LUND UNIVERSITY
DEPARTMENT OF LINGUISTICS
General Linguistics
Phonetics

WORKING PAPERS

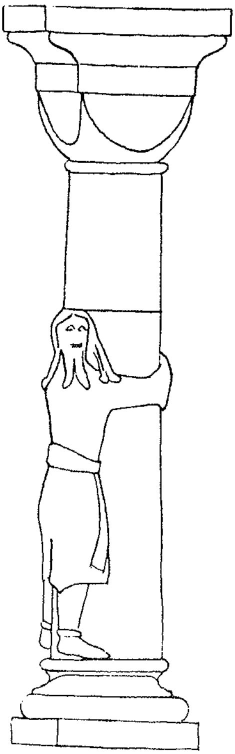
25 • 1983

LUND UNIVERSITY
DEPARTMENT OF LINGUISTICS
General Linguistics
Phonetics

WORKING PAPERS

25 • 1983

ISSN - 0280 - 526X



Department of Linguistics and Phonetics
Helgonabacken 12
S-223 62 LUND

This issue has been edited by Gösta Bruce and Eva Gårding

ANNUAL REPORT

Our present interests and activities are best represented by a list of the titles of our two seminars. One of them is called Language Speech Sound Hearing and is held in cooperation with the Dept. of Logopedics and Phoniatrics. The other one, simply called the Friday Seminar is a free-wheeling research seminar mainly for the benefit of the graduate students.

Language Speech Sound Hearing

20.9	On musical therapy	Olu-Birgit Jeppson, Dept. of Music, Lund University
4.10	How should we define phonetics as a subject?	Sven Öhman, Dept. of Linguistics, Uppsala University
18.10	Presentation of the project, Immigrants and communication: How to understand and be understood	Robert Bannert, Hans Dahlbäck, Gisela Håkansson, Kerstin Tevajärvi
8.11	Treatment of functional voice disorders	Peter Kitzing, Malmö Hospital
15.11	Phonological universals and phonological development	Charles Ferguson, Stanford, USA
6.12	Report from a project: Phonetical descriptions of some languages in the third world	Mona Lindau-Webb, Kjell Norlin, Jan-Olof Svantesson
31.1	The concept of selforganizing systems in physics, biology, sociology and linguistics	Björn Lindblom, Dept. of Linguistics, Stockholm University
14.2	Dialects in Southern Sweden	Bengt Pamp, Dept. of Dialects and Placename Research, Lund University
28.2	Sound and sound effects	Bertil Johansson, Dept. of Technical audiology, KTH, Stockholm
21.3	Some contributions to the Speech Sound Hearing symposium in Stockholm	Dept. of Logopedics and Phoniatrics

28.3	How is a vocabulary organ-nized?	Åke Viberg
11.4	Reading and writing diffi-culties in the light of present research	Ingvar Lundberg, Dept. of Psychology, Umeå Univer-sity
25.4	Some contributions to the Speech Sound Hearing sym-posium in Stockholm and the Prosody, Normal and Abnormal Conference in Zürich	Phonetics Dept., Lund Uni-versity
9.5	The new training program for audiologopeds at Copenhagen University	Børje Frøkjaer Jensen, Uni-versity of Copenhagen
16.5	Therapy by music	Grete Lund, Viborg

<u>The Friday Seminar</u>

10.9	Report from the Linguistics Congress in Tokyo	Gösta Bruce
8.10	Register in the Austroasiatic language Va	Jan-Olof Svantesson
	The aerodynamics of phonation	Anders Löfqvist
15.10	Velopharyngeal plastic sur-gery and speech	Kjerstin Becker, Eva Borell
22.10	How can we measure hoarse-ness?	Ann-Christine Ohlsson
29.10	Tests on comprehension of Swedish spoken with a foreign accent	Robert Bannert, Kerstin Tevajärvi
12.11	Project report: Language and brain	Christina Hedqvist-Dravinš
19.11	Transitions and perception	David House
26.11	The pronunciation of French	Paul Touati
3.12	Perception of interrogative intonation	Anne-Christine Bredvad-Jensen
10.12	Intonation and pivots, syntax and semantics	Eva Gårding
4.2	A visit to the institute of phonetics in Aix en Provence	Paul Touati
11.2	The intonation of Finnish	Kjell Weimer
18.2	Invariance and variability	Anders Löfqvist

25.2	The outline of a thesis on prosodic interference in the Swedish of Finns	Kjell Weimer
4.3	Rehearsal for the Speech Sound Hearing Symposium in Stockholm 10-11 March	
18.3	Invariance and variability	Robert Bannert, Sidney Wood, Gabriella Koch
25.3	Tone and intonation in standard Chinese	Eva Gårding, Jan-Olof Svantesson, Jialu Zhang
15.4	Evaluation of voice quality	Ann-Christine Ohlsson
21.4	Parameters in intonation analysis	Robert Ladd, Giessen University
29.4	The prosodic structure of Greek	Antonis Botinis
6.5	The importance of short-term memory for the faculty of reading and writing	Gunnel Gahne, Lili-Ann Rudberg, Elsa Åström
20.5	On rhythmic alternation	Gösta Bruce
	A prosody test for brain damaged patients	Christina Hedqvist-Dravinš
27.5	Rhythmical patterns in Swedish compounds	Olle Persson
30.5	Rehearsal for the Nordic Prosody Congress III Umeå	

By the end of the fall, 23 logopedics students had completed the second term of their phonetics courses. During the spring, alas, we have had only a few new students. This seems to be something we have to live with as long as phonetics is not a part of the regular courses for other vocations. Next term we are looking forward to a new group of logopedics students. There are at present 11 third-term students and 13 graduate students.

Two graduate courses were given, one on phonology by Gösta Bruce and one on technical acoustics by Bengt Mandersson.

I had a sabbatical during the fall and Gösta Bruce substituted for me.

Eva Magnusson defended her PhD thesis, The Phonology of Language Disordered Children - Production, Perception and Awareness, on April 22. Professor Ragnhild Söderbergh, Stockholm, was the faculty opponent. The dissertation has appeared as no 17 of Travaux de l'Institut de Linguistique de Lund.

Jan-Olof Svantesson's PhD thesis in general linguistics Kammu Morphology and Phonology, which will be defended on June 8, has been published as no 18 of the same series.

Professor Ragnhild Söderbergh, who holds a special chair at HSFR for childlanguage acquisition (sponsored by HSFR, Swedish Council for Research in the Humanities and Social Sciences) is moving to our department on July 1 1983. We are looking forward to fruitful cooperation with her.

Gunnar Fant and Jan Gauffin visited us in November and professor Fant talked about his research concerning source-filter interaction. Professor Jialu Zhang from the Acoustics Institute of the Academia Sinica in Peking was a guest researcher here for three months. In particular he made valuable contributions to the third world languages project.

Dr Edda Farnetani, Padua, spent one week with us in February.

Our department was visited by professor Sven Erik Back, director of HSFR and Bo Öhlund of FRN (Swedish Council for Planning and Coordination of Research).

The University Study Program Committee held one of their sessions at our department in April and were shown our research and teaching facilities.

Professors Steinar Engen and Knut Fintoft (Trondheim), Ernst Håkon Jahr (Tromsø), and Stig Eliasson (Uppsala) gave an extra seminar in connection with a visit to our department.

Our laboratory has installed a new VAX 11/730 computer thanks to a grant from FRN and we are awaiting delivery of a new digital spectrograph donated by the Wallenberg Foundation.

The Swedish Prosody project (HSFR) was completed and three new projects have started:

Language and Brain (HSFR)

Phonetic descriptions of some important languages in the third world (HSFR)

Phonological and phonetic studies of vowel reduction in Bulgarian (HSFR)

Lund in May 1983

Eva Gårding

CONTENTS

Working Papers 25 1983
Linguistics-Phonetics
Lund University, 1-34

SOME PHONETIC CHARACTERISTICS OF A MODEL FOR GERMAN PROSODY

Robert Bannert

ABSTRACT

In this paper some phonetic characteristics of a model for German prosody are described. The acoustic data form the most important basis for the outline of the model that is presented in Bannert (1983). The variables investigated are number of accents (sentence length), intonation type, position of phrase boundaries, and speaker. Aspects of the basic temporal and tonal components are illustrated. Variations of sentence and stress group durations do not show any clear and consistent pattern. The fundamental building blocks of the intonation algorithm, namely the basic tonal points, the tonal floor and the tonal movements in the accents, are treated in more detail. The effects of different factors on these tonal movements become evident.

1. BACKGROUND AND AIM

Compared to other languages and to my knowledge, no model existed until recently which was able to generate the prosody, i.e. the rhythm and the melody, of German utterances. Lately, however, an attempt has been made to fill this gap. An outline of an intonation model for German was presented in Bannert (1983a) and a model for German prosody was sketched in Bannert (1983b).[1]

The basic concept of the model for German prosody, which in its main components may also be valid for other languages, is shown in Figure 1. The model which is meant to generate the temporal and tonal structure of utterances consists of three main components, namely the basic temporal and tonal components and the modification component. The input to the

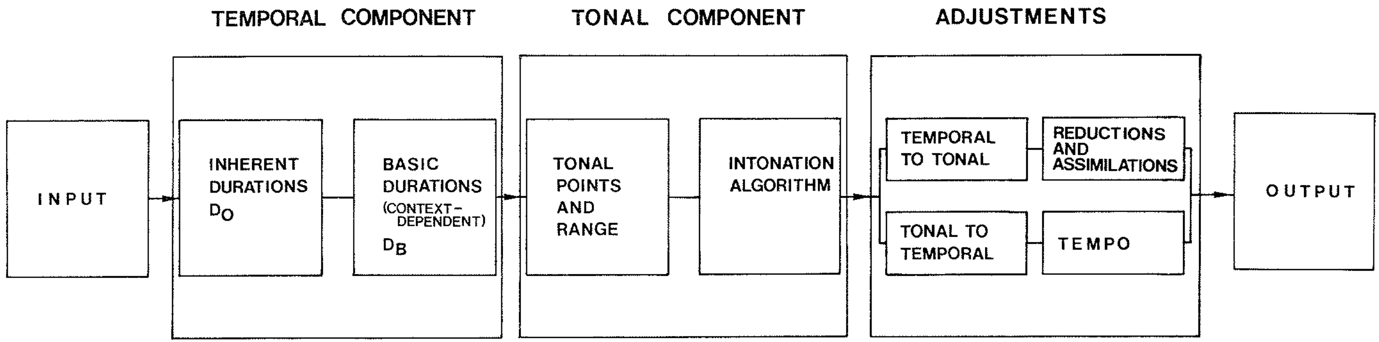


Figure 1. The three main components of a model for German prosody.

model is an abstract, linguistically completely specified structure. The output is the corresponding concrete temporal and tonal structure, i.e. the durations of segments and the Fo-contour.

The outline of the prosody model is based on, among other things, a relatively large acoustic material. However, in the papers mentioned above, the acoustic data could not be treated. Therefore, the aim of this paper is to account in some detail for the phonetic characteristics of the prosody model, especially the duration and intonation components.[2] This will be done by presenting the observed German data against the background of some general and important problems of a prosody model.

In order to provide the phonetic frame, the investigation is presented first. Secondly, the temporal structure on the sentence and stress-group level is illustrated. Finally, the tonal data are treated in more detail.

When evaluating the data an attempt is made to draw a distinction between linguistic features which are obligatory for all speakers of the language and individual features which are typical of a certain speaker and which therefore are optional. In the basic components of the model only such features should appear which are elements of the common linguistic code. Speaker-specific features may be incorporated into the model on the individual level. As the criterion to determine the limit between common linguistic and individual prosodic features in the material, the claim is made that a certain observation has to be found with all three speakers in order to qualify as an obligatory feature.

2. THE INVESTIGATION

The investigation was confined to the unit of the prosodic phrase, i.e. the utterances had to be spoken as one breath group and the accents should receive equal weight. Starting from this restriction, it seemed desirable to consider very different and almost extreme conditions in order to obtain basic acoustic data to build the model upon. Therefore, the following variables were chosen:

1. Length of utterance (prosodic phrase): number of accents 1 - 8, total number of syllables 3 - 29.
2. Intonation type: falling vs. rising melody at the end of the utterance, here called statement vs. echo and information question for the sake of simplification.
3. Speaker's involvement: normal, calm, detached (statement, information question) vs. involved, surprised, excited speech (echo question).
4. Syntactic structure: position of syntactic boundaries between accents (number of accents per syntactic phrase).
5. Number of speakers: three.

Table 1 shows the material which consisted of 14 utterances each of which was elicited as the intonation types statement, echo question, and information question.[3] Utterances 2-13 are pairs containing the same number of accents, but differ with regard to the position of the syntactic phrase boundaries. Thus, for instance, in sentence 4 (with 3 accents) the three accents make up one syntactic phrase. In sentence 5, however, the three accents are to be found in three syntactic phrases.

The sentences were constructed with regard to phonetic features which, of course, very often results in utterances that do not belong to every day language use. Nevertheless, these utterances are possible choices in the language (as to the conditions of the choice of the material, cf., for instance, Bruce 1977). Each accentuated vowel of the sentences is phonologically short and not open. Most of the consonants are sonorants, and thus hardly affect the Fo-curve.

The sentences were read by three female speakers from Northern Germany (speaker B, born 1956 in Lübeck, speaker K, born in 1955 in Uetersen, near Pinneberg, and speaker E, born in 1957 in Meldorf in Dithmarschen). The speakers, who were enrolled at the University of Kiel, were instructed to read the sentences without hesitation or break or pause and to give equal weight to each content word. The speakers chose their own individual speech tempo.

The material was recorded in the anechoic chamber of the Institute of Phonetics at the University of Kiel[4] on a Revox tape recorder, operating speed 9 ips. The three intonation

Table 1. The test sentences containing from 1 to 8 accents. Sentences with 2 to 7 accents appear as pairs differing in the position of the syntactic phrase boundaries which are indicated. The test sentences were read as the intonation types statement (A), echo question (E), and information question (I) with inverted word order.

number of accents	sentence number															
						#						#				
1	1							Die		Männer.						
2	2							Die	längeren	Männer.						
	3							Die		Männer	in der Menge.					
3	4							Die	längeren	Männer	in der Menge.					
	5	Der		Müller			will	die		Männer			immer		Lümmel	nennen.
4	6	Der		Müller			will	die	längeren	Männer	in der Menge					nennen.
	7	Der		Müller	in Lingen		will	die		Männer			immer		Lümmel	nennen.
5	8	Der	lullende	Müller	in Lingen		will	die		Männer			immer		Lümmel	nennen.
	9	Der		Müller			will	die	längeren	Männer	in der Menge		immer		Lümmel	nennen.
6	10	Der	lullende	Müller	in Lingen		will	die	längeren	Männer	in der Menge					nennen.
	11	Der	lullende	Müller	in Lingen		will	die	längeren	Männer			immer		Lümmel	nennen.
7	12	Der	lullende	Müller			will	die	längeren	Männer	in der Menge		immer	lungernde	Lümmel	nennen.
	13	Der	lullende	Müller	in Lingen		will	die	längeren	Männer			immer	lungernde	Lümmel	nennen.
8	14	Der	lullende	Müller	in Lingen		will	die	längeren	Männer	in der Menge		immer	lungernde	Lümmel	nennen.
						#						#				

types statement, information question, and echo question were presented in appropriate contexts and treated in three blocks in the order mentioned. The informants knew the little story about an article in a newspaper which contained all the 14 test sentences.[5] They were also well acquainted with the chamber and the recording procedure.

The method of eliciting the three intonation types is illustrated by the following examples (cf. Bruce 1977):

1. Statement. As the answer to an appropriate question, the statement was written on a card, e.g.

> 31
> Welche Nachricht aus unserer Stadt brachte der Kurier heute?
>
> Der Müller will die Männer immer Lümmel nennen.

In a kind of dialogue, the speaker would first read the reference number, then the question, and finally the answer (the statement).

2. Information question. The question was written on cards as a response by the speaker to a statement which referred to the known article in the newspaper "Kurier", e.g.

> 31
> Die Geschichte in der Zeitung machte die Leute neugierig.
> Will der Müller die Männer immer Lümmel nennen?

3. Echo question. The story was read aloud by the test leader. The speaker pronounced each test sentence which was underlined in the text by repeating it in the same word order as the statement in the text, but, at the same time, expressing involvement, namely surprise or astonishment.

Each sentence was spoken in five series. The recorded material was checked auditorily as to the correct realisation of the intonation type and the accents. For the acoustic analysis the material was processed by a Frøkjaer-Jensen pitch meter in the laboratory of the Department of Linguistics and Phonetics at Lund University. Duplex-oscillogrammes and Fo-curves were analysed and measured by hand. Each value to be presented later represents the arithmetic mean of five measured values.

3. DURATIONS

In the second step of the basic temporal component (cf. Figure 1) the inherent segment durations are lengthened or shortened due to various contextual effects. In the following account of some temporal aspects it will be shown how the durations of sentences and stress groups behave in the intonation types statement, information question, and echo question. Clear patterns of temporal variation will indicate the need to incorporate the context factor "intonation type" into the second step of the basic temporal component of the model. Thus the following questions will be asked:

1. Do the three intonation types show their own characteristic durations? Are there any clear and consistent differences between the sentence durations of the intonation types?
2. How do the stress group durations vary as a consequence of intonation type? And if they do, is this difference located at the end of the sentences (the last stress group) where it could be a consequence of the opposed Fo-contours (fall vs. rise)?

An account of the temporal features on the levels of syllables or segments will not be given in this paper.

3.1 Sentences

Means and standard deviations (in ms) of the duration of each of the 14 sentences according to intonation type and speaker are given in Table 2. An examination must be made as to whether the sentence durations of the three intonation types are different from each other or not. As a simplified measure, the threshold value of the durational difference ΔD_S is chosen according to the following definition: The sentence duration of two intonation types is clearly different if the durational difference ΔD_S is greater than the sum of their standard deviations s_1 and s_2, thus

$$\Delta D_S > s_1 + s_2$$

If the duration of the statement is taken as the reference of this comparison, the two following conditions hold:

Table 2. Mean sentence durations and standard deviations (ms) for each intonation type (I,A,E) and for each speaker (B,K,E). First line $\bar{x}$, second line s. A clear difference between durations is indicated by an asterisk (see text).

sentence number	number of accents	speaker:	B			K			E		
		intonation type:	I	A	E	I	A	E	I	A	E
1	1		448	* 498	* 608	470	448	* 590	378	* 440	410
			19.2	23.9	39.6	10.0	19.2	49.0	11.0	24.5	25.5
2	2		828	* 890	* 998	798	822	* 940	780	838	842
			22.8	35.4	56.8	25.9	22.8	62.0	51.0	34.2	28.6
3	2		966	1008	* 1170	862	928	* 1050	888	* 1006	986
			39.7	22.8	27.4	37.0	31.1	67.8	40.9	36.5	27.0
4	3		1368	1376	* 1478	1228	* 1298	* 1487	1270	* 1394	1384
			42.1	27.0	40.9	37.7	27.7	89.6	32.4	36.5	64.3
5	3		1884	* 1972	1962	1654	1700	1813	1694	1764	1764
			27.0	19.2	55.0	28.8	30.8	102.1	32.1	68.0	45.6
6	4		2250	2282	* 2438	2008	2036	2210	2058	-	2174
			43.6	42.1	76.3	35.6	69.1	116.8	55.6		46.2
7	4		2294	2350	2404	2000	* 2138	2250	2036	* 2234	2192
			29.7	51.5	110.6	46.4	56.3	195.7	61.1	65.8	49.2
8	5		2700	2764	* 2900	2322	* 2604	2648	2514	* 2660	2574
			79.1	65.0	56.6	53.6	129.7	149.6	75.0	60.4	61.1
9	5		2798	2860	* 3014	2496	* 2608	* 2756	2606	2716	2798
			64.2	68.6	74.0	32.1	39.6	86.8	139.6	82.6	62.2
10	6		3116	3112	3172	2764	2856	3002	2850	-	3078
			71.6	74.0	108.3	58.6	51.3	132.2	31.6		61.4
11	6		3156	3132	* 3286	2826	2918	2980	2944	3044	3108
			59.4	13.0	129.9	65.0	106.6	137.8	89.9	62.3	62.6
12	7		3694	3636	* 3898	3318	* 3536	3642	3490	3594	3654
			88.2	43.4	66.5	50.7	99.6	149.7	70.7	85.3	62.7
13	7		3578	3566	* 3810	3340	3376	3510	3398	3476	3526
			51.7	38.5	145.3	56.1	65.0	187.7	53.6	143.3	101.4
14	8		4084	4010	* 4270	3780	3796	4028	3874	4008	4098
			76.4	58.3	59.2	100.0	32.1	128.9	81.1	69.1	55.4

$$|D_A - D_I| > s_A + s_I$$

$$|D_E - D_A| > s_E + s_A$$

where A, E, I means the intonation type statement, echo question, and information question respectively, D is the sentence duration, and s the standard deviation.

All the cases, where the information question has a shorter duration than the statement, and the echo question has a longer duration than the statement, are marked with an asterisk between the duration values of the two intonation types which are compared in Table 2. It may be seen very clearly that no regular pattern among the three speakers is to be found. In total, clear differences of sentence duration between intonation types show up in only 35% of the 84 comparisons possible. Furthermore, the share for each speaker is very different: the echo question of speaker B nearly always has a longer duration than the statement. However, both have equal durations for speaker E. For the remaining comparisons, clear differences are to be found only in one third or one fifth of all cases. In order to simplify further, the relationship of the sentence durations between the three intonation types for each speaker may be expressed like this:

speaker	intonation type
B	I = A < E
K	
E	I = A = E

An irregular pattern of temporal behaviour of sentence durations is also to be found when another measure is used, namely the mean percentage of durational differences between each sentence and the intonation types. The mean values (%) of the 14 sentences are as follows:

speaker	A - I	E - A
B	0.07	7.4
K	3.6	8.6
E	6.3	-0.3

The mean durational differences between the intonation types compared do not exceed 10% and vary considerably between speakers and intonation types. According to this measure, the relationship between the sentence duration of the intonation types is expressed as follows:

speaker	
B	I = A < E
K	I < A < E
E	I < A = E

Figure 2 shows the pattern of variation of sentence durations between the intonation types for each speaker with the duration of the statement as the reference. For the sentences containing 2 to 7 accents, two curves are shown (cf. Table 1). It is evident from Figure 2 that no overall pattern of relative sentence duration is to be found. As the only common feature, it can be observed that the durational differences between intonation types are largest among the sentences containing the smallest number of accents.

These results concerning the sentence durations of the intonation types are relevant for the basic temporal component of the model. As the basic components should contain linguistic and general features only, i.e. features that are shared by all speakers, the second step of the basic temporal component will not include a context factor "intonation type" obligatory for all speakers.

3.2 Stress groups

As was the case with sentence duration, the durations of the stress groups do not show any clear pattern of variation either. Due to word inversion in the information question, the stress group durations of this intonation type could not be calculated. Thus stress group durations of statements and echo questions only are considered.

Figure 3 shows the stress group durations of statement and echo question of sentence 5 with 3 accents, sentence 9 with 5 accents and, sentence 14 with 8 accents (cf. Table 1) for each speaker. It can be seen that the pattern of durational variation is similar for all speakers. Thus two things be-

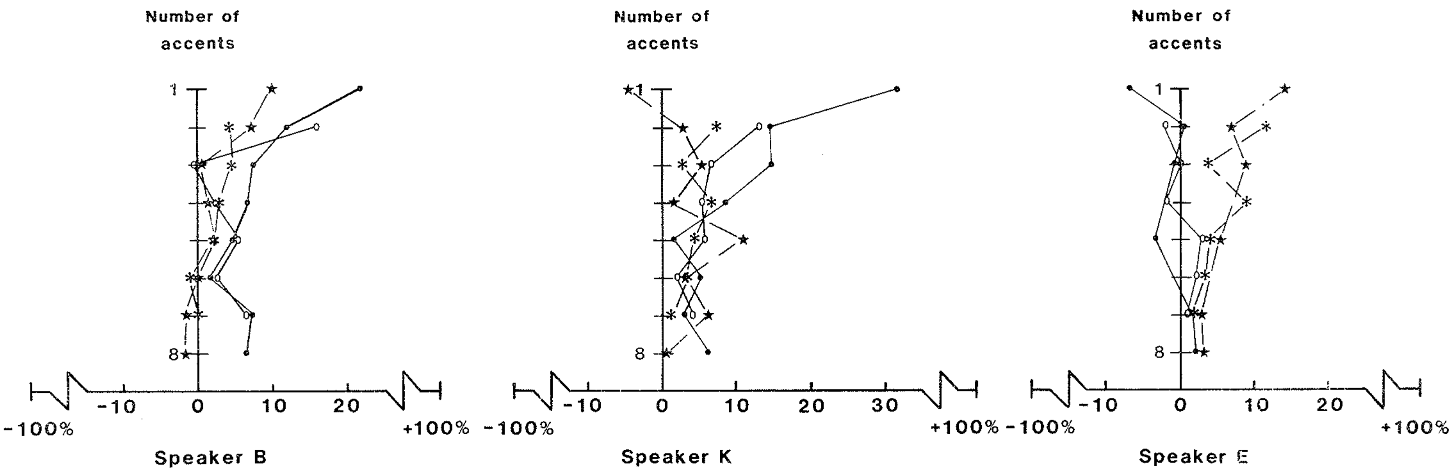


Figure 2. Relative mean differences of sentence duration between statement - information question (★,✻) and echo question - statement (●,o) for each speaker, the durations of statements being the reference.

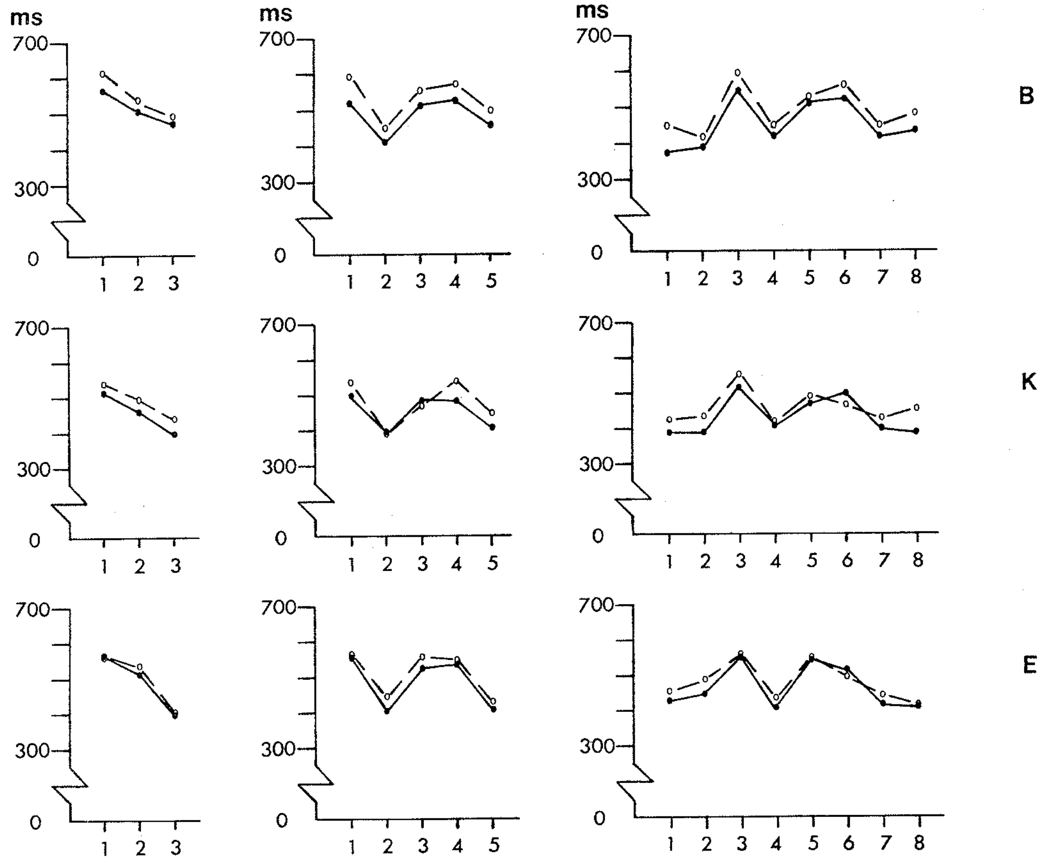


Figure 3. Mean durations of stress groups in sentences 5, 9, and 14 with 3, 5, and 8 accents respectively for each speaker (B,K,E). ●——● STATEMENT, ○— —○ ECHO QUESTION

come evident: First, stress group duration is determined by the duration of their elements, namely the syllables and segments. Evidence for this interpretation is to be found in the parallel variation of stress group durations for all speakers. Therefore it is taken to reflect a linguistic, i.e. general and obligatory feature of the temporal structure of utterances. Secondly, the difference in sentence duration reported above seems to be a global change of duration, i.e. it is spread over each stress group, and thus is not a consequence of the final tonal contours or the varying range of the tonal movements of the whole utterance (cf. Figure 5).

Figure 4 shows the relative differences of stress group duration between echo question and statement (= reference) of sentence 8 with 5 accents and sentence 14 with 8 accents. The total mean of each speaker is also included.

In parallel with the relative differences of sentence duration between the intonation types (cf. Figure 2), no regular pattern of durational variation can be discerned with the stress groups either. As a consequence, therefore, the basic temporal component of the prosody model will not contain any factor that may determine differences between various intonation types. The temporal component operates with one basic temporal structure, irrespective of intonation type. The durational differences, however, are not skipped altogether. They are seen as individual, and thus optional, features and are accounted for in the modification component by a device generating speaker-specific variation.

4. INTONATION

The basic tonal component of the prosody model consists of two parts, the tonal re-writing rules and the intonation algorithm (cf. Figure 1 and Bannert 1983a). Before data concerning the algorithm will be presented, some basic information on the Fo-contours will be given.

4.1 Fo-contours

Figure 5 shows the normalized Fo-contours of the utterances 1, 2, 5, and 14 containing 1, 2, 3, and 8 accents defined by the initial Fo, the Fo-minima, the Fo-maxima, and the final

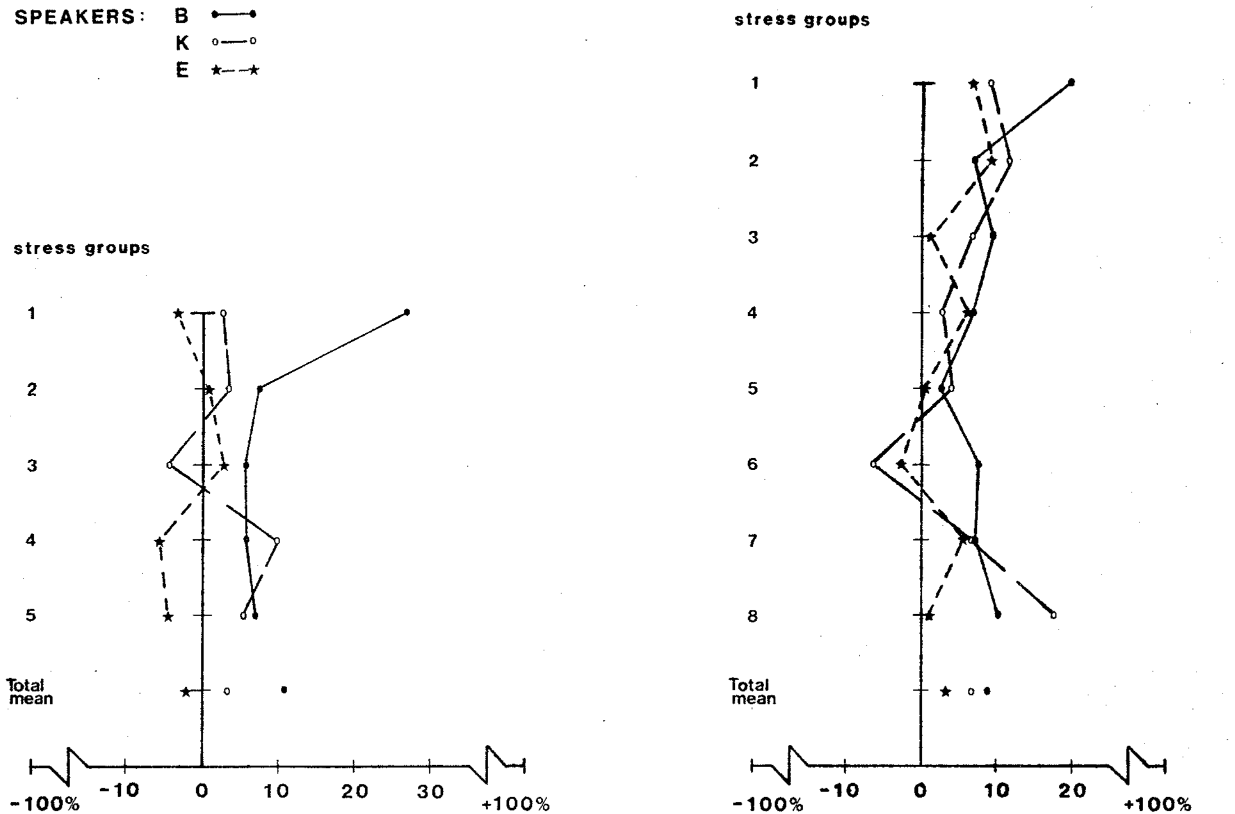


Figure 4. Relative mean differences of stress group durations and total mean in sentences 8 and 14 with 5 and 8 accents (stress groups) respectively between echo question - statement for each speaker, the durations of stress groups of statements being the reference.

Fo. The curves of the three intonation types are superimposed, the VC-boundary of the accentuated syllable being the temporal reference. It is easy to imagine the remaining contours with 4, 5, 6, and 7 accents by a simple operation of interpolation. The horizontal time axis is normalized and determined by the number of accents with equal distances between them. Comparing the different Fo-contours, some characteristic features of German intonation may be summarized as follows: Apart from one exception, each accent is manifested tonally as a rise starting from an Fo-minimum in the initial consonant of the accentuated syllable and ending in an Fo-maximum in the following, unstressed syllable. The only exception to this is the final accent of the statement which is tonally expressed as a fall. As a consequence of this difference in the final part of the utterances, the tonal contrast between the statement with a final fall, signalling terminal intonation and the questions with a final rise, signalling non-terminal intonation is established. The echo question, which also expresses surprise and astonishment, is characterized by at least a somewhat larger tonal range of the tonal movements. The tonal peaks show higher values. Except for the Fo-minimum of the first accent, the bottom points of the accents in all three intonation types show approximately the same value for all speakers. It seems as if the tonal movements of the different utterances started from a common bottom line or as if the tonal contours of the intonation types rested on a common tonal floor.

4.2 The intonation algorithm

In the second step of the basic tonal component of the model,[6] an algorithm generates the tonal structure of a given utterance in five steps. The working of the algorithm in the time-fundamental frequency-field for a sentence with four accents of equal weight, to be spoken as a statement and an information question, is illustrated in Figure 6.

