:	"Consonant types and tone in Siamese", J.Ph. 2, p. 337-350
Halle, M. and K.N. Stevens 1971:	"A note on laryngeal features", MIT QPR 101, p. 198-213
Han, M.S. and R.S. Weitzman 1970:	"Acoustic features of Korean /P,T,K/, /p,t,k/ and /p^h,t^h,k^h/", Phonetica 22, p. 112-128
Holtse, P. and J.H. Stellinger 1976:	"A system for computer aided processing of phonetic measurements", ARIPUC 9, p. 201-219

Hombert, J.-M. 1976: "Phonetic explanation of the development of tones from prevocalic consonants", *UCLA WPP* 33, p. 23-39

Hombert, J.-M. and P. Ladefoged 1977: "The effect of aspiration on the fundamental frequency of the following vowel", *UCLA WPP* 36, p. 33-40

Hombert, J.-M., J.J. Ohala, and W.G. Ewan 1976: "Tonogenesis: Theories and queries", *Report of Phonology Laboratory* 1 (Berkeley), p. 48-92

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

Hutters, B. 1978: "The glottal gesture in some Danish consonants - preliminary observations", (this issue)

Jeel, V. 1975: "An investigation of the fundamental frequency of vowels after various consonants, in particular stop consonants", *ARIPUC* 9, p. 191-211

Johansson, I. 1976: *Inherenta grundtonsfrekvenser hos svenska vokaler och deras inflytande på satsintonationen*, *Stadsmål i övre Norrland* 9 (Umeå University)

Kagaya, R. 1974: "A fiberscopic and acoustic study of the Korean stops, affricates and fricatives", *J.Ph*. 2, p. 161-180

Kagaya, R. and H. Hirose 1975: "Fiberoptic, electromyographic and acoustic analyses of Hindi stop consonants", *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics* 9 (Tokyo), p. 27-46

Lea, W.A. 1973: "Segmental and suprasegmental influences on fundamental frequency contours", *Consonant types and tone*, (ed.: L.M. Hyman), *Southern California Occasional Papers in Linguistics* 1, p. 15-70

Lehiste, I. and G. Peterson 1961: "Some basic considerations in the analysis of intonation", *JASA* 33, p. 419-425

Löfqvist, A. 1975: "Intrinsic and extrinsic F_0 variations in Swedish tonal accents", *Phonetica* 31, p. 228-247

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

Ohala, J.J. 1973: "The physiology of tone", Consonant types and tone (ed.: L.M. Hyman), Southern California Occasional Papers in Linguistics 1, p. 1-14

Ohala, J.J. 1978: "The production of tone", Report of the Phonology Laboratory 2 (Berkeley), p. 63-117

SLIT HEIGHT AND BLOWING PRESSURE AS PARAMETERS CONDITIONING THE FUNCTION OF THE CLARINET AS A QUASI-QUARTER-WAVE RESONATOR[1]

Niels Bak

Abstract: Using mechanical blowing applied both to a conventional clarinet and to experimental devices with the same reed excitation mechanism, the author has investigated the relationships referred to as "clarinet embouchure". The results confirm John Backus' statement that "Apparently, the intonation characteristics of the clarinet (and, perhaps, the other woodwinds) depend on more than just the geometry of the instrument" (Backus 1964, p. 2014).

1. Purpose of the investigation

The primary purpose of this author's research project dealing with the clarinet has been to throw light on those properties of tone production which are influenced by the player's blowing technique - his embouchure. The present paper deals with but a single aspect of this complex of problems, viz. the question why the clarinet has resonant properties typical of a tube closed at one end, even in such, frequently occurring, cases where by witness there is only a transient closure or no closure at all in the vibratory cycle of the reed during tone procution. It is the aim of the present paper to contribute to an explanation of this crux.

---

1) I am indebted to professor Eli Fischer-Jørgensen for her encouragement and her kind permission to use the facilities of the Institute of Phonetics. My thanks are extended to the staff of the Laboratory, and to my two excellent assistants, Ole Birk Wulff, M.Sc., and Jan Nicolas Fredholm, M.Sc., for their enthusiasm and diligence in contributing to complete the various tasks within this project. Moreover, I wish to thank Dr. Johan Sundberg of the Speech Transmission Laboratory, Royal Institute of Technology, Stockholm, for fruitful and inspiring discussions. I am particularly indebted to Professor Jørgen Rischel, now at the Institute of Linguistics, University of Copenhagen, for his advice and extensive help in bringing the text of this report into its present shape.

2. Presentation of the basic clarinet

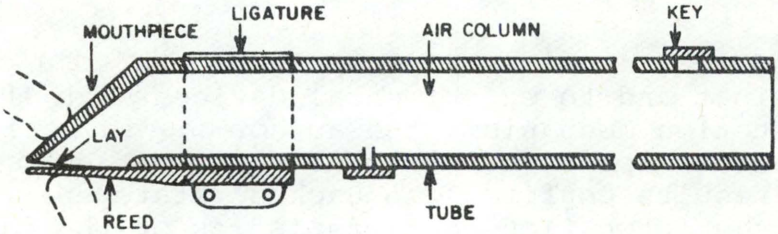


Figure 1

The primary components of the clarinet, viz. the excitation mechanism and the resonator tube. (From Backus 1969, p. 189.)

In a somewhat simplified exposition the clarinet may be described as a woodwind instrument consisting of two parts:
(i) the resonator, i.e. the (quasi-)cylindrical tube whose acoustic length can be varied by the fingering.
(ii) the excitation mechanism, i.e. the mouthpiece with reed, which is attached to the "upper" end of the tube and functions as a valve-like regulator of the stream of exhaled air forced into the instrument by the player. The outermost part of the reed, which is thin and elastic, performs oscillatory movements up and down during tone production and thus varies the slit height at the tip of the mouthpiece. This action of the reed is a crucial factor in the excitation and maintenance of vibrations of air in the tube.

Although those acoustic properties of the clarinet which are associated with the geometry of the tube are quite well understood, there are still many unsettled questions concerning the excitation mechanism. The research project of which some results are presented in this paper is intended to counteract this bias in the research. It focuses primarily on acoustic properties which - although they are inseparable from the functioning of the instrument as a totality - are specifically associated with the mouthpiece and with the blowing technique used in clarinet playing.

3. Aims and purposes of recent research on the acoustics of wind instruments, and the need for research of a complementary kind

During the last couple of decades improvements in measuring techniques have made it possible to gather highly reliable data on the acoustic behaviour of wind instruments, and thanks to a research activity of high quality, significant advances have been made towards an understanding and description of the functioning of these instruments.[1]

The following passage from Nederveen (1969) will serve to illustrate what kinds of issues have been predominant in previous research (according to the literature accessible to me): "The aim of the investigations was to find a method for calculating the right position and size of the holes of a woodwind instrument. This can be useful for designing a new instrument or changing an existing one." Research with this expressed goal may first and foremost be useful to the professional musician as a performer, since he may expect new insights to assist the manufacturer in designing instruments for easy intonation and with the timbre desired by the expert player. It is different for the clarinet teacher, however. Among his tasks is the very essential one of guiding his pupils in their struggle to establish the <u>embouchure</u>,[2] which is so crucial in clarinet playing. The scholarly literature referred to above hardly offers much information that will assist him in accomplishing this.

---

1) Henri Bouasse's works, which appeared some decades earlier, were a valuable source of inspiration (Bouasse 1929; 1930). One of the landmarks of this research is A.H. Benade (cf. Benade 1959; 1960; and also his book, <u>Horns, Strings and Harmony</u>, Anchor, Garden City, N.Y., 1960, which is intended as a textbook for the common reader). In more recent years, John Backus and C.J. Nederveen have likewise contributed significantly along much the same lines as Benade. A thorough and lucid survey of the major advances due to the research of the last few decades is available in Benade's book from 1976.

2) The term <u>embouchure</u> is here used in the wide sense associated with it in Webster's Dictionary: "(...) the mouthpiece of a wind instrument; also, the fitting of the lips and tongue to the mouthpiece in playing a wind instrument".

The organization of the present project reflects the author's opinion that conditions are favourable now for attempts to arrive at a better understanding of the rôle of the embouchure as a component of clarinet playing. The possibility of arriving at results which may be applicable in a pedagogical framework was an essential driving force.

The approach chosen for this purpose consists in a continuation and further development of the kind of experimental work that is represented so instructively in Backus' pioneer work (Backus 1961, p. 806-809, and 1963, p. 305-313) with the goal of making it possible to throw light on various aspects of the acoustics of the clarinet which still await elucidation.

Since the investigation dealt with in this paper is specifically directed towards the excitation mechanism, it seems useful to start by presenting the following list of symbols to be employed in the remainder of the paper.

H_I: the constant slit height (i.e. the height of the slit between the reed and the lay at the tip of the mouthpiece) with no load on the reed,

H_{II}: the constant slit height if the reed is loaded by lip pressure only,

H_{III}: the constant slit height just before the reed begins to oscillate,

P: the excess air pressure on the outer side of the reed compared to a mean value of air pressure inside the mouthpiece during excitation (this corresponds to "blowing pressure" or "mouth pressure" in the case of natural blowing),

P_{thr}: the lower threshold for excitation of tone at a given embouchure.