As the first step, the intonation algorithm defines the four basic tonal points or levels which constitute the frame of the whole tonal structure of the utterance. This is done using the given temporal structure as the time reference.
In the second step, the Fo-minima of the accents and the end

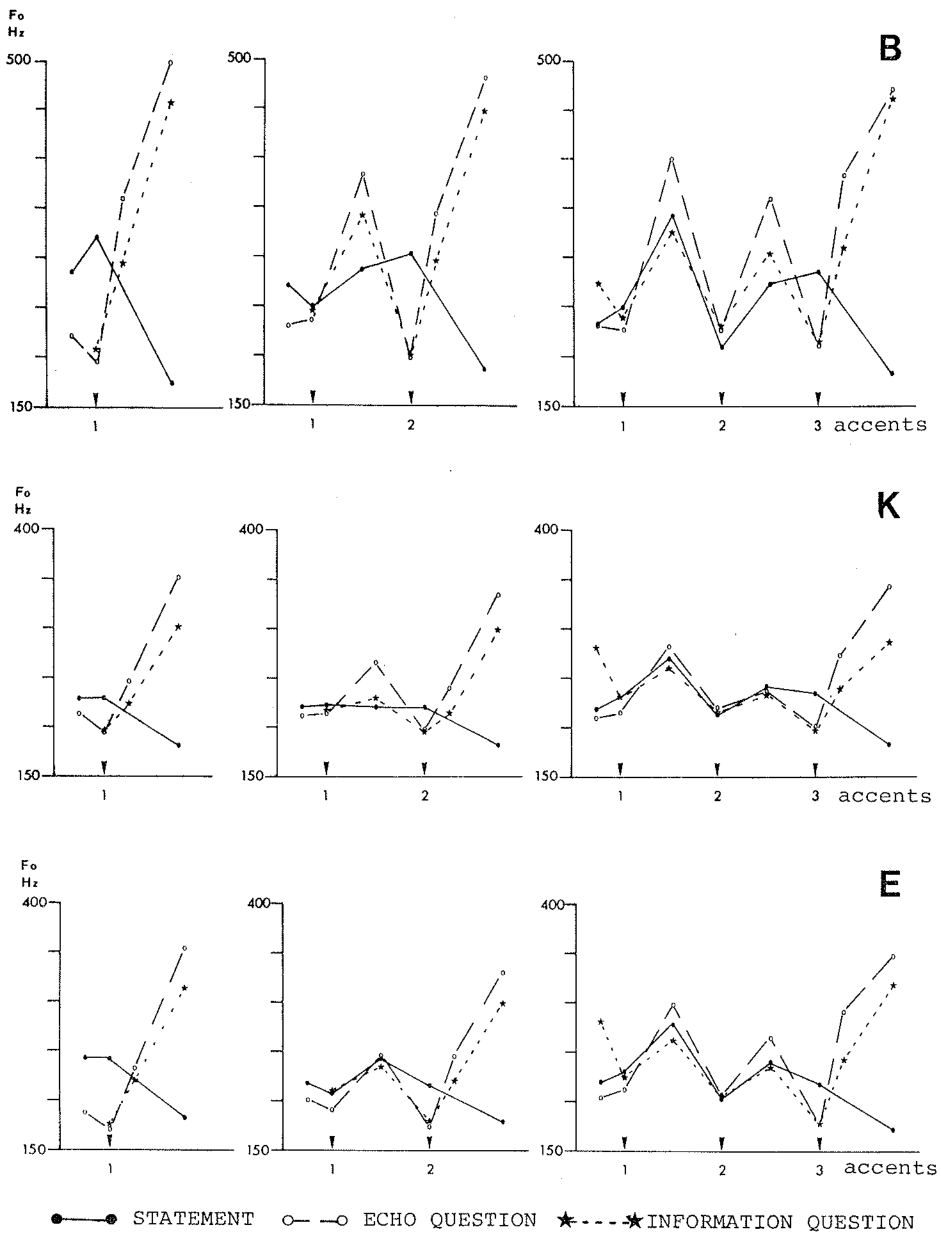


Figure 5. Normalized and superimposed Fo-contours of the three intonation types shown for the sentences with 1,2, and 3 accents (sentences no.1,2,5) for each speaker.

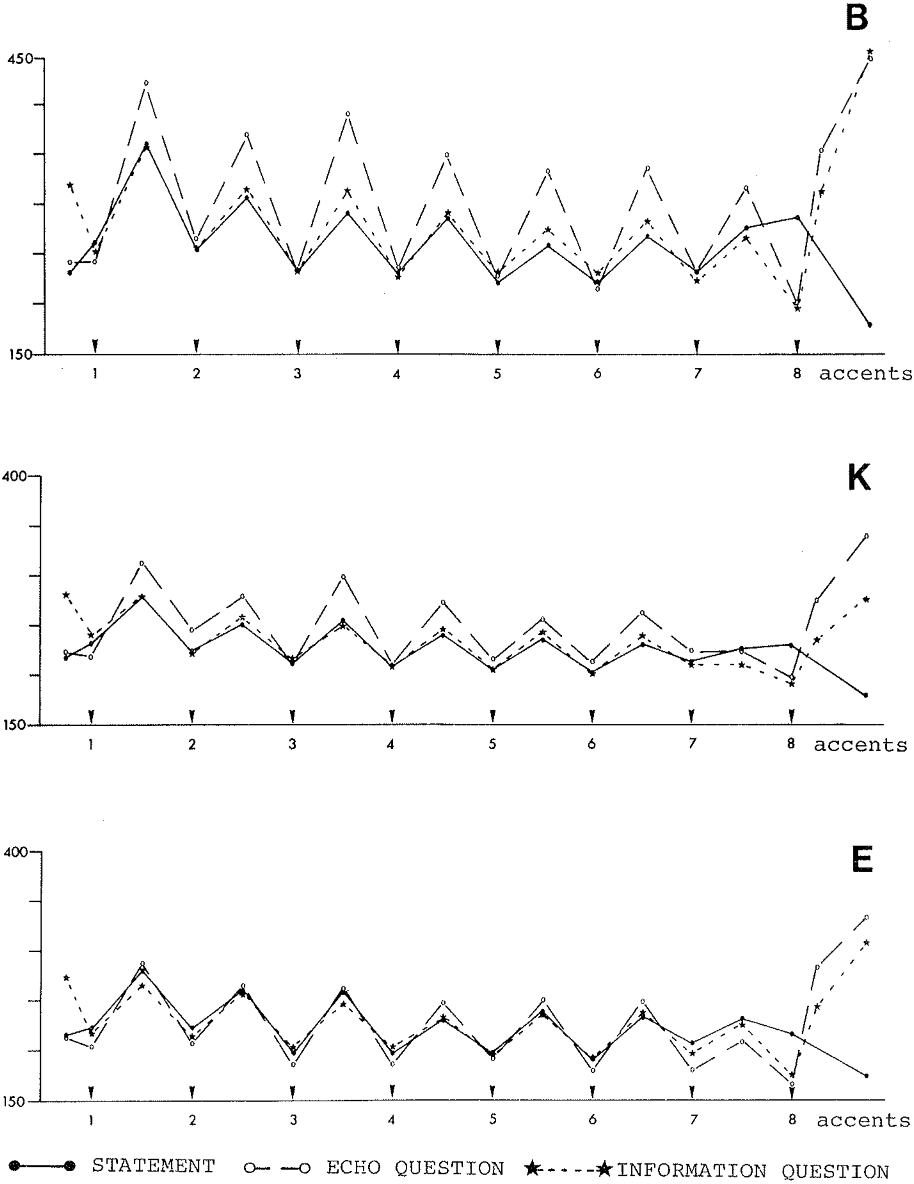


Figure 5. Normalized and superimposed Fo-contours of the three intonation types shown for sentence 14 with 8 accents for each speaker.

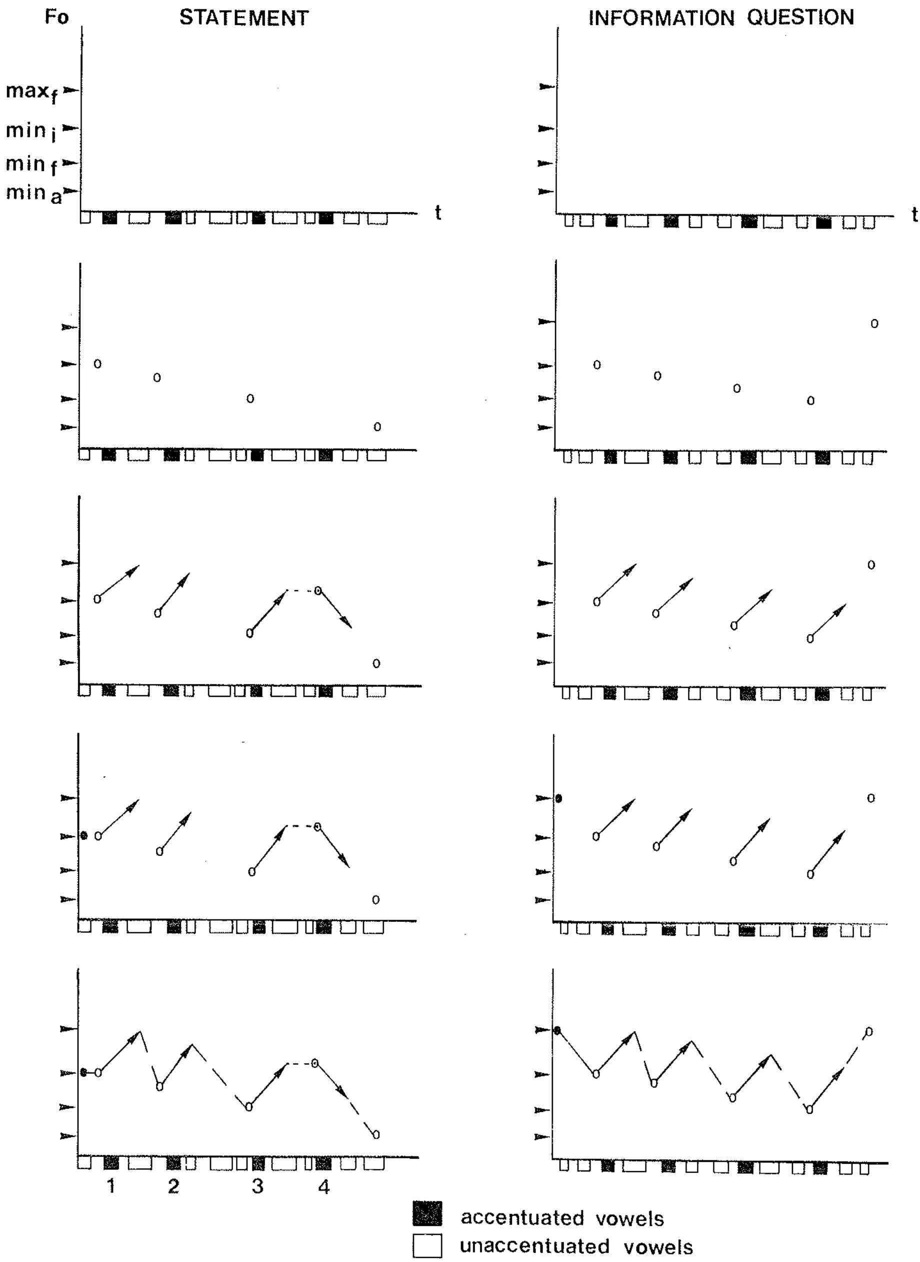


Figure 6. The intonation algorithm illustrating the generation of the tonal structure for a sentence with 4 accents (1,2,3,4) as statement and information question.

point of the utterance (the absolute Fo-minimum of the statement and the final Fo-maximum of the questions respectively) are inserted. Third, the tonal accent movements are generated, starting from each tonal point (Fo-minimum) and thus must be rising by nature. The tonal rises of the accents are of equal size which is defined as a certain tone interval. They end in the vowel of the following unstressed syllable. Therefore, the steepness of the tonal rise may vary according to the temporal conditions. In the final accent of the statement, the tonal accent movement must go down because, as a consequence of the falling contour of the sentence intonation, the final accent of the statement is characterized by a high point. Fourth, the starting points of the Fo-contours are inserted. In the fifth and last step the existing points and fragments of the tonal structure are interconnected either by straight lines or by a cosine function. Now the basic tonal structure of a given utterance is present.

The design of the intonation algorithm clearly shows that the tonal structure is generated from below, i.e. from the tonal bottom or tonal floor which is defined by the Fo-minima. Considering the position of the lowest points of the Fo-contours of the three intonation types (Figure 5), it is quite evident that the corresponding minima of each intonation type have rather similar values. It seems as if the Fo-contours rested in the lowest points, while the peaks varied more freely. Therefore, the main elements of the intonation algorithm are: 1) the Fo-minima of the accents, 2) the final end points of the utterance (low with the statement, high with the questions) which are the most important expressions of the intonation type or the sentence intonation, and 3) the tonal movements of the accents. By utilizing the feature of the tonal accent movements instead of a tonal accent point HIGH, as is the case in the Swedish and American intonation models, the dynamic character of the accents and the tonal changes associated with them is taken into due consideration. The size of the tonal movements varies due to different factors, the most important of which are syntactic, semantic, and pragmatic ones. After all, intonation is no more than kind (timing of tonal changes with reference to temporal units like segments or syllables and thus resulting in a fall or a rise) and size of the tonal changes resting in the tonal floor and their interrelationships.

When constructing the intonation algorithm, several important questions have to be answered:

1. How are the beginning and end of the Fo-contours related to the length of utterances?
2. How do the tonal differences of the accents behave as a function of length of utterance?
3. What does the bottom line, i.e. the distribution of the Fo-minima between starting and end point, look like as a function of length of utterance?
4. What is the tonal expression of the accents and the intonation types? Are the intonation types manifested only locally at the end of an utterance or also globally over the whole phrase?

In the following sections, an attempt will be made to answer these questions. First, the five basic tonal points that define the frame of the whole intonation contour in the first step of the algorithm will be presented and discussed. Secondly, the intermediate tonal points of the floor, i.e. the course of the bottom line, will be investigated and finally, the tonal difference of the accents, the tonal accent rise, will be considered.

4.3 Basic tonal points

The most important tonal points in the intonation algorithm are in a falling scale according to their importance: the Fo-minimum (Min_i) of the first and last accent (Min_f), the end points of the tonal contours, namely the absolute Fo-minimum at the end of statements (Min_a) and the final Fo-maximum (Max_f) of the questions, the starting point of the tonal contour ($Fo_{initial}$), and the peak of the first accent (Max_i). Although this latter point does not belong to the floor of the tonal structure, it may be important for the declination or the global expression of the intonation type (cf. Thorsen 1980, 1981; Bruce 1982) and therefore will be treated together with the other basic tonal points.

When constructing the intonation algorithm, it is essential

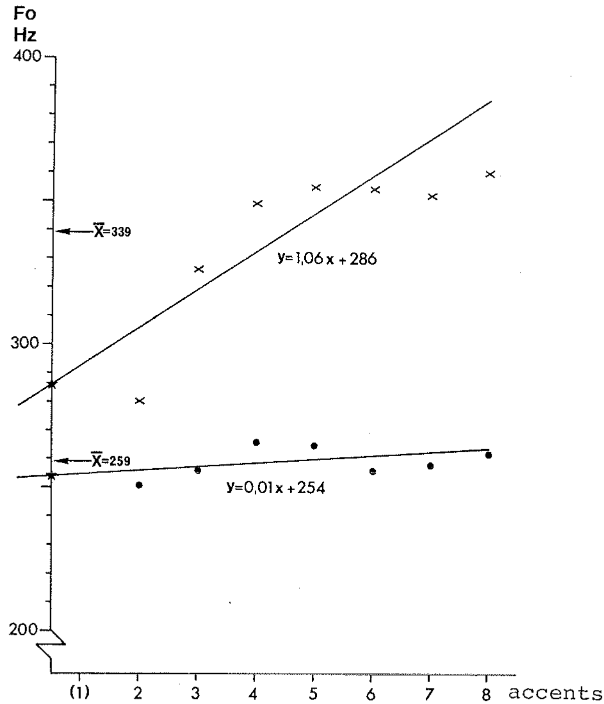


Figure 7. Illustration of the statistical procedure of determining constancy and variation of tonal points.

to know whether or not the most important tonal points, which, as it were, function as the hinges of the tonal structure, are dependent on the factors of sentence length (number of accents) and intonation type. If the points are independent of these two factors, only one value has to be specified in the algorithm. However, if the points are dependent on these two factors (and others), the algorithm has not only to account for the variation per se but has also to determine their kind and size.

In order to motivate the choice of the fixed tonal points of the intonation algorithm, the behaviour of the most important tonal points will be described by statistical measures. By way of illustration, Figure 7 elucidates the statistical method using the points of the initial Fo-minimum (Min_i) and the initial Fo-maximum (Max_i) of speaker B. Each point in the two series of points represents the mean of the Fo-values for the sentences containing 2 - 7 accents. The points pertaining to the sentence with only one accent are not taken into consideration. This is due partly to the fact that one of the two points is missing in the contour (cf. Figure 5), and partly to tone assimilation. As a simple statistical measure, the straight line $y = Ax + B$ is fitted to each series of points. Comparing the total mean $\bar{x}$ of one series of tonal points to the intercept B, i.e. calculating the difference $(\bar{x} - B)$, some conclusions concerning the size of the variation of the given tonal point as a function of the utterance length can be drawn. More information is to be found in the slope A of the straight line concerning the size of the variation and in the correlation coefficient r concerning the strength of the relationships.

The one extreme case is illustrated by the points Min_i showing very little variation as a function of utterance length $(\bar{x} - B = 5$ Hz, $A = 0.01)$. Thus these points represent tonal constancy. In contrast to this, the other points Max_i vary rather considerably $(\bar{x} - B = 53$ Hz, $A = 1.06)$. They represent the case where utterance length exerts a clear effect on the tonal point Max_i, at least up to the length of 4 accents.

The statistical values of the five tonal points for each speaker and each intonation type are presented in Table 3. The following values are shown which are based on the mean for each sentence: The total mean $\bar{x}$ and its standard deviation s in Hz, the intercept B of a straight line fitted to the series of points in Hz, the slope A, the correlation coefficient r, and the difference $(\bar{x} - B)$ in Hz as a measure of the variation of the 7 tonal points as a function of utterance length (number of accents). If it should be the case that a certain tonal point, e.g. $Fo\text{-}Min_i$, has the same value, irrespective of the increasing number of accents and thus utterance length, for instance 200 Hz, then it follows that $\bar{x} = 200$ Hz, $s = 0$ Hz, $B = \bar{x} = 200$ Hz, $A = 0$, $r = 0$, and $(\bar{x} - B) = 0$ Hz. In this case, utterance length has no effect at all on the tonal point $Fo\text{-}min_i$. It has a constant value. This point, however, can not be considered to be constant when $(\bar{x} - B)$ and A increase because then the variation of each point as a function of utterance length also increases.

Table 3. Constancy and variation of five basic tonal points for each speaker and intonation type (see text).

1) Fo-Min$_i$

	B			K			E		
	A	I	E	A	I	E	A	I	E
$\bar{x}$	259	248	233	229	233	220	224	221	204
s	5.4	4.9	5.5	4.9	6.0	5.5	6.4	6.7	6.6
B	254	245	228	220	220	214	218	220	194
A	0.096	0.05	0.11	0.18	0.25	0.11	0.12	0.03	0.20
r	0.38	0.22	0.42	0.76	0.90	0.45	0.40	0.08	0.66
$\bar{x}$-B	5	3	5	9	13	6	6	1	10

2) Fo-Min$_{a,f}$

$\bar{x}$	181	204	203	180	197	198	177	178	172
s	3.8	5.2	4.7	1.6	2.6	2.5	2.9	1.9	3.2
B	182	212	203	182	199	196	181	178	176
A	-0.008	-0.18	0.007	-0.048	-0.06	0.05	-0.09	-0.01	-0.09
r	-0.05	-0.86	0.04	-0.73	-0.55	0.51	-0.77	-0.18	-0.69
$\bar{x}$-B	1	8	0	2	2	2	4	0	4

3) Fo-Max$_f$

$\bar{x}$		454	465		290	341		312	341
s		4.7	15.3		13.1	6.3		8.4	9.9
B		457	488		313	350		312	350
A		-0.07	-0.50		-0.50	-0.19		0.006	-0.21
r		-0.35	-0.80		-0.94	-0.73		0.02	-0.52
$\bar{x}$-B		3	23		23	9		0	9

4) Fo-initial

$\bar{x}$	235	298	237	217	278	215	220	279	207
s	18.7	15.3	5.2	1.2	6.1	4.3	2.4	4.1	3.4
B	237	256	226	218	269	204	225	284	198
A	-0.03	0.76	0.21	-0.02	0.15	0.21	-0.09	-0.10	0.17
r	-0.23	0.93	0.76	-0.32	0.47	0.92	-0.70	-0.47	0.95
$\bar{x}$-B	2	42	11	1	9	11	5	5	9

5) Fo-Max$_i$

$\bar{x}$	339	342	410	265	266	295	276	264	288
s	28.4	10.8	7.9	20.3	14.9	21.2	12.2	9.9	14.9
B	286	317	396	228	238	249	259	250	266
A	1.06	0.49	0.28	0.74	0.56	0.91	0.35	0.27	0.44
r	0.81	0.98	0.77	0.79	0.81	0.92	0.62	0.59	0.64
$\bar{x}$-B	53	25	14	37	28	46	17	14	22

If a large and systematic variation is observed, the tonal point has to be made flexible in the algorithm.

Table 3 is examined for speaker independent patterns of variation and differences. This means that all three speakers have to show the same behaviour in certain respects. Thus considering the five tonal points and their relationship to the variables utterance length and intonation type, the following statements can be made:

A. Utterance length

Only a slight variation as a consequence of utterance length is to be found with the basic tonal points Fo-Max_f and $FO_{initial}$, especially with Fo-Min_i and Fo-$Min_{a,f}$. Therefore, they are considered to be constant and do not receive a correction factor for utterance length in the model. The slight variation which can be observed in the material and which is quite natural may be achieved in the model not by assigning a fixed value to a given tonal point in each utterance but rather by choosing a varying value. Appropriate values may be chosen from the interval defined as the double standard deviation 2s for each speaker. This choice does not apply to the tonal points only but also to the tonal movements of the accents and the inherent segment durations D_o.

However, there is one point, namely Fo-Max_i, which obviously varies considerably as an effect of utterance length with all speakers. It should be remembered that this variation need not be characteristic of all the utterances with an increasing number of accents (cf. Figure 7). Although the significance of the differences measured in Hz is not clear for perception (it must be examined in listener tests), these differences can be accounted for in the model. However, this will not be done in the intonation algorithm, i.e. in the basic tonal component but rather in the modification component. Utterances with one accent and only a few syllables like sentence 1 (cf. Table 1) have to be modified tonally in different ways because they represent the prosodically most complex cases. All the prosodic features of the sentence have to be manifested in these few syllables and, in a sentence consisting of only one syllable, in just this very syllable.

B. Intonation type

In most cases where the points in the intonation types with regard to the position and to the range are not totally different from each other, the same values are to be taken, irrespective of intonation type.

Identical points are the Fo-Min_i and Fo-Max_i of statements and information questions, the Fo-Min_f of the questions, the $Fo_{initial}$ of statements and echo questions. An obvious difference is to be found with echo questions compared to the two other types, where the points Fo-Min_i and Fo-Max_i are concerned. If we assume the enlarged tonal movement of the accents to represent the tonal expression of the echo question, i.e. the tonal range of the movements is increased, the larger value of the Fo-Max_i is given as the sum (v + w) of the accent rise (cf. Bannert 1983a). This enlarged tonal range of the rise of the echo question, however, is to be found with all accents, not only with the first one. The lower value of the Fo-Min_i of the echo question compared to the two other intonation types may then be explained as a secondary effect of the enlarged tonal range, assuming the general principle of optimal effect and minimal energy. According to this principle, an optimal tonal difference (the largest tonal range possible) is reached, if both the starting point and the end point of the movement are shifted, which means that the starting point is lowered, while the end point is raised.

A clear difference is to be found between the starting points $Fo_{initial}$ of information questions opposed to statements and echo questions. As this difference results from a different course of the tonal contour at the very beginning, it is accounted for in the algorithm (cf. Figure 6, step 4).

There is furthermore a difference between the high final points (Fo-Max_f) of information questions vs. echo questions with speakers K and E. It should be remembered, however, that the preceding Fo-minima (Fo-Min_f) have the same value. As this difference is not valid for all the speakers, only one point for Fo-Max_f, independent of utterance length and question type, is assumed in the algorithm. However, if it should turn out that the difference observed for Fo-Max_f of the questions would constitute a characteristic and necessary perceptual feature of questions, and thus would represent a linguistic demand, it may be accounted for in the intonation algorithm of the basic tonal component.

4.4 Intermediate tonal points

After having established the beginning and end and the frame of the tonal structure in terms of tonal points (or levels), the question has to be asked how the intermediate points of the tonal floor, i.e. the Fo-minima, should be inserted. In intonation models, a declination or base line is usually constructed which has varying slopes for different intonation types. It should be emphasized, however, that in this outline of a model for German prosody no lines whatsoever are utilized. The reason for this is connected with the view that the tonal structure of an utterance should not be considered as a visual or geometric construction. On the contrary, in this model the dynamic character of the tonal events and their interrelationships are stressed.

As the tonal structure is generated starting from the tonal floor in the intonation algorithm, the position of the Fo-minima forming the tonal bottom is essential. Therefore, their position and distribution between the starting point and end point of the sentences with increasing number of accents and in relation to syntactic phrase boundaries will be shown in some detail.

The position of the Fo-maxima, i.e. the top line, is not considered to be independent and thus a linguistic target per se. The top line is here viewed as a result of the tonal accent movement starting from the tonal floor. Of course, the tonal movements tied to the accents may reflect more prosodic features than just the accents. Other prosodic features concomitant with the accents and superimposed on the tonal movement may be the features CONTRAST, EMPHASIS or PHRASE INTONATION which are expressed in the tonal range feature WIDE and accounted for in the tonal re-writing rules by the factor <u>w</u> (cf. Bannert 1983a). The size fo the tonal movements is determined by the sum of all contributions of tonal features, thus total size of the tonal movement $V_G = (v + w + a + ... + n)$.

In Figure 8 the positions of the Fo-minima in the sentences 4 and 5 with 3 accents, 8 och 9 with 5 accents, and 12 and 13 with 7 accents of the three intonation types of speaker B are shown graphically. A very similar picture is to be found with the two other speakers. The pairs of sentences differ with regard to the position of the syntactic phrase boundary which is indicated in Figure 8 (cf. also Table 1). An assumed "declination line", defining the positions of succeeding Fo-minima with equally decreasing frequency values, is drawn for the sake of comparison. This line is defined by the means of the Fo-Min_i and Fo-Min_f of the information question. The reason for choosing these reference points lies in the fact that the final accent has a minimum with questions only and that the point Fo-Min_i of the information question is situated between that of statements and echo questions. In most cases, however, it is closer to the mean Fo-Min_i of statements (cf. Table 3).

In Figure 8 it can be seen clearly that the Fo-minima between the starting points and the end points of the floor in the sentences 5 with 3 accents and 9 with 5 accents occupy positions which show approximately equal frequency differences. This is also true of the sentences with 2 and 4 accents which cannot be shown due to lack of space. In other words, the Fo-minima are distributed equally over the frequency interval between the beginning and the end of the tonal floor. However, the last but one Fo-minimum in longer sentences (with 6, 7, and 8 accents) always has a higher value than the last but two, the highest value is to be found in statements. In spite of this, the final Fo-minimum of the questions (Min_f) and the absolute Fo-minimum (Min_a) of the statements show values similar to those in shorter sentences. It is not so easy to find an explanation for the increased value of the last but one Fo-minimum in the longest sentences. This upstepping of the tonal floor which is shared by all speakers appears independent of the syntactic phrase structure (cf. Table 1). Neither can it be conceived as the tonal assimilation to the Fo-maximum of the final accent of statements because this

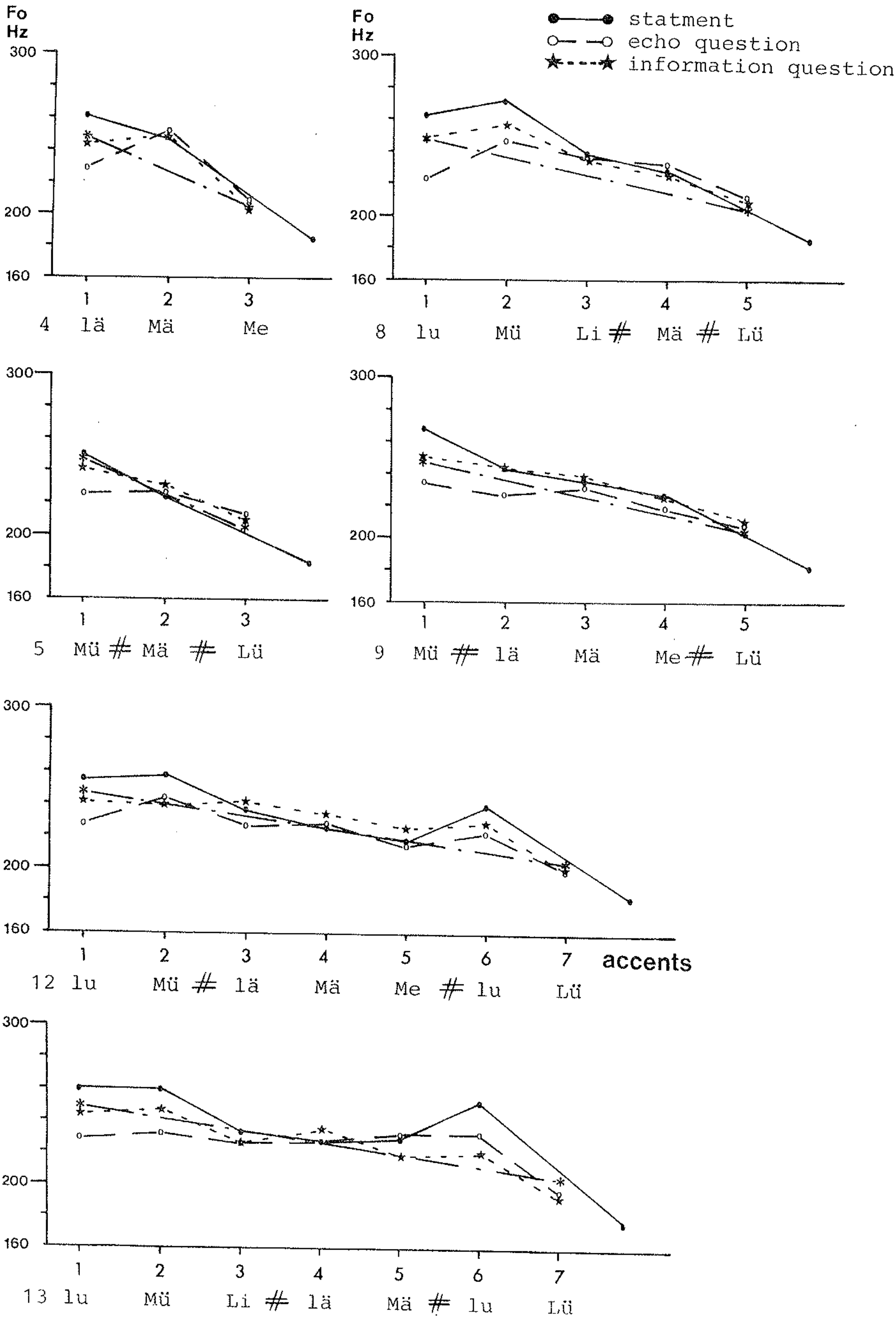


Figure 8. Positions of Fo-minima in sentences with 3,5,and
7 accents and in the three intonation types.
Speaker B. A "declination line" is drawn (see text).

upstepping is also found with questions. Tonal assimilation may perhaps account for the higher value in statements compared to questions. One explanation may be presented, though: Due to the large numbers of accents (6, 7, and 8), the frequency intervals of successive Fo-minima become rather small, i.e. the perceptual effect of the successively falling Fo-minima (the downstepping) diminishes, and therefore the speaker has to compensate for this loss of signalling power in order to signal the end of the utterance. As a kind of re-setting, the speaker increases the Fo-minimum of the last but one accent, in order to sharpen the jumping down to the final Fo-minimum of the questions. In statements, the tonal demand of a clearly audible final fall is therefore met by this device.

The successive Fo-minima, however, do not always show diminishing values. The declining course of the tonal floor is broken sometimes. This seems to be the case when there is an effect of syntactic phrase boundaries on the Fo-minima. A comparison between the sentences 5 and 4, 9 and 8 respectively will make this effect clear. While the values of the Fo-minima in sentences 5 and 9, the first phrase of which contains only one accent, decrease gradually, the second Fo-minimum in the sentences 4 and 8 breaks the falling course of the Fo-minima. In both these cases, the first syntactic phrase contains three accents. The value of the Fo-minimum of the second accent is higher than the first value. Thus the second accent in the middle of the three-accent group (Männer in sentence 4 and Müller in sentence 8) appears to be weakened and therefore less prominent than the first and last accent of the group. As an explanation, the principle of tonal grouping of successive accents presents itself. It says that three (or more) successive tonal movements of equal size are not permitted and therefore the second (fourth, etc.) movement will be weakened. This phenomenon might be called the principle of inequality, because it aims at breaking the chain of potentially equal events and thus introduces variation into the tonal contours. This principle may be considered to have its parallel where rhythm is concerned. Rhythmic alternation which means that successive syllables of equal strength are avoided, i.e. one of two equal syllables in succession are weakened, is well known to be characteristic of the rhythmic structure of phrases.

A second explanation may be found on semantic grounds. Due to their meaning, the accents in the syntactic phrases under consideration can not be equal. The place adverbial (Lingen) and the attribute (lullende) which are expressed in the third and first accent respectively, may have a semantically larger weight than the person (Müller) which appears in the second accent. Therefore, the person is moved into the background by using a weakened accent. However, these explanations presuppose that the tonal movement in the second accent, too, is smaller than in the adjacent accent. In the following section (4.5) it will be shown that this is actually the case.

In summary, then, it can be stated that the Fo-minimum of the tonal floor are not always spaced equally between the starting point and the end point of the tonal bottom. In some cases one of the successive Fo-minima is shifted upwards and thus breaks the declining line of the tonal minima. Therefore, it is concluded that various factors, especially semantic features, determine the position of the Fo-minima of the tonal floor.

4.5 Tonal accent movements

The element of tonal movements associated with the accents, which is one of the basic features of the intonation algorithm, expresses the dynamic character of intonation. However, the tonal movements found in the accents, namely rises and falls, do not represent the expression of the accents alone, i.e. the prominence feature of the most important syllable of the lexical units, but also contain reflexes of other prosodic tonal features, like CONTRAST, EMPHASIS, and PHRASE INTONATION which are expressed by the tonal feature WIDE. Finally, the tonal movements are also determined by semantic and pragmatic features.

Figure 9, which is constructed in parallel with Figure 8, shows the tonal movements found in the accents, defined as the tonal difference ΔFo = (Fo-Max - Fo-Min), in the sentence pairs with 3, 5, and 7 accents. The final accent is omitted

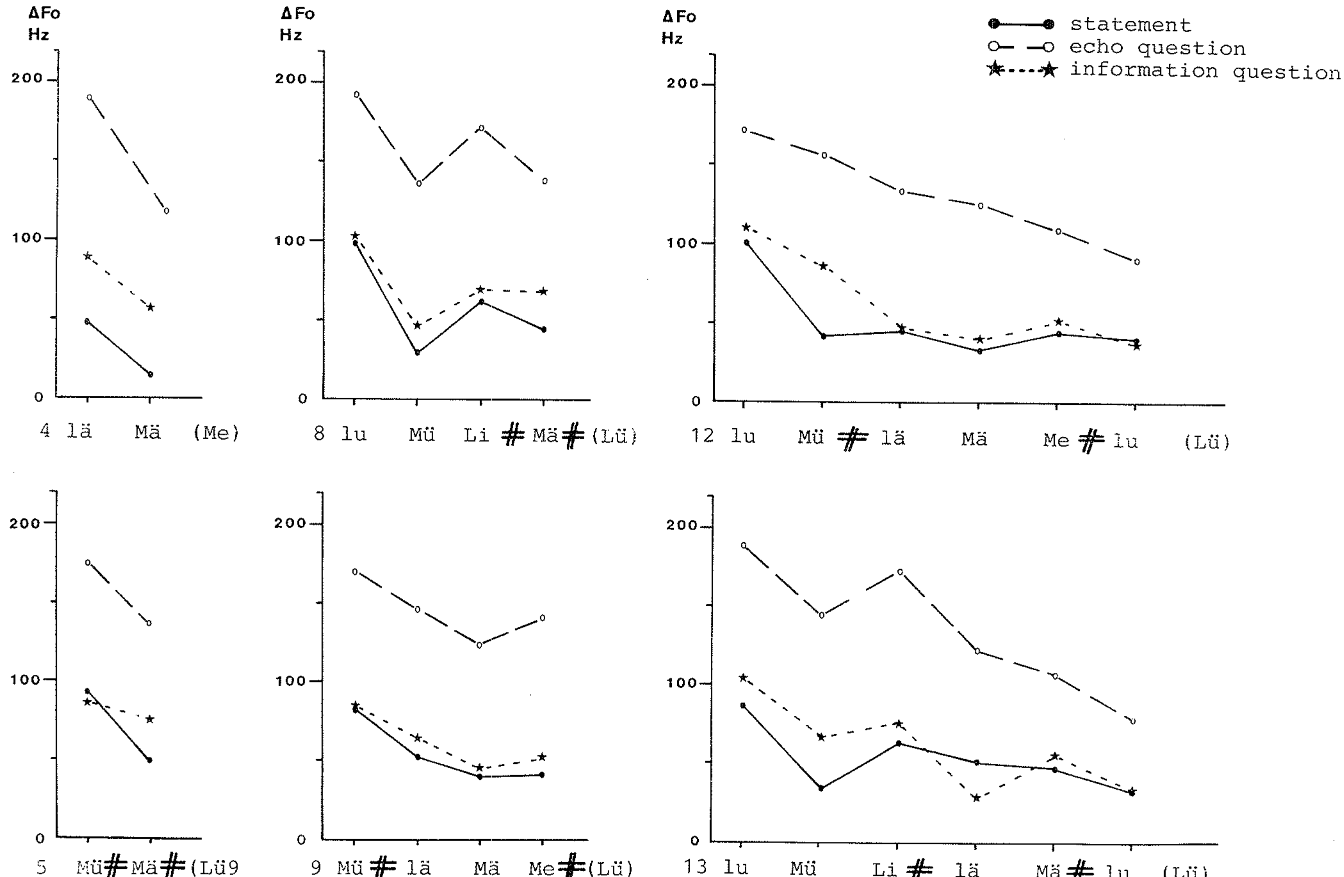


Figure 9. Tonal accent movements in sentences with 3, 5, and 7 accents and in the three intonation types. Speaker B. Sentence numbers, accentuated syllables, and phrase boundaries are indicated.

because it is also affected by the sentence intonation.