4. Experimental setup

The basic components of the experimental setup is shown in fig. 2a. In a number of cases it proved advantageous to produce

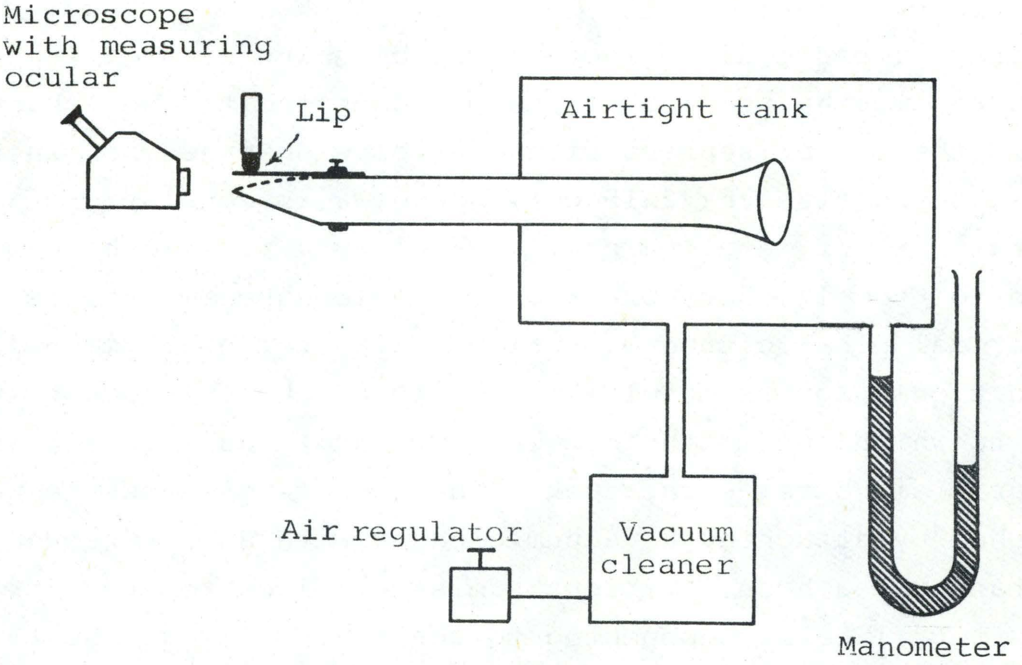


Figure 2a

Simplified presentation of the experimental setup used for a detailed investigation of the excitation mechanism.

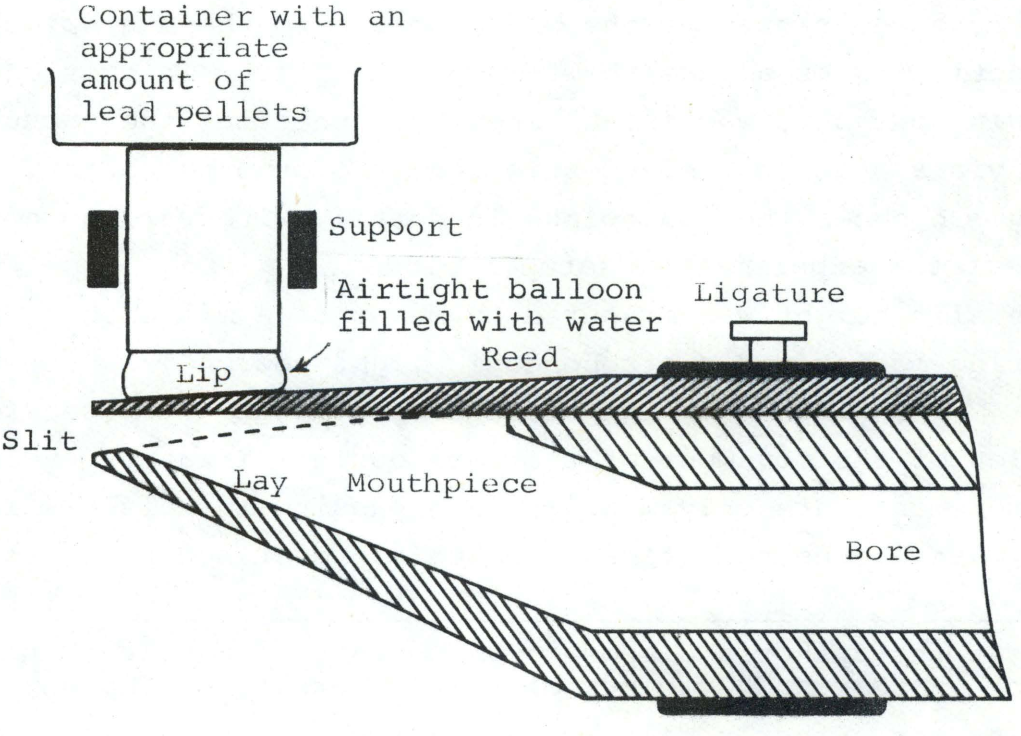


Figure 2b

An enlarged section of fig. 2a, presenting the embouchure end of the clarinet in more detail.

the desired pressure difference by establishing a "negative pressure" in the mouthpiece and resonator, compared to the ambient pressure (the latter representing the "blowing pressure" in this case). A tank (size 78x32x15 cm), whose walls consisted of a rigid framework of metal threads covered by a plastic bag, was attached to the clarinet bore - or (in some experimental series) a cylindrical plastic tube of an inner diameter of 17 mm - in such a way that only approximately one fourth of the instrument (viz. the end to which the mouthpiece was attached) was situated in the atmospheric air outside the tank. The "negative" pressure was established by attaching a vacuum cleaner with an air regulator to the tank via a hose. Various admission pipes for registration purposes were likewise connected to the cavity inside the tank. When using pressure above the atmosphere instead of "negative" pressure, the bag functioning as a tank for the air supply was fitted around the mouthpiece leaving the other, open end of the resonator free (in this case, the plastic bag was placed inside the metal framework for obvious reasons).

In one specific experiment (see fig. 8) the above mentioned plastic tube was closed at the end opposite to the mouthpiece, thus functioning as a quasi-half-wave resonator. In this case, the blowing pressure was established by connecting the vacuum cleaner via a hose to the acoustic center of the resonator.

Fig. 2b shows the mouthpiece in detail, with various components of the experimental setup included.

The slit height was measured by means of a microscope[1] with a measuring ocular and a stroboscopic light source[2], as shown in fig. 2a.[3] The flashes of the stroboscopic light were used for registration of the movement of the reed during vibratory cycles (these movements are only mentioned in passing in this preliminary report), but at the same time a number of reference points for the

---

1) Put at the author's disposal by courtesy of the Institute of Plant Anatomy and Cytology of the University of Copenhagen.

2) Put at the author's disposal by courtesy of the Audiologopedics Research Group at the University of Copenhagen.

3) The actual setup included a considerable amount of instrumentation and acoustic shielding not shown in the schematic drawings. (An essential part of this extra outfit was required for the recording of data which are not included in the present paper.)

movements of the reed were established.[1] These included the position of the edge of the reed when there is no load (in terms of air pressure or mechanical pressure) on the reed (H_I); further, the position of the reed loaded with lip pressure but without any difference in air pressure between the outer side of the reed and the cavity inside the mouthpiece (H_{II}); and, finally, the position of the reed when lip pressure is applied and there is, furthermore, a difference of air pressure that is just below the value making the reed begin to oscillate (H_{III}).

Almost all the experiments were made employing a geometry of the tube corresponding to the excitation of tones in the low register,[2] viz. D_3, A_3 and D_4.

The measurements testify to the existence of the following tendencies, which as such are generally accepted: rising pressure (in the case of alternating pressure) inside the mouthpiece contributes to a tendency toward greater reed aperture, and falling pressure contributes to a tendency toward closing of the aperture. However, the results yield a picture of the functioning of the embouchure which deviates on some points from the traditional view. For example, it has been known for a long time that there is no complete closure when notes are played piano, but authorities have been convinced that in mezzoforte, and under any circumstances in forte, there is an interval of complete closure which may last for as much as half of the duration of a complete cycle. This is not confirmed by my measurements, and other such details might be mentioned.

It should be noted in this context that the basic parameters - which fortunately are easily checked - such as P, P_{thr}, and the possibilities of varying the sounding frequency and the spectral composition of the tone, are in good agreement with measurements

---

1) The use of stroboscopic light was not essential for measuring these positions, since the reed was stationary in each position. However, the measurements were made in connection with measurements serving to determine the movement of the reed, and hence the stroboscopic light was used throughout.

2) The choice of these particular notes was primarily due to the fact that the literature accessible to me fails to give detailed information on the conditions with respect to the embouchure for intonation in the deep register (see, e.g., Backus 1963, p. 311).

made by others. Impressionistically, the timbre of the tone was adequate, disregarding the necessarily artificial character of a completely stationary tone deprived of transient phenomena (cf. Backus 1969, p. 193, and Coltman 1973, p. 418 on the character of mechanically blown tones).

Analyses of the spectrum of the tones produced by mechanical blowing show that the spectral composition can be varied apparently over the total range of variation possible with natural blowing, by manipulation of the parameters referred to above.

Thus it can be concluded that the artificial embouchure used in this investigation was sufficiently good.

5. Presentation of some experimental data with interpretation[1]

5.1 Threshold blowing pressure

One way of approaching the many phenomena associated with the embouchure is to gather experimental data on parameters which influence P_{thr}, i.e. the minimum blowing pressure necessary for the starting transient of the clarinet tone to appear (provided that certain other conditions are met). Diagrams figs. 3a, 3b, and 3c exemplify this approach by presenting data from a pilot experiment.

A setup of the type shown in fig. 2a above was used. In this case the resonator was the above mentioned cylindrical plastic tube adjusted for intonation of the note A_3. The clarinet reed was made of cane and of medium stiffness. During the experiments no moisture was applied to the reed (with the exception of a single experimental

---

1) In this project a new type of pick-up: the flow-meter, was introduced and, furthermore, the number of parameters to be studied was augmented by the addition of lip function, which does not appear in analogous fashion as a parameter in previous research. These novel features of the experimental approach entail a widening of scope of the research. As might be expected under such circumstances, part of this research must be characterized as pilot experiments. For this same reason, some of the results must be tentative, since further research may reveal relations and show the operation of factors which cannot be grasped at present. These reservations, however, apply specifically to results which are not included in the present paper. The relationships reported on in this paper can be interpreted with certainty.

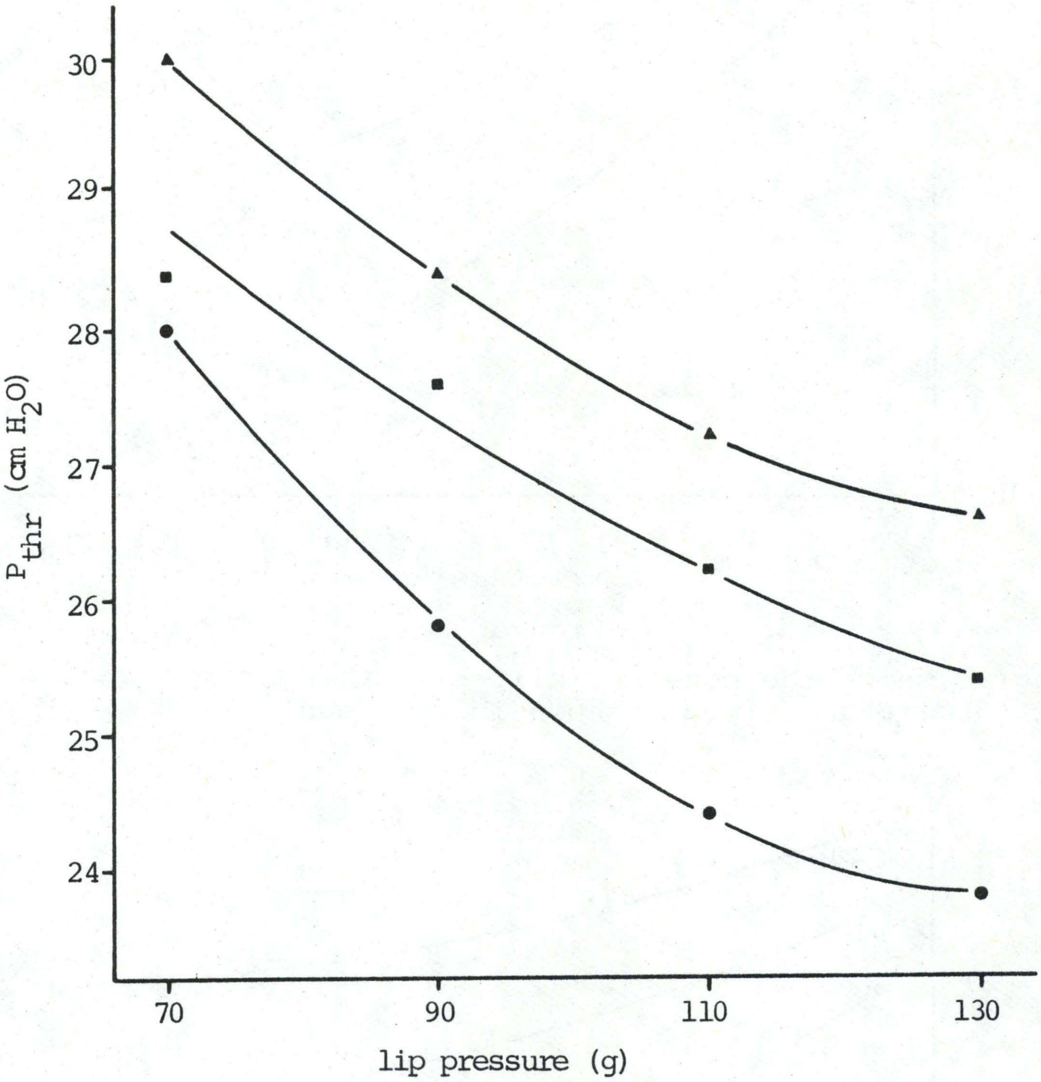


Figure 3a

Curves showing threshold blowing pressure versus lip pressure. The three curves represent different lip positions. In the case of the lowest curve the artificial lip was placed so as to leave the tip of the reed free, the area not covered by the lip extending 1 mm inwards from the edge; in the case of the next curve, this free area extended 2 mm inwards; in the case of the uppermost curve, it extended 3 mm inwards.

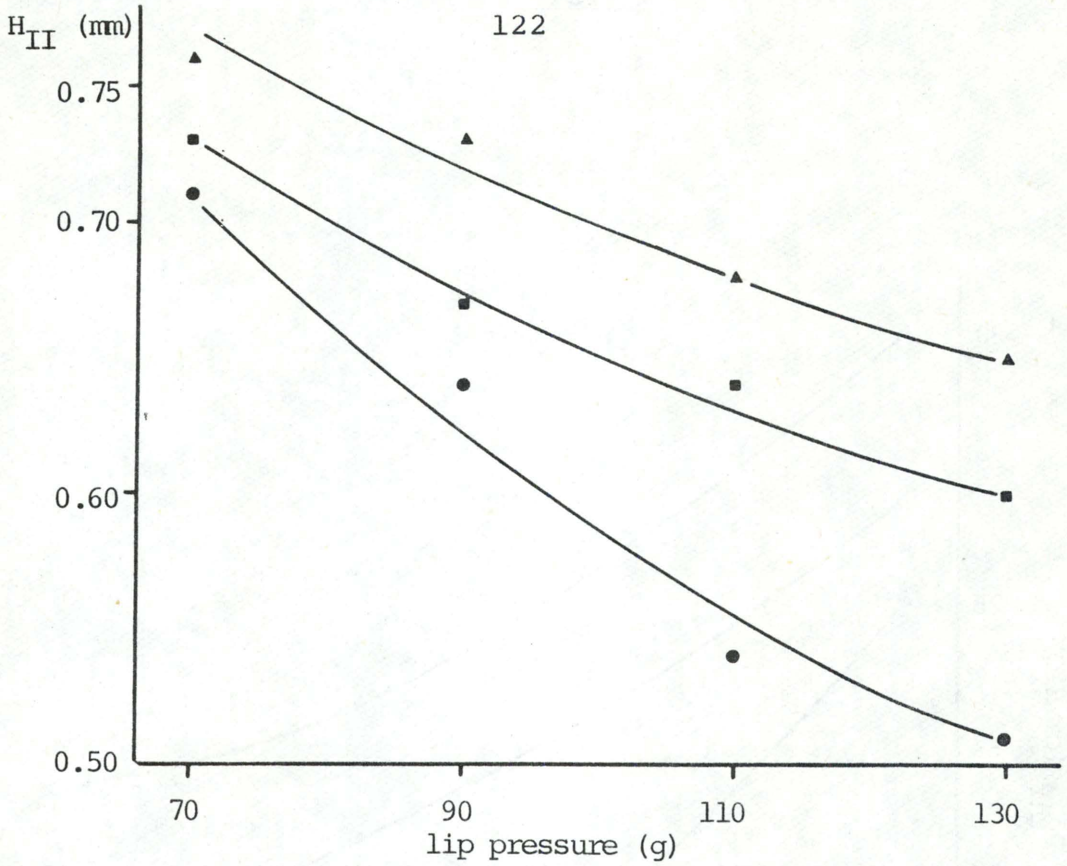


Figure 3b

The diagram shows the constant slit height (H_{II}) in mm as a function of the lip adjustments shown in fig. 3a.

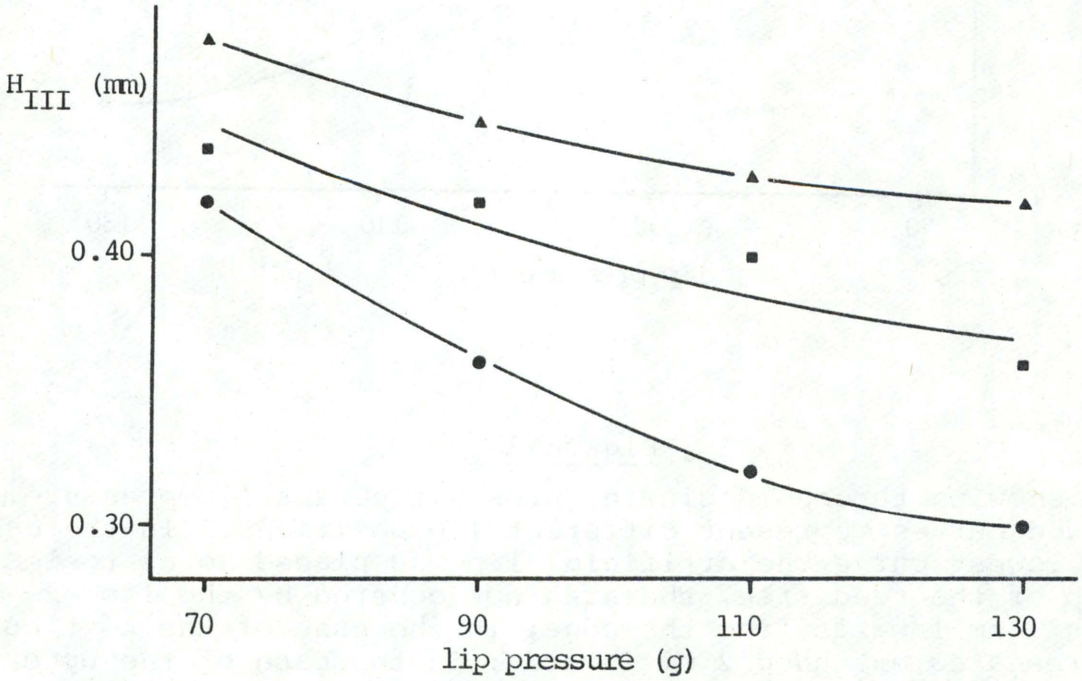


Figure 3c

This diagram shows the size of H_{III} in mm, i.e. the slit height measured just before the blowing pressure, which was gradually increased, reached the value $P = P_{thr}$ (the level at which periodical oscillations of the reed will begin). H_{III} is depicted as a function of the lip adjustments shown in fig. 3a.

series not accounted for in this report). Thus the moistness of the reed was conditioned by the natural humidity of the surrounding air only. This solution was chosen in order to keep the elasticity and the weight of the reed as constant as possible during measurements.