The differences of the tonal movements of the accents appear very clearly in the sentences 8 and 9 with 5 accents. The second accent (Müller) of the first syntactic phrase with 3 accents has a clearly smaller tonal rise than the adjacent accents. Except for the statement, this smaller tonal movement is also to be seen in sentences 12 and 13.

Leaving out this irregularity which may be a consequence of the semantic inequality of the three accents of the first syntactic phrase, the tonal movements of the accents, in general, show gradually falling values. In the intonation algorithm, I assume as a rough simplification that the tonal movements of the accents, provided that the accents are of equal weight and the tonal movements do not reflect other tonal features, be given the same value in another scale doing more justice to pitch or the perception of Fo in general. The size of the tonal movements is defined as a constant tone interval. As a consequence of this, the slope of the tonal movement will vary in accordance with the time limitation which is given by the temporal structure. Some evidence for this assumption is to be found in Bannert (1982, 12ff.) where the effect of vowel quantity on the rising Fo-movements of the accents was investigated. The varying slope of the tonal movements is a natural consequence of the definition of the starting point and the end point of the movements. The starting point of the accent, i.e. the Fo-minimum is located in the consonant preceding the accentuated vowel; its end point is located in the following unstressed vowel or the CV-boundary of the post-accentuated syllable.

5. CONCLUSIONS

The acoustic data concerning temporal and tonal aspects of German represent the phonetic basis for the outline of a model for German prosody. It goes without saying, though, that considerable work remains to be done before the model for German prosody will reach the level of sophistication that models for other languages have reached. Some of the most urgent areas of progressing research are linguistic variables of various kinds, the number of speakers and different speech styles, i.e. directed and free spontaneous speech.

Linguistic and phonetic analysis alone, however, will not suffice. The quality of the model for German prosody will heavily depend on investigating how prosodic features of utterances are perceived. As an excellent means for finding out the relevant prosodic features of the signal, given the acoustic description of the temporal and tonal structures of utterances, speech synthesis has to be used. It will be readily agreed upon that it is the aim of our endeavours to understand prosody and its role and function in speech communication as used by both speaker and listener.

FOOTNOTES

1 I would like to point out the preliminary and rather restricted character of this model. Much work remains to be done in order to arrive at a more satisfactory understanding of German prosody. Nevertheless, it can be assumed that some general and language-specific features of German prosody are already captured in this outline.

2 A comparison of the prosody model with those of other languages, a discussion and an assessment of the elements chosen for the model, and other discussions and comments fall outside the scope of this paper. They will need to be saved for future work.

3 The longest sentence 14 reads in English: The miller in Lingen doing pee-pee will always call the taller men in the crowd loitering louts.

4 I am deeply indebted to the director of the Institute, Prof.Dr. Klaus J. Kohler.

5 The story "Bericht über einen geheimnisvollen Zeitungsartikel" (Report on a mysterious newspaper article) was written by Ursel Krützmann and Micheal Weinhold who participated in the seminar "Rhythm and Intonation".

6 Whereas the model for German prosody shown in Figure 1 is also applicable to other languages, the specific features of German intonation are brought out in the intonation algorithm (and the re-writing rules). See also footnotes 1 and 2.

REFERENCES

Bannert, R. 1982. Temporal and tonal control in German. Working Papers 22:1-26. Department of Linguistics and Phonetics, Lund University

Bannert, R. 1983a. In print.Modellskizze für die deutsche Intonation. Zeitschrift für Literaturwissenschaft und Linguistik 49

Bannert, R. 1983b. Towards a model for German Prosody. Paper presented at the Interdisciplinary Symposium "Prosody - Normal and Abnormal", Zürich, April 6-8, 1983

Bruce, G. 1977. Swedish word accents in sentence perspective. Lund: Gleerup

Bruce, G. 1982. Developing the Swedish Intonation Model. Working Papers 22:51-116. Department of Linguistics and Phonetics, Lund University

Gårding, E. and Bruce, G. 1981. A Presentation of the Lund Model for Swedish Intonation. Working Papers 21:69-75. Department of Linguistics and Phonetics, Lund University

Thorsen, N. 1980. Intonation Contours and Stress Group Patterns in Declarative Sentences of Varying Length in ASC Danish. Annual Report of the Institute of Phonetics 14:1-29. University of Copenhagen

Thorsen, N. 1981. Intonation Contours and Stress Group Patterns in Declarative Sentences of Varying Length in ASC Danish-Supplementary Data. Annual Report of the Institute of Phonetics 16:19-36. University of Copenhagen

Working Papers 25 1983
Linguistics-Phonetics
Lund University, 35-52

ON RHYTHMIC ALTERNATION

Gösta Bruce

ABSTRACT

Rhythmic alternation among sequences of unstressed syllables in Swedish utterances was examined from both phonological and phonetic starting points. This alternation between strong and weak syllables was found to be determined by the location of the actual stressed syllables. Generally, the unstressed syllables will be alternatively weak and strong starting from the upcoming stress and counting backwards. Phonetically this is reflected as an alternation in relative durations between successive syllables. I suggest that the division into weak and strong unstressed syllables in addition to the division into stressed and unstressed syllables is an important aspect of speech rhythm in Swedish.

1. PROBLEM

The present paper is a preliminary study of speech rhythm in Swedish where the objective is to explore the nature of rhythmic alternation.

Speech rhythm seems to be one of the most problematic topics in contemporary phonetics and phonology. There is far-reaching agreement about the importance of rhythm in spoken language, but when it comes to more precise statements about what speech rhythm really is and where it is, the disagreement among scholars in the field becomes evident.

The phonetic study of speech rhythm has led to the postulation of isochrony as one expression of speech rhythm and the division of languages into syllable-timed and stress-timed (see e.g. Pike 1945, pp. 34 ff.). The maintenance of (nearly) constant inter-stress intervals independent of the number of intervening unstressed syllables defines isochrony in stress-timed languages, e.g. Germanic languages.

Most production studies of isochrony show, however, at best a tendency towards isochrony which has led to the assumption that isochrony is primarily a perceptual phenomenon (Lehiste 1973,1977, Allen 1975); cf. e.g. Lehiste's observation "that listeners tend to hear utterances as more isochronous than they really are,..." (Lehiste 1980, p. 252) and the discussion of the isochrony issue in the temporal symposium at the phonetics congress in Copenhagen 1979.

My own approach to the study of speech rhythm is somewhat different. I have chosen to study what is known as rhythmic alternation, trying to relate this concept to a general definition of rhythm such as the one formulated by the psychologist Woodrow (1951, p. 1232):

> "By rhythm, in the psychological sense, is meant the perception of a series of stimuli as a series of groups. The successive groups are ordinarily of similar pattern and experienced as repetitive. Each group is perceived as a whole and therefore has a length lying within the psychological present."

For spoken language I assume that a series of syllables will be arranged according to a principle of rhythmic alternation, alternation between strong(er) and weak(er) syllables. This will create groups and subgroups of syllables of similar structure, which is thought to constitute an important aspect of speech rhythm. It is my bias in thinking about rhythm in spoken language that the leading principle is alternation and not isochrony.

The goal of the present study is to examine rhythmic alternation from both phonological and phonetic starting points, i.e. to try to find out what rules govern the occurrence of rhythmic alternation and what are its phonetic reflexes.

2. THE PHONOLOGY OF RHYTHMIC ALTERNATION

The rhythm of an utterance in Swedish is determined mainly by the stress patterns of the actual, prominent words, i.e. how stressed and unstressed syllables are distributed in these words. In the presentation here I will make a distinction between lexical and actual stress. The lexical stress, attributed to each lexical item, may or may not be actualized in an utterance depending on intricate context factors. When nothing else is indicated, stress will mean actual stress. It is a well-known fact that it is primarily the content words that will carry stress in an actual utterance, while the form words often appear unstressed.

The rhythm of an utterance is, however, partly determined by other factors as well. The tendency towards a rhythmic alternation is assumed to be one such factor.

Rhythmic alternation seems to occur in two kinds of situations:

(1) One situation is when two or more stresses would clash in an utterance. Under certain conditions one of the stresses will yield resulting in a rhythmic alternation among the syllables involved. A stress shift is said to take place in order to avoid a clash of stress. A classical example of the avoidance of stress-clash in English is the word thirtéen with final stress in isolation but with the opposite stress pattern in certain contexts, e.g. thírteen mén (cf. e.g. Liberman & Prince 1977). A corresponding Swedish example, also classical, contains the word kaptén 'captain' with final stress in isolation, but with the stress shifted in the context káptan Àndersson 'captain Andersson' (cf. Malmberg 1966, p. 95).

Although a clash of two stressed syllables may result in a rhythmic alternation, as above, the normal case, at least for Swedish, is to still maintain an unchanged sequence of two stressed syllables.

In a stress-clash situation consisting of three consecutive syllables, a rhythmic alternation is more common. While in the example Jan såg 'Jan saw' we usually have two consecutive stressed syllables, it is at least unusual that in Jan såg Bo 'Jan saw Bo' såg will retain its stress. Normally a rhythmic alternation will occur.

It is possible that in certain sequences of three (or more) stresses like tripp, trapp, trull 'tick-tack-toe' or en, två, tre 'one, two, three', where no syntactic groupings will elicit specific subphrases, you may find three equally prominent syllables. But also for these kinds of constructions I believe it is customary to have a rhythmic alternation, where the middle stress is weakened.

(2) The other situation which calls for rhythmic alternation is when several unstressed syllables form a sequence. In this case a rhythmic alternation among the unstressed syllables as strong and weak will occur forming groups consisting of two (or three) syllables with one more prominent syllable per group. This tendency towards a rhythmic alternation among unstressed syllables is known from long, oxytonic words, e.g. Swedish <u>d$\overset{1}{e}$m$\overset{0}{o}$kr$\overset{0}{a}$t$\overset{4}{i}$</u> 'democracy', <u>k$\overset{1}{o}$nst$\overset{0}{i}$t$\overset{1}{u}$ti$\overset{0}{o}$n$\overset{4}{e}$ll</u> 'constitutional', where the integers 1 and 0 indicate an alternation between strong and weak unstressed syllables and 4 denotes primary stress (cf. Elert 1970, p. 37).

It is reasonable to believe that this alternation is typical not only of single words but of any succession of several unstressed syllables that forms a unit of speech, more or less independent of word boundaries.

The principle of rhythmic alternation entails a conspiracy against sequences of several equally prominent syllables. This applies to both unstressed and stressed syllables.

In the present paper I will concentrate on (2), the rhythmic alternation among sequences of unstressed syllables.

In the following I will try to make the simplest possible assumptions to account for rhythmic alternation. I assume that rhythmic alternation is a process that is triggered by the immediate prosodic context, i.e. no access to a hierarchical tree structure à la Liberman & Prince (1977) is needed. This also seems to be in line with the reasoning by Liberman & Prince, who described the rhythm rule for clash of stress involving a rhythmical alternation in terms of a metrical grid, which is based on a metrical tree structure but does not in itself encompass any constituent structure.

Thus, the relevant input knowledge for the rhythmic alternation rule is the location of the stresses in the actual prominent words. For the present study I will simply assume that the location of these stresses is known. The division of successive syllables into stressed and unstressed is considered part of the rhythmical organization of an utterance. The phonological matrices below used to illustrate the operation of the rhythmic alternation rule also contain information about rhythmical strength. As part of the input knowledge I assume that stressed is rhythmically strong and unstressed rhythmically weak. It should be noted that strong and weak as used in metrical theory (Liberman & Prince 1977) do not have these implications. I will use strong and weak to describe the rhythmic alternation among unstressed syllables.

A preliminary formulation of the rhythmic alternation principle is the following: Give the value strong to every second unstressed syllable counting from the closest stressed syllable.

For the two examples below with an odd number of unstressed syllables between the stresses, we have the following rhythmical structure:

(1)	input string		output string	
	kompisar me tjeckerna		kompisar me tjeckerna	'friends with the Czechs'
stress	+ – – – + – –	→	+ – – – + – –	
strength	+ – – – + – –		+ – + – + – +	

(2)	input string		output string	
	kompisar me kapital		kompisar me kapital	'friends with capital'
stress	+ – – – – – +	→	+ – – – – – +	
strength	+ – – – – – +		+ – + – + – +	

For an example with an even number of unstressed syllables between the stresses, e.g. four as in the example below, the preliminary rule will generate the following structure:

(3)	input string		output string	
	kompisar me soldater		kompisar me soldater	'friends with soldiers'
stress	+ – – – – + –	→	+ – – – – + –	
strength	+ – – – – + –		+ – + + – + –	

Although this seems to be a possible rendition of the actual utterance, it is according to my analysis not the most normal one. Instead the normal rendition of (3), at least in running speech, seems to be:

(3')		output string
		kompisar me soldater
stress	→	+ – – – – + –
strength		+ + – + – + –

That the actual location of word boundaries is irrelevant to the rhythmic pattern and that it is the number of unstressed syllables between the stresses that is crucial is evidenced by the following example, also with four syllables, but with another location of word boundaries:

(4)	input string		output string	
	soldater me kapital		soldater me kapital	'soldiers with capital'
stress	- + - - - - +	→	- + - - - - +	
strength	- + - - - - +		- + + - + - +	

Therefore the rhythmic alternation principle has to be reformulated as: Give the value strong to every second, unstressed syllable counting backwards from the upcoming stress to the preceding stress. Give also strong to every second, unstressed syllable after the final stress of the actual speech unit.

This rule will introduce a rhythmical shift after each stressed syllable which is followed by an even number of unstressed syllables before the following stress. Thus, the reformulated rule will give the following structure to a sequence of two unstressed syllables between the stresses as in the following example:

(5)	input string		output string	
	attentat me soldater		attentat me soldater	'attempt with soldiers'
stress	- - + - - + -	→	- - + - - + -	
strength	- - + - - + -		+ - + + - + -	

A retention of the preliminary rule would have created a conflict of adaptation for the two unstressed syllables in (5), i.e. whether to adapt to the preceding or the following stress, which does not occur with the reformulated rule.

Sequences containing an uneven number of unstressed syllables between the stresses, e.g. one, three or five, seem to be less problematic from a rhythmical point of view. They fit rhythmically both to the preceding and the following stress, while sequences with an even number of unstressed syllables, e.g. two or four, are rhythmically uneven and necessitate a rhythmical shift.

In this preliminary investigation I have only studied the most simple structures in the rhythmic alternation, the subgroups consisting of one strong and one weak syllable, while the rhythmical grouping of two consecutive weak syllables between two strong syllables remains to be taken into consideration.

One interesting implication of the alternation principle, as it has been formulated here, is that the same unstressed syllable may appear strong or weak depending on the actual context, i.e. its position in relation to the upcoming stress. This observation was an unexpected finding for me, but having gone through a number of similar examples, I have become convinced of its correctness. So in the two examples repeated here for convenience, we find that the syllable sar is strong in (1) and weak in (3):

	(1) output string	(3) output string
	kompisar me tjeckerna	kompisar me soldater
stress	+ − − − + − −	+ − − − − + −
strength	+ − + − + − +	+ + − + − + −

So far we have been discussing examples where the stressed syllables belong to separate words. If we instead turn to compound words, it appears that the rhythmic alternation principle works in exactly the same way.

Although a compound may be composed of several elements, each containing a lexical stress, I assume that in the input string of the actual compound word accessible to the rhythmic alternation rule only the first and the last stress of the compound will be present.

While in the example Sw. hov-rätt 'court of appeal' the second element of the compound has stress, this stress will not necessarily surface in a longer compound such as hovrätts-president 'president of a court of appeal' or hovrätts-fiskal 'public prosecuter of a court of appeal'. Although rätt will count as unstressed in both these compounds, it will appear weak in the first and strong in the second example according to the rhythmic alternation rule:

(6)	input string		output string
	hovrättspresident		hovrättspresident
stress	+ − − − +	→	+ − − − +
strength	+ − − − +		+ − + − +

(7)	input string		output string
	hovrättsfiskal		hovrättsfiskal
stress	+ − − +	→	+ − − +
strength	+ − − +		+ + − +

Another similar pair of examples but with another location of the internal word boundary is utbildnings-ledare 'education leader' and utbildnings-minister 'minister of education' with the following rhythmical structures:

(8)	input string		output string
	utbildningsledare		utbildningsledare
stress	+ − − + − −	→	+ − − + − −
strength	+ − − + − −		+ + − + − +

(9)	input string		output string
	utbildningsminister		utbildningsminister
stress	+ − − − + −	→	+ − − − + −
strength	+ − − − + −		+ − + − + −

A final example, demonstrating that the rhythmical structure of the compund elements in isolation does not necessarily determine the resulting rhythmical pattern in the compound is utrikes-departementet 'State Department'. Here the elements as independent words would come out as:

	(10) output string	(11) output string
	utrikes 'foreign'	departementet 'Department'
stress	+ + −	− − − + −
strength	+ + −	+ − − + −

In the compound word, however, the location of strong and weak syllables between the stresses will be shifted relative to the above rhythmical structure in the compound elements:

(12)	input string		output string
	utrikesdepartementet		utrikesdepartementet
stress	+ − − − − − + −	→	+ − − − − − + −
strength	+ − − − − − + −		+ − + − + − + −

It is even possible that the so-called clash-of-stress phenomenon described above is not distinct from the rhythmical alternation among unstressed syllables discussed here. At least in the Swedish examples it is tempting to conceive of kaptén with lexical final stress in the context kápten Àndersson as occurring unstressed and subject to the principle of rhythmic alternation. Accordingly, in the actual example the kap-syllable will appear strong, while in kaptén Molín 'captain Molin' - Molín having final stress - the ten-syllable is the strong one.

Summing up so far, we have demonstrated that the rhythmic alternation among sequences of unstressed syllables appears to be determined by the location of the stressed syllables in the actual prominent words of an utterance. A stressed syllable functions as a line-up point in the rhythmical structure, after which a rhythmical shift may take place. The alternation seems to be adapted to the upcoming stress so that going backwards from a stressed syllable the syllables will be alternatively weak and strong up to the preceding stress or to the beginning of the utterance.

Thus the phonological analysis of rhythm presented here gives us three types of syllables from a rhythmical point of view: $\begin{bmatrix} +\text{ stress} \\ +\text{ strong} \end{bmatrix}$, $\begin{bmatrix} -\text{ stress} \\ +\text{ strong} \end{bmatrix}$, $\begin{bmatrix} -\text{ stress} \\ -\text{ strong} \end{bmatrix}$. This means that the basic rhythmical division into stressed and unstressed syllables forming successive stress groups is supplemented by a division of the unstressed syllables into subgroups consisting optimally of one strong and one weak syllable.

3. THE PHONETICS OF RHYTHMIC ALTERNATION

The next issue to account for is how the rhythmical structure of an utterance, the rhythmic alternation, is expressed phonetically. My expectation is to find reflexes on the phonetic surface of the rhythmic alternation between strong and weak unstressed syllables as alternating relatively longer and shorter syllable durations.

A pilot experiment was conducted to give a preliminary and partial answer to the question of the phonetic reflex of rhythmic alternation.

3.1 Test material

In the test material used for this pilot experiment the distance between two stressed syllables with accent 1 in terms of the number of unstressed syllables was varied from one to five, while the total number of syllables per utterance was held constant. The test material contains the following original sentences:

1. kompisar me tjeckerna - ∪ ∪ ∪ - ∪ ∪ 'friends with the Czechs'
2. kompisar me soldater - ∪ ∪ ∪ ∪ - ∪ 'friends with soldiers'
3. kompisar me kapital - ∪ ∪ ∪ ∪ ∪ - 'friends with capital'
4. soldater me kapital ∪ - ∪ ∪ ∪ ∪ - 'soldiers with capital'
5. attentat me kapital ∪ ∪ - ∪ ∪ ∪ - 'attempt with capital'
6. attentat me soldater ∪ ∪ - ∪ ∪ - ∪ 'attempt with soldiers'
7. attentat me tjeckerna ∪ ∪ - ∪ - ∪ ∪ 'attempt with the Czechs'

Thus, the test material contains one sentence with one, two with three, two with four and one with five unstressed syllables between the stresses.

3.2 Method

Each test sentence was first recorded in its original form and then in so-called reiterant speech form immediately afterwards (Liberman & Streeter 1978), matching the syllables of the original sentence by the iteration of stereotypic syllables.

The problem of studying speech rhythm in real speech is that it is not directly accessible from the physical signal. In order to be able to extract the reflexes of speech rhythm in the consecutive syllable durations of an utterance, the segment structure of these syllables has to be controlled. In normal speech there is considerable variation in the segmental structure of the actual syllables (cf. e.g. Klatt 1976). There may be differences in such things as open vs. closed syllables, cluster size, vowel quality with different inherent length, voiced vs. voiceless consonants, etc., all contributing to the bias of the internal make-up of the syllables which may obscure the speech rhythm in the physical signal. It is commonly assumed that the human mind can compensate for such intrinsic phenomena.

In order to avoid problems of interpretation of this sort, I chose reiterant speech with stereotypic, open syllables. In my experiment I used as alternation between ba- and da-syllables replacing the syllables of the original sentence in the reiterant version of the sentence. The choice of two alternating instead of one truly reduplicating syllable was motivated by the assumed possibility of maintaining a more normal speech tempo in the former case.

The rhythm of natural speech is considered to be maintained in reiterant speech and to appear more naked there.

Each test sentence both in its sense and non-sense, reiterant form was recorded a minimum of eight times by the present author.

3.3 Results

As noted above I assume that the temporal relations of successive syllables among other things reflect the speech rhythm. Measurements of segment durations were made on the reiterant versions only.

In the following I will concentrate on the temporal relations among unstressed syllables.

Figure 1 compares the temporal relations of consecutive syllables in test sentences 1-3 (means of 5-9 repetitions).

It should be noted that the use of alternating ba- and da-syllables creates a bias so that every odd syllable (ba) in Figure 1 has an absolute duration which is slightly overestimated and every even syllable (da) has a slightly underestimated duration. This is due to the inherently longer occlusion of b than of d. The size of this inherent difference is minute, however, approximately on the order of 10-15 ms (cf. Löfqvist 1976, p. 5).

If we take this small bias into consideration, we get the following picture.

The rhythmic alternation among unstressed syllables is reflected in the temporal domain as an alternation in syllable duration. This alternation between shorter and longer syllables can be seen as a reflex of the postulated rhythmic alternation between weak and strong syllables (see the preceding section). The temporal relations displayed here also seem to be in accord with the hypothesis that the alternation between weak and strong syllables is governed from the upcoming stress, so that the pre-stress syllable is relatively shorter, the preceding one is relatively longer and so on by way of alternation.

Looking at the two unstressed syllables following the first stressed syllable, we find a very similar pattern for sentences (1) and (3), those with an odd number of syllables

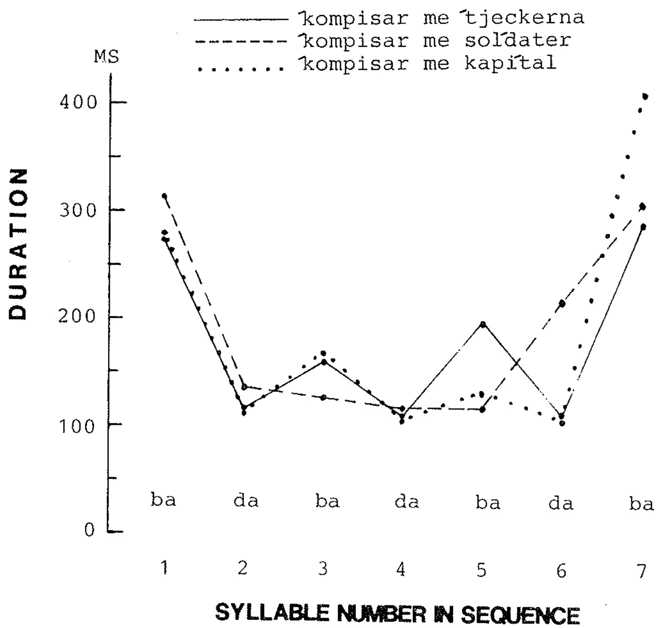


Figure 1. Rhythmic alternation. Durations of consecutive syllables in the reiterant versions of test sentences 1-3 (means of 5-9 repetitions).

between the stresses: the first unstressed syllable corresponding to pi is clearly shorter than the second unstressed syllable corresponding to sar reflecting the alternation. For sentence (2) with an even number of syllables between the stresses the situation is reversed; the first unstressed syllable pi is longer than the second one sar. This is exactly what is predicted by the above hypothesis. An uneven number of syllables between the stresses fits neatly into an alternating pattern from the point of view of both stresses, while an even number of syllables between the stresses necessitates a rhythmical resetting, which appears to be governed from the second of the two stresses.

If we look at the temporal pattern of the whole succession of unstressed syllables, four in sentence (2) and five in sentence (3), we observe that the alternation is not phonetically uniform. The first, strong unstressed syllable after the first of the two stresses is clearly longer than the second, strong unstressed syllable, while the difference between the corresponding weak unstressed syllables is very small.

Figure 2 shows the temporal relations for the whole test material displayed in a different way. The values are means of 5-9 repetitions.

Looking at test sentences 4-6 the alternating temporal pattern recurs for the unstressed syllables between the stresses. In sentences (4) and (5) with four and three unstressed syllables respectively between the stresses the same proportions hold as for a comparison between sentences (2) and (1) also containing four and three unstressed syllables respectively between the stresses.

It is notable that in sentence (6) with two unstressed syllables between the stresses the same relation holds as for the corresponding, last syllables in sentences (2) and (4) having four unstressed syllables between the stresses. My interpretation of this is that even with two unstressed syllables between the stresses, these two syllables are not equally prominent but have a strong to weak relationship. This is in agreement with the hypothesis that the alternation is governed from the following stressed syllable.

Looking also at the unstressed syllables in initial and final position (before and after the stresses) we find reflexes in the temporal pattern of the postulated strong-weak alternation.

In reasoning so far about rhythmic alternation I have been regarding the CV-syllable as a unit. If instead we look at C and V separately, it appears that the rhythmic alternation is reflected in both C's and V's duration. Judging from the present pilot experiment the relative difference in duration of a weak C compared with a strong C is greater than a corresponding comparison between a weak and a strong V. It is therefore a possible hypothesis for further experimentation that C's strength expressed in relative duration is as important as (more important than?) V' strength expressed correspondingly.

Finally it is worth noting that no tendency towards isochrony between the stressed syllables is apparent from the present investigation.

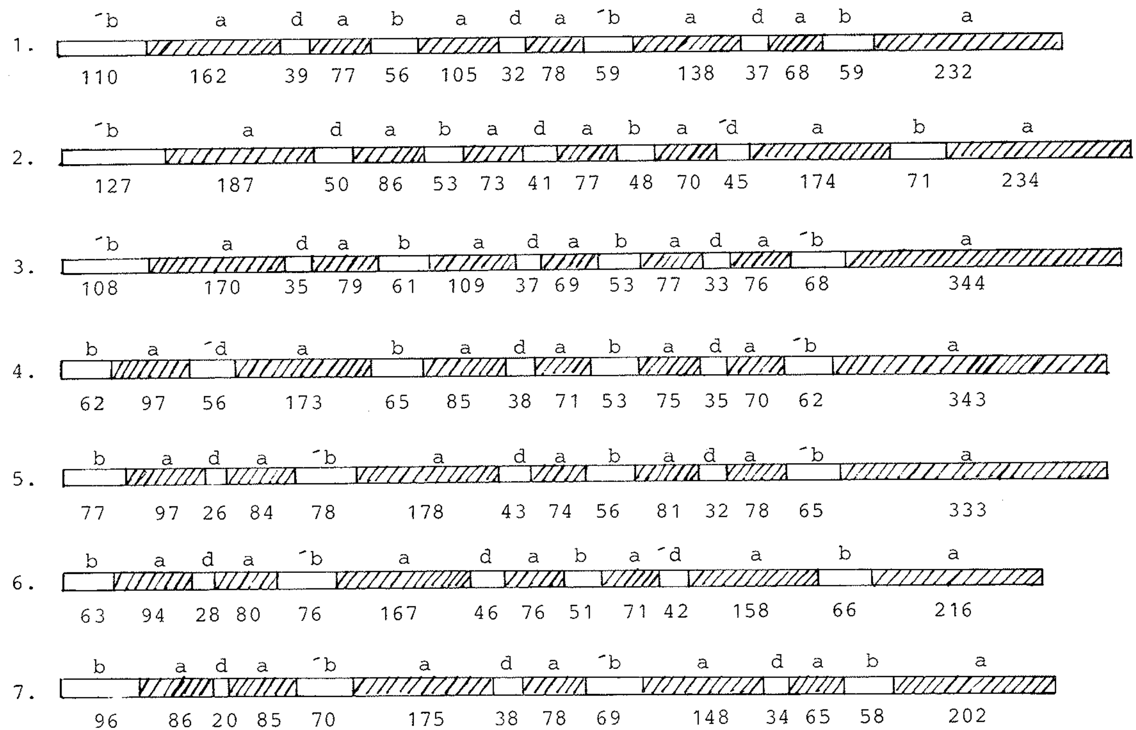


Figure 2. Rhythmic alternation. Durations in ms of consecutive segments in the reiterant versions of the test sentences (means of 5-9 repetitions).

4. DISCUSSION

In this paper I have tried to demonstrate that besides the alternation between stressed and unstressed syllables in Swedish utterances there is normally also an alternation among the unstressed syllables described as strong and weak, which is not so widely recognized. See, however, suggestions of a similar kind for English by Giegerich (1978, 1980), and also for Swedish by Strangert (1981). I suggest that this division into groups and subgroups of syllables in terms of prominence relations is an important aspect of speech rhythm in Swedish.

I have argued that the simplest kind of weak-strong alternation among unstressed syllables is related to the actual stressed syllables so that the syllable adjacent to a stress is weak, the next one strong and so on by way of alternation. An even number of unstressed syllables between two stresses creates a conflict, however. In this situation the alternation appears to be governed from the second of the two stresses, thereby introducing a rhythmical resetting after the first of the two stresses.

This suggests the existence of a pre-planning device or look-ahead several syllables in advance. In fact, the plan for the rhythmic alternation up to the next stressed syllable must exist at the point in time where a stressed syllable is being executed.

It is reasonable to ask why in situations of rhythmical conflict there should be an adaptation to the following and not to the preceding stress.

One way of answering this question is to ask, if a sequence, for example, of four syllables between two stresses realized as weak-strong-weak-strong, i.e. adapted to the preceding stress, would be impossible. The probable answer is that such a rhythmical sequence would be perceived as having yet another stress on the syllable preceding the final stress, as in <u>kómpisar me tvǻ tjécker</u> ' friends with two Czechs'. The same sequence realized as strong-weak-strong-weak between the stresses has no such implication, although the first strong in this sequence may coincide with a lexical secondary stress of a compound as in <u>útbildningen på sö́der</u> 'the education on the south side' comparable to <u>kómpisar me soldáter</u>.That this lexical secondary stress is not necessarily rhythmically strong is evidenced by the fact that it may be subject to rhythmic alternation in <u>útbildning på tómbola</u> ' education on the tombola' corresponding to <u>kómpisar me tjéckerna</u>.

It is possible, however, that a motivation for the observed rhythmical asymmetry can be formulated in more general, psychological terms.

I have found it possible to account for rhythmic alternation among unstressed syllables in Swedish, i.e. rhythm at a micro level, without recourse to metrical constituent structure. In an extended study of speech rhythm including also rhythm at a macro level the relation between underlying metrical patterns and the rhythmical structure of an utterance has to be taken into consideration.

A problematic issue is the division of consecutive syllables into stress groups and rhythmical subgroups within stress groups. The principle of rhythmic alternation suggested above can be interpreted as counterevidence against the traditional division into stress groups, where a stress group is said to contain the stressed syllable and consecutive unstressed syllables up to the next stressed syllable or to the end of the speech unit. The implication of the principle of rhythmic alternation, which shows a right-to-left planning for the unstressed syllables between two stresses, is to put the stress group boundary immediately after a stressed syllable. Thus a stress group would instead in most cases begin with unstressed syllables and end with a stressed syllable. Also among the unstressed syllables there is a corresponding problem of division into rhythmical subgroups.

Related to the issue of division into rhythmical groups is the question of the strong-weak relationship. It is assumed that in the case of rhythmic alternation, as it has been described here, strong-weak is a local relation between two consecutive syllables, i.e. a certain syllable is strong or weak only in relation to its closest neighbouring syllable. This means in phonetic terms that a certain weak syllable can be physically longer than or as long as a strong syllable in the same sequence of syllables, provided it is shorter than its strong pair mate.

In this paper I have looked for reflexes of rhythmic alternation mainly in the relative durations of successive syllables. I have also suggested that the length contribution to the rhythmical signalling may not be evenly distributed over a syllable but may be confined to a certain portion of the syllable, for example the C preceding the vowel (syllable onset) or to a portion across the CV-boundary. This proposed difference in the distribution of rhythmical signalling seems to reflect the two different conceptions of temporal events, such as rhythm, as durations (additive) versus temporal points (divisive) discussed in Liberman (1975, pp. 271 ff.).

Although I have been looking here for reflexes of rhythmic alternation exclusively in the time dimension, this does not mean that I conceive of rhythm merely as a length phenomenon. It is clear that the kind of rhythmic alternation found for example in situations of clash of stress may involve alternation of pitch. In cases of three consecutive stresses it is my impression that the weakening of the middle stress is expressed phonetically not primarily by reduced relative length but by the absence of pitch correlates characteristic of the surrounding stresses.

REFERENCES

Allen, G. 1975. Speech rhythm: its relation to performance universals and articulatory timing. Journal of Phonetics 3: 75-86

Elert, C.-C. 1970. Ljud och ord i svenskan. Stockholm: Almqvist & Wiksell

Giegerich, H. 1978. On the rhythmic stressing of function words: a modest proposal. Work in progress 11: 42-51. Department of Linuistics, Edinburgh University

Giegerich, H. 1980. On stress timing in English phonology. Lingua 51: 187-221

Klatt, D. 1976. Linguistic uses of segmental duration in English: acoustic and perceptual evidence. Journal of the Acoustical Society of America 63: 1208-1221

Lehiste, I. 1973. Rhythmic units and syntactic units in production and perception. Journal of the Acoustical Society of America 54: 1228-1234

Lehiste, I. 1977. Isochrony reconsidered. Journal of Phonetics 5: 253-263

Lehiste, I. 1980. Symposium no. 5: temporal relations within speech units. Proceedings from the 9th International Congress of Phonetic Sciences 1979 vol.III:247-265. Institute of Phonetics, University of Copenhagen

Liberman, M. 1975. The intonational system of English. Unpublished doctoral thesis. Massachusetts Institute of Technology

Liberman, M. and A. Prince. 1977. On stress and linguistic rhythm. Linguistic Inquiry 8: 249-336

Liberman, M. and L. Streeter. 1978. Use of nonsense syllable mimicry in the study of prosodic phenomena. Journal of the Acoustical Society of America 59: 231-233

Löfqvist, A. 1976. Closure duration and aspiration for Swedish stops. Working Papers 13: 1-39. Phonetics Laboratory, Department of General Linguistics, Lund University

Malmberg, B. 1966. Svensk fonetik. Lund: Gleerups

Pike, K. 1945. The intonation of American English. Ann Arbor: The University of Michigan Press

Strangert, E. 1981. Rhythmic patterns of Swedish in a cross-linguistic perspective. Publication 18. Department of Phonetics, Umeå University

Woodrow, H. 1951. Time perception. Handbook of experimental psychology. S. S. Stevens (ed.): 1224-1236. New York: Wiley

Working Papers 25 1983
Linguistics-Phonetics
Lund University, 53-65

A generative model for TONE AND INTONATION IN STANDARD CHINESE based on data from one speaker

Eva Gårding, Zhāng Jiālǔ* and Jan-Olof Svantesson

In most descriptions of Chinese prosody, the basic units are the morphemes with their tone patterns, and these patterns are then modified by sandhi, stress, intonation, and tempo (see e.g. Wú 1982a, 1982b). Figure 1, borrowed from Kratochvíl's well-known manual <u>The Chinese Language today</u> (1968) will serve as an example. In the figure the author shows how a sequence of morphemes, the basic units, develops into a sentence which is supposed to be part of continuous speech. At the first stage we have the individual morphemes and their tone patterns, i.e. from left to right, high, atonic, dipping, dipping, falling, dipping and high tones. The second stage shows the results of the sandhi rules. For instance, in a sequence of two dipping tones, the first is turned into a rising tone. At the next three stages, the effect of stress is exemplified, e.g. a reduction of duration and tone in the pronoun <u>wǒ</u> 'I', and the quantifier <u>běn</u> 'a volume of'. The last stage shows the workings of sentence intonation. The most conspicuous effect in this case is a fall over the last syllable of the sentence.