The diagram fig. 3a exemplifies how the parameters lip position and lip pressure influence the value of P_{thr}. It is seen that there is a trading relationship among these parameters: with a given lip position (measured as a displacement from the extreme position at the tip of the reed) and a given lip pressure, a certain blowing pressure is required to just excite a tone, viz. the P_{thr}; the further the lip is moved inwards, and the lesser lip pressure one employs, the higher is the value of P_{thr}, and vice versa.

On closer inspection several factors conditioning this trading relation can be singled out. One obvious cause is that if the lip position and the lip pressure are such that the slit height is reduced only very slightly, a much higher static pressure difference, P, between the outer side and the inner side of the reed is required in order to force the reed inwards toward the lay. Hence, under such conditions of lip position and lip pressure, P_{thr} is correspondingly higher. However, this interpretation is an oversimplification of the causal relations.[1] This is because any change in either P or the slit height, H, has a multiple influence on the entire sound generating mechanism (with reed and resonator as integral parts). Because of this complexity it is useful to sort out as a separate intermediate step a series of experiments serving to establish the influences of variations in H and in P on the entire oscillatory system.

---

1) To exemplify the relationships, it may be pointed out that H_{II} assumes practically the same value, according to figs. 3b and 3c, under the following three conditions: (i) Lip position A + 1 mm, lip pressure 90 g; (ii) Lip position A + 2 mm, lip pressure 110 g; (iii) Lip position A + 3 mm, lip pressure 130 g. However, the size of H_{III} is different, viz. 0.36 mm, 0.40 mm and 0.42 mm, respectively, see fig. 3c. - It may be added that the performance of the reed as a variable valve is highly dependent upon the size of H_{III}.

5.2 The dependence of resonance conditions on slit height[1]

In most of the series of experiments dealt with in this paper a test instrument, Buffet-Lyre S 1, model no. F 152064, was employed.[2] As mentioned already, the clarinet bore was replaced by a cylindrical tube (of an inner diameter of 17 mm) in some of the series. The mouthpiece was a Vandoren 5 R V.

As shown in the schematic drawings of the experimental setup for resonance measurements (figs. 4 and 5), the resonator was excited by an external sound source, viz. a Philips loudspeaker, type RH 541 MFB. Measurements were made in an anechoic chamber.

The resonance conditions were analyzed by a sweep-tone technique. The analyzer, B&K, type 2010,[3] was manually operated. By using this approach under otherwise favourable conditions, the author succeeded in acquiring more precise information than previous investigations did about the values of the resonant frequencies.

---

[1] In the following the frequency measurements are converted into values of frequency deviation, viz. the deviation of a resonant frequency from the frequency of the corresponding component in the complex tone in question. (100 cents = 1 tempered semitone.) The deviations of the resonant frequencies of a half-wave resonator can be calculated by means of this formula:

$$\text{deviation} = \log \frac{f_{\text{res n}}}{n \times f_{\text{blow}}} \times (1200/\log 2) \text{ cents}$$

With reference to a quarter-wave resonator, the formula is like this:

$$\text{deviation} = \log \frac{f_{\text{res n}}}{(2n-1) f_{\text{blow}}} \times (1200/\log 2) \text{ cents}$$

where $f_{\text{res n}}$ is the n'th resonant frequency, and f_{blow} is the blowing frequency of the tone in question.

This paper deals with measurements made with the test instrument adjusted for the notes D_3 and A_3. In the calculation of frequency deviations it is assumed that the blowing frequencies of the tones in question are 146.7 Hz and 220 Hz, respectively.

[2] The test instrument was placed at the author's disposal by courtesy of Marno Sørensen Instrument Dealers, Copenhagen. The staff of this firm kindly assisted the author in modifying the clarinet by closing the radial holes (which are otherwise available for changing the acoustic length of the clarinet by means of the fingering) and in other ways.

[3] Put at the author's disposal by courtesy of the Acoustics Laboratory of the Danish Technical University and Brüel & Kjær, Nærum.

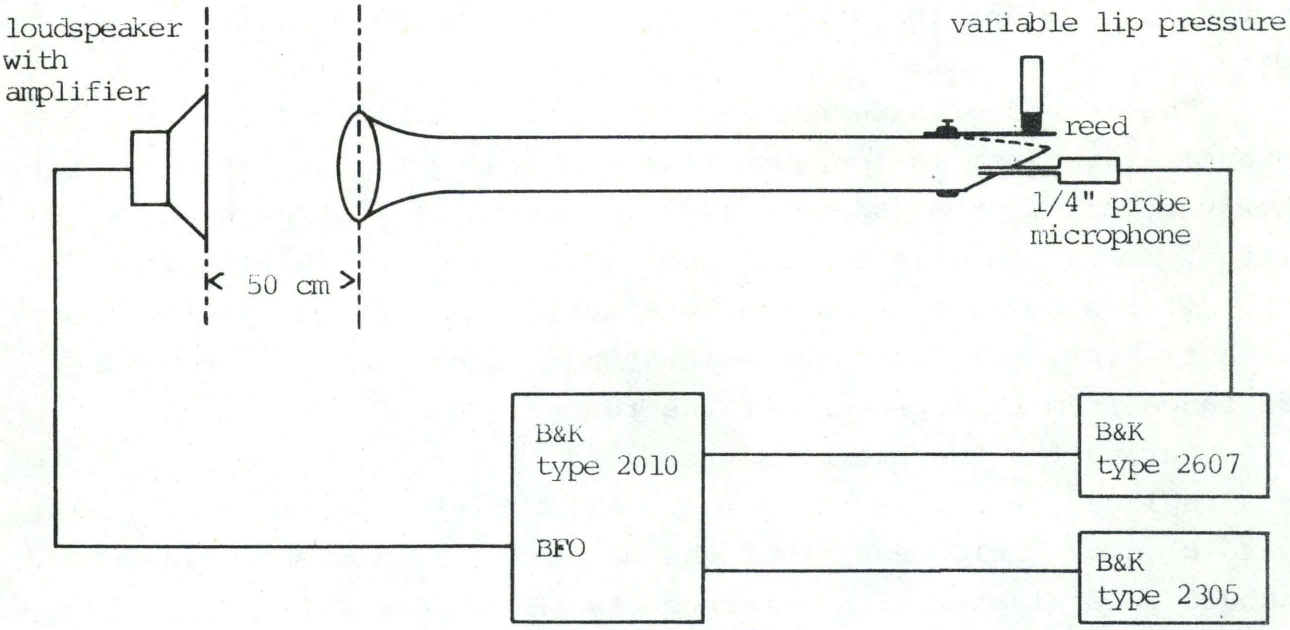


Figure 4

Schematic drawing of the experimental setup employed in measurements of the resonant frequencies of the resonator with external excitation.

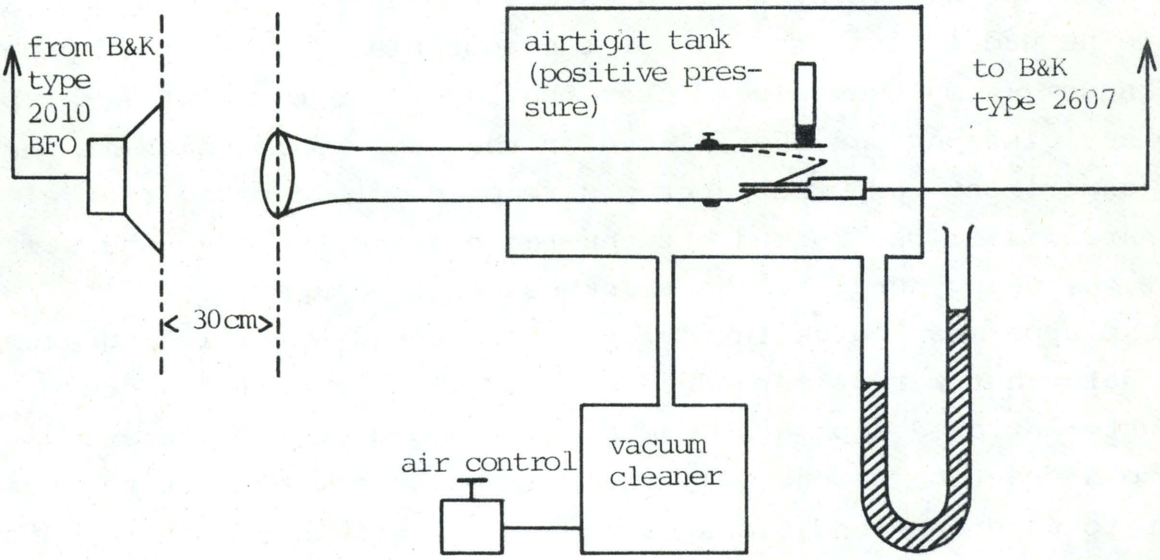


Figure 5

Schematic drawing of the setup employed in measurements of resonant frequencies (cf. the setup in fig. 4) with P added as a variable. In this case a "positive" pressure was established in the tank which included the upper end of the clarinet. Note that in the case of resonance measurements P did not represent a "blowing pressure" in the traditional sense, since the slit height was kept constant for such measurements.

(In return, the measurements had to be confined to relatively few notes, as against the greater number of frequencies investigated with the automatic technique of Backus.)

The method of external excitation cannot be used to obtain quantitative information about the amplitudes of the resonant peaks, however (cf. Backus 1968 and 1974). But as will be pointed out later, it is possible to get some qualitative information about the differences as long as the excitation is made in exactly the same fashion, viz. with the same sound source placed at the same distance from the open end of the resonant tube.

During the measurements presented in fig. 6, the clarinet reed was replaced by a plasticine pad. All radial holes in the clarinet were closed. This means that the measured resonant frequencies should be evaluated with reference to the note D_3 (fundamental frequency 146.7 Hz).