What we will do here is to show how we have tried to handle all these phenomena within the framework of a generative intonation model developed earlier for Swedish (Bruce and Gårding 1978) and later expanded to account for the intonation of other languages as well (Gårding 1981, 1983). The main feature of this model is that it separates lexical prosody from phrase and sentence prosody. Table 1 shows examples of the sentences we have worked with. In Figure 2, two examples of the intonation

* Institute of Acoustics, Academia Sinica, Beijing

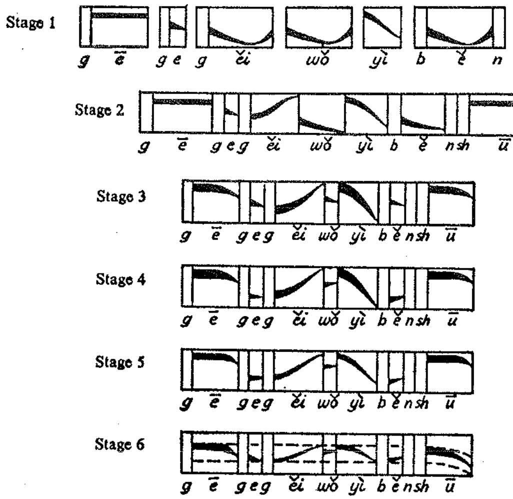


Fig. 17 Modifications of suprasegmental features in the sentence *gēge gěi wǒ yi běn shū*

Figure 1. From Kratochvīl 1968.

Table 1. Examples of test material.

STATEMENTS

I. Focus free	1,2 :	Wāng Yī chōu xiāngyān. Wāng Yī smokes cigarettes.
	3,4 :	Sòng Yán mài niúròu. Sòng Yán sells calf meat.
	5,6 :	Wāng Lǐ chuān yǔyī. Wāng Lǐ wears a raincoat.
II. Focus left	1,2 :	Shì, Wāng Yī chōu xiāngyān.
	3,4 :	Shì, Sòng Yán mài niúròu.
	5,6 :	Shì, Wāng Lǐ chuān yǔyī.
III. Focus right	1,2 :	Wāng Yī chōu xiāngyān.
	3,4 :	Sòng Yán mài niúròu.
	5,6 :	Wāng Lǐ chuān yǔyī.

QUESTIONS

Focus free	1,2 :	Wāng Yī chōu xiāngyān?
	3,4 :	Sòng Yán mài niúròu?
	5,6 :	Wāng Lǐ chuān yǔyī?

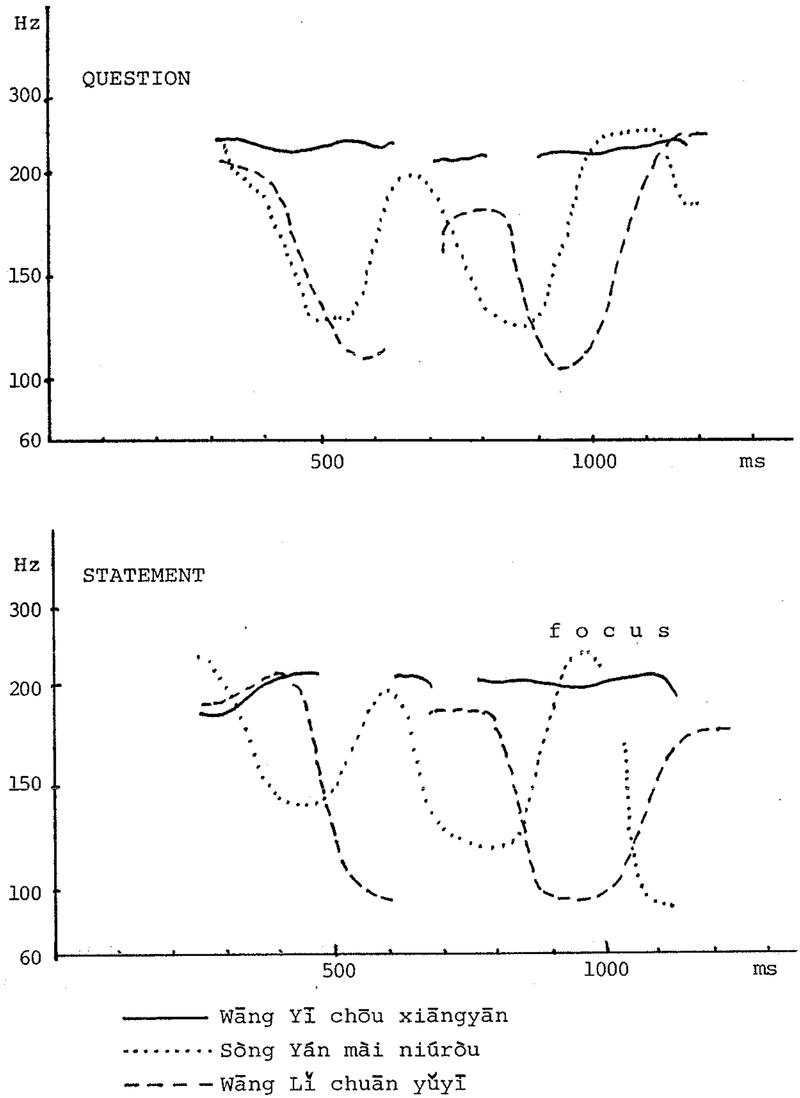


Figure 2. Three sentences with different tone patterns under two different intonations. Focus-free question above, statement with focus on the last word below.

contours are given, each with three different tone patterns.

Figure 3 shows the main concepts and principles of the model. It is here demonstrated how a typical intonation contour has maxima and minima, turning points, which are part of a global pattern, the grid, which can be rising, falling or level. The grid can be expanded or compressed. The part of the grid where its range or direction changes is called a pivot.

The local turning points, as well as the direction and range of the grid and the pivots are associated with communicative events in the following way:

Local turning points	lexical tones, accents (words, morphemes)
Pivots	syntactic boundaries
Direction of grid	speech act (statement, question, etc.)
Range of grid	prominence (focus, out of focus)

When a pitch curve is generated by our model, the grid, that is the global expression of sentence intonation, is generated first. Then the local maxima and minima pertaining to lexical items are inserted as points into the grid according to specific rules which state how the points are aligned relative to the segments. The fact that the location of these points is practically independent of sentence intonation and tempo makes this arrangement natural. In the last step, the pitch curve is obtained by smooth interpolation between the points over the voiced segments.

Figure 4 shows the different stages of the model of which the pitch algorithm is the final part. The stages are the same as for the other languages we have investigated (cf. Gårding 1981). The input is a sentence in which the morphemes are equipped with markings for tones, phrase accent and sentence accent, syntactic boundaries and speech act. It may also be marked for focus. The first three stages of the model take care of the duration and the last three of pitch. So far, this is a practical arrangement which was motivated by the fact that, in an accent language which was behind the original model, syllables could carry accentuation by means of duration only, without accentual pitch movements. The order of the stages for a tone

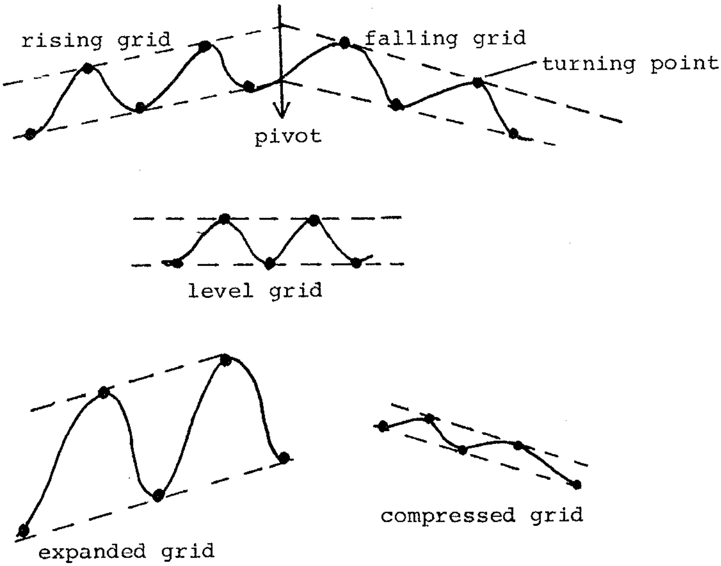


Figure 3. Concepts of the model

---

language must be regarded as very tentative.

We shall here concentrate on the pitch generating part, of which Figure 5 is a close-up. Among the intermediate phonological rules for Chinese we have the sandhi rules, e.g.

T3 ⟶ T2 / ___ T3

which means that a dipping tone becomes a rising tone before another dipping tone.

We also have tone reduction rules. For instance, in a phrase which consists of three morphemes (such as jiàoyùbù 'ministry of education' in the example sentence on Figure 4), the middle one may be reduced in duration and become atonic.

At the next level, the intermediate pitch representations, the abstract tone symbols are converted into more concrete ones, Lows and Highs, and combinations of Lows and Highs. These symbols are rough approximations of the tone contours of the dialect we are analyzing. It should be noted that they are not

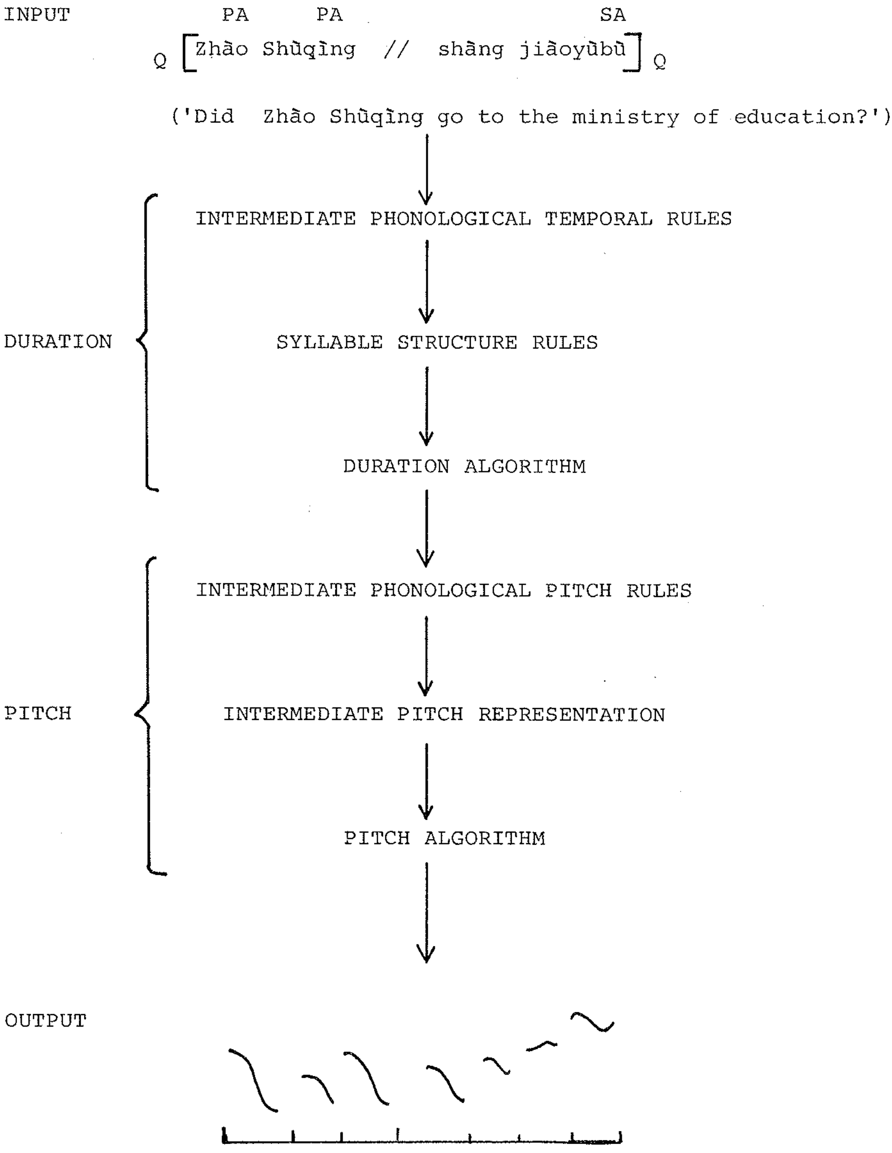


Figure 4. Model for prosody.

Intermediate phonological rules (symbol level)

Sandhi rules, e.g. T3 → T2 / ___ T3
Tone reduction rules

Intermediate pitch representations (symbol level → more concrete level)

Tones: T1 → HH
T2 → LH
T3 → LL
T4 → HL

Speech acts:
Statement: Global fall and/or local fall
Question: Global rise and/or local rise
Focus: Part in focus expanded and/or
Part out of focus compressed

Boundaries: Pivots of various kinds
Time position predictable from syntactic structure mainly

Pitch algorithm (cf. Figure 6)

Figure 5. Close-up of the pitch-generating part of the model.

---

equivalent to citation forms. To obtain a citation form, one of of these tone symbols has to pass through the pitch algorithm with a grid expressing statement.

Markings for speech act and boundaries are turned into concrete representations bearing on the grid:
- A statement has a global fall over the final phrase
- A question has a global rise over the final phrase
- Focus is represented by an expanded grid
- Out of focus is represented by a compressed grid
- The boundaries are converted into pivots

All these global features seem to be next to universal.

The final stage is the pitch algorithm (Figure 6). The first rule concerns the grid, i.e. the global frame for sentence intonation. This grid is drawn using information about speech act and syntax.

The second rule inserts the tone marks.

The third rule adjusts the tone marks according to the context.

The last rule takes care of interpolation.

Attached to the pitch algorithm there are prescriptions for determining the phonetic values needed to generate the output signal. These are exemplified in Figure 7.

First we need to know the relative levels that the pitch moves between, the floor (L) and the ceiling (H) of the normal voice register, and the levels of the highs (h) and the lows (l). We call this scheme of levels the stave. For our speaker, the range between the floor and the ceiling is about one and a half octaves, and the range between high and low tones one octave. Visual inspection of the intonation contours from our other informants shows that this is not an individual characteristic. Our Swedish speakers have a range of about an octave. For our Chinese speaker, the range narrows in fast, informal speech.

We also need prescriptions for the grid. In the exemplified case, which is a question, we have a pivot in the middle, a grid rising to the ceiling at the end of the final phrase, and a falling grid for the preceding phrase.

We need to know how the grid is expanded and compressed for the part which is in, or out of focus.

Only one example of an alignment rule will be given here, the one for the dipping tone, T3. It says: reach level L in the middle of the vocalic segment and stay there for a short interval. This lingering is behind our representation, LL. It is the only tone that always comes down to this level and it seems to be a necessary condition, as shown informally by experiments using the LPC technique.

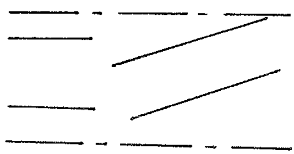
There are at least three kinds of context rules. They express undershoot (assimilation), sharpening of a contrast (dissimilation), and priority given when there is a clash between commands at phrase level and lexical level.

PA PA SA

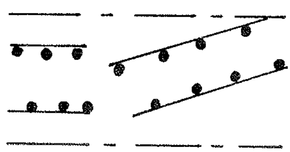
INPUT [Zhào Shùqìng / shàng jiàoyùbù?]

RULES

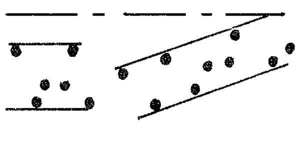
1. Draw the tonal grid:



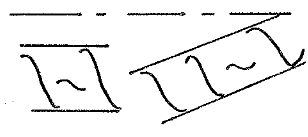
2. Insert the tone marks:



3. Context rules:



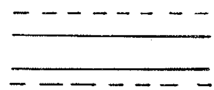
4. Interpolation and smoothing:



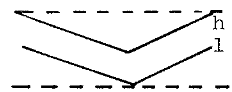
OUTPUT

Figure 6. Application of the pitch algorithm

Levels in the register: the stave



How to draw a grid



Alignment rules for tones in relation to segments

Context rules: Assimilation
Sharpening

Smooth interpolation

Figure 7. Prescriptions and conventions of pitch algorithm

We shall give one example of each:

Undershoot: in a sequence consisting of high tones, these tones tend to undershoot Level h.

Sharpening: in a sequence of high tones, there are, particularly in slow speech, small dips between the highs as if to emphasize the highs. There is a corresponding sharpening of the lows in a sequence of low (dipping) tones, by introducing rise-falls. Only in this case, the effect is constant and has a larger range. It is a phonological sandhi rule (T3 → T2 / ___ T3), and has its place among the intermediate phonological rules.

Priority: the constant feature of focus is an expanded pitch range. The frequency position of this range is determined by the tone. For a rising tone, the ceiling will be given priority to achieve the expansion, and for a falling tone the floor.

Interpolation: cosine functions may be used for the interpolation between the generated points.

The earlier shown figure, Figure 6, shows the workings of the four rules for a particular input sentence.

Figure 8 gives an idea of the algorithm generating our demonstration sentences of Figure 2. The existence of the pivot is strengthened by the contour of the long sentence below (Figure 9).

Some well-known physiological and psychological principles are at work in the rules of the model:

The inertia principle is reflected in the context rules which take into account the difficulty of large movements over a short interval of time.

The least effort principle is also apparent in the context rules in various short-cuts of F_0-movements which do not lead to an impairment of communication.

The contrast principle may be at work in the rules of sharpening and in the rules of priority.

Finally, the look-ahead principle finds its expression in the grid.

This broad-brush analysis of one speaker's Standard Chinese intonation will serve as a model for the treatment of our remaining data, collected from five more speakers of Standard Chinese.

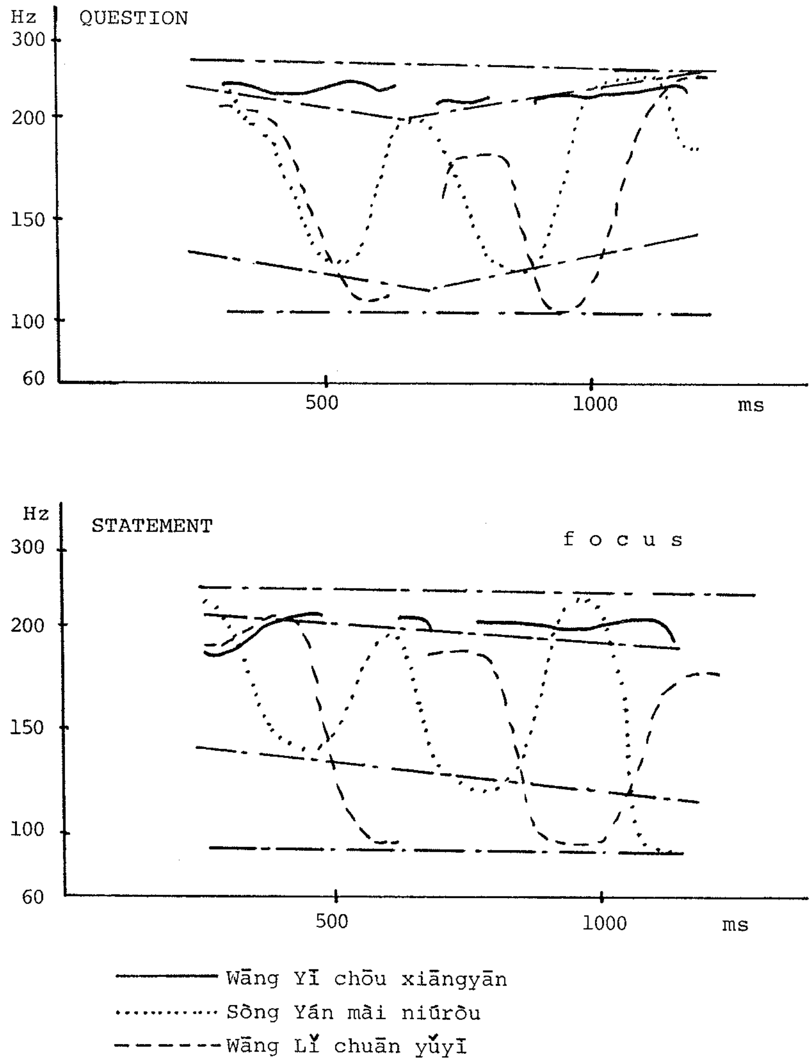


Figure 8. Grid superimposed on intonation contours of Fig. 2

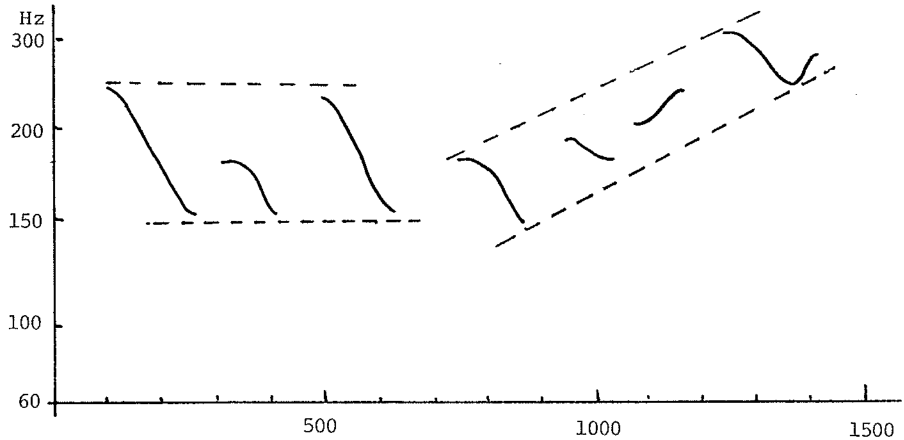


Figure 9. Pivot, illustrated by the sentence <u>Zhào Shūqìng shàng jiàoyùbù?</u> 'Did Zhào Shūqìng go to the ministry of education?'

REFERENCES

Bruce, G and E Gårding. 1978. A prosodic typology for Swedish dialects. In Gårding, E., G Bruce and R Bannert (eds.) Nordic prosody. Travaux de l'Institut de Linguistique de Lund 13, pp. 219-28. Lund:Gleerup

Gårding, E. 1981. Contrastive prosody: a model and its application. Studia Linguistica 35,1-2:146-65.

Gårding, E. 1983. A generative model of intonation. In Cutler & Ladd (eds),Models and measurements. Springer. In press.

Kratochvíl, P. 1968. The Chinese language today. London.

Wú, Z. 1982a. Pǔtōnghuà yǔjù zhōng de shēngdiào biànhuà (Tone changes in Standard Chinese sentences). Zhōngguó yǔwén 1982:6,439-50.

Wú, Z. 1982b. Rules of intonation in Standard Chinese. Preprints from the 13th International Congress of Linguistics, Tokyo.

This paper was presented by Gårding at a conference, Prosody normal and abnormal, arranged by the Association Européenne de Psycholinguistique in Zürich.

Our work is part of a project, Fonetisk beskrivning av några viktiga språk i tredje världen, sponsored by the Research Council for Humanistic and Social Sciences.

Working Papers 25 1983
Linguistics-Phonetics
Lund University, 67-74

PERCEPTUAL INTERACTION BETWEEN FO EXCURSIONS AND SPECTRAL CUES

David House

ABSTRACT

Questions and problems are formulated concerning perception of Fo movement in the frequency domain and possible perceptual relationships between such movement and spectral cues, particularly vowel formant transition place-of-articulation cues for stops. These issues are to provide the basis for experimental thesis work in speech perception involving both hearing impaired listeners and listeners with normal hearing. A preliminary experiment using band-pass filtered speech and normal listeners where the task was to identify VCV words with varying Fo movement indicated the possibility that a certain improvement in stop identification might be related to increased movement of Fo.

INTRODUCTION

Fundamental frequency excursions in running speech are generally considered to play an important role in signaling sentence focus, i.e. key words or phrases which contain new information important for sentence comprehension. Listeners can be forced to rely heavily upon "bottom-up" speech processing when perceiving segmental characteristics of words in focus where little or no contextual information is present to assist in presupposive "top-down" speech processing (Marslen-Wilson & Tyler, 1980). This type of new information processing therefore places greater demands on the segmental resolution of words in focus than does the processing of words outside focus. How then does the perception mechanism process the Fo excursions related to focus? Is it simply a matter of contrasting the focus word to the other words and, in effect, causing a sharpening of attention, or can we find relationships between

Fo movement and spectral cues which could provide direct assistance to the perceptual mechanism in performing the task of segmental resolution?

Fo excursions can be thought of as perceptually significant from the point of view of dynamic perception in that they comprise a movement or change in frequency over time. This movement could be registered as an event which would sharpen attention and aid in short-term memory retrieval of spectral cues. Hypothetically, Fo movement could also directly interact with spectral cues such as a timing relationship between V.O.T. and an Fo rise or an interaction between Fo movement and vowel formant transitions. Also, can Fo movements alter the spectrum, e.g. causing F_1 to rise, and if so, what perceptual significance is found in such alteration? Finally, is there an optimum Fo excursion in scope and range for facilitating segment perception?

The purpose of this report is not to provide answers to these questions but rather to formulate some problems which are to at least serve as the initial basis for my thesis work in speech perception and to report the results of a preliminary listening test.

FO AND FORMANT TRANSITIONS

As both Fo movement and vowel formant transitions involve a change in frequency over time, these two aspects of the speech wave can serve as a point of departure in the investigation of perceptual interaction. Resonance induced vowel formant transitions are generally considered to be important perceptual cues for stop consonant place-of-articulation identification. Formant frequency movement can be seen as rising or falling shifts in intensity through successive layers of harmonics of the fundamental. As long as a steady fundamental frequency is maintained, the formant transition patterns comprise a single movement. During an Fo excursion, however, the formant transitions will be altered by rising and falling harmonics. We are presented then, during focus, with a more complex frequency movement, the question being whether this added complexity is perceptually significant and if so, can this movement facilitate stop consonant identification?

FREQUENCY MOVEMENT AND HEARING LOSS

If we can establish some type of perceptual interaction between Fo and formant transitions, what could the possible implications be for listeners with hearing disabilities? It is well documented that place-of-articulation identification causes difficulties for individuals with moderate sloping sensorineural hearing losses. These difficulties, however, do not necessarily relate directly to specific frequency intensity attenuation. Van de Grift Turek, et al. (1980) demonstrated that subjects having similar audiometric configurations differed radically in their performance in synthetic stop identification tests.

Similar performance differences have also been reported by Risberg and Agelfors (1978) concerning the identification of intonation contours. This brings us back to the question of frequency movement processing. Can individual differences be accounted for by variations in capacity to perceive and process various types of movements in frequency and their possible interactions? If so, can this information be of any help to individuals with hearing disabilities?

PRELIMINARY TEST METHOD

To begin testing some of these questions, a preliminary perception experiment was devised using band-pass filtered speech in an attempt to roughly simulate a hearing loss and to essentially confine formant transition information to F_1 while varying Fo excursions by using different forms of presentation. The following test material was randomized and recorded by a native speaker of American English. Three different presentation types were used to vary Fo: 1) neutral, 2) question, and 3) emotional-emphatic. The capitalized words indicate sentence focus. Each sentence was read twice, once with "fifteen" and once with "fifty".

Test material:

She said to KEEP fifteen. (fifty)
She said to TUCK fifteen. (fifty)
She said to PIT fifteen. (fifty)
She said to put FIFTEEN. (FIFTY)
She said to tip FIFTEEN. (FIFTY)
She said to cut FIFTEEN. (FIFTY)

He said to DUB fifteen. (fifty)
He said to BAG fifteen. (fifty)
He said to GOAD fifteen. (fifty)
He said to dab FIFTEEN. (FIFTY)
He said to bid FIFTEEN. (FIFTY)
He said to dig FIFTEEN. (FIFTY)

The test was presented through a Fonema filter (390-820Hz, -36dB/octave) to two listeners with normal hearing and no prior knowledge of the test material. They were instructed to write the word and number which followed "He said to". They were also instructed to indicate if they heard "She" instead of "He" at the beginning. This was to provide a rough test of the filter function.

Selected spectrograms and mingograms were then made from the material (Fig. 1, 2).

RESULTS

Sentence focus had no significant effect upon word identification in neutral and question presentation. Correct identification remained about 1 in 3 (Fig. 3). With emotional-emphatic presentation, however, word identification was greater in focus (1 in 2) than outside focus (1 in 3) (Fig. 3). The numbers, "fifteen, fifty" were correctly identified in all cases, and the sibilant in "she" was effectively filtered.

DISCUSSION

Although this preliminary experiment contains a number of uncontrolled variables such as intensity and task word frequency, a relationship can be seen between increased word identification and Fo movement. Fo movement is greatest in the task word

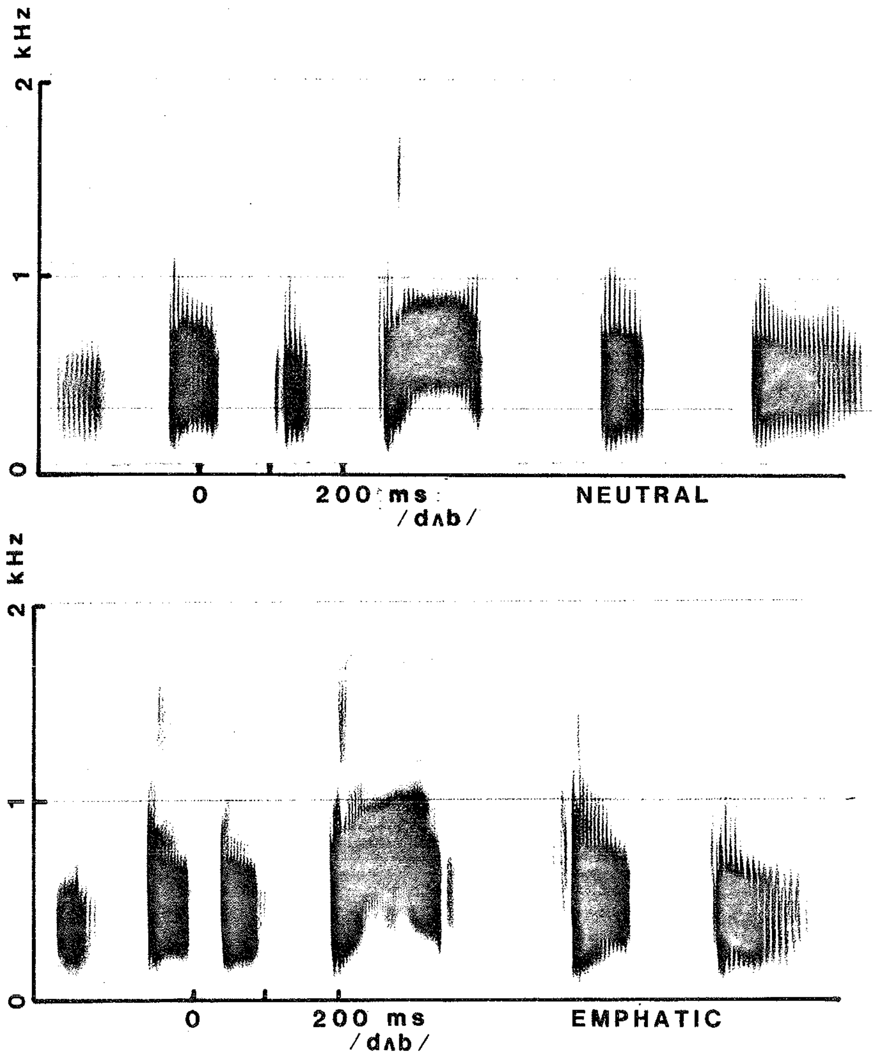


Figure 1. Spectrograms of the filtered test sentences, "He said to DUB fifteen," in neutral presentation (top) and "He said to DUB fifty," in emphatic presentation (bottom).

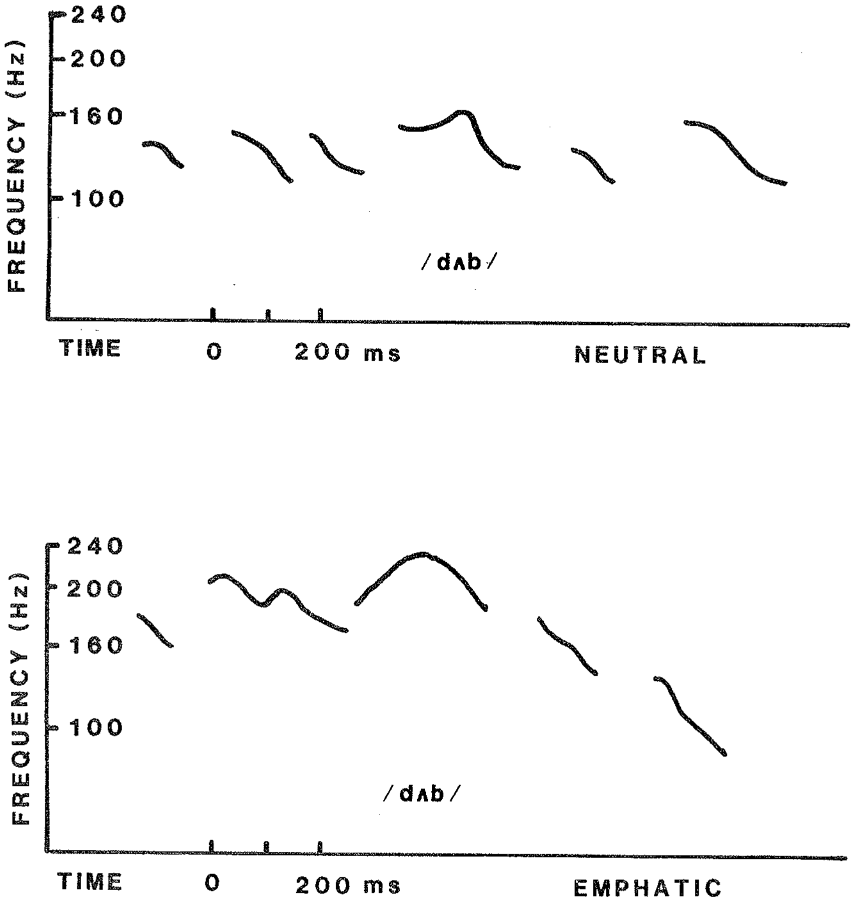


Figure 2. Fo tracings of the test sentences, "He said to DUB fifteen," in neutral presentation (top) and "He said to DUB fifty," in emphatic presentation (bottom).

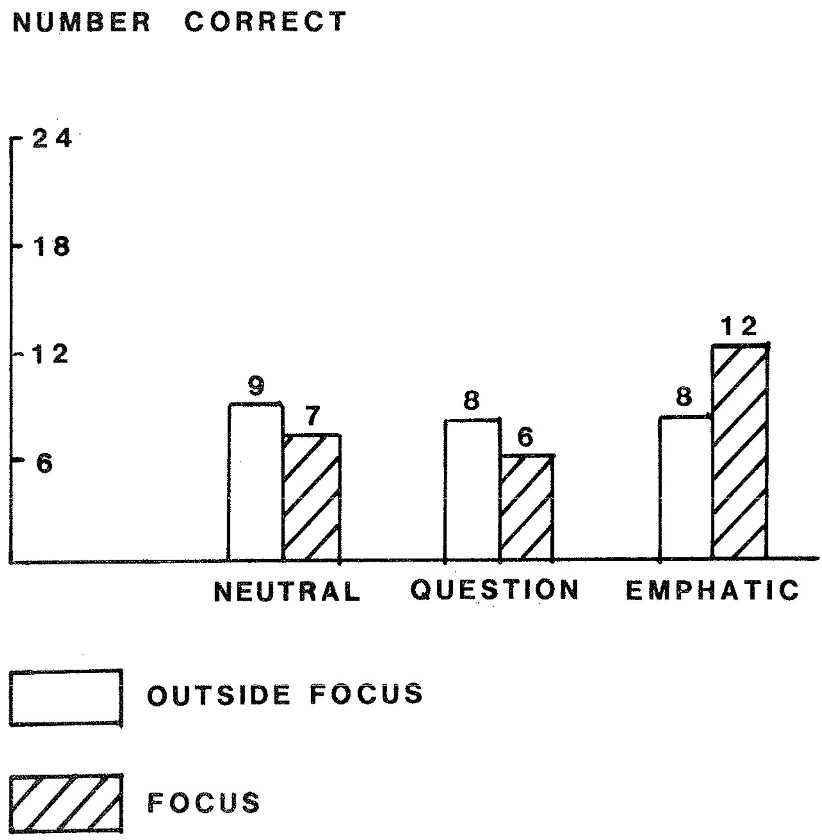


Figure 3. Results of the listening test showing number of task words correctly identified in the three presentation types divided into focus and outside focus.

in focus in the emotional-emphatic presentation (Fig. 2) where identification was also greatest (12 of 24). Intensity, of course, may have been a major factor although length was not appreciably different between words in and outside of focus or between words in different presentations. There was also a spectral raising in the emphatic presentation (Fig. 1) which could have contributed to identification.

The correct identification of "fifty-fifteen" can be attributed to length and intonation differences which are readily perceived since the same numbers are used throughout the test. This demonstrates the power of these cues in primarily "top-down" processing.