The lowest curve in fig. 6 (represented by a filled circle) shows the resonant frequency under conditions of total closure at the embouchure end of the clarinet, the plasticine pad forming an airtight closure with the lay.

The next lowest curve in fig. 6 (represented by filled triangles) shows data that were recorded after the conditions at the embouchure end had been modified as follows: before placing the plasticine pad against the lay of the mouthpiece, a sheet of paper (thickness 60 my) was placed over the lay so as to cover its tip. The plasticine pad was then placed in the same position as in the first experiment, and the paper was removed with care so that a very narrow slit was formed between the pad and the lay, the slit height, H, being equal to the thickness of the paper.

The uppermost curve in fig. 6 (represented by filled squares) shows data that were recorded after the slit height (adjusted by the just mentioned approach) had been enlarged from the previous size to a fraction of one millimeter, H being now similar to average height found under conditions of mechanical blowing, according to the author's estimate.

In fig. 6 the two uppermost curves have been cut off (for considerations of space), so that the plots of the frequency deviation of the first harmonic are missing. The values were as follows: for the smaller value of H (the data plotted with filled

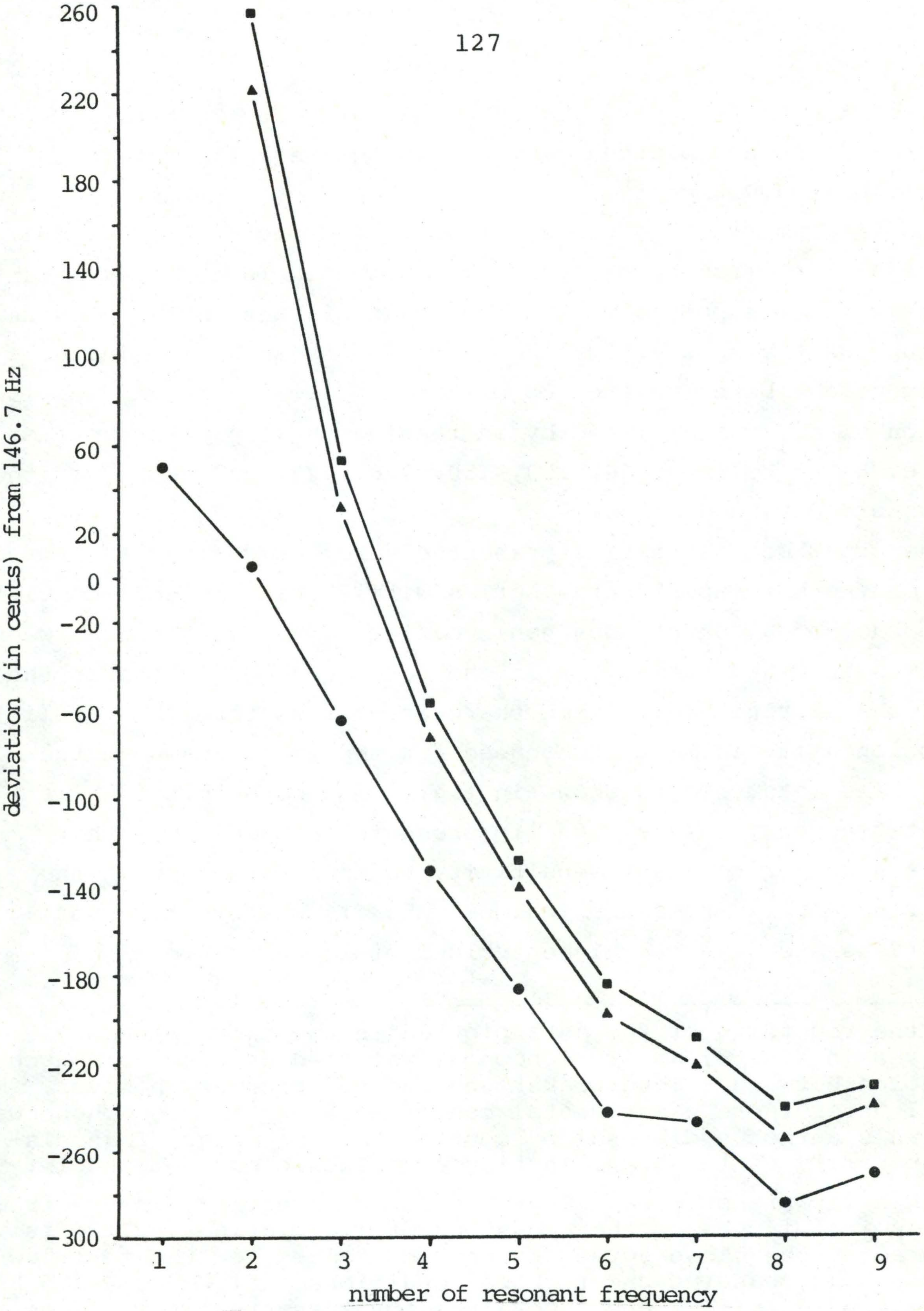


Figure 6

Diagram showing the result of three series of resonance measurements made with the setup shown in fig. 4, but with a pad of plasticine replacing the reed. This diagram shows the deviation (in cents) of the actual resonant frequencies from those of an ideal quarter-wave resonator tuned for a first resonant frequency of 146.7 Hz (corresponding to the note D3). The curves represent different degrees of aperture at the embouchure end (see text for details). - For limitations of space the deviation of the first harmonic is omitted from two of the curves; the values are respectively 880.8 cents (for the next highest value of H) and 937.4 cents (for the highest value of H).

triangles) the deviation was 880.8 cents, and for the higher value of H (the data plotted with filled squares) the deviation was as much as 937.4 cents.[1]

The diagram fig. 7 gives a more detailed picture of the dependence of the first harmonic on slit height. In this experiment the clarinet was supplied with a reed made of cane, but otherwise the setup was the same as for the measurements shown in fig. 6. The slit height being controlled by the artificial lip, a stepwise reduction of H_{II} was achieved by increasing the lip pressure from 0 g (i.e. $H_{II} = H_{I}$) over 50, 100, 150, 200, 223, 249 to 269 g (the lip pressure giving $H_{II} = 0$).[2]

The experimental results presented here, as well as the results of the many other experimental series within this project, confirm the presence of a conspicuous nonlinearity in the relation between variations in H and variations in the resonance conditions in the bore of the clarinet (even when there is no flow through the slit). This nonlinearity is most pronounced for the first resonant frequency. The measurements shown in fig. 7 indicate that the first resonant frequency varies with lip pressure in such a way that there is a rather constant sensitivity to this parameter in the lowest range of lip pressure (ca. 0 - 150 g), whereas the sensitivity rises steeply with higher values of lip pressure.

---

1) In the recording of the data plotted in the two uppermost curves in fig. 6, an STL-Ionophone was used as a sound source (the choice between a loudspeaker and a Ionophone was, however, immaterial under these experimental conditions). - The Ionophone was put at this author's disposal by courtesy of the Speech Transmission Laboratory of the Royal Institute of Technology, Stockholm.

2) With a lip pressure of 269 g, there was effective closure from the point of view of the acoustics of the clarinet. This is documented by the close agreement between values for the four lowest resonances measured under these conditions (151.2 Hz, 441.5 Hz, 706.0 Hz, and 951.7 Hz) and the values for these resonances measured (two days earlier) with the plasticine pad replacing the reed and forming a tight closure with the lay (resonant frequencies: 151.1 Hz, 441.7 Hz, 706.9 Hz, and 951.9 Hz). (The ambient temperature was 25^{o} C in both experiments.)