Additional experiments are planned enlisting the help of speech synthesis and listeners with hearing disabilities in an attempt to improve our understanding of cue interaction in speech perception.

REFERENCES

Marslen-Wilson, W. & Tyler, L. K. 1980. The temporal structure of spoken language understanding. Cognition 8: 1-71.

Risberg, A. & Agelfors, E. 1978. On the identification of intonation contours by hearing impaired listeners. Quarterly Progress and Status Report of the Speech Transmission Laboratory, Royal Institute of Technology, Stockholm 2/3: 51-61.

Van de Grift Turek, S., Dorman, M. F., Franks, J. R., & Summerfield, Q. 1980. Identification of synthetic /bdg/ by hearing-impaired listeners under monotic and dichotic formant presentation. Journal of the Acoustical Society of America 67: 1031-1039.

Working Papers 25 1983
Linguistics-Phonetics
Lund University, 75-83

INTONATION CONTOURS IN DIFFERENT REGISTERS
A pilot study.

Gabriella Koch and Anne-Christine Bredvad-Jensen

1. INTRODUCTION

The work presented here deals with how two different intonation contours, a statement and an engaged question, appear in different voice registers. The registers were obtained in two ways.
1. The intonation contours were produced by a man, a woman and a child, respectively, using their natural voice register.
2. Three different voice registers were produced by one person, an actress-singer (GK) using first her natural voice register and then a higher and a lower register, imitating a child's and a man's register, respectively.
In this paper we will concentrate on the natural registers used by the two women and the man for our two intonation contours. The results of the imitation will be presented later. In analysing the data, we addressed in particular the following question: Does a change from one register to another imply that all frequency values can be transposed by using a certain factor or is the transposition between different registers nonuniform?

2. PROCEDURE

The material used was a short SVO-sentence, Å mamma nannar Malla 'And mummy puts Malla (a girl's name) to bed' pronounced both as a statement and as an engaged question. Thus the sentence intonation is the only interrogative cue in the question.
The sentences were produced six to ten times by the two women and the man. The speakers were instructed to deliver focus-free productions. One of the women (KL) had some difficulty in producing focus-free questions. Examples of tracings of typical tonal curves, which are judged to represent the average case, are

shown in figure 1. Mean frequency values were calculated for tonal peaks, valleys, starting points and endpoints, see tables 1, 2 and 3. The mean frequency values are presented in a logarithmic scale in figures 2 and 3. The frequency values are plotted equidistant in the time domain. Successive medium-sized peaks are connected with a broken line, constituting the topline of the tonal grid. A baseline is drawn in a corresponding manner and the two lines obtained constitute the grid which encloses a major part of the tonal curve. See Gårding (1983) for a more thorough description of the tonal grid concept.

3. RESULTS

3.1 The effect of sentence intonation.

The declination of the grid lines represents a global fall of the tonal contours for all the statements. The questions are represented by rising grid lines except for one of the women (KL) who had some difficulty in producing this kind of question intonation. Still, there is a clear difference between her statement grid and her question grid. KL's global fall is much less pronounced for the question as compared with the statement and the frequency range between the topline and the baseline of the grid is markedly expanded in the question. A frequency range expansion of the grid in the questions can also be seen for the other persons.

3.2 The effect of register change.

A comparison between the male grid and the female ones displays a difference in the slope of the lines. The topline in the male grid has a steeper slope than the baseline, which makes the grid funnel-shaped. This is not the case for the two women where the topline and the baseline are approximately parallel. On the other hand, the male and female toplines have approximately the same slope. The distinguishing factor between male and female grids seems to be the different slopes of the baseline. These grid differences imply that register changes cannot be described by a simple transposition of the frequency values using a certain factor. Parallel grid lines have also been found for other female speakers of Swedish. A funnel-shaped grid is found for a Chinese male speaker (Jialu Zhang, personal communication).

3.3 The effect of syntactic boundary.

A tonal signalling of a major syntactic boundary, NP-VP, can be seen in figures 2 and 3, where the fall-rise in connection with the NP-VP boundary is more prominent than the following fall-rise, verb-object boundary (except for KL's statement). In the former case it reaches the baseline, in the latter case it does not. In connection with the NP-VP boundary the tonal configurations in statements and in questions behave differently. In questions the rise is more prominent than the preceding fall. In statements the fall is more prominent than the following rise.

4. CONCLUSIONS

The material presented here is rather limited, but if the results hold for more extensive material it will have importance for the understanding of the function of the topline and the baseline in intonation. It will also have implications for text-to-speech systems and the generating of natural-sounding male and female voices.

REFERENCE

Gårding, E. 1983. Intonation units and pivots, syntax and semantics. Paper to be presented at the International Congress of Phonetics in Utrecht 1983.

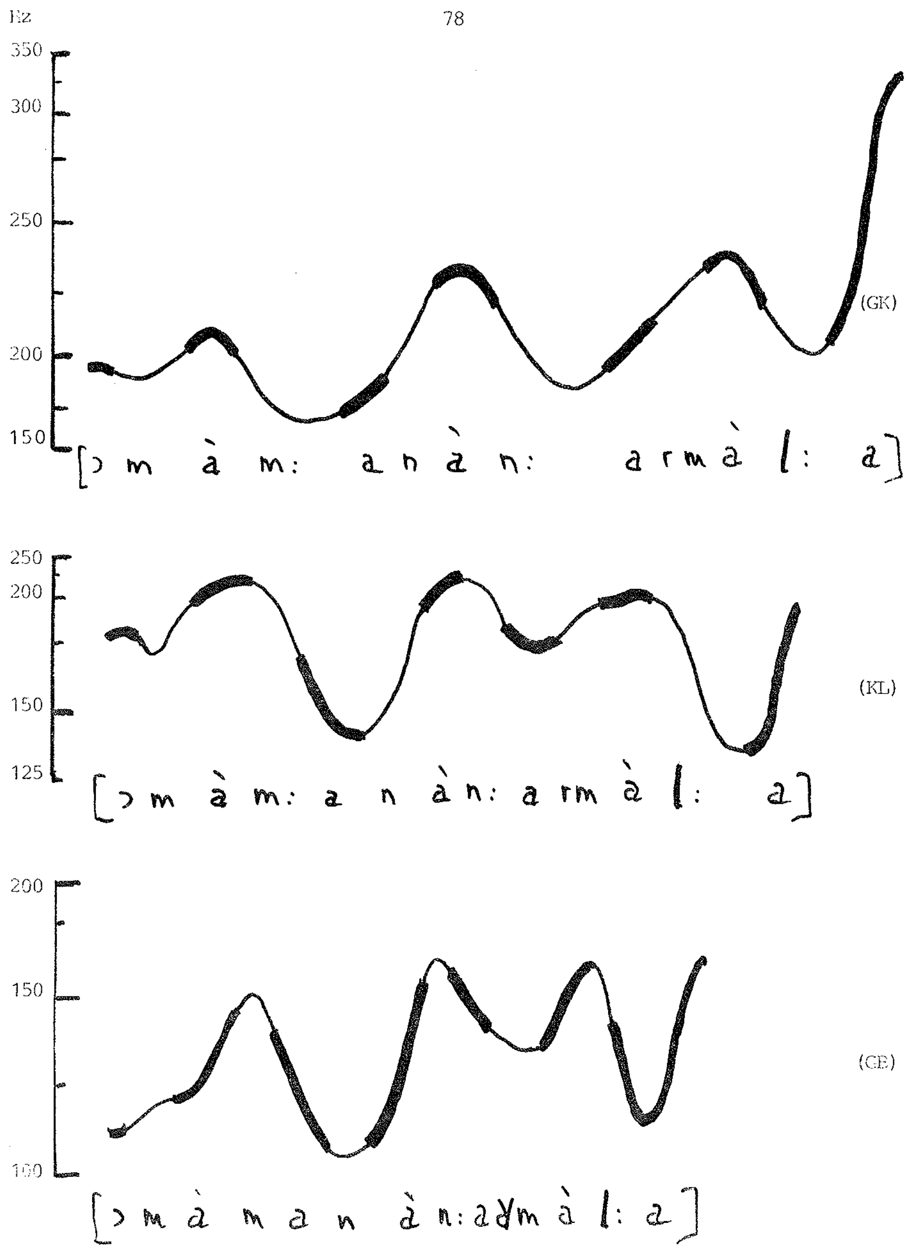


Figure 1. Tracings of typical interrogative tonal curves for two women (GK,KL) and a man (GB).

——— consonant ▬▬▬ vowel

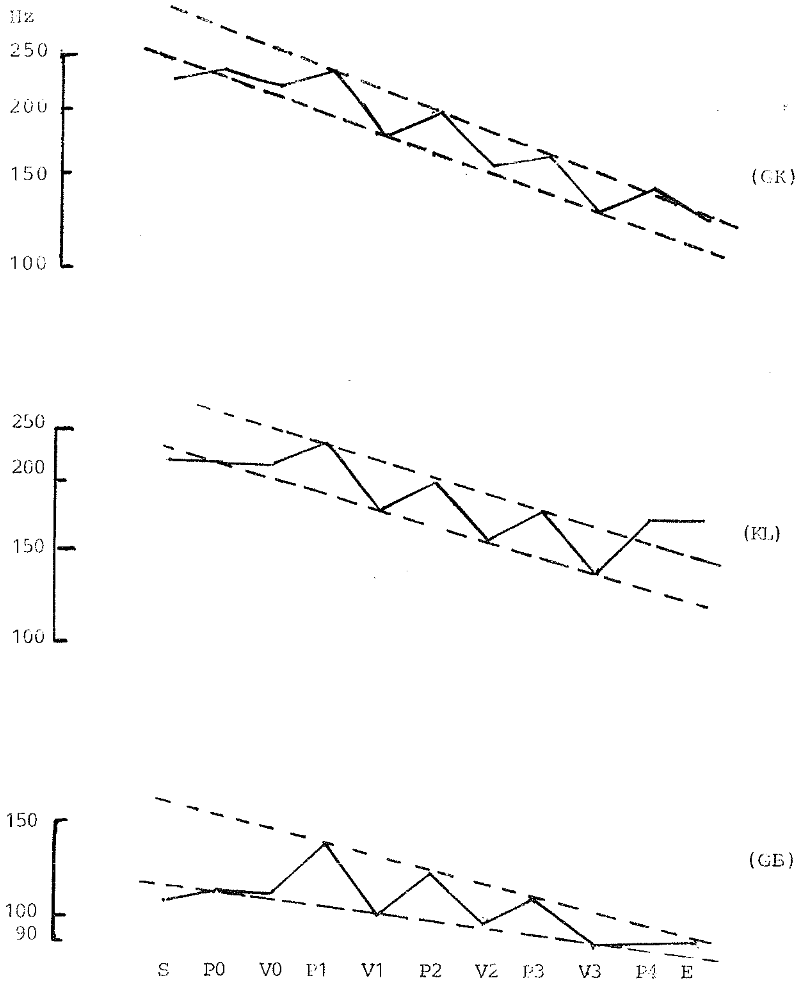


Figure 2. Mean frequency curves for statements. Two women (GK, KL) and a man (GB). S = starting point, P = peak, V = valley, E = endpoint. Logarithmic scale.

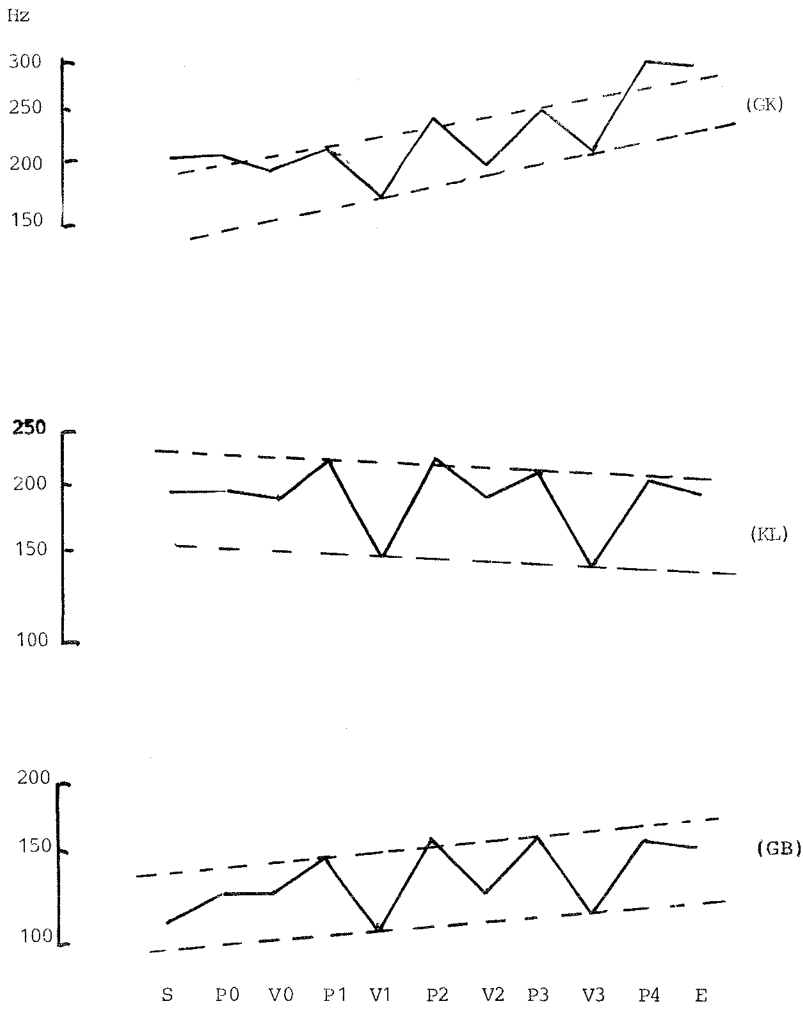


Figure 3. Mean frequency curves for questions. Two women (GK, KL) and a man (GB). S = starting point, P = peak, V = valley, E = endpoint. Logarithmic scale.

Table 1. Frequency values in Hz for the different productions, and mean values. Questions above and statements below. GK, woman. S = starting point, P = peak, V = valley, E = endpoint.

Productions	S	P0	V0	P1	V1	P2	V2	P3	V3	P4	E
1	200	204	190	198	168	212	180	225	180	250	250
2	200	200	180	200	165	222	185	242	210	300	300
3	200	210	195	212	170	230	197	243	205	280	272
4	210	210	198	212	175	233	202	245	215	286	280
5	200	200	190	205	170	240	195	245	193	275	267
6	200	205	195	205	165	242	204	250	210	300	300
7	180	185	183	207	170	240	190	247	203	310	310
8	210	205	200	210	167	238	197	245	210	312	312
9	197	195	190	210	165	233	182	238	198	320	320
10	200	200	185	200	165	238	180	252	205	318	318
Mean values	200	201	191	206	168	233	191	243	203	295	293

Productions	S	P0	V0	P1	V1	P2	V2	P3	V3	P4	E
1	227	235	223	227	177	195	157	163	123	140	130
2	230	235	215	230	190	205	160	168	125	152	130
3	210	227	215	230	183	190	160	158	142	140	125
4	225	230	223	245	190	205	153	154	120	135	115
5	227	240	225	240	190	200	153	162	125	140	120
6	235	240	225	246	190	195	160	163	130	142	120
7	230	240	215	226	177	198	155	160	130	140	130
Mean values	226	235	220	235	185	198	157	161	128	141	124

Table 2. Frequency values in Hz for the different productions, and mean values. Questions above and statements below. KL, woman. S = starting point, P = peak, V = valley, E = endpoint.

Productions	S	P0	V0	P1	V1	P2	V2	P3	V3	P4	E
1	195	195	190	225	150	210	170	196	141	190	187
2	176	180	170	217	142	220	173	201	139	197	195
3	210	212	210	240	154	240	210	222	148	210	195
4	190	190	190	220	140	212	163	192	130	185	166
5	195	195	190	220	146	225	220	228	151	228	215
6	220	220	210	225	157	223	185	210	145	200	198
7	178	178	170	198	140	230	185	209	136	190	180
8	174	175	175	220	147	238	212	220	149	212	210
Mean values	192	193	188	221	147	225	190	210	142	202	193

Productions	S	P0	V0	P1	V1	P2	V2	P3	V3	P4	E
1	245	242	240	260	176	202	148	176	126	185	178
2	215	218	226	245	195	205	158	175	143	191	186
3	235	235	232	242	196	208	175	188	138	184	178
4	240	235	230	240	178	193	158	168	130	156	152
5	175	174	170	210	164	191	154	185	140	172	163
6	178	178	175	203	159	195	150	174	139	177	174
Mean values	215	214	212	233	178	199	157	178	136	172	172

Table 3. Frequency values in Hz for the different productions, and mean values. Questions above and statements below. GB, man. S = starting point, P = peak, V = valley, E = endpoint.

Productions	S	P0	V0	P1	V1	P2	V2	P3	V3	P4	E
1	113	120	124	152	106	163	135	159	115	163	158
2	98	118	120	143	104	156	120	163	118	170	162
3	123	125	120	138	107	155	125	155	110	156	154
4	120	135	135	155	112	162	133	162	117	155	155
5	115	128	130	152	110	162	133	160	122	150	150
6	110	127	133	152	113	160	130	162	127	157	157
7	105	120	120	146	107	165	118	160	125	160	150
Mean values	112	125	126	148	108	160	128	160	119	159	155

Productions	S	P0	V0	P1	V1	P2	V2	P3	V3	P4	E
1	108	114	115	138	105	122	96	110	90	--	90
2	111	111	113	138	105	125	98	111	90	--	90
3	102	110	110	137	104	122	98	110	93	--	93
4	112	111	112	137	102	123	100	110	90	--	90
5	106	115	115	136	100	118	97	111	90	--	90
6	101	107	105	133	102	124	97	102	90	--	90
7	105	107	104	134	104	126	98	103	90	--	90
Mean values	106	111	111	136	103	123	98	108	90	--	90

Working Papers 25 1983
Linguistics-Phonetics
Lund University, 85-98

GLOTTALIC CONSONANTS IN HAUSA AND SOME OTHER LANGUAGES

Mona Lindau

The Hausa sound system includes one rather diverse class of consonants that is usually referred to as GLOTTALIC. Different places of articulation in this consonant class are associated with different glottal mechanisms. Labial and alveolar glottalic consonants have been described as voiced implosives, /ɓ, ɗ/. These sounds are produced by a rapid downward movement of the glottis. A palatal glottalic consonant is realized as a palatal glide with creaky voice, /'y/. The velar glottalic consonant is a voiceless ejective, /k'/. Ejectives are produced by a rapid upward movement of the glottis. These Hausa sounds were investigated acoustically. In addition, implosives and ejectives in some other languages were studied in order to highlight the language specific properties of glottalic consonants in Hausa.

The data consist of tape recordings of several speakers for each language. Most of the recordings were made on a good reel to reel tape recorder in the field. The glottalic consonants were in intervocalic position between a-type vowels in real words, said in a frame. These consonants were analyzed using a computer system for displaying the waveforms and spectra.

The palatal glide /'y/.

This sound is unusual in Hausa, but it clearly contrasts with a regular palatal glide /y/. Compare for example /'ya'ya/ "children" with /yaya/ "how?".

The acoustic analysis of the /'y/ was done in a qualitative way, inferring the glottal activities from the wave form. Figure 1 shows a typical wave form of a medial /'y/ in /'ya'ya/. One major characteristic is the very low fundamental frequency, about 35 Hz in this particular sound. Another characteristic is the high damping of the sound wave. This usually indicates a large amount of high frequency components in the sound. Emphasis of higher frequencies in the sound spectrum will occur when the glottal wave form of the sound source contains an angle approaching ninety degrees in the closure phase. A square wave contains an infinite number of high frequencies. The closer the closure angle of the glottal wave form approaches a square wave, the more higher frequencies will be emphasized in the sound spectrum. This fairly sharp angle of the closure phase in the glottal wave is a result of the vocal cords coming very rapidly and tightly together when vibrating. In this type of creaky voice in Hausa the vocal cords are vibrating very slowly, with a relatively high closing rate.

Implosives.

The list below states the languages concerned, their linguistic classification, and the number of speakers from whom data was collected. According to Maddieson (1981) about 10 % of the world's languages have implosives, and they are very common in certain geographical areas, like West Africa. All the languages in this list are spoken in Nigeria.

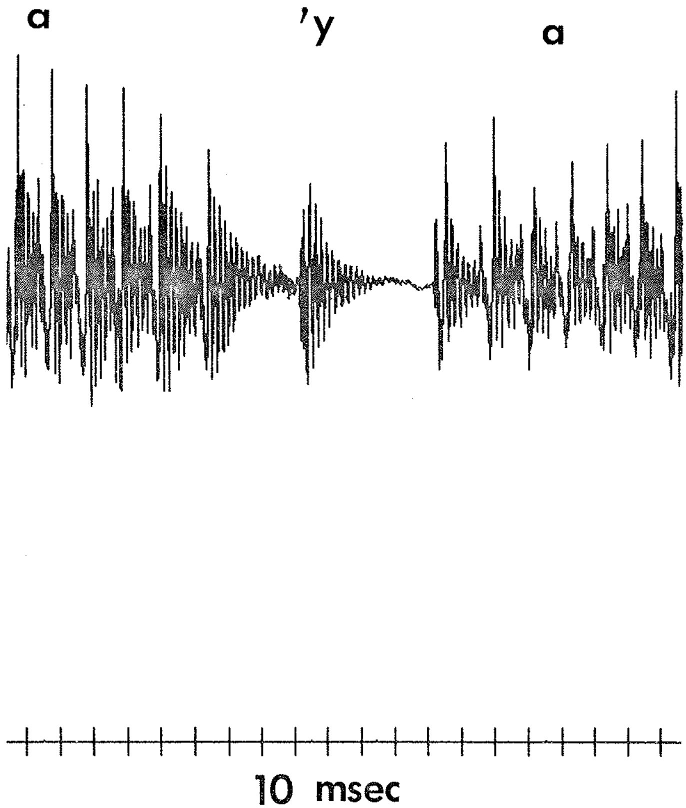


Figure 1. Waveform of the laryngealized palatal glide in Hausa.

	Languages		Number of subjects
AFRO-ASIATIC	Chadic	Hausa	14
NIGER-CONGO	Ịjọ	Kalaḅari	7
		Okrika	4
		Bụmọ	5
	Edoid	Degema	4

Figure 2 shows the waveforms of a regular plosive [b] and an implosive [ɓ] spoken by a speaker of Degema.

In the plosive, the amplitude of the vocal cord vibrations decreases gradually throughout the closure. As the supraglottal pressure increases, airflow through the glottis decreases, and eventually voicing dies out. Implosives, on the other hand, typically have either a gradual increase in the amplitude of the voicing - as in the figure - or, in other cases, a level, fairly large amplitude. This is due to an increase in the size of the vocal tract, as the larynx is lowered, and, also, the tongue body behind the place of articulation is typically lowered. The enlarged vocal tract volume keeps the supraglottal pressure from increasing, and voicing can be maintained at the same amplitude or at an increasing amplitude throughout the closure.

Note, too, that the first part of the implosive sound wave contains a certain amount of higher frequencies.

To describe some of the differences between the implosives in the selected languages, voicing amplitude and closure duration were measured, and characteristics of the waveform were examined. Peak-to-peak amplitude of voicing was measured in the middle and at the end of the closure at points marked by arrows on the figure. A ratio of these two amplitudes was calculated by dividing the final amplitude by the medial one. An implosive with level or increasing amplitude will have a ratio of 1 or more. In a regular plosive this ratio will be much less than 1. This measurement provides an indirect indication of the amount of cavity expansion in a voiced implosive.

Closure duration was also measured between points on the waveform display which, because of sharp changes in amplitude, were taken to be the offset and onset of the surrounding vowels. Thirdly, periodicity of the waveform and spectra at selected points were studied as an indication of phonation type.

Five out of the fourteen Hausa speakers produced implosives with voiceless closures, so these were excluded in the measurements of amplitude ratios and closure duration.

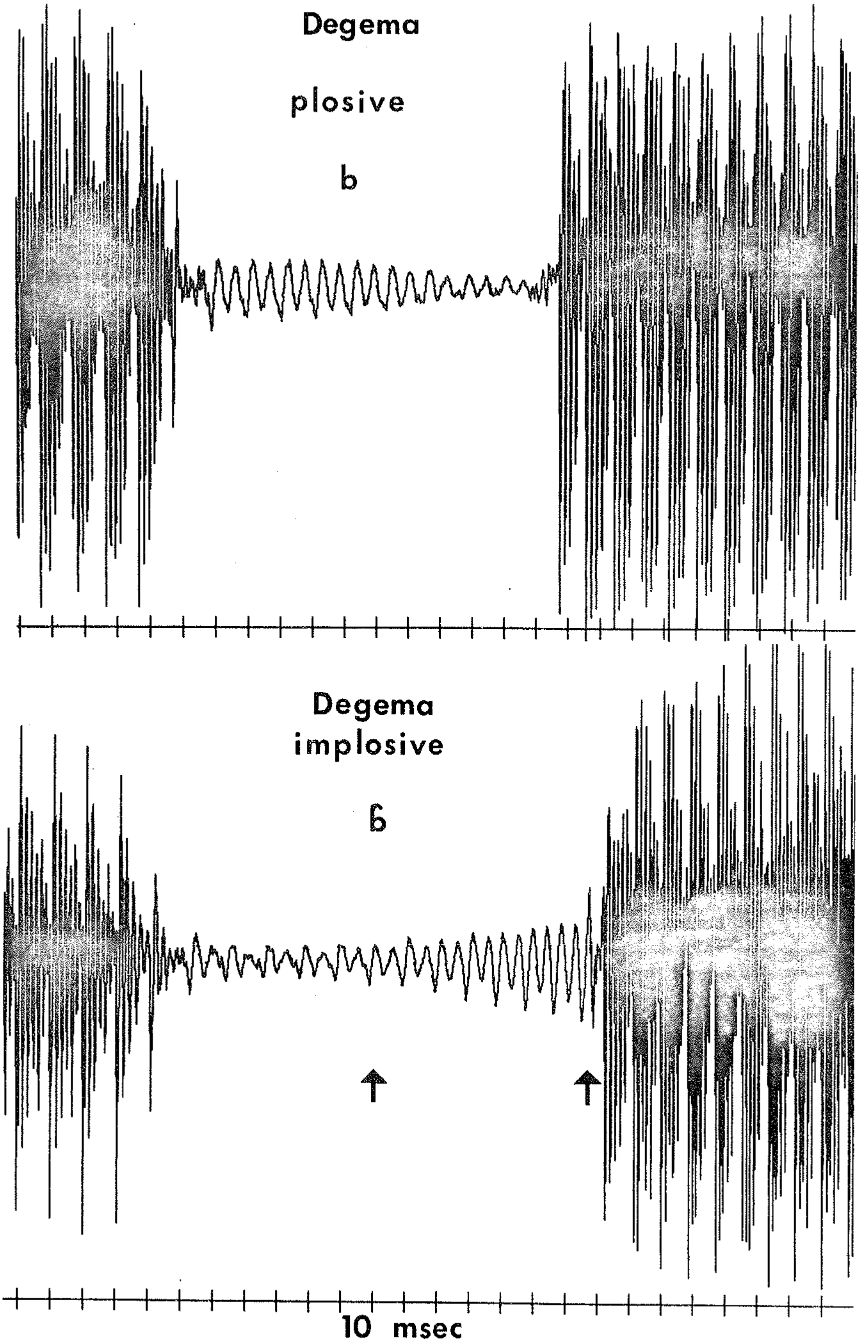


Figure 2. Waveforms of the medial plosive [b] and implosive [ɓ] in Degema.

The amplitude ratios and closure durations were averaged for all the speakers of each language. T-tests were used to assess the significance levels of the differences between the languages.

The results are illustrated in figures 3-7. Figure 3 shows histograms of the means of the amplitude ratios for the bilabial and alveolar implosives in each of the five languages. The histograms have been plotted in an order of increasing ratios, that is in an order of increasing amount of cavity expansion. The bars indicate one standard deviation above and below the mean.

For both the labial and the alveolar implosive Hausa has the highest amplitude ratio, indicating that it has the highest degree of implosion.

For the bilabial implosive the amplitude ratios are very significantly different between Hausa, on the one hand, and Degema, Kalaḅari, and Okrika on the other. Okrika is also significantly different from Bụmọ and Degema. For the alveolar implosive this measurement is significantly different only between Hausa on the one hand, and Kalaḅari and Degema on the other.

Figure 4 shows histograms of mean closure durations in the five languages in increasing order. Note that this measurement orders the languages in the same way for both bilabial and alveolar implosives. Hausa has significantly shorter closure durations than the other languages.

The Hausa implosives from the nine speakers are thus produced with both a relatively short closure duration and an amplitude ratio indicating a higher degree of implosion than occurs in the other languages.

However it is not true that a shorter closure always implies a greater degree of cavity expansion. For the languages apart from Hausa there is a tendency towards the opposite relationship between amplitude ratio and closure duration. Particularly for the bilabials, shorter closure durations are associated with lower amplitude ratios. Thus closure duration and degree of cavity expansion are independently variable phonetic parameters of voiced implosives. As to these two parameters Hausa behaves opposite to the Niger-Congo languages.

In addition phonation types were studied. Figure 5 shows a waveform of the bilabial implosive /ɓ/ in Bụmọ. It is typical of the voiced implosives in the four Niger-Congo languages. The first part of the closure displays a considerable amount of high frequency energy. Below the waveform there is a spectrum of the first 50 milliseconds of the closure, centered at the arrow, showing a clear formant structure. The high frequency energy in the waveform is thus an indication of the upper formants. This is probably due to the vocal cords vibrating with a relatively sharp closure while they are being held tightly together in the descending larynx, and this results in cavity resonances. The first part of the closure in these languages is thus typically produced with a form of laryngealization.

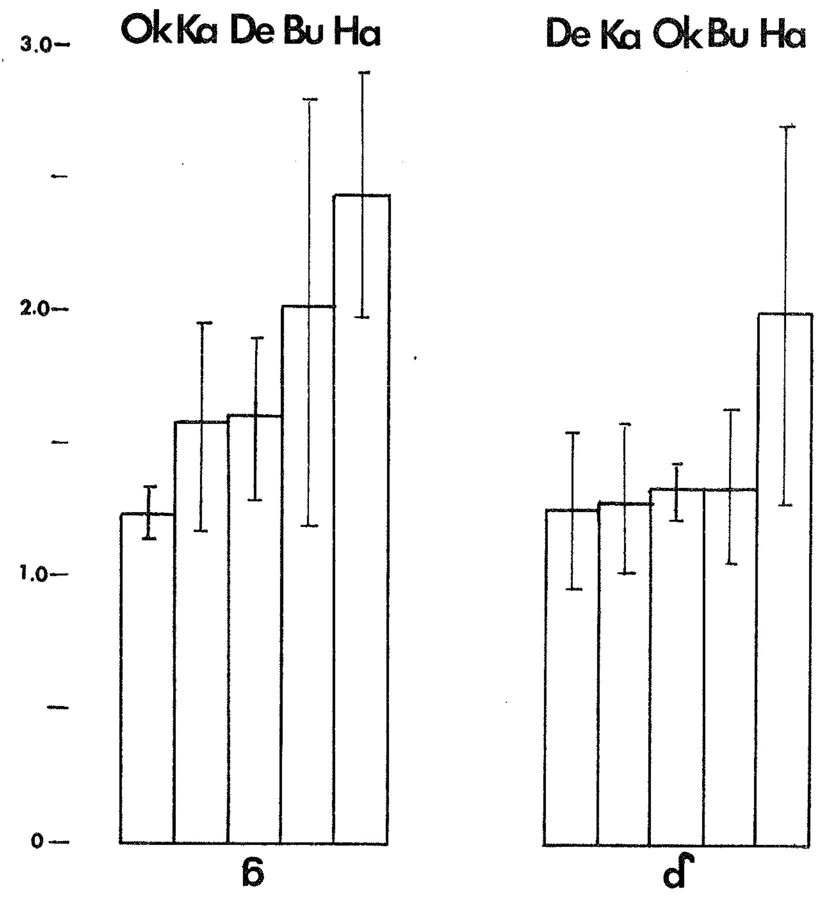


Figure 3. End/middle amplitude ratios of the labial and alveolar implosives in Okrika, Kalaḅari, Ḅụmọ, Degema, and Hausa.

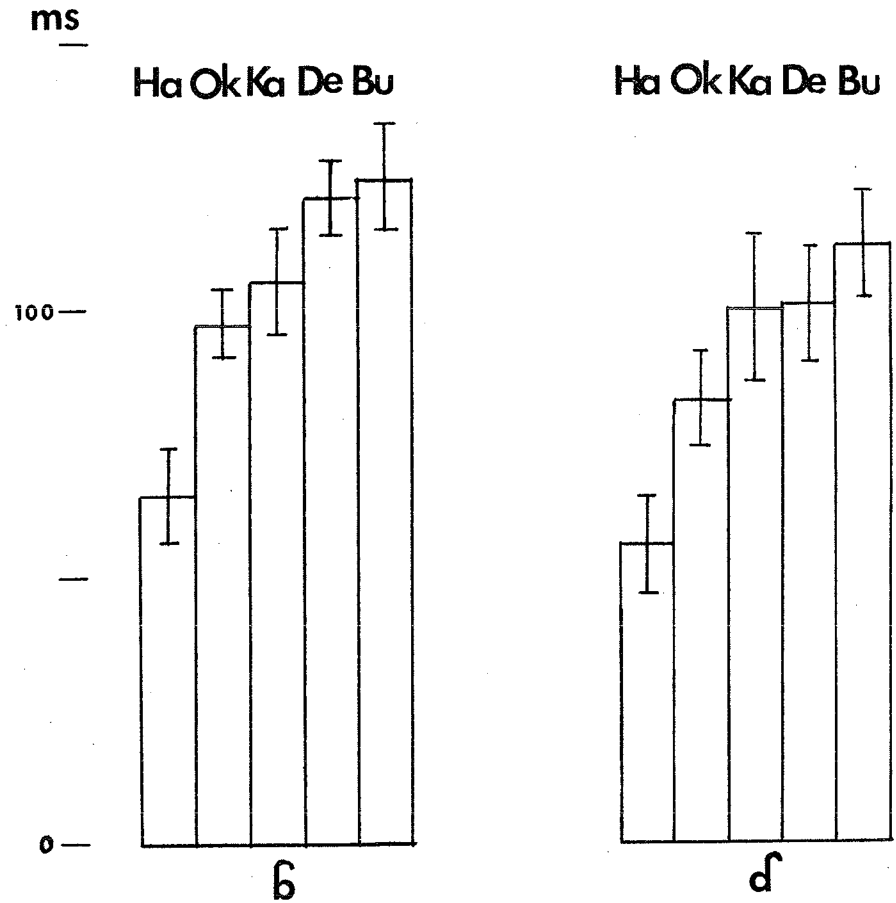


Figure 4. Mean closure durations of the labial and alveolar implosives in Hausa, Okrika, Kalaḅari, Degema, and Bụmọ.

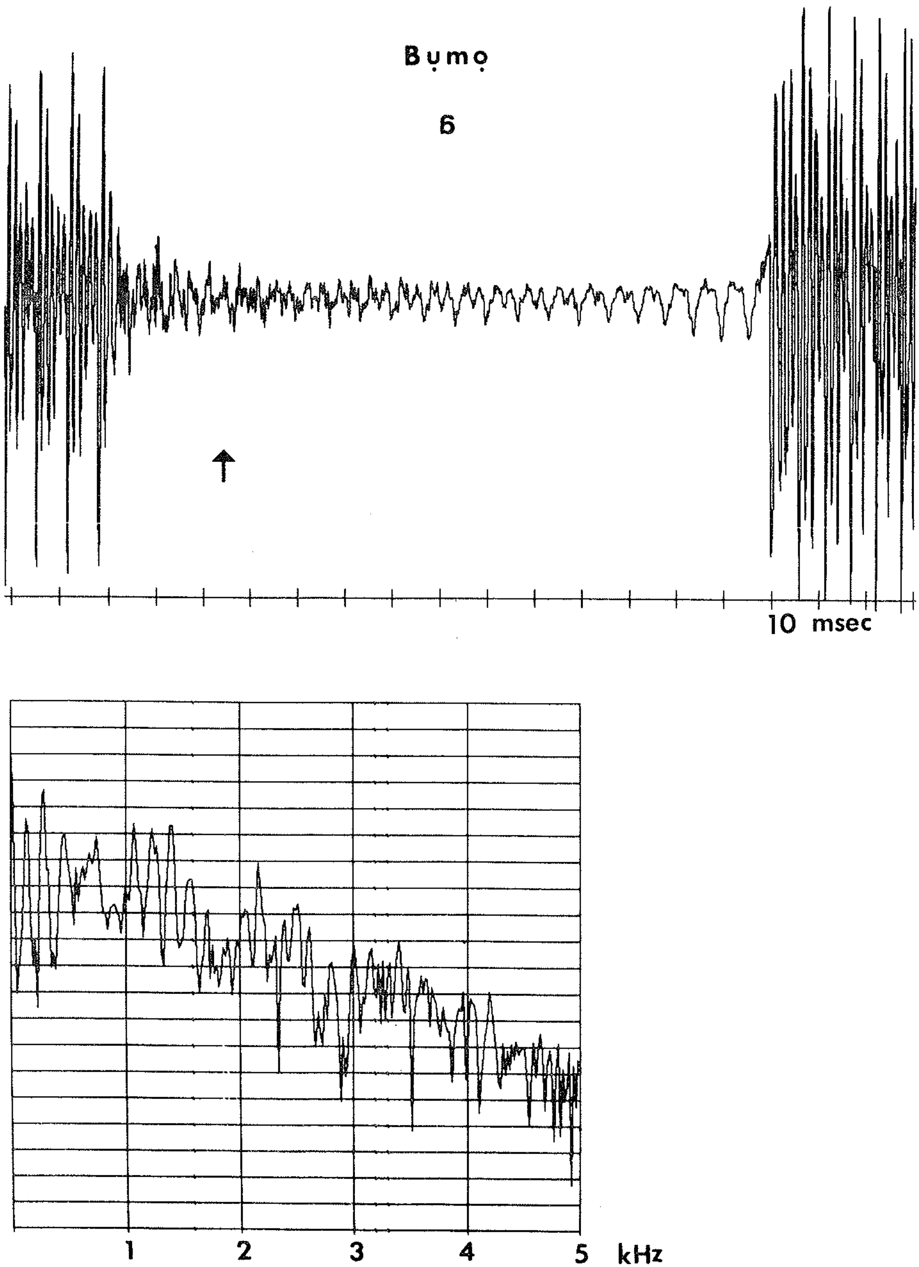


Figure 5. Waveform and spectrum of the labial implosive in Bụmọ.

Again, Hausa implosives differ considerably from the Niger-Congo languages in the waveform pattern during the closure (cf. Ladefoged 1964). There is also considerable individual variation in Hausa. Five out of the fourteen speakers produce a voiceless beginning of the closure, presumably from a glottal closure as the larynx descends. One speaker has an implosive just like those in the Niger-Congo languages. The eight remaining speakers produce an implosive as seen in figure 6. Hausa implosives display highly aperiodic vocal cord vibrations during the closure. The spectrum shows no clear formant structure but there is a peak around 3500 Hz. This peak cannot be due to cavity resonance from sharp closures in the vocal cord vibrations. If it were, the lower formants would be apparent as well as one at this high frequency. The peak is possibly instead due to noise from INcomplete closures in the vocal cord vibrations, and possibly also to noise generated by perturbations of the vocal tract walls as the larynx descends. The incomplete closures would also explain why the Hausa implosives last a shorter time, as there will be leakage of air through the descending glottis.The Hausa implosives are thus produced with aperiodic, inefficiently closing vocal cord vibrations. This is usually also labelled "laryngealization". Apparently, what we label laryngealization may involve several different mechanisms.

The voiced plosive [b] in Hausa is illustrated in figure 7. Typically it has periodic voicing vibrations with decreasing amplitude during the closure phase, so the voicelessness or aperiodicity in Hausa implosives may serve to keep them apart from the voiced plosives.

Thus, Hausa implosives are characterized by a type of creaky phonation, implosion, and relatively short duration. The Hausa implosives differ considerably from those in the Niger-Congo languages. This supports the reconstruction of implosives in Proto-Chadic (Newman and Ma 1966).In case they had been borrowed from neighbouring Niger-Congo languages, one would have expected closer agreement in their phonetic properties to Niger-Congo type implosives.

Summarizing the data on implosives, it is apparent that there are several independent phonetic parameters distinguishing segments that have been called implosives in different languages. Some languages use one combination of values of these parameters and others another combination. It is also evident that there is considerable speaker to speaker variation between implosives in languages, and that languages may differ in the way that they maintain distinction between implosives and the corresponding voiced plosives.

Ejectives.

Ejectives are more common in the languages of the world than implosives, about 18% of the world's languages have ejectives. These stops are produced with a closed glottis moving rapidly upwards, followed by glottal and oral releases. My data on ejectives consist of velar ejectives from twelve Hausa speakers and nine speakers of Navaho, an Athabascan language spoken in the Southwestern United States. These data were also analysed from displays of the waveform.

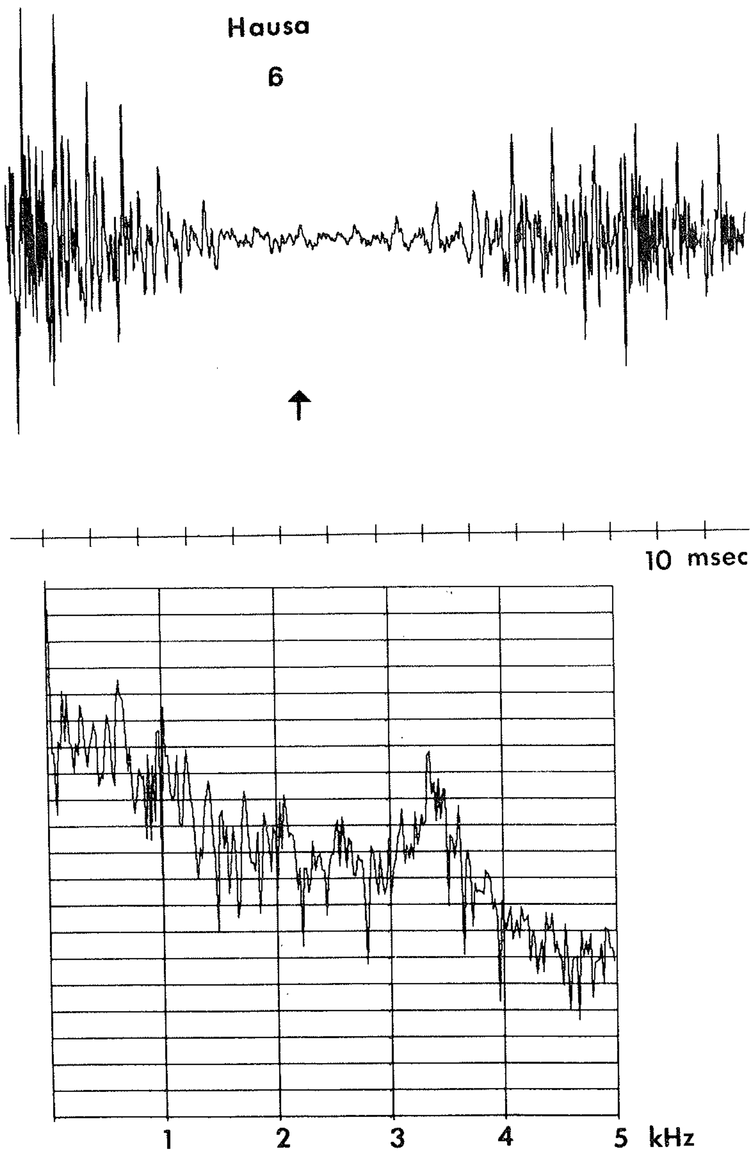


Figure 6. Waveform and spectrum of the labial implosive in Hausa.

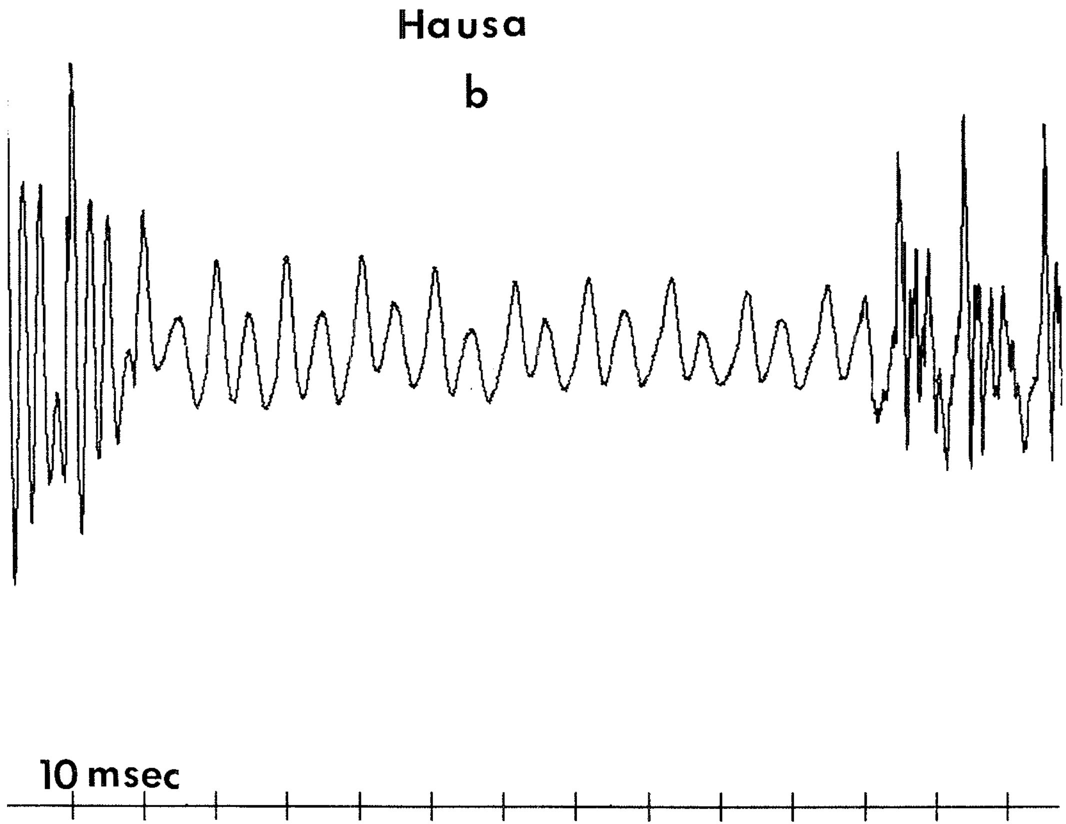


Figure 7. Waveform of the voiced plosive [b] in Hausa.

For each ejective, the total duration - that is, the sum of the closure duration and the VOT - was measured and a ratio of the closure duration to the VOT was calculated. A qualitative examination of the waveform was made to reach conclusions about phonation type.

The results show that here too variation between speakers was prominent in Hausa. Four out of the twelve Hausa speakers realise the ejective /k'/-phoneme, not as an ejective but as an unaspirated [k] or as a voiced [g]. All nine Navaho speakers have a true ejective [k'].

Figure 8 shows histograms of the means of the total duration and the closure duration/VOT ratio. Standard deviations are shown by the bars. For both these measures, the differences between Hausa and Navaho are highly significant. The ejective in Navaho has more than twice the total duration of the one in Hausa. This difference is not due to a slower speaking rate in Navaho. The rate of speaking in both languages was measured as number of syllables per second, and the difference was non-significant (both 4.3 - 4.5 syllables per second).

As the difference in the ratio indicates, these languages also differ in relative durations of the different parts of the ejectives. The closure duration in Hausa is about twice that of the VOT part, while the closure duration in Navaho is only 1.5 times as long as the VOT.

In addition, the vowel onset differs considerably in the two languages. This is illustrated in figure 9. In Navaho the glottal release coincides with the vowel onset, so the vowel starts with a sharp, large amplitude.

In Hausa, the glottal release occurs together with the oral release, and the vowel begins gradually, with aperiodic vibrations. This aperiodicity of the vowel onset after the ejective could be an important cue for differentiating voiceless plosives and ejectives in Hausa, since the Hausa velar plosive is followed by periodic onset of the vowel.

In conclusion, this study shows that both implosives and ejectives can vary in a number of ways. Some of this variation is reliably associated with the particular language concerned. From this it follows that contrasts between similar pairs of segments, such as implosives and voiced plosives, may be maintained by different values of particular phonetic parameters in different languages. These facts suggest that the phonetic component of the grammars of languages must be much more specific and detailed than is provided for in most current theories.

References.

Ladefoged, P. 1964. Phonetic differences in West African languages. Cambridge: Cambridge University Press.

Maddieson, I. 1981. UPSID. UCLA Phonological Segment Inventory Database: Data and Index. UCLA Working Papers in Phonetics 51.

Newman, P. and R.Ma 1966. Comparative Chadic: Phonology and Lexicon. Journal of African Languages 5.218:251.

Many, many thanks to Peter Ladefoged and Ian Maddieson for all their helpful comments. Thanks are also due to Kay Williamson, C. Jenewari, S.Okpara, and M.Garba for their help in collecting the data.

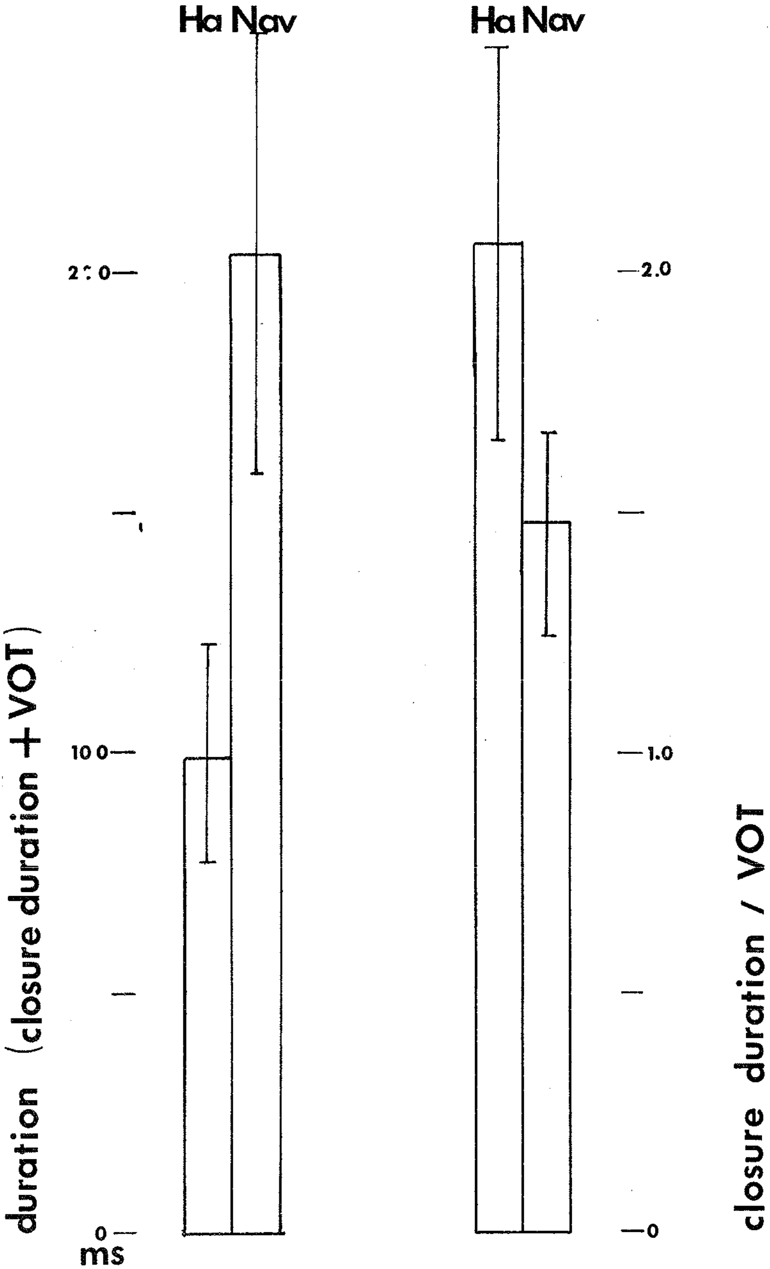


Figure 8. Means of the total duration and the ratio closure duration/VOT in Hausa and Navaho.

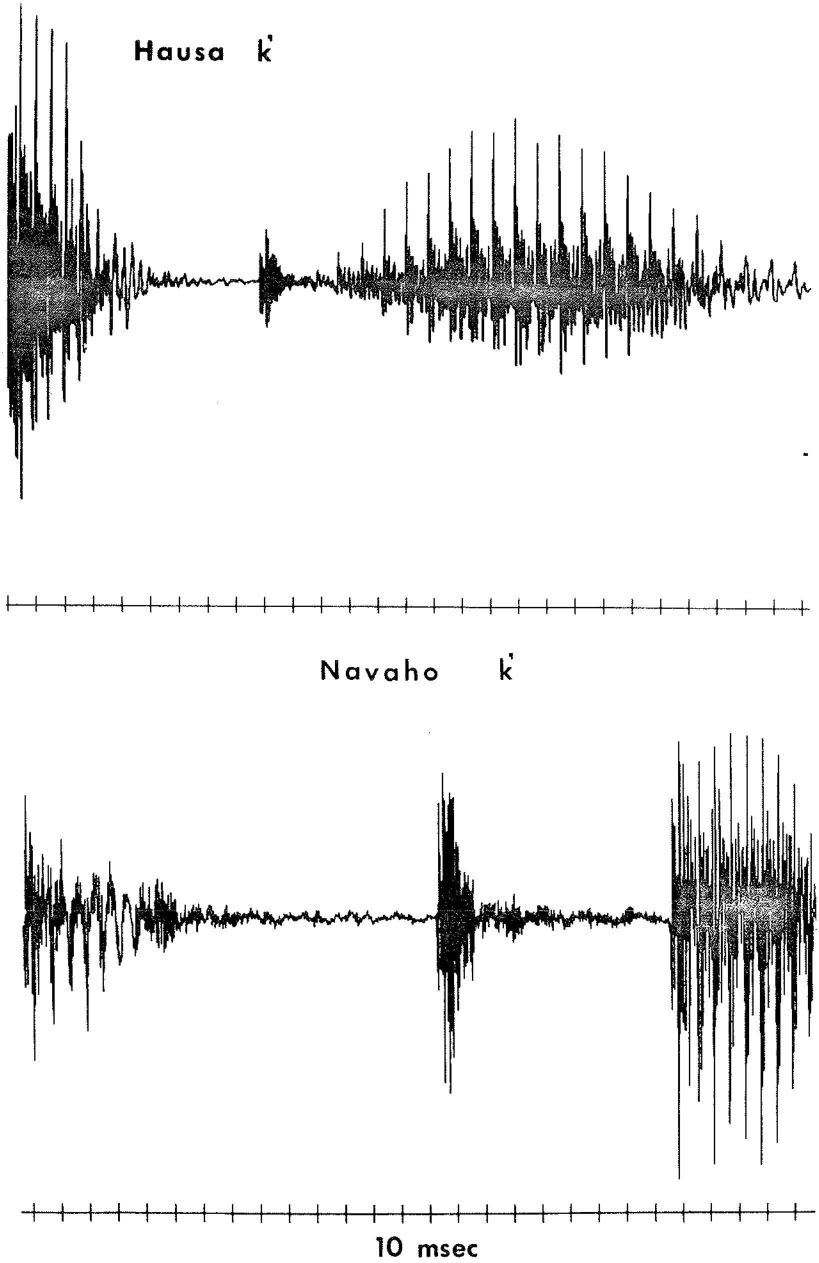


Figure 9. Waveforms of the velar ejectives in Hausa and Navaho.

Working Papers 25 1983
Linguistics-Phonetics
Lund University, 99-103

PhD THESIS SUMMARY: THE PHONOLOGY OF LANGUAGE DISORDERED CHILDREN

Eva Magnusson

INTRODUCTION

Not all children learn to speak at the same age and in the same seemingly effortless way. Some children have greater difficulties than others in acquiring language though there is no apparent reason for why this should be the case, such as defective hearing, mental subnormality, neurological dysfunctions, or emotional disturbances. When referred to a speech clinic, their problem is diagnosed as a functional articulation disorder, retardatio loquendi idiopathica, indicating that there is no known etiology for the difficulties.

Within the group of language disordered children, variation is considerable. In phonology, for instance, where the same patterns of substitution and deletion have been identified for many of these children, there is interindividual variation both in terms of phonological rules and in the conditions for their application. The children appear to differ not only in terms of how much and in what way they respond to language therapy. These observations suggest that children diagnosed as exhibiting a functional articulation disorder do not constitute a homogeneous group and that not all members of the group have the same degree or type of language difficulties.

In order to examine the variation within the group a study was undertaken with the aim of identifying clinically relevant subgroups on linguistic criteria. The study focuses on the following questions as a possible basis for such a subclassfication:

1. What characterizes language disordered children's produced forms and how do they differ from the normal adult forms?

2. To what extent can these children's produced forms be explained by limited production, perceptual, or processing abilities and to what extent by a different organization of the phonological system?

3. What kind of awareness do language disordered children show of phonological forms?

PROCEDURE

In order to examine these questions, an investigation was made based on the following kind of data:

Speech production Children's speech was recorded during conversation and the naming of pictures. The elicitation material had been prepared so that the items were varied as to number of syllables, syllables structure, and stress patterns. All Swedish vowels and consonants were represented in pre-, inter-, and postvocalic position under different stress conditions. Frequent consonant clusters were also included.

Auditory discrimination Sounds which the children did not use correctly were tested in minimal pairs where the two members of a pair were differentiated by a contrast made up of the sound that the children did not manifest and the sound they used as a substitution. If children were saying e.g. [sø:k] for /çø:k/, (kitchen), i.e. were using [s] as a substitution for /ç/, they were asked to identify a pair like [su:l] - [çu:l], (sun, skirt), two words that were homophones in the children's production. They were required to do this by pointing to one of two pictures when someone else was uttering the words and when they were listening to their own recorded version of the same words.

Imitation The children were asked to imitate the minimal pairs and to imitate the substituted sound and its substitute in nonsense syllables where both position and vowel contexts were varied.

Rhyming The children were asked to choose rhyming word pairs out of sets of rhyming and non-rhyming words.

Longitudinal data The longitudinal data consist of information about the children's phonological development during the 12 months subsequent to the investigation as described in their hospital records.

SUBJECTS

The subjects were 32 children with the diagnosis retardio loquendi idiopathica in the age range from 3;9 years to 6;6 years, 10 girls and 22 boys. At the time of the investigation, none of them had yet been enrolled in therapy programs.

RESULTS

The children's spontaneous speech production is described primarily in relation to the normal forms, i.e. the description is an account of the regularities in the differences between the child forms and the forms in the target language. The relations between child forms and adult forms are analysed in terms of processes, here defined as procedures or strategies that children use when modifying normal adult forms so that they be processable for an individual with a child's perceptual, productive, and cognitive capacity. In cases where the child's forms deviate markedly from the norm, a description of conditions exclusive to the system of that particular child is also made. The description is made both on a group level in order to bring out general tendencies and on an individual level in order to show individual preferences and patterns in the application of processes.

On the basis of individual patterns in the production data, four subgroups can be indentified, namely

1. Children whose speech is nearly mormal (4 subjects)
2. Children in whose speech implicational patterns can be found (17 subjedts)
3. Children who have one dominating segmental problem (5 subjects)
4. Children whose speech is characterized by a restricted number of word patterns (6 subjects)

1. Nearly normalized group Since the subjects had no consistent substitutions, auditory discrimination testing was not motivated in more than one case. Results on the rhyming test vary, indicating that normal or nearly normal speech does not guarantee an ability to rhyme.

2. Implicational pattern group The degree to which the children in the group deviate from the norm differs but they are classed together since the same implicational ordering, with only a few

exceptions, can be observed in the patterns they exhibit Implicational patterns are found only if different types of processes, substitutions, cluster reductions, and word structure processes are considered separately. The children also vary in terms of auditive discrimination, self-discrimination, and rhyming and this variation can be seen as a consequence of the children's differing developmental levels.

3. Special problem group The children in this group all have one dominating problem which almost exclusively characterizes their speech. They substitute one segment or one type of segment and not the types that are the latest to be acquired by normally developing children. All subjects manage auditive discrimination and four out of the five children in the subgroup can discriminate forms in their own speech which are perceived as homophones by an adult listener. Three children can rhyme.

4. Word pattern group The children in this subgroup show a preference for a small number of word patterns. Restrictions of word structure seem to be a more important determinant of their speech than substitution or cluster reduction patterns. All except one of the children are unsuccessful on auditory discrimination and half of them identiy homophones in their own speech correctly. Only one of the children showed some understanding of rhyming.

CONCLUSIONS AND CLINICAL IMPLICATIONS

The conclusions that can be drawn from this study of language disordered children's phonology are that they follow a generally recognized developmental order with some individual variation. This variation is not totally random but of a kind where similarities emerge between children which make a subclassification possible based on the individual children's developmental patterns. One group largely follows what is considered a normal phonological development, though slower. In the other two groups the children's patterns show chronological mismatch in that their phonologies have characteristics typical for both early and late phonological acquisition. For one of these groups, one segment or one class of segments is problematic and for the other, a small number of preferred word patterns place restrictions on their forms.

The motivations for the children's forms may differ depending on the degree and type of deviance. Perceptual problems are more apparent among children early in development, motor production problems predominate in the special problem group while processing problems and phonological organization are more important as motivations for forms in the word pattern group. In the implicational pattern group, all types of motivations may exist in varying degrees and combinations.

The acquisition of the phonology of a language is not only evidenced in production and perception of word forms but also by the ability to consciously manipulate phonological forms and to reflect upon language i.e. linguistic awareness. Linguistic awareness is not directly related to phonological ability as it is evidenced in speech production. A nearly normal speech does not guarantee a linguistic awareness, which might be quite developed in a very deviantly speaking child. To include linguistic awareness in the assessment of phonological disability increases the possibility of further differentiation within the group of disordered children and is also of interest because of its importance for learning to read and write.

For a valid assessment of phonological ability, it is thus not sufficient to study children's production. Such data need to be supplemented with information about possible motivations for the forms and about the children's linguistics awareness. Increased knowledge about the variation and patterning within the group may contribute to the further development of differential methods in clinical assessment and intervention. More specifically, descriptions of subgroups may be helpful in discovering patterns for individual children and, especially, patterns in seemingly unsystematic children where the organizing principles may not be immediately apparent. A differential diagnosis may provide guidelines for the planning of therapy directed towards overcoming the individual child's problems by concentrating on modifications of word pattern and/or the establishment of phonemic contrasts and their phonetic manifestations, as well as for allowing for different learning styles and motivations for the forms.

Working Papers 25 1983
Linguistics-Phonetics
Lund University, 105-111

DURATION OF STANDARD CHINESE WORD TONES IN DIFFERENT SENTENCE ENVIRONMENTS

Magnus Nordenhake and Jan-Olof Svantesson

1. BACKGROUND

In Standard Chinese (pǔtōnghuà) there are four tones: 1. high (denoted ¯); 2. rising (´); 3. low (dipping, ˇ) and 4. falling (`). Apart from fundamental frequency, two other acoustic correlates have been deemed important in the manifestation of the Standard Chinese word tones: intensity and duration.

Kratochvīl (1968) refers to tones produced in isolation when he describes the length of the first tone as "slightly above average"; the second tone as "slightly below average"; the third tone as "well above average"; and the fourth tone as "far below average". As far as we are aware, there is no published investigation of the duration of Standard Chinese tones in controlled sentence environments.

In this paper the duration of Standard Chinese tones in different environments (sentence medial and final), and under different focus conditions (focus or non-focus) is investigated. It is shown that the durations of the four tones are affected differently by different environments. In particular, the fourth tone shows the shortest duration of the four tones in sentence final position, but is the longest tone in sentence medial position. All four tones are lengthened in focus position.

2. PROCEDURE

The speaker, a male resident of Běijīng in his forties, was asked to read at normal speed a number of questions and answers at the sound treated studio at the Institute of Phonetics in Lund. The material was recorded on a Studer tape recorder.

The analyzed sentences were of the type (the standard pīnyīn transcription is used):

Wǒmen V le Wáng N 'We V-d Wáng N'

Here, V is a monosyllabic verb, and N is a given name. The studied syllables (i.e. V and N) were chosen so that all four tones were represented in both positions, while these syllables were all of the type Ca. By necessity, different consonants had to be chosen. In the analyzed material, the medial syllable V is followed by an unstressed syllable in order to minimize the influence of tone sandhi.

The total material consisted of the following sixteen sentences:

Wǒmen	{ shā / chá / dǎ / mà }	le Wáng	{ Bā / Pá / Mǎ / Chà }
We	{ kill / examine / beat / scold }	-d Wáng	{ Bā / Pá / Mǎ / Chà }

Each sentence was recorded twice, once with focus on the sentence medial verb, and once with focus on the final noun. This was achieved by recording each sentence as the answer to two different questions: Nǐmen gēn Wáng N zuòle shénme? 'What did you do with Wáng N?' the answer of which focussed on the sentence medial verb; and Nǐmen V le shéi? 'Who did you V?' which brought the focus to the name. The speaker read each of the 32 questions and answers twice, and wideband spectrograms were made of the answers. On the spectrograms, the durations of the vowels (a) in the studied syllables were measured.

3. RESULTS

Table 1 shows the duration of the vowel in each of the 128 analyzed syllables. On Figures 1 and 2, each point represents the average of 8 measurements with the same tone, position and focus condition. The data are arranged in different ways on the two figures in order to illustrate the effects of the factors tone, position and focus.

The results show that tones 1, 2 and 3 are longer in final position than in medial position. In fact, in our material, all

Table 1. Duration of the vowel in Standard Chinese syllables of the type Ca (ms).

	Medial		Final	
	Focus	Non-focus	Focus	Non-focus
Tone 1	166	158	205	217
	166	181	213	205
	174	150	185	205
	181	166	229	197
	158	148	221	205
	174	158	221	197
	166	150	213	197
	181	158	229	213
mean:	171	159	214	204
Tone 2	142	103	252	221
	174	142	252	205
	150	95	221	213
	174	142	237	221
	158	142	229	205
	189	150	252	221
	142	118	237	205
	166	142	237	197
mean:	162	129	240	211
Tone 3	189	158	315	237
	174	166	284	205
	189	166	292	260
	189	174	292	197
	205	174	308	252
	174	181	300	252
	197	150	308	270
	189	174	276	260
mean:	188	168	297	242
Tone 4	181	184	166	142
	197	181	158	142
	197	166	166	158
	213	189	174	181
	181	174	150	150
	197	181	174	174
	181	174	158	158
	221	189	174	110
mean:	196	180	165	152

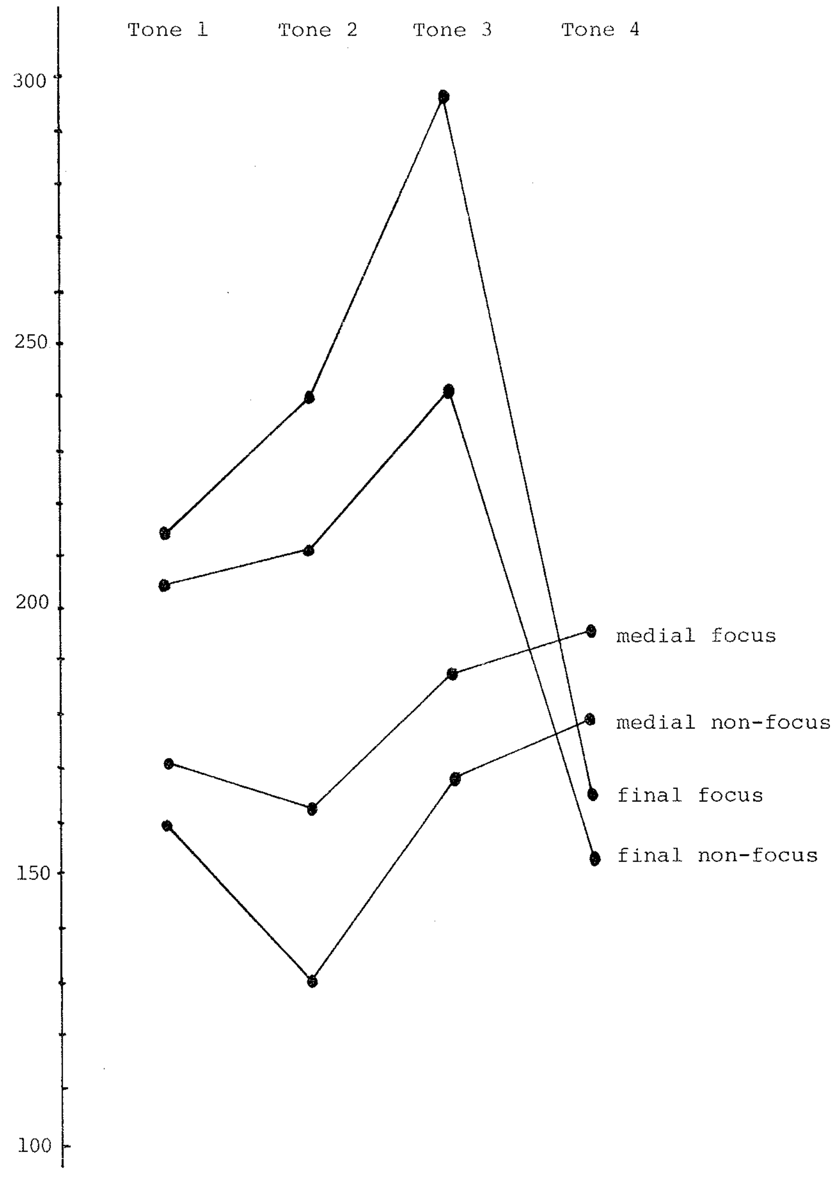


Figure 1. Average duration for each combination of position and focus conditions over the four tones.

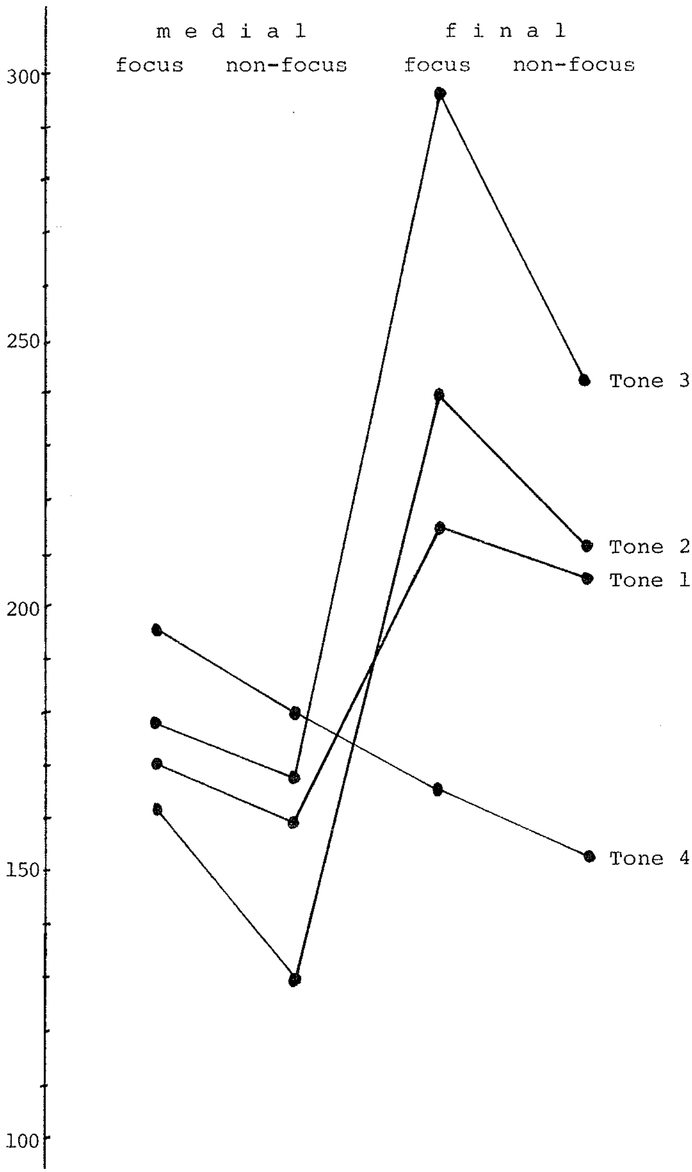


Figure 2. Average duration of the four tones over each combination of position and focus conditions.

finals are longer than all medials with the same focus condition for these three tones. Tone 4, however, is on the average shorter in final than in medial position (see Fig. 1).

Thus the factor "position" influences the different tones in different ways. The factor "focus", however, always has the same effect: a focussed vowel becomes longer irrespective of its tone.

In order to test statistically the interaction between the factors tone, position and focus, a three-way analysis of variance was performed with the following result:

Factor	Sum of squares	Degrees of freedom	Test quantity	
tone	45 565	3		
position	69 425	1		
focus	17 743	1		
tone-position	72 327	3	115.0	($p \ll 0.001$)
tone-focus	3 904	3	6.2	($p < 0.001$)
position-focus	328	1	1.6	($p > 0.05$)
tone-position-focus	2 165	3	3.4	($0.01 < p < 0.05$)
residual	23 490	112		

Thus there is no significant interaction between position and focus. The interaction between tone and focus and between tone and position are both significant, but the tone-focus interaction is much smaller than that between tone and position. The average effects of focus and of final position are (cf. also Figure 2):

	Focus	Final position	(ms)
Tone 1	11.1	44.9	
Tone 2	30.6	79.8	
Tone 3	37.8	91.2	
Tone 4	14.7	-29.4	

Thus, tones 2 and 3 are lengthened more than tone 1 and 4 in focus, and as already said, the effect of final position (as compared to medial position) is to lengthen tones 1, 2 and 3, but to shorten tone 4.