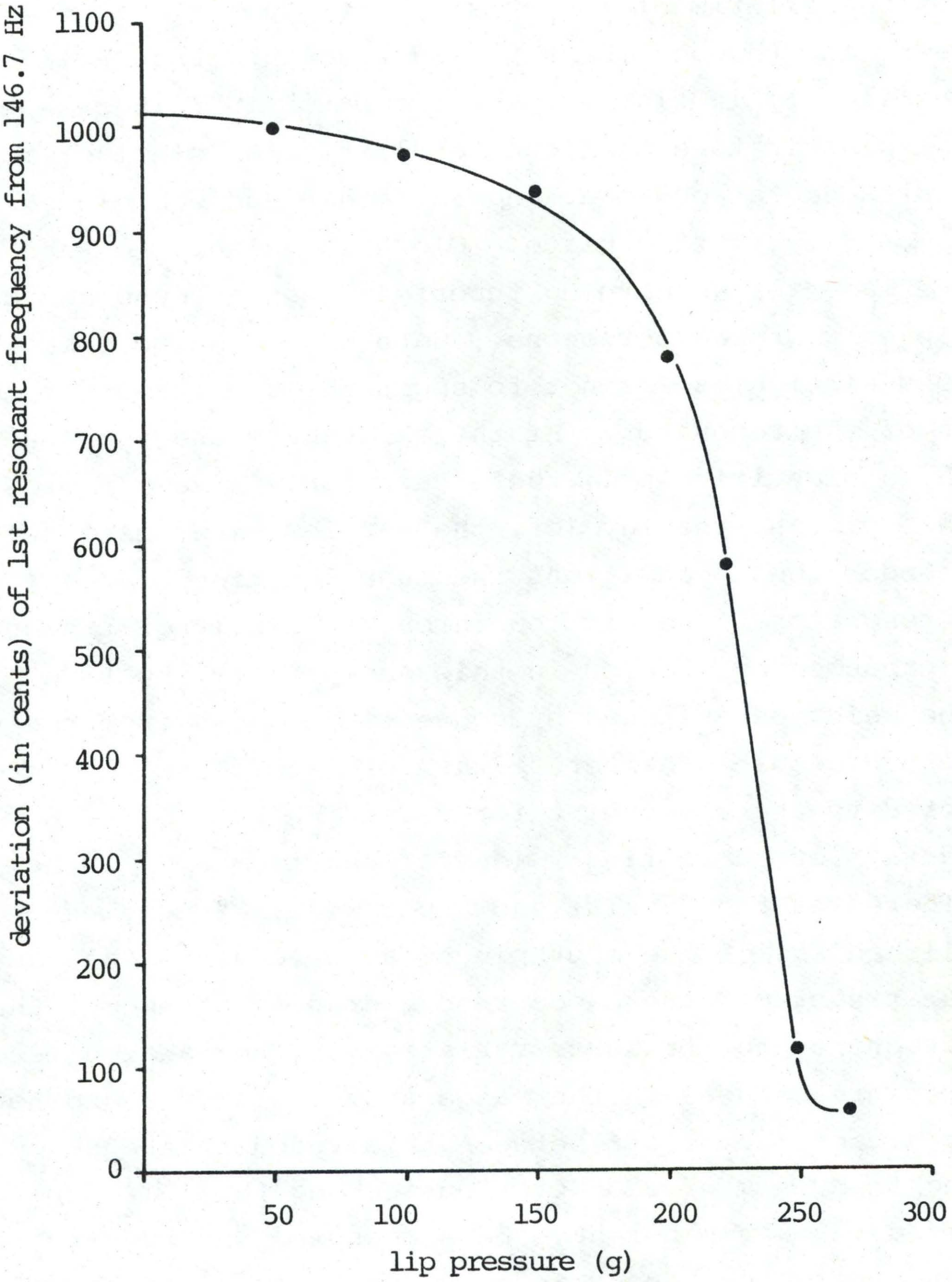


Figure 7

Diagram showing how much the first harmonic deviates (in cents) from the blowing frequency of 146.7 Hz (the note D3) as a function of lip pressure. Measurements were made at lip pressures of 0, 50, 100, 150, 200, 223, 249 and 269 g. Experimental setup as in fig. 4.

5.3 The dependence of resonance conditions on flow through the slit

One of the experimental setups used to investigate the relation between flow through the slit and resonant frequencies is shown schematically in fig. 8. The resonator in this case was the cylindrical plastic tube mentioned earlier. The clarinet mouthpiece was attached at one end, and the other end was closed by means of a closely fitting plate. Close to this latter end admittances were made for an earplug functioning as a sound source, and for the tip of a probe microphone. Care was taken to make these perforations airtight so as not to spoil the conditions at this closed end of the resonator. At the embouchure end the reed was replaced by a plasticine pad, there being only a very narrow slit between this pad and the lay (cf. the earlier experiment described in 5.2). Under these conditions the tube functioned as a (quasi-) half-wave resonator. The airflow through the slit of the mouthpiece was produced by suction in this setup, the "negative pressure" in the tube being established by connecting a hose from the vacuum cleaner to the acoustic center of the resonator. An U-tube manometer was connected to the same point for measuring P.

The data plotted in fig. 9 (deviations from F_{blow}, the latter being indicated on the Y-axis) show that when a "negative pressure" was established inside the mouthpiece (as well as in the rest of the resonator system), there occurred a drastic change of the first resonant frequency for pressure variation in the range 0-10 cm H_2O, whereas pressure changes in the range above 10 cm H_2O did not substantially affect the offset between this resonant frequency and the blowing frequency of 220 Hz. Considering that all experimental data reported in this paper have been measured in order to throw light on the parameters determining P_{thr}, it is evident that this is a crucial point.

5.4 A qualitative assessment of the resonator Q as a function of (i) flow and (ii) slit height

As mentioned in section 5.2, the external excitation does not make it possible to elicit quantitative information about the Q of the bore of a wind instrument. An arrangement that would make

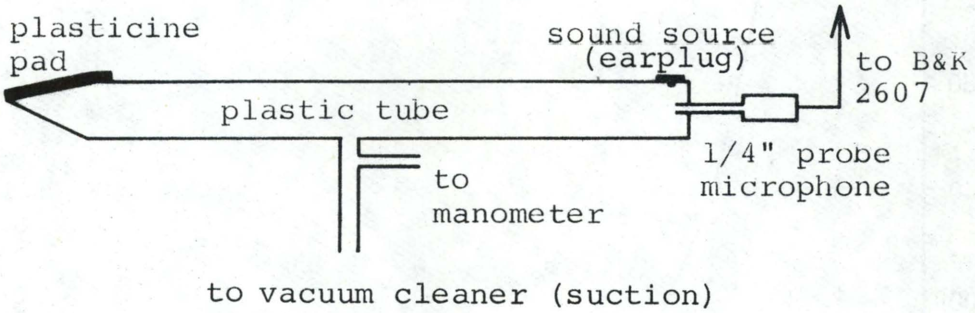


Figure 8

Schematic drawing of a half-wave resonator with a mouthpiece attached at one end, and an earplug, used as a sound source, and a probe microphone inserted at the opposite, closed end. At the mouthpiece end a tiny slit was formed between the lay and the plasticine pad which otherwise covered the opening. The hose attached to the acoustic center of the tube served to remove air by suction so as to cause a flow inwards through the slit at the mouthpiece end because of the "negative pressure" in the system.

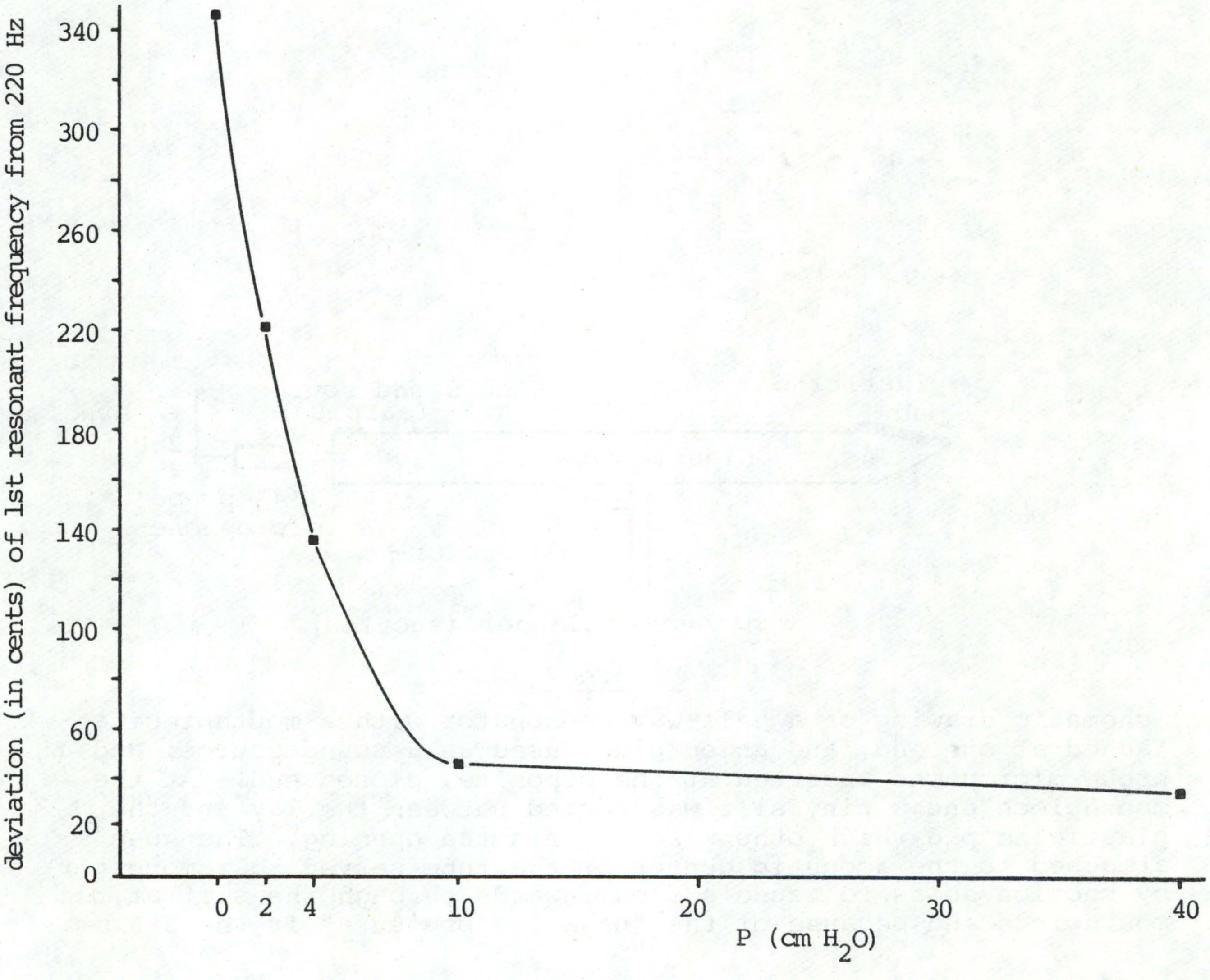


Figure 9

Diagram showing the deviation of the first resonant frequency from a reference blowing frequency, as a function of the "negative pressure" inside the system. The reference, viz. 220 Hz, represents the fundamental frequency of the tone produced by the setup with an artificial embouchure (as mentioned earlier), and a suitable "negative pressure" inside the system. (The blowing frequency had been determined in some earlier experiments.) - The setup shown in fig. 8 was employed.

such measurements possible is described by Backus (1974, p. 1266ff). It is, however, possible to acquire certain qualitative information. Figs. 10 and 11 show the results of some series of measurements illustrating how the Q varies as a function of lip pressure or of blowing pressure. (Cf. figs. 4 and 5.)

Each figure presents a series of curves recorded one after another on a B&K Level Recorder type 2305 attached to the analyzer (B&K type 2010). The curves are placed with an offset of 10 dB between each pair of curves in order to keep them visually apart. In each figure vertical (broken) lines serve to mark the harmonic frequencies for a hypothetical, naturally blown tone with a fundamental frequency of 220 Hz. The uppermost curve in fig. 10 as well as fig. 11 is a reference curve showing the frequency response of the clarinet with fingering for the note A_3 and complete closure at the embouchure end. The bottom curve shows the frequency response of the loudspeaker used as a sound source. These measurements of the frequency response were performed in the anechoic chamber, without the clarinet present and with the microphone placed at a distance from the loudspeaker corresponding to the length of the clarinet plus 50 cm. The other curves (i.e. the resonance curves) were measured with the same setup except that the clarinet was placed adjacent to the probe microphone (thus with a distance of 50 cm between the open end, the flare, of the clarinet and the loudspeaker). The analysis was made automatically, using the sweep tone from the tone generator of the analyzer, geared to the level recorder.[1]

---

1) The determination of frequencies from these automatically recorded curves is inherently less accurate than the results obtained with manual operation of the analyzer, as used for the other experimental series reported on here (cf. section 5.2). Backus has estimated that frequencies determined from automatically recorded curves "should be accurate to within some 20-30 cents" (Backus 1968, p. 1276). As for the frequency axis, Backus' curves and the curves shown in this paper differ very much. (Backus used a paper speed three times lower than the one employed by the present author, but it seems reasonable to assume that the accuracy of frequency measurement does not differ markedly for that reason, the limitations of accuracy having to do with the use of automatic registration rather than with the paper speed. Thus the determination of frequencies from the curves in the present case may be assumed to exhibit the same degree of accuracy as Backus' measurements.)

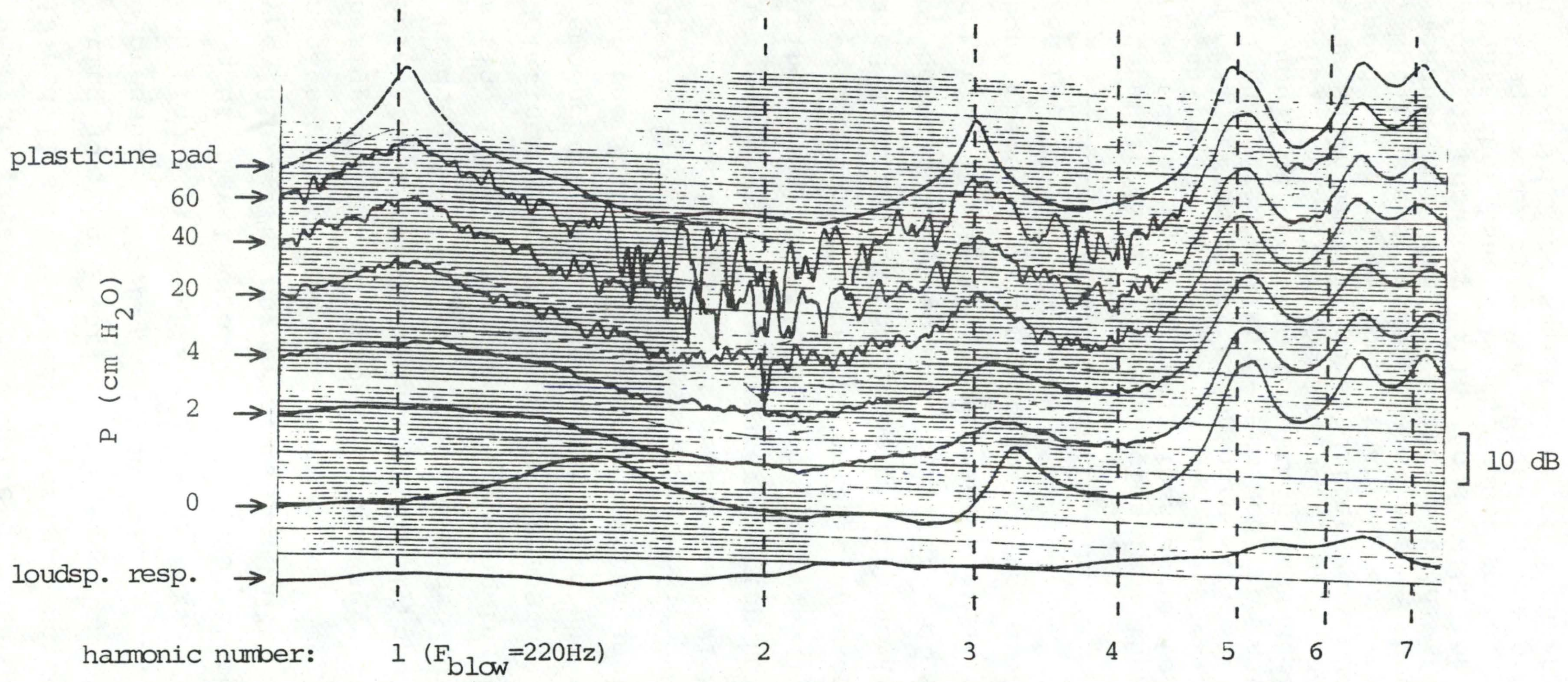


Figure 11

Automatically recorded impedance curves made with the setup shown in fig. 5 and fingering as for the note A_3. The uppermost and the lowest curves are reference curves corresponding to those of fig. 10. The intermediate curves were recorded with the following values of blowing pressure: 0, 2, 4, 20, 40 and 60 cm H_2O. A hard-walled slit was used in order to keep H constant during the recording of the impedance curves. - The skewness of the graph is due to optical distortion in the copying process.

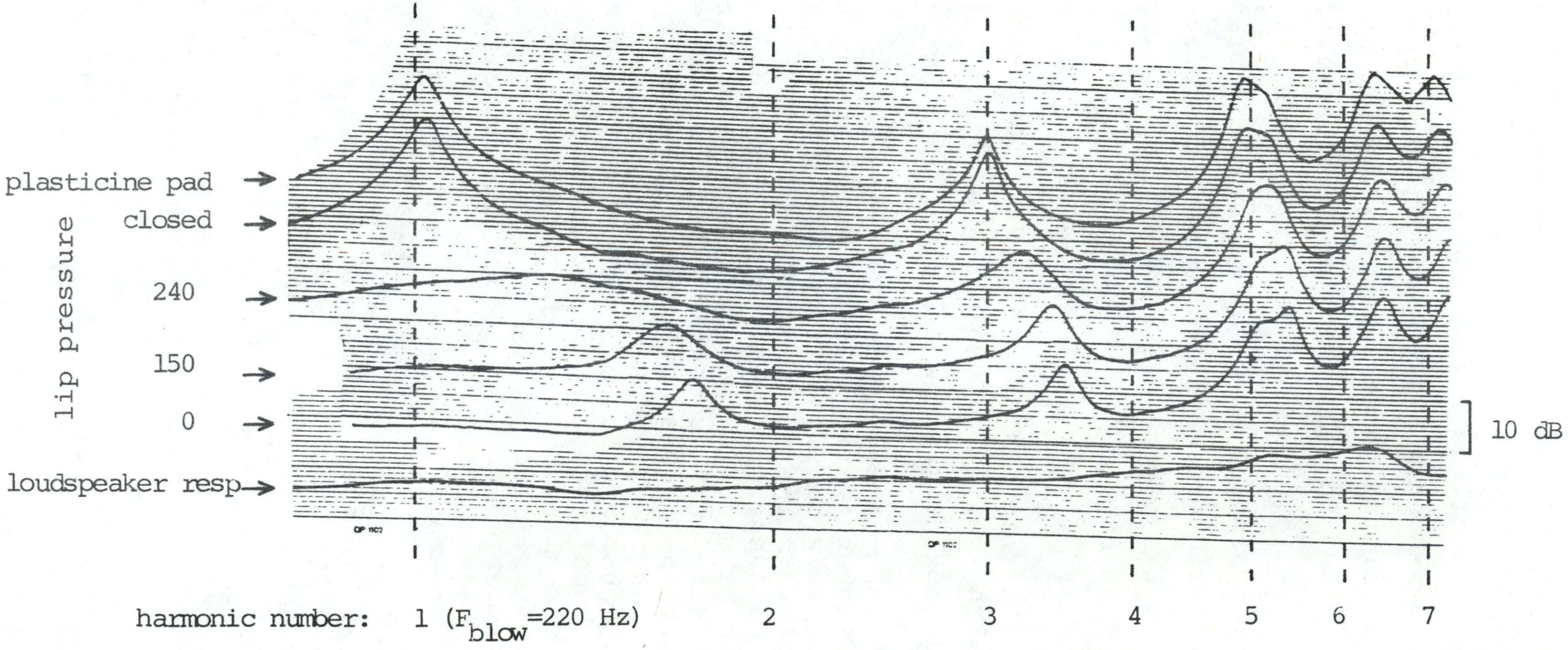


Figure 10

Automatically recorded data produced with the setup shown in fig. 4, and with fingering as for the note A_3. The lowest curve shows the amplitude response of the loudspeaker with the clarinet removed from the setup in the anechoic chamber. Curves nos. 2 to 4 from bottom are impedance curves for lip pressures ranging from 0 g over 150 g and 240 g to "high lip pressure". The uppermost curve, which was recorded on a different occasion, is an impedance curve showing the performance of the system with a plasticine pad replacing the reed and forming an airtight closure with the lay. - Here and in fig. 11 the curves are placed with a mutual offset of 10 dB. The vertical broken lines in figs. 10 and 11 mark the frequencies of the harmonics corresponding to a fundamental of 220 Hz. - The skewness of the graph is due to optical distortion in the copying process.