Because of these interactions, it is hardly meaningful to give the average durations of the four tones without stating the environment. In medial position the tones are ordered in increasing duration as: T2 < T1 < T3 < T4, and in final position the order is T4 < T1 < T2 < T3. The order given by Kratochvīl 1968 for citation forms (T4 < T2 < T1 < T3) is the same as that for final position, except that the order between T2 and T1 is reversed (the difference between T2 and T1 as found here is small, however).

4 DISCUSSION

One might speculate why the falling fourth tone is shortened in final position. One possible reason is that because of sentence intonation downdrift in statements (cf Gårding, Zhāng and Svantesson 1983), the starting point for a sentence-final falling tone is rather low, so that it takes comparably short time to reach the bottom of the voice register. Perception experiments by Kratochvīl 1970 show that the fourth tone is associated with shortness. He found that shortened tones, may they be tone 1, 2, 3 or 4, are almost exclusively identified as tone 4.

Tone 3 is also somewhat exceptional since it is lengthened more than the other tones both by focus and by final position. It seems that staying for some time at a low level is an essential property of this tone, and this property is strengthened by focus and by final position. The dipping citation form of this tone is obtained by combining the intrinsic low tone with sentence intonation (cf. Gårding, Zhāng and Svantesson 1983).

REFERENCES

Gårding, E., Zhāng, J and Svantesson J-O. 1983. A generative model for tone and intonation in Standard Chinese. In this volume.

Kratochvīl, P. 1968. The Chinese language today. London: Hutchinson.

Kratochvīl, P. 1970. An experiment in the perception of Peking dialect tone. In I-L Hansson, ed., A symposium on Chinese grammar. London: Curzon Press, pp. 7-31.

Working Papers 25 1983
Linguistics-Phonetics
Lund University, 113-137

ACOUSTIC ANALYSIS OF FRICATIVES IN CAIRO ARABIC

Kjell Norlin

1. INTRODUCTION

This article describes the acoustical properties of fricatives in Cairo Arabic (CA). With its eleven fricatives, CA, like all forms of Arabic, belongs to the small group of less than one per cent of the languages of the world with such a large inventory of fricative phonemes (Nartey 1979). Acoustical investigations of these do not abound. On the contrary it is easy to find somewhat fanciful descriptions of many Arabic phonemes in the literature. One author states that /ḍ/ is "un peu comme le 'dang' sonore et prolongé, qui veut imiter le son d'un cloche de cathedral", (Jomier 1964), and that /ḥ/ reminds you of "la respiration d'un chien haletant après une course" (Jomier 1964). Still another one finds that /ʕ/ "is made far back in a tightened throat, and sounds and feels rather like being sick"(Scott 1962), /ɣ/ "is the noise which the camel makes when growling at being loaded" (Scott 1962), /q/ "resembles more than anything else the 'kok-kok-kok' sound made by liquid being poured out of a full bottle". (Scott 1962).

This investigation presents a more data-oriented approach and a method of analyzing fricatives which provides the means of describing the fricative acoustic space of a language. Parameters for characterizing fricative spectra are given and applied to CA fricatives. This method makes comparison possible between different speakers and also between CA and other languages, for example Chinese, which has been investigated by the same method (Svantesson 1983).

This method has made it possible to quantify data and state the difference between pharyngalized fricatives and their non-pharyngalized counterparts.

1.1 The fricatives of CA

f s z ṣ ẓ š x ɣ ḥ ʕ h

2. PROCEDURE

Six informants from Cairo were recorded on a Studer tape-recorder in the studio of the Phonetic Department in Lund. The fricatives were pronounced in real words in a sentence frame, preceded by a# and followed by ā, (a# Cā). The word list was read twice. /s/ and /ṣ/ followed by all long vowel phonemes of CA, /ā/, /ī/, /ū/, /ē/, /ō/, in minimal or near-minimal pairs, were read twice by three speakers.
Analysis was made from the second reading.
Sampling started after the first third of the fricative for 25.6 ms.

Some difficulties were experienced in finding the boundaries of /ɣ/ and /ʕ/ because they more closely resembled approximants than fricatives as revealed by their wave-form on duplex oscillograms.

3. ANALYSIS

1. FFT spectra up to 10 kHz of the middle of the 25 msecs of all the fricatives were made. These spectra were converted to critical band spectra (Schroeder et al. 1979) and analyzed in terms of the spectral center of gravity and dispersion in the manner described by Svantesson (1983).

2. Spectrograms were used to measure and compare formant transitions and vowel formants after /s/ and /ṣ/.

3. Duplex oscillogram along with intensity and F_0 curves were used for analyzing non-spectral properties for pharyngalized versus non-pharyngalized fricatives, as well as for the pharyngals.

4. RESULTS FROM CRITICAL BAND SPECTRA

Figure 1:1-22 shows oscillograms of sound waves, FFT spectra in linear scale, spectra in logarithmic units (dB) and critical band spectra of one speaker (speaker 6). /š/ of speaker 4 is missing in the data. The critical band units were in some cases measured twice, the result of the check being practically the same as the first measurement. Table 1 gives the center of gravity of the critical band spectra, measured in critical band units and also given in Hz, the dispersion and the mean intensity level (dB). The mean intensity level are given as deviations from the average for each series of fricatives read on the same occasion. This makes them roughly comparable also between other speakers.

In figure 2 the center of gravity for each fricative is plotted against the dispersion, thus representing the fricative space of CA in a way which enables comparison with other languges. Figure 3 gives the mean values of the six speakers in figure 2. In figure 4 the fricative space is represented in another form. The center of gravity is plotted against the mean intensity level over the critical bands.
Figure 5 gives the mean values of the six speakers in figure 4. In figures 2 and 4 the individual fricatives are rather well kept apart even if the distance between /s/ and /ṣ/ on one hand and /z/ and /ẓ/ on the other is usually small with some outstanding exceptions. There is overlapping, especially between individual speakers and particularly between /s/ and /z/ and their pharyngalized counterparts. One must suppose that perception of fricatives involves normalization between different speakers.

/s/ and /ṣ/ are characterized by a sharp peak in the higher frequency ranges, band 21-23, and an abrupt fall towards the lower ranges. Figure 1:15 shows that the peak of /ṣ/ is somewhat broader than that of /s/ (figure 1:13). This difference is obvious here, but is usually impossible to notice when comparing all the critical band histograms of the informants. This difference is possible to quantify, however, by measuring the center of gravity. Table 1 shows that /ṣ/ has the center of gravity in lower frequency ranges than /s/ with one negligible exception , speaker 1. It also has a greater range of dispersion. The difference is not excessively large, but corresponds well to the slight, but quite noticeable perceptual impression of these sounds.

/z/ and /ẓ/ both have a substantial peak of energy in the bands 3-6 in addition to the high frequency peaks of /s/ and /ṣ/, although the former have lower intensity. The concentration of

energy in both ends of the spectrum, together with a cut in the top frequency range as compared to /s/ and /ṣ/ make their centers of gravity more or less coincide with their voiceless counterparts. There is a strong tendency for /z/ and /ẓ/ to have greater dispersion than /s/ and /ṣ/. For speakers 4 and 5 the relationship between center of gravity and dispersion of /z/ and /ẓ/ is reversed as compared to all others.
The difference between centers of gravity in critical bands has been the only criterion found in this investigation to measure the difference between pharyngalized and non-pharyngalized fricatives. Inspections of intensity and wave-form on mingograms has not revealed any obvious differences between these sounds, nor can anything be seen on spectrograms.

/š/ is characterized by a broader peak than /s/ and /ṣ/, ending in two smaller peaks on top. This is the typical shape of all histograms except one. The fall towards lower frequencies is as sharp as for /s/ and /ṣ/. The top covers the range from bands 16-21, showing lower frequency, but not greater dispersion.

/f/ has a spectrum falling much more slowly from high frequencies than the sibilants and has also much lower mean intensity. Spectra of /f/ have roughly the same centers of gravity as /š/ and is distinguished from the latter by greater dispersion.

/ḥ/ has the center of gravity roughly in the center of the critical band spectrum, with a fairly steep slope in the lower ranges and a more gradual slope in the higher frequencies. The main contour of the spectrum is rather dome-shaped.

/h/ has energy spread over the whole frequency range, giving the contour a roughly flat shape, /h/ and /ḥ/ intermingle in figure 2, showing no absolute contrast in either center of gravity or dispersion. There is a tendency, however, for /h/ to have less dispersion.
For contrasts in wave-form, see below.

/x/ has energy evenly spread from band 7 upwards, with a slowly graded descent towards the lower ranges.

/ɣ/ and /ʕ/ are not sharply divided either in centers of gravity or dispersion, but there is a tendency for /ʕ/ to have its center of gravity in the somewhat lower frequencies. /ʕ/ generally has lower mean intensity (figures 4 and 5).

5. RESULTS FROM SPECTROGRAMS

Measurements on spectrograms of locus and formants of vowels following /s/ and /ṣ/ show great impact on formant transitions after pharyngalization (figure 6:1-5). F2 is particularly affected and locus is drastically lowered for all vowels except /ū/ where locus is raised by 100-150 Hz (figure 6:5). Changes of F1 and F2 are also noticeable, but on a very small scale and of no consistent pattern. The raising of F2 of the vowel /ū/ after /ṣ/ in contrast to the falling of F2 of all other vowels is due to the successive movement backwards of the tongue constriction area as shown in the nomograms by Fant (1968).

6. RESULTS FROM MINGOGRAMS

Inspection of duplex oscillograms along with intensity and F_o curves revealed no difference between pharyngalized and non-pharyngalized sibilants. Duplex oscillograms of /ɣ/ and /ʕ/ (figure 7:1-2) show that these sounds are phonetically realized as approximants in CA, rather than as fricatives. However, they are always classified as fricatives in phonetic descriptions (eg. Harrell 1957, Abdel-Massih 1975).

/ḥ/ and /h/ differ in two ways. Both sounds stand in intervocalic position in the sentence frame, but /ḥ/ is always voiceless, /h/ is voiced (figure 8:1-2). The vowel following /ḥ/ always has hard onset, after /h/ soft onset.

7. DISCUSSION

The place of each fricative within the fricative space of a language can be defined by the method of making critical band spectra and deriving the suggested parameters from them. The method also makes it possible to compare fricatives of different languages.
Owing to the great number, the fricatives of CA are fairly well spread within the fricative space, defined by the center of gravity and dispersion (figures 2 and 3). The difference between pharyngalized fricatives and their non-pharyngalized counterparts has formerly been investigated on spectrographic evidence (Obrecht 1968).

The critical band method has made it possible to quantify the difference. The pharyngalized fricatives nearly always have a lower center of gravity and greater dispersion. The difference is not necessarily very great. Exceptional cases even show non-pharyngalized fricatives having a lower center of gravity and greater dispersion (/s/, /ṣ/ of speaker 1, /z/, /ẓ/ of speaker 4 and 5). The described difference is therefore neither a sufficient, nor reliable cue to discrimination between these pairs of sibilants. Pharyngalization, or emphasis to use another current term, "never occurs as a feature of a single segment. The minimum range of emphasis is V(:)C or CV(:)" (Harrell 1957). The perceptual cue of importance to discriminate pharyngalization seems to be in the lowering of the locus (slight raising for C̣ū), while the steady-state portion is much less affected, except for C̣ā. Perceptual tests with synthetic speech give ample evidence to the fact that F2 transitions are the most important factors in identification of pharyngalized fricatives (Obrecht 1968).

Since it is generally assumed, however, that the pharyngalization gives another set of fricative phonemes, producing vocalic allophones in the following segments, it is important to quantify the difference between pharyngalized and non-pharyngalized fricatives in terms of their place in the fricative space.

ACKNOWLEDGMENTS

The work reported here is supported by the Swedish council for Research in the Humanities and Social Sciences.

REFERENCES

Abdel-Massih, E. 1975. An introduction to Egyptian Arabic. Ann Arbor: The University of Michigan, pp. 2, 5

Fant, G. 1968. Analysis and synthesis of speech processes. In Bertil Malmberg, ed. Manual of Phonetics. Amsterdam: North-Holland Publishing Company, p. 214

Harrell, R.S. 1957. The phonology of colloquial Egyptian Arabic. New York: American council of learned societies, pp. 25, 78

Jomier, J. 1964. Manuel d'arabe égyptien. Paris: Librairie C. Klinksieck, p. 1

Obrecht, D.H. 1968. Effects on the second formant on the perception of velarization consonants in Arabic. The Hague, Paris: Mouton, p. 41

Schroeder, M.R., B.S. Atal and J.L. Hall. 1979. Objective measure of certain speech signal degradations based on masking properties of human auditory perception. In Björn Lindblom and Sven Öhman, ed. Frontiers of speech communication research, London: Academic Press 1979, pp. 217-229

Scott, G.C. 1962. Practical Arabic. London: Longmans, pp. 2, 6, 7

Svantesson. J-O. 1983. Acoustic analysis of Chinese fricatives and affricates. To be published in Working Papers, Phonetics Laboratory, Lund University

Table 1. Center of gravity, dispersion and mean intensity level of the critical band spectra of fricatives in Cairo Arabic.

	speaker	center of gravity crit.band units	Hz	dispersion crit.band units	mean intensity level, dB
/f/	1	18.42	4179	4.10	-4.54
	2	20.53	5665	3.54	-12.94
	3	17.75	3793	3.76	-8.85
	4	18.44	4191	4.77	-9.90
	5	20.88	5957	2.14	-8.39
	6	22.21	7210	2.16	-7.50
/s/	1	20.55	5681	2.15	4.24
	2	21.18	6219	1.48	7.18
	3	21.53	6540	1.80	4.91
	4	20.83	5914	1.75	13.06
	5	21.83	6827	0.95	7.74
	6	21.98	6976	1.09	6.36
/ṣ/	1	20.59	5714	2.63	2.55
	2	20.90	5974	1.56	1.44
	3	20.99	6052	2.31	1.22
	4	20.71	5813	1.90	13.66
	5	21.39	6410	1.46	7.88
	6	20.80	5889	1.77	6.73
/z/	1	18.36	4143	4.61	-5.84
	2	21.51	6521	1.23	4.50
	3	21.85	6847	1.18	8.80
	4	19.87	5151	2.90	0.70
	5	19.62	4969	5.44	-5.67
	6	21.74	6740	1.62	5.73
/ẓ/	1	15.22	2622	6.49	-2.35
	2	20.52	5656	1.97	2.49
	3	20.66	5771	3.57	-0.71
	4	20.76	5855	2.17	7.84
	5	21.11	6157	1.39	1.60
	6	19.45	4849	5.73	-5.23
/š/	1	18.47	4209	3.16	2.77
	2	18.38	4155	2.43	12.48
	3	16.45	3140	2.02	6.15
	4				
	5	18.85	4447	2.27	7.86
	6	19.21	4684	2.36	9.34
/x/	1	14.05	2205	3.17	0.94
	2	16.02	2948	4.08	1.84
	3	15.83	2868	3.68	0.17
	4	15.95	2918	4.00	2.93
	5	16.15	3005	3.31	0.21
	6	18.51	4234	3.92	2.51

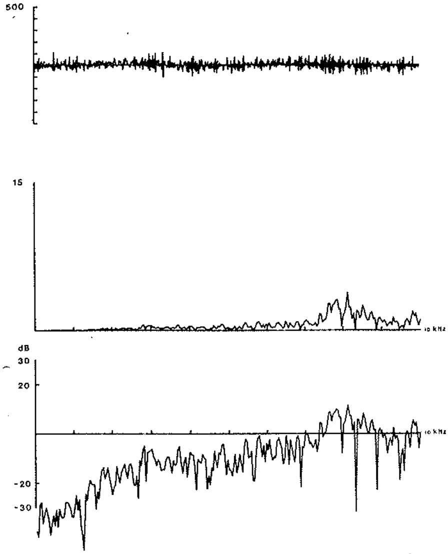
Table 1 cont.

	speaker	center of gravity crit.band units	Hz	dispersion crit.band units	mean intensity level, dB
/ɣ/	1	11.00	1384	4.41	-3.78
	2	11.80	1568	3.30	-1.45
	3	9.97	1173	4.34	-2.48
	4	12.21	1670	3.86	8.47
	5	11.75	1556	6.24	-0.17
	6	9.98	1175	4.79	-4.98
/ḥ/	1	13.22	1947	2.67	-0.79
	2	11.58	1516	3.95	-4.63
	3	13.75	2108	3.37	-0.82
	4	13.22	1947	3.86	-7.00
	5	15.64	2789	2.23	-0.75
	6	11.60	1520	1.83	0.36
/ʕ/	1	10.95	1373	3.96	4.30
	2	9.78	1137	2.90	3.57
	3	8.77	959	3.01	-7.15
	4	9.97	1173	3.74	-15.44
	5	11.91	1595	3.18	-5.76
	6	13.24	1953	4.31	-7.51
/h/	1	12.94	1867	3.35	2.47
	2	10.91	1364	5.07	-10.89
	3	13.81	2127	2.62	-1.20
	4	14.28	2282	3.45	-14.28
	5	15.43	2704	2.84	-4.53
	6	14.70	2428	3.78	-5.78

<u>Mean values</u>

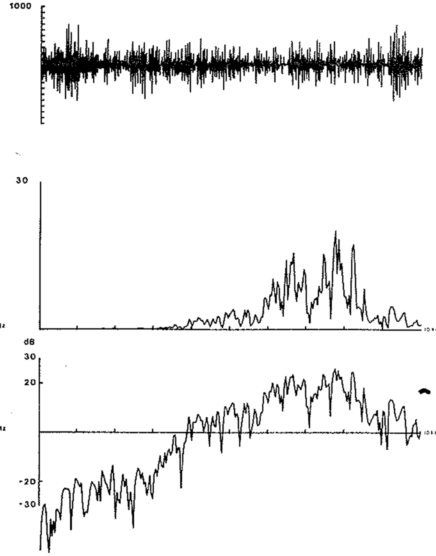
/f/	19.71	5034	3.41	-8.68
/s/	21.32	6345	1.54	7.25
/ṣ/	20.90	5974	1.94	5.58
/z/	20.49	5632	2.83	1.37
/ẓ/	19.62	4969	3.55	0.60
/š/	18.27	4089	2.45	6.43
/x/	16.09	2979	3.69	1.43
/ɣ/	11.14	1415	4.49	-0.73
/ḥ/	13.17	1933	2.99	-2.27
/ʕ/	10.77	1335	3.52	-4.66
/h/	13.68	2087	3.52	-5.70

figure I:1



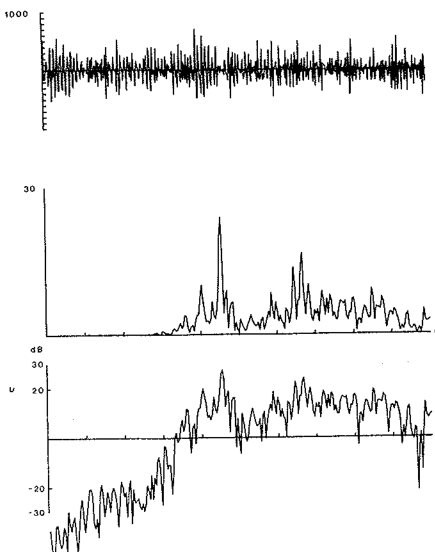
f6

figure I:2



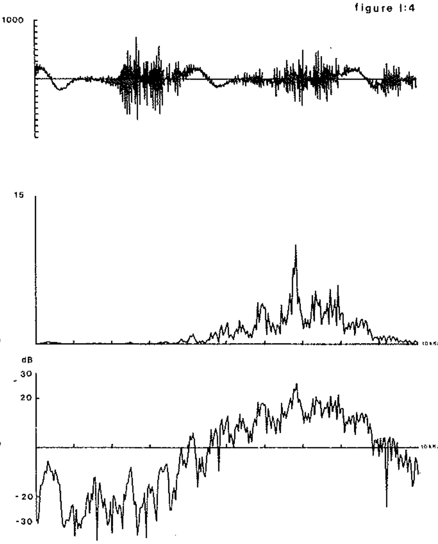
s6

figure I:3



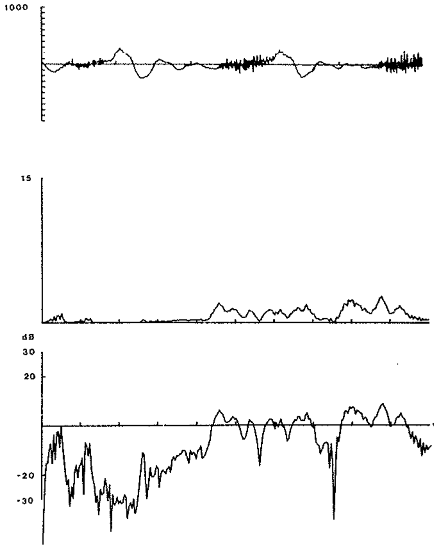
ṣ6

figure I:4



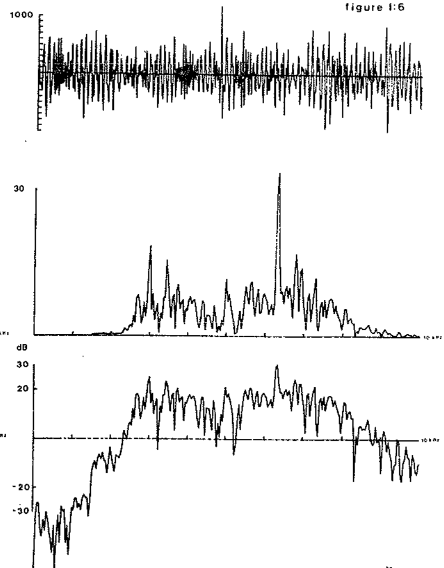
z6

figure I:5



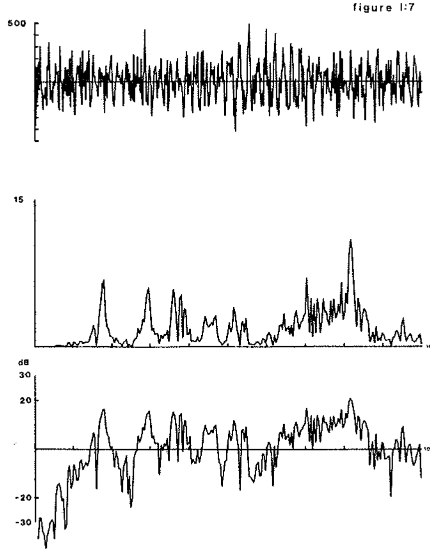
ẓ6

figure I:6



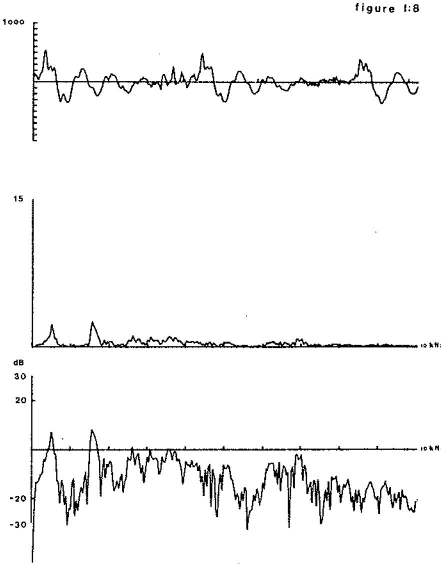
š6

figure I:7



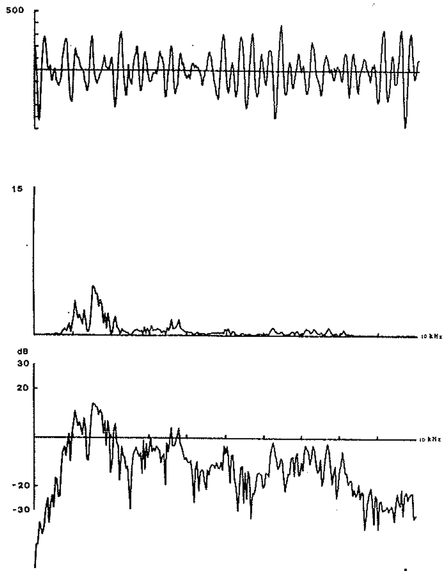
x6

figure I:8



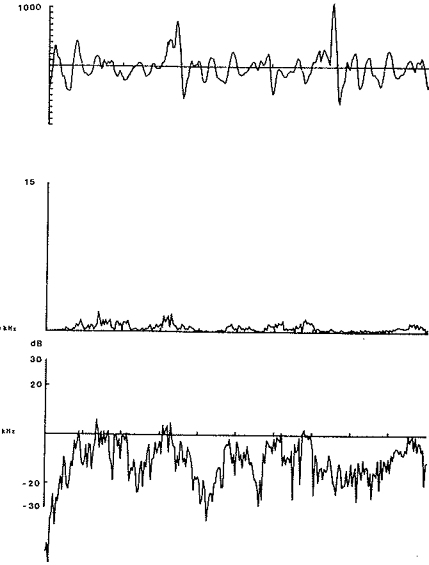
ɣ6

figure I:9



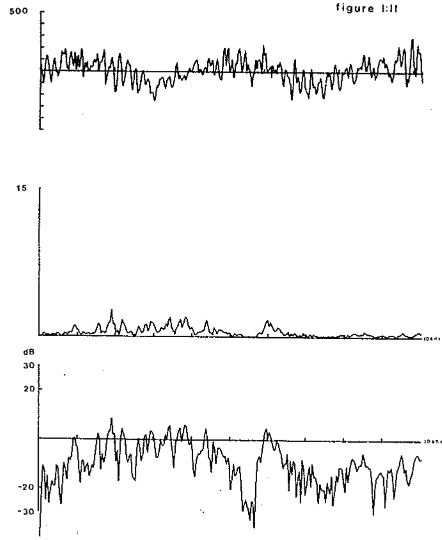
ḥ$_6$

figure I:10

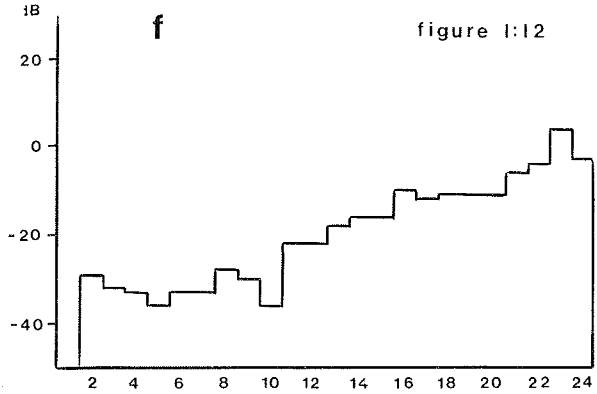


ʕ$_6$

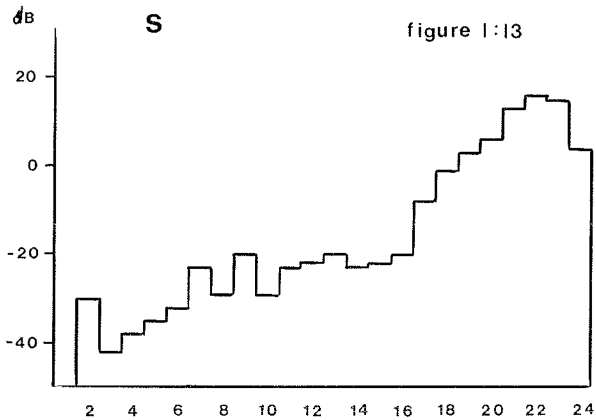
figure I:11



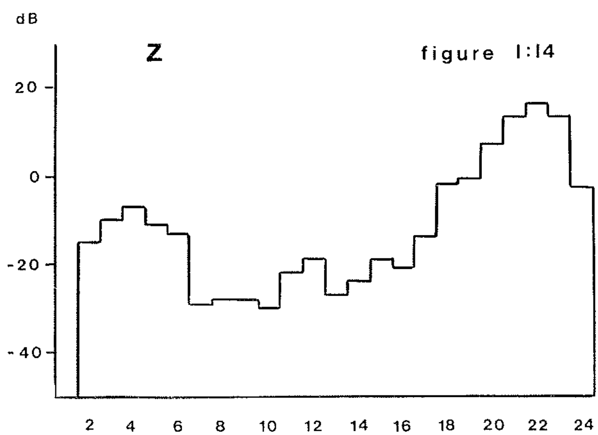
h$_6$



dB
f
figure I:12
20
0
-20
-40
2
4
6
8
10
12
14
16
18
20
22
24



dB
S
figure I:13
20
0
-20
-40
2
4
6
8
10
12
14
16
18
20
22
24



dB
Z
figure I:14
20
0
-20
-40
2
4
6
8
10
12
14
16
18
20
22
24

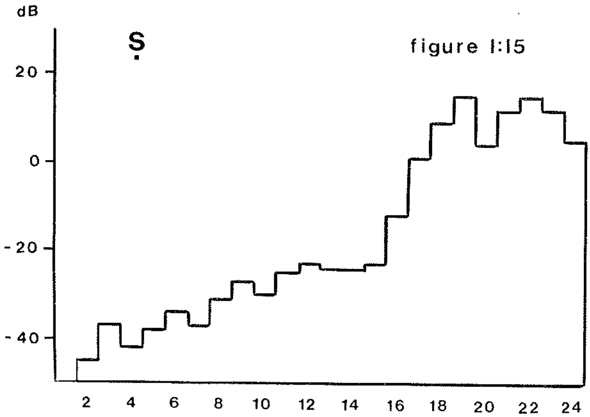


figure I:15

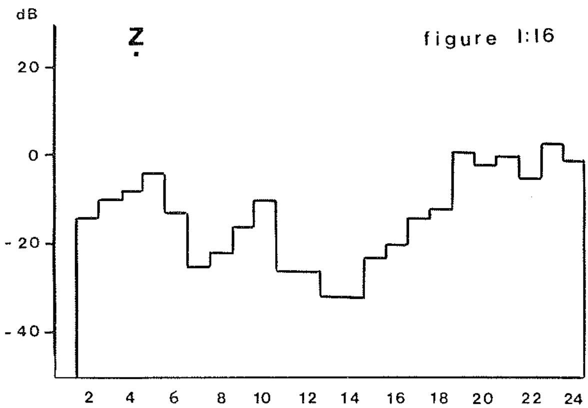


figure I:16

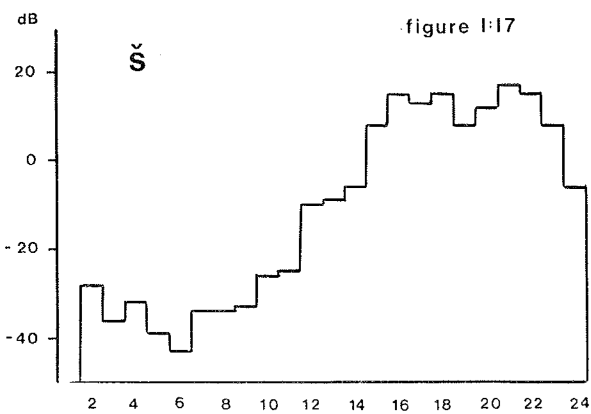


figure I:17

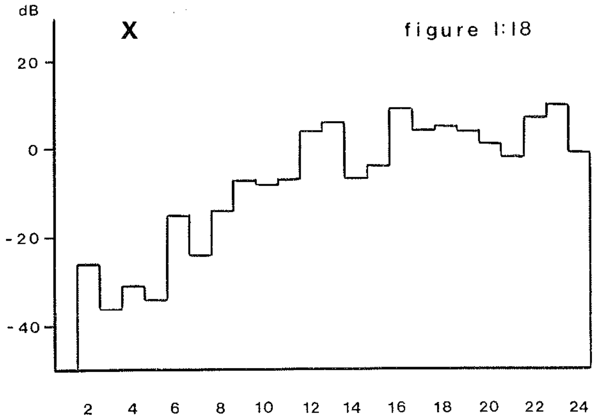


figure I:18

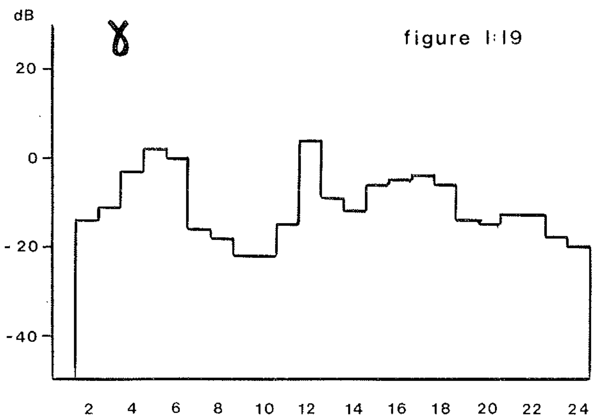


figure I:19

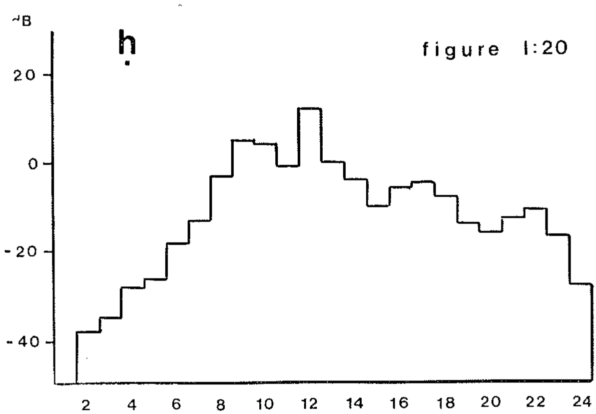
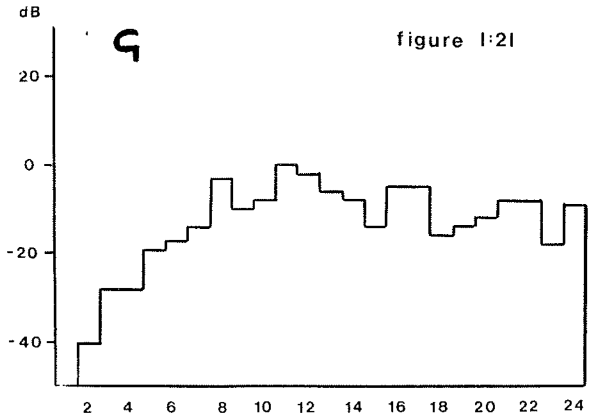
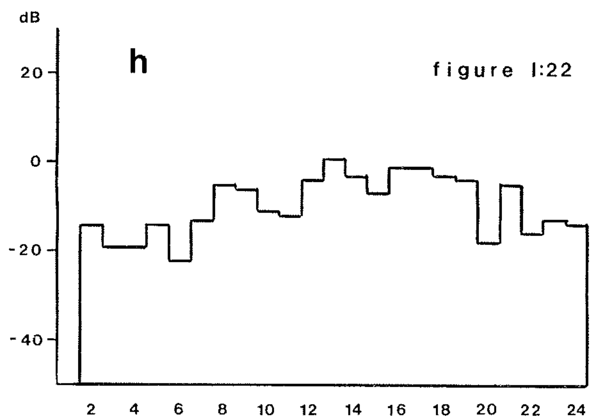


figure I:20


dB
ɥ
figure I:21
20
0
-20
-40
2
4
6
8
10
12
14
16
18
20
22
24


dB
h
figure I:22
20
0
-20
-40
2
4
6
8
10
12
14
16
18
20
22
24

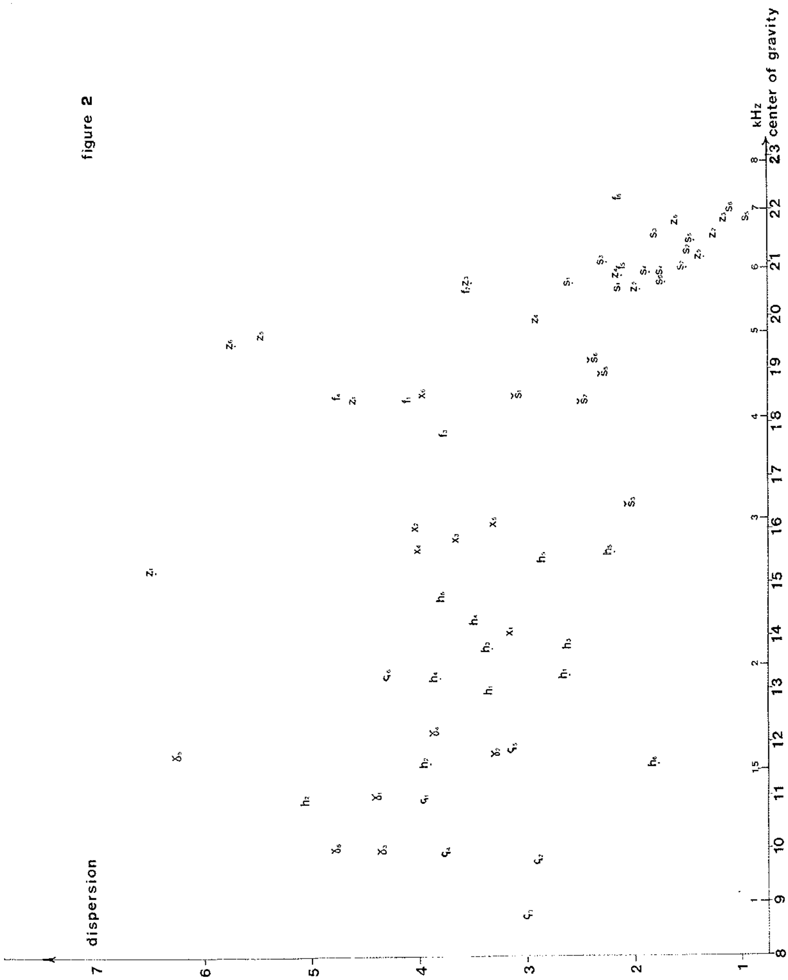

dispersion
7
6
5
4
3
2
1
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
kHz
center of gravity
1
1.5
2
3
4
5
6
7
8