The difference between the two series of experiments was that in the case of fig. 10 the variable parameter was lip pressure (or the resulting slit height, which is dependent upon lip pressure), there being no suction, whereas in the case of fig. 11 the reverse obtained: the slit height was fixed, the reed being replaced by a metal strip leaving only a gap with a height of a small fraction of a mm at the tip of the mouthpiece, and the pressure drop across the slit was varied instead. In the first series the lip pressure was chosen so as to produce a variation in H_{II}/H_I ranging from 1 (at 0 g lip pressure) to 0 (at maximum lip pressure). In the second series, i.e. with pressure drop across the slit as parameter, the following values of the pressure difference were used: 0, 2, 4, 20, 40, and 60 cm H_2O.[1]

Although the independent variable controlled in the experimental series presented in fig. 10 was lip pressure, the parameter of direct interest is slit height. The two being obviously directly correlated (see fig. 3b concerning the general relationship

---

[1] In reports on related investigations of the impedance as a function of airstream it is mentioned that turbulence in the airstream causes trouble for the measurements (Backus 1963, p. 311; Coltman 1973, p. 418). The question, then, is why the present series of recordings could be made in the presence of flow without posing serious problems. Note in this context that there were no sudden increases of the internal cross-sectional area downstream, as seems to have been the case with the setup used by Coltman for his measurements. On the contrary, the cross-sectional area inside the mouthpiece expands in such a way that it performs like an effective diffuser of the air streaming through the mouthpiece (cf. Ower and Pankhurst 1969, p. 110ff.). Thus it is very possible that there may, in specific circumstances, have been turbulence affecting Coltman's measurements. Presumably, in designing the mouthpiece, instrument designers have attempted to minimize such turbulence by giving the mouthpiece a quasi-conical shape. This explanation of the shape of the mouthpiece is corroborated by the fact that it is crucial to have laminar flow in the vicinity of the slit, even at high rates of flow. The better the flow passes straightly through the slit and onwards along the inner side of the reed, the more effectively will the (alternating) height of the slit be influenced by the Bernouilli force during tone production.

If the present measurements were less affected by noise from turbulent air than seemed to be the case in the experiments made by Backus under otherwise identical conditions, the reason probably is that the present author used a sweep tone with a very high sound pressure level (the loudspeaker producing its maximum effect according to specifications), because this had proved expedient. Thus the signal passing via the bore of the clarinet into the microphone probe in the mouthpiece has been so intense that noise from the flow has not been able to override it completely and hence to upset the recording of the response curve of the instrument.

between lip pressure and H_{II}), it may be concluded that these curves give just as reliable information on the impedance characteristics of the resonator as a function of slit height.

The most interesting feature of figs. 10 and 11 is that essentially the same changes occur, both with respect to resonant frequencies and with respect to the Q of the resonator at the various resonances (as indicated by the peakedness of the curves), irrespective of whether one or the other parameter is chosen as the independent variable. This similarity is most conspicuous for the first resonant mode. Note that for this mode one finds the highest resonant frequency when H is at its maximum, or P_{blow} at its minimum). Under these conditions one also observes a resonant peak indicating a fairly high Q. When the influence of the parameter in question is increased, changes occur which exhibit two distinct phases. In the first phase the resonant frequency is shifted downwards by a substantial amount, and the flattening of the peaks indicates a very considerable reduction of the Q (see the first resonant peak at 240 g lip pressure in fig. 10 and, correspondingly, at a P_{blow} of 4 cm H_2O in fig. 11). The second phase (i.e. the situation when the influence of the parameter in question is increased further) is characterized by two features: (i) the Q is increased (i.e., the influence on Q is reversed), and (ii) the resonant frequencies asymptotically approach a set of values which are only slightly higher than the ones measured with complete closure at the mouthpiece end (see the uppermost curve in figs. 10 and 11).

Since the presentation of data in this section is confined to analyses of the conditions for the note A_3, it is essential to emphasize that tracings made for other notes recorded under identical conditions, exhibit quite analogous shifts in resonant frequencies and Q as described for the note A_3.

Returning to fig. 9 it may be added that a similar effect - with two distinct phases - was found in experiments using the half-wave resonator and producing a stepwise increase in the flow through the slit by varying P. There was, however, a difference of degree in that the aperture occurring at the tip of the mouthpiece under these conditions was so small that the effect of the flow on the acoustic impedance and hence on the resonant properties was much less pronounced. Therefore, the first resonant frequency and the Q (which is not documented in this report) changed less as a function of variations in the parameter involved than in the experiments to which fig. 11 refers.

6. Discussion

Curves such as the one in fig. 7 and the next lowest curve in fig. 10 (lip pressure: 0 cm H_2O) suggest that as long as the combined reed-resonator system of the clarinet is not influenced by the player's embouchure, the tube acting as a resonator is a hybrid transitional type, being something in between a half-wave resonator and a quarter-wave resonator, although it is more closely related to the latter. It is seen from the same figures that the characteristics of the resonator become successively more like those of a quarter-wave resonator the more the player reduces the cross-sectional area of the aperture at the embouchure end (i.e. the slit) by means of his lip pressure.

It should be mentioned in passing that the blowing pressure applied by the player will contribute further to move the reed inwards toward the lay because of the pressure difference between the layers of air at the outside and at the inside of the reed. At the flexible tip of the reed the pressure drop is, furthermore, influenced by the Bernouilli effect which varies during a vibratory cycle because of the flow through the slit (cf. Benade 1976, p. 438). The issue here is, however, the acoustic effect of flow on the characteristics of the resonator as documented by the curves in fig. 11 and the curve in fig. 9. In agreement with results from numerous other similar measurements performed as part of the present project, these curves show that the clarinet bore behaves like a quasi-quarter-wave resonator if the blowing pressure exceeds a certain value, even though the clarinet is not totally closed at the embouchure end. This relationship is, of course, particularly important for the interpretation of the behaviour of the clarinet when there is not a complete closure between the reed and the lay during any interval within a vibratory cycle.

We return to the issue raised in section 1 of this paper. For obvious reasons a closure of the slit during tone production cannot be accomplished by using lip pressure (which would prevent the reed from oscillating). Thus the only type of sustained closure available is "acoustic closure".

The curve in fig. 9 may seem to indicate that a blowing pressure slightly exceeding 10 cm H_2O would suffice to establish this "acoustic closure". The values measured for P_{thr} spread evenly from some 10 to some 30 cm H_2O. It should be noted, however, that the clarinet reed used in these experiments had not been exposed to humidity except for that of the atmospheric air. In natural blowing the air exhaled by the player will cause the reed to absorb humidity because of condensation. If the reeds had been exposed to a similar absorption of humidity in connection with the present measurements, all measures of P_{thr} given in this report would have been slightly lower (the difference being of the order of a few cm H_2O, according to this author's estimate).

It may be added that in laboratory experiments involving artificial embouchure it is not very difficult to adjust the embouchure to a much lower P_{thr}. Thus, by using a soft plastic reed it proved possible to get a P_{thr} of no more than 6.6 cm H_2O. (In that extreme case the excitation was of no importance from the point of view of musical applications since no audible tone was produced. Vibrations of the air were, however, recorded via the probe microphone picking up the alternating pressure inside the mouthpiece, and were visible on the oscilloscope screen.)

This concludes the report on experiments serving to determine the influence of the parameters in question on P_{thr}.

7. Goals of the research project in its totality

As stated in the introductory section, this paper is confined to one aspect of the complex of relationships which must be studied in order to arrive at a complete description of clarinet embouchure. The results presented here are useful as a basis for further research on the acoustics of the clarinet and the conditions obtaining in blowing. This further research has been undertaken on the basis of a wealth of experimental data gathered during the work with the project but in part still awaiting processing.

One of the obvious tasks for further investigation is to examine what happens if the blowing pressure is increased successively as the only independent parameter, for given values of the other two parameters: lip position and lip pressure. The present author has found that this is a useful way of acquiring insight into the influence of the just mentioned parameters on the spectral composition of the tone, as well as on the dynamics of clarinet playing. It would lead too far to give details here, but it may be mentioned that the data clearly indicate that it is quite crucial both for the spectral composition of tones and for the dynamics how the two parameters conditioning H_{II} (cf. the graph fig. 3b) are weighted in relation to one another. Thus the following passage from Nederveen (1969, p. 36): "... the player has a wide choice of lip positions. So the mere excitation of the instrument is not very critical" can hardly be considered generally valid. It must be understood in the context of the specific assumptions on which it was based.

According to the experience gained from the experimental data of the present project it must be the case that a shift in lip position of the order of 1 mm or less may have a significant effect, e.g. on the timbre of the tone and on the dynamic variations which the clarinet player is able to perform.

Acknowledgements

This project was financed by a grant from the Danish Research Council for the Humanities (Statens humanistiske Forskningsråd).

References

Backus, John 1961:	"Vibrations of the reed and the air column in clarinet", *JASA* 33, 6, p. 806-809
Backus, John 1963:	"Small-vibration theory of the clarinet", *JASA* 35, 3, p. 305-313
Backus, John 1964:	"Effect of steady air flow on the resonance frequencies and Q's of the clarinet", *JASA* 36, p. 2014 (paper S 4)
Backus, John 1968:	"Resonance frequencies of the clarinet", *JASA* 43, 6, p. 1272-1281
Backus, John 1969:	*The acoustical foundations of music*, New York: Norton
Backus, John 1974:	"Input impedance curves for the reed woodwind instruments", *JASA* 56, 4, p. 1266-1279
Benade, A.H. 1959:	"On woodwind instrument bores", *JASA* 31, 2, p. 137-146
Benade, A.H. 1960:	"On the mathematical theory of woodwind finger holes", *JASA* 32, 12, p. 1591-1608
Benade, A.H. 1976:	*Fundamentals of musical acoustics*, Oxford Univ. Press, New York
Bouasse, H. 1929, 1939:	*Instrument à vent*, Paris: Librairie Delagrave
Coltman, John W. 1973:	"Mouth resonance effects in the flute", *JASA* 54, 2, p. 417-420
Nederveen, C.J. 1969:	*Acoustical aspects of woodwind instruments*, Amsterdam: Fritz Kunt
Ower, E. and R.C. Pankhurst 1969:	*The measurement of air flow*, Oxford: Pergamon Press