figure 2

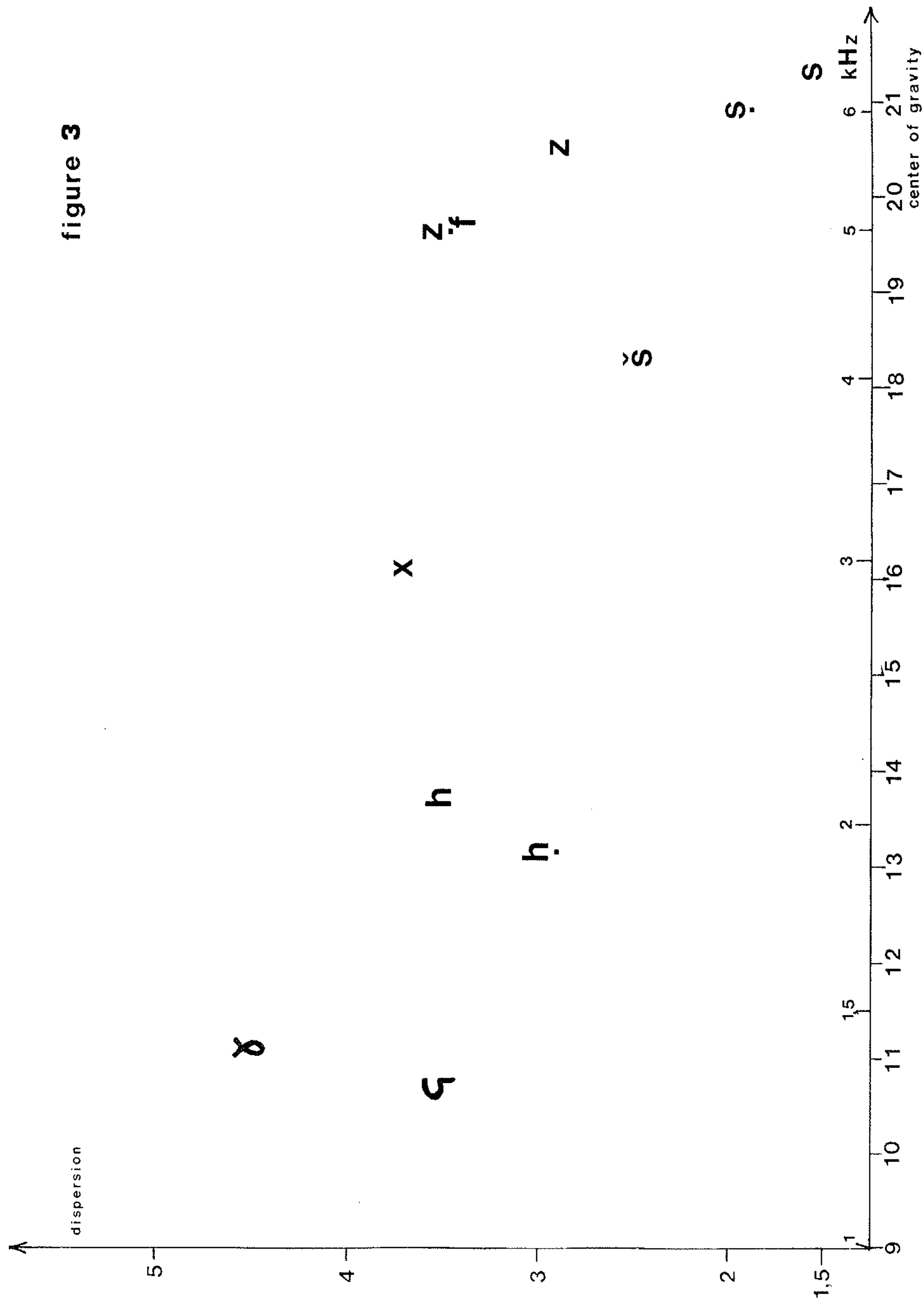

dispersion
5
4
3
2
1,5
center of gravity
kHz
9
10
11
12
13
14
15
16
17
18
19
20
21
1
1,5
2
3
4
5
6

figure 3

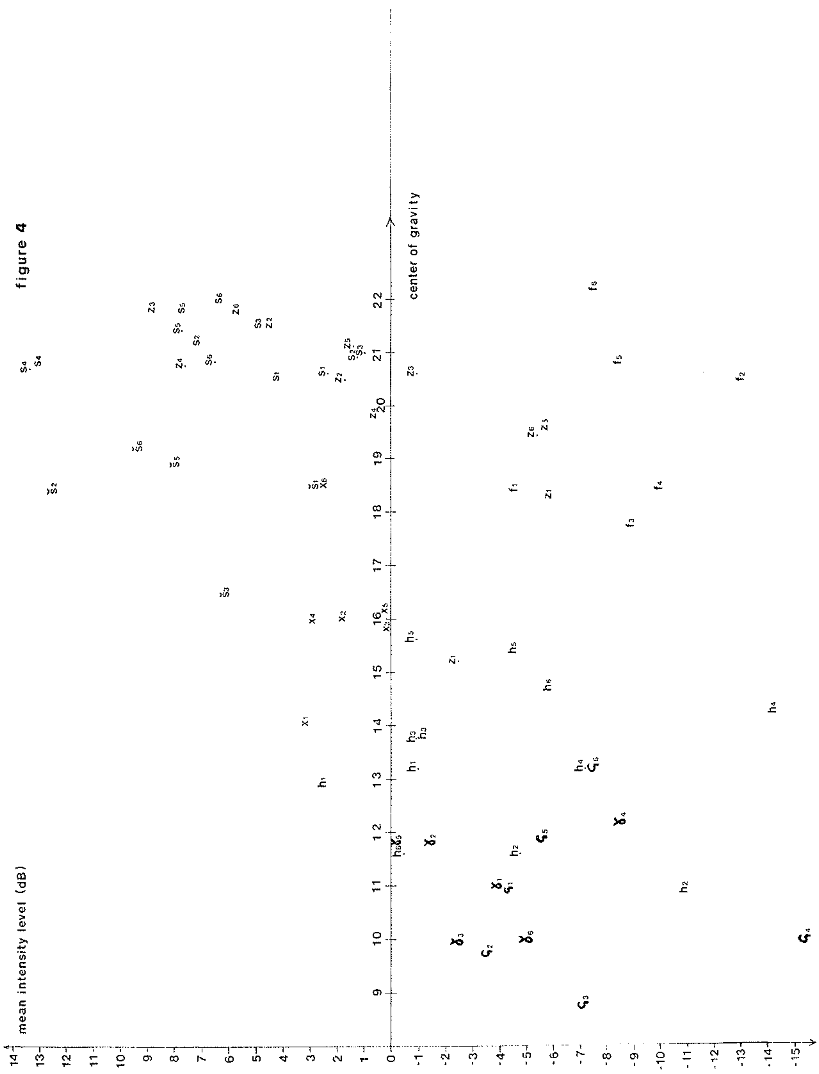
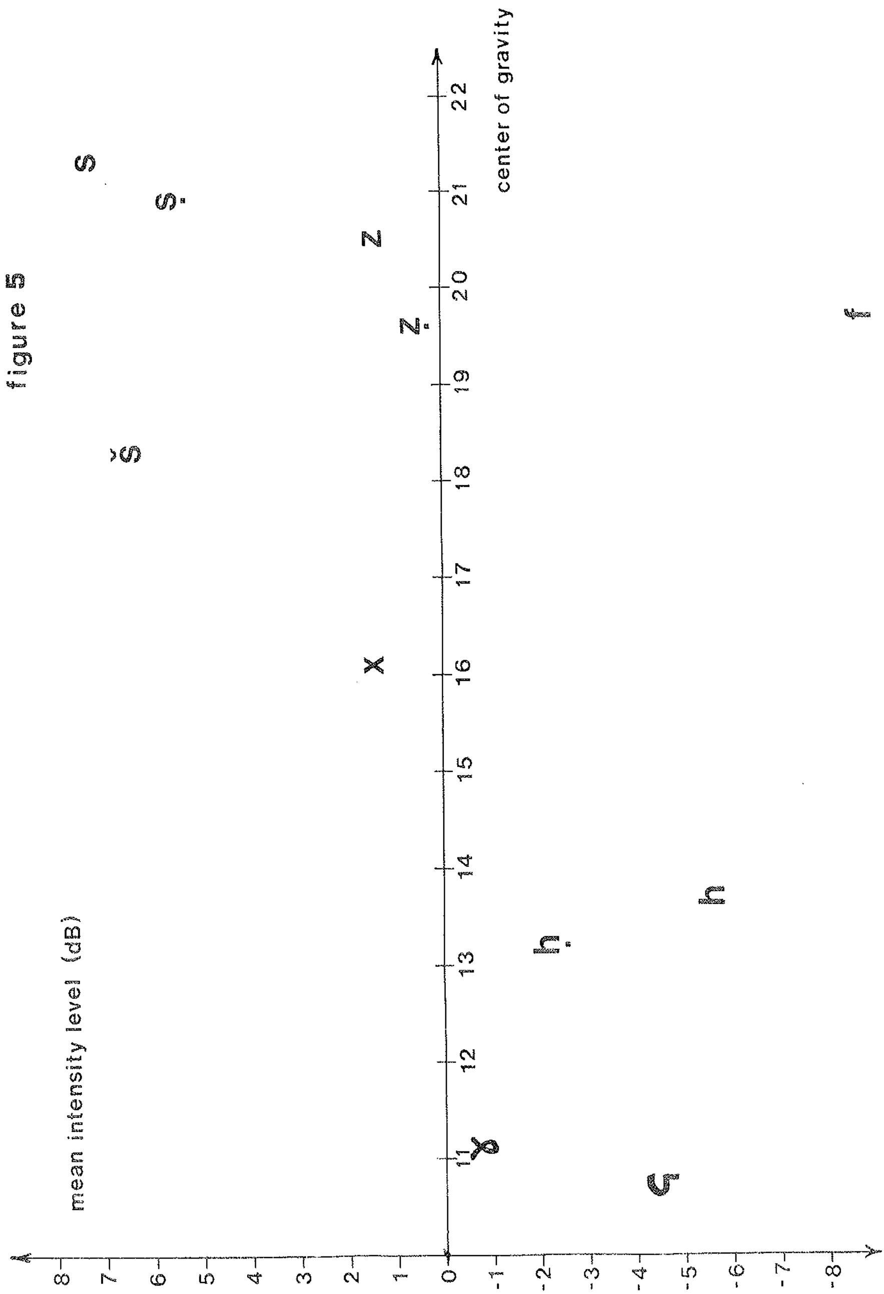


figure 4

figure 5


mean intensity level (dB)
center of gravity
8
7
6
5
4
3
2
1
0
-1
-2
-3
-4
-5
-6
-7
-8
11
12
13
14
15
16
17
18
19
20
21
22
s
ṣ
š
z
ẓ
x
ḥ
h
ɣ
ʕ
f

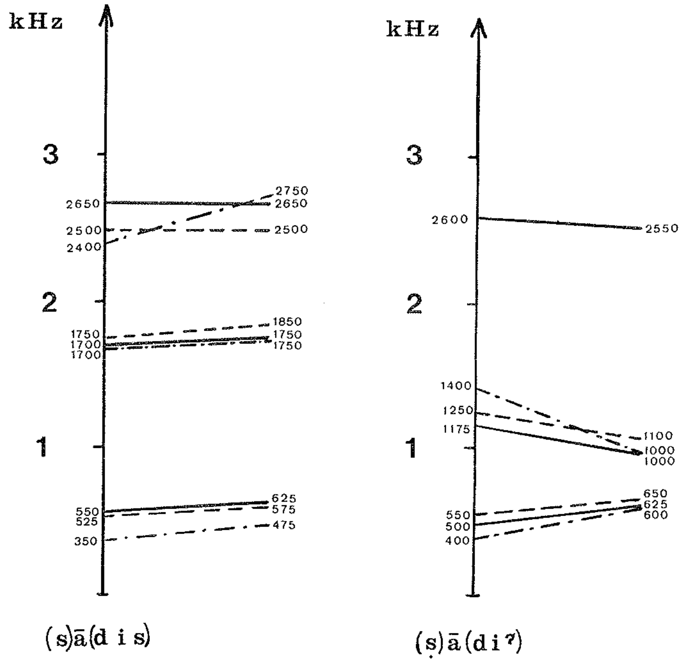


figure 6:1

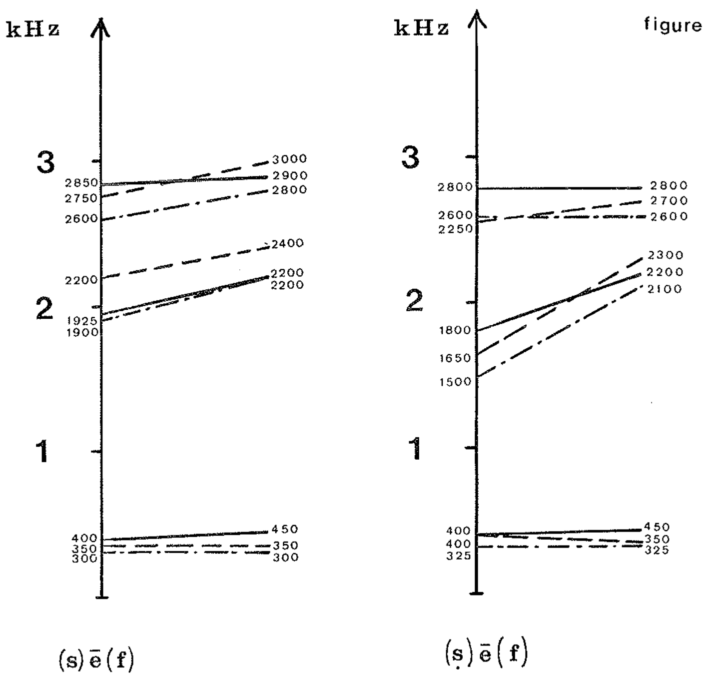
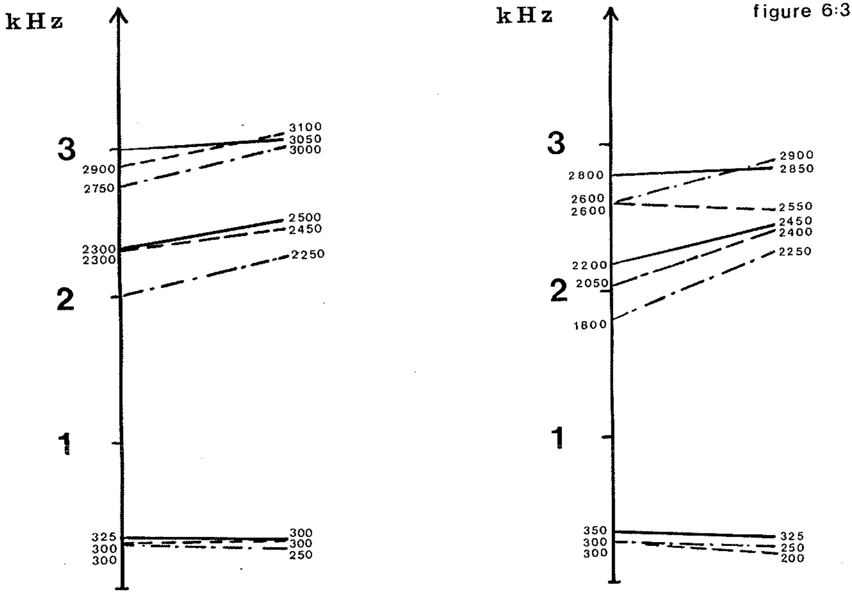


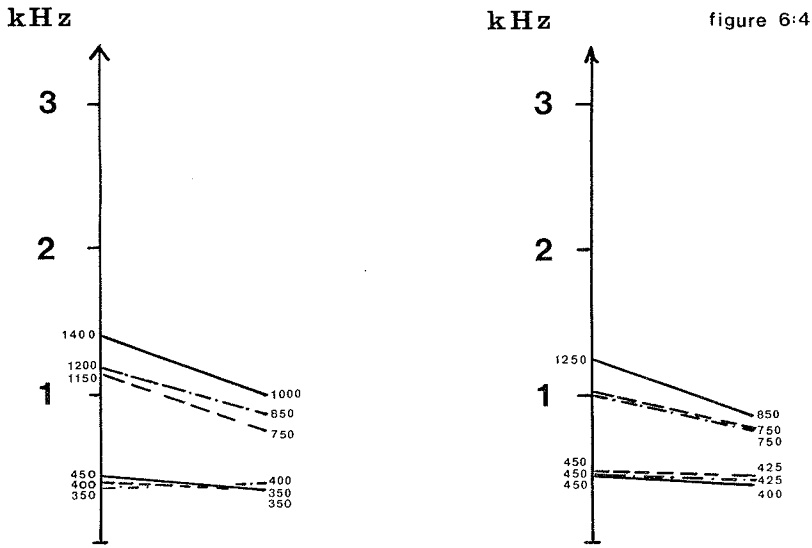
figure 6:2

figure 6:3



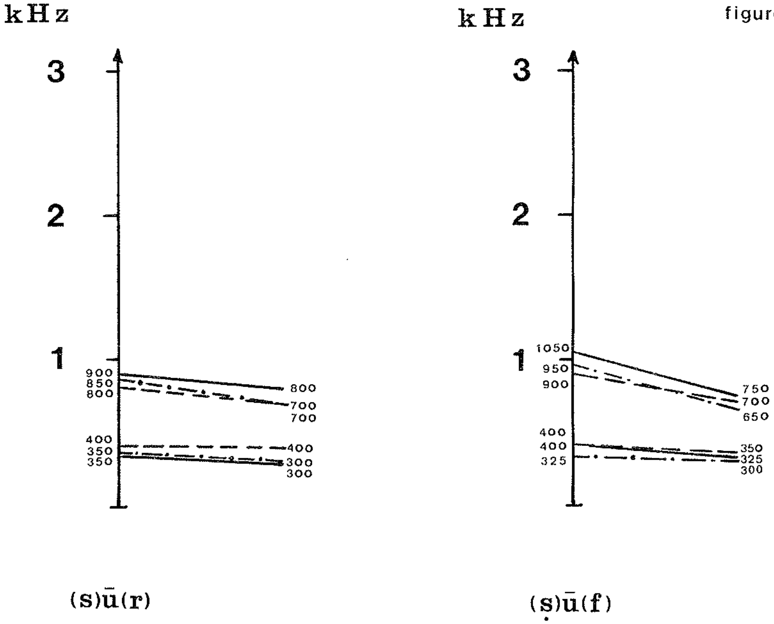
(s)ī(di) (ṣ)ī(ni)

figure 6:4



(s)ō(da) (ṣ)ō(t)

figure 6:5


kHz
3
2
1
900
850
800
400
350
350
800
700
700
400
300
300
(s)ū(r)
kHz
3
2
1
1050
950
900
400
400
325
750
700
650
350
325
300
(ṣ)ū(f)

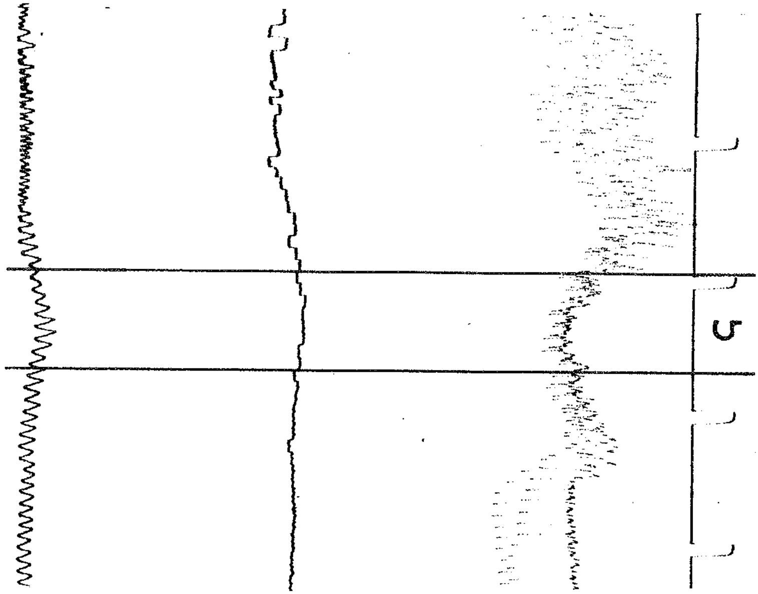


figure 7:2

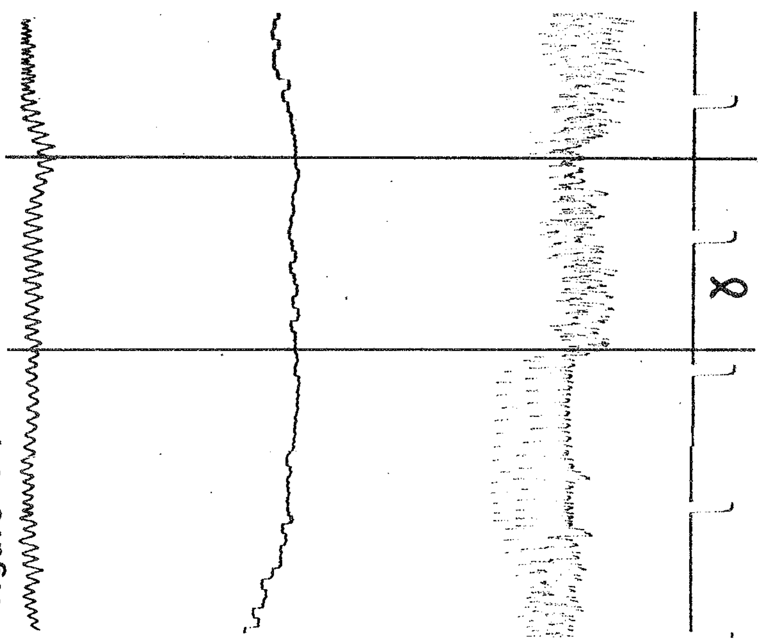


figure 7:1

figure 8:1

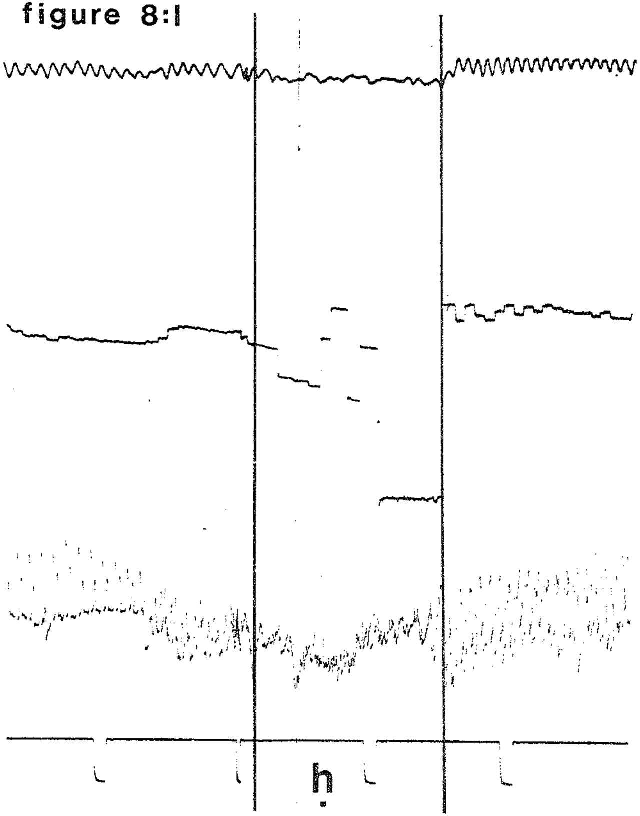
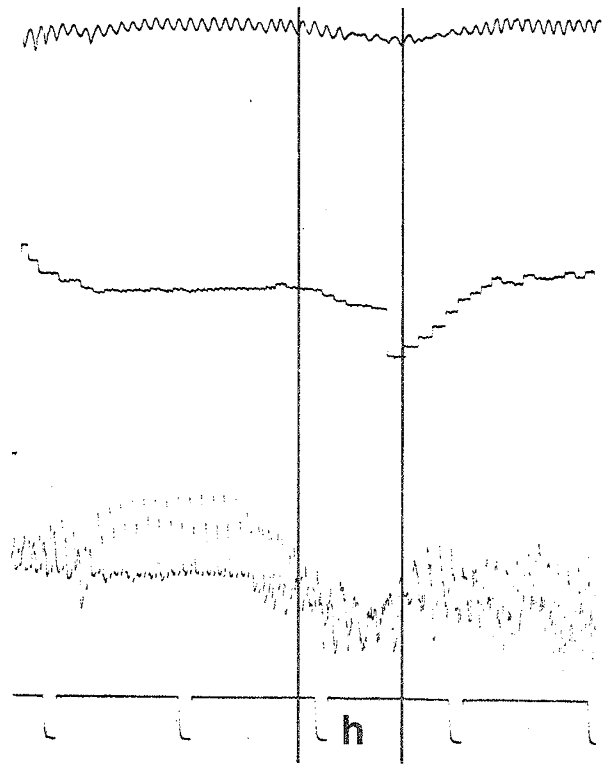


figure 8:2



Working Papers 25 1983
Linguistics-Phonetics
Lund University, 139-150

VOICE PROBLEMS AND OCCUPATIONAL ENVIRONMENT

Ann-Christine Ohlsson

ABSTRACT

Knowledge of environmental causes of voice disorders is scanty and is mainly based on the clinical experience of speech therapists. However, there is evidence in occupational research that factors in the physical environment, such as air pollution and chemical substances, might harmfully influence the mucous membrane of the voice organ. This paper presents the outlines of a study in progress where the aim is to describe voice quality in two groups of shipyard workers working in different environments regarding amount of air pollution and chemical substances. Within this framework a preliminary study is presented indicating that perceptual analysis of voice quality differentiates significantly between organic and non-organic voice disorders. The report on voice problems in shipyard workers will be published later.

1. INTRODUCTION

Voice problems constitute one of the largest groups of disorders at departments of logopedics and phoniatrics. The waiting list for voice therapy is consequently very long and reflects an unsatisfied need for preventive work in the area. Knowledge of the organic and non-organic causes of voice disorders in the environment and behaviour has mostly been gained from clinical experience. Little research has been done in the field of occupational environment and voice behaviour, but the subject has attracted growing interest in recent years.

In Foniatri för medicinare (1973) Fritzell writes (in my English translation) "We do not know how many persons there are who have problems with their voices or how many who have deviating voices without experiencing discomfort... No attempts have been made to define the prevalence of voice deviations in the general population". Since this was written voice studies have been carried out on teachers (Johansson and Södersten 1977; Eklund, Ludwigsson and Åstrand 1977), which confirm the clinical consensus that persons working in occupations where the voice has to be used frequently, such as teachers, telephone operators and politicians, run the risk of acquiring voice problems. These are also the kinds of professions that are typically found among patients seeking help for non-organic voice disorders at speech therapy clinics. Voice studies on other occupations than those mentioned above have not been reported. Occupations associated with environmental factors that might be expected to influence the voice organ harmfully are those in industrial environments with noise, air pollution and chemical substances.

2. THEORETICAL BACKGROUND

Earlier work on voice occupational environments in industry is scanty. However, results from research on occupational hazards related to the upper respiratory tract and larynx underline the probability of persons working in polluted environments developing laryngeal diseases.

H von Zenk (1968) points out that changes in industrial production and new insights in phoniatrics should result in increased interest in occupational effects on the larynx. He cites results from various studies indicating that lesions in the larynx have been caused by factors in the occupational environment. He classifies the factors into three groups: Mechanical, chemical and physical. Examples of mechanicla factors are dust from asbestos, aluminium, fibre glass and chalk. Chemical factors are water soluble gases, superphosphates, cadmium, vanadin, sulphur, salt solutions and salt compounds. Zenk is of the opinion that these materials cause chronic laryngitis. Examples of physical factors are heat and alternations between dry and humid air. Zenk also points out substances that are carcinogenic to the larynx such as asbestos, chromium and hydrocarbons.

Volney Bulteau (1975) cites Fabricant (1963) who points out that chemicals are the most frequent cause of industrial diseases in the throat. He lists various chemical components that can cause irritations to the mucous membrane in the upper respiratory tract. These can be found in the form of gas, fumes, mist, vapour, dust or smoke. He writes "...fumes from metals and alloys used in welding or burning are capable of producing severe irritation in the postnasal space".

G B Glasenapp (1975) describes a case of cancer of the larynx that developed from chronic laryngitis owing to mechanical and physical conditions in the air at the working place.

G P Moore (1971) lists the same substances as does Frabricant, above, and adds "Considerable impairment to the respiratory and vocal mechanisms of persons in the general population probably occurs from inconspicuous inflammatory substances that exist in occupational or home environments and which exert their influences cumulatively over considerable periods of time. This possibility demonstrates that the speech clinician needs to investigate these potentially irritating factors along with the pattern of vocal use, particularly when the amount and kind of speaking or singing does not support a diagnosis limited to vocal abuse. With suitable mucosal conditions, a mild amount of vocal misuse could cause excessive vocal deviation. The presence of inflammation and associated edema in the vocal folds provides an ideal basis for vocal deterioration". Moore is of the opinion that the irritants described by Fabricant often produce chronic laryngitis.

The quatotions above motivate this forthcoming study.

3. QUESTIONS AND GOAL

The aim of this study is to investigate the vocal behaviour and vocal experience of welders as compared to non-welders.

Questions.

Are there any connections between vocal behaviour and occupational environment?

What methods are optimal in measuring vocal behaviour in this situation?

4. PLAN

4.1 Data collection

4.1.1 Material

The subjects of this study are shipyard workers. Two groups will be studied, one exposed to welding and the other not. All subjects are non-smoking men. Mean noise levels at the working places are known. The size of the sample has not yett been decided.

4.1.2 Methods

The informants will be interviewed individually, using questionnaires that highlight voice problems and vocal use.
The questionnaire that will be used has been refined several times and will now be tested in its sixth version.

Voice recordings will be made in a soundproof room with a two-channel tape recorder (Revox A77). The airborn signal will be recorded on one channel from a microphone mounted on a spectacle frame worn by the subject. The other channel will record the signal from a contact microphone attached to the neck, below the cricoid cartilage. The subjects will be asked to read aloud a standardized text and thereafter to produce a vowel, [ɑ:], three times. The subject will be instructed to prolong the vowel maximally in each trial. Before the vowels are recorded I will demonstrate this task to the subject.

4.2 Data work

The replies to the questionnaires will be evaluated with regard to voice problems and vocal use.

Auditory analysis of the voice recordings will be carried out by a group of logopeds. The voice diagnosis scheme that will be used is similar to the one used by Hammarberg et.al. (1981) for their study of waveform perturbations in patients with voice disorders. The scheme will be presented later in this paper in the section dealing with the auditory analysis of organic and non-organic voice disturbances.

The harmonic-to-noise ratio will be quantified with the aid of sound spectrograms of the vowels (Yumoto, Gould and Baer, 1982).

The waveforms of parts of the contact microphone signal will be analysed. Many investigators have studied the cycle-to-cycle variations in the speech waveform and the results indicate that these variations occur more frequently in organically caused voice disturbances than in others (Iwata and Leden, 1970; Lieberman, 1963; Askenfelt and Hammarberg, 1981).

The fundamental frequency distributions will be analysed. It is known from clinical experience and from the research literature that voice disorders caused by mass lesions of the vocal folds result in restricted fundamental frequency distribution (Hecker and Kreul, 1970; Kitzing, 1979).

Maximum phonation time will be measured on the longest of the three vowel productions. A decrease in maximum phonation time is expected in cases where the glottal closure is incomplete (Neiman and Edeson, 1981; Hirano, 1981). The vital capacity of the lungs will be known for all of the informants.

In order to answer the main questions that this study addresses the data will be mutually correlated. The data on the subjects that are exposed to welding will be compared to the data on the subjects that are not.

5. AUDITORY ANALYSIS OF VOICE QUALITY IN ORGANIC AND NON-ORGANIC VOICE DISORDERS

5.1 The aim of this study is to test whether the voice diagnosis scheme proposed for the study on welders can differentiate between

a) Hoarse voices and non-hoarse voices
b) Organic voice disorders and non-organic voice disorders

5.2 Material

A sample of voices previously diagnosed as to organic or non-organic disorders was selected from the tape archives of the Department of Logopedics and Phoniatrics in Gothenburg. The recordings were checked to make sure that the voices chosen for this study really were representative of their diagnosis, i.e. that no one voice would sound too good or too bad compared with the average voice of that diagnosis. Fifteen voices were selected. The diagnoses were chronic laryngitis (5), paralysis of the recurrent nerve (3), vocal nodules (3) and phonasthenia (4). The examples of the last diagnosis come from patients with healthy vocal folds. All the other patients had pathological changes of their vocal folds. The organically caused voice deviations were all recorded before voice therapy.

5.3. Method

The recordings were copied twice, onto two separate tapes. The voice productions were ordered randomly, differently on each tape. The listener group consisted of speech therapists and students of speech therapy. They were all unaware of the fact that the two tapes were composed of the same voices. The number of listeners available for the first tape was 13 and for the second tape 11. Each tape was listened to on a separate occasion with an interval of at least fourteen days between them. When judging the first tape the listeners rated the degree of hoarseness for every voice on a 4-point scale. When judging the second tape the listeners rated their impression of voice quality on a 4-point scale. The voice quality terms rated on every voice were those in daily use in the phoniatric clinic: tense, lax, creaky, hard glottal attacks, breathy, rough, gratings, intermittent aphonia and diplophonia. It is assumed in clinical experience that hoarseness corresponds best to the terms breathy, gratings, rough and intermittent aphonia.

5.4 Results and discussion

Mean scale degrees (MSD) and standard deviations (SD) were computed for voice productions made by the patients with organic voice disorders (ORG) and compared to the voice productions made by the patients with non-organic voice disorders (NON-ORG). The results are presented in Table 1. ORG stands for voice productions from the patients with chronic laryngitis, paralysis of the recurrent nerve and vocal nodules, NON-ORG stands for voice productions from the patients with phonasthenia. MSD and SD for every voice quality term are listed. The listeners' interreliability was significant for W =.728 at the .001 level (tested with Kendall's Coefficient of Concordance).

Table 1 shows significantly higher scores on the the vocal terms breathy, gratings and rough for the organic voice disorders. The differences between the samples on intermittent aphonia and diplophonia are also large but not quite significant. These are all terms used in the every-day clinical language to describe "organic hoarseness". The topmost terms tense and lax in the table (and also on the diagnosis sheet) express, regardless of the causes, the function of the vocal folds, which could explain the non-significant difference between these small samples. The terms creaky and hard glottal attacks have scores that differ only marginally between the two samples. This is probably due to the small samples. The masking effect of roughness and gratings on these qualities in the organic voice disorders should result in higher scores for the non-organic voice disorders. The high scores on breathiness, diplophonia and intermittent aphonia in the voice disorders with organic

Table 1. Statistical comparison of summated diagnoses of the nine voice quality terms for 11 subjects with organic voice disorders (ORG) and 4 subjects with non-organic voice disorders (NON-ORG). Significance tested with Mann - Whitney U Test.

Vocal quality term	ORG MSD_o	SD	NON-ORG MSD_n	SD	Diff. MSD_o-MSD_n	Level of signif. .01	.05
Tense	1.34	.67	1.02	.45	.32	NS	NS
Lax	.68	.69	.38	.24	.30	NS	NS
Creaky	.89	.58	.88	.49	.01	NS	NS
Hard glott. attacks	.68	.30	.73	.26	-.05	NS	NS
Breathy	1.50	1.07	.54	.22	.53		x
Gratings	.73	.62	.17	.11	.45	x	
Rough	1.12	.77	.15	.11	.97	x	
Interm. aphonia	.53	.67	.08	.06	.45	NS	NS
Diplophonia	.45	.64	.04	.05	.41	NS	NS

causes are inconsistent with hard glottal attacks and creaky voice, and should therefore also contribute to high scores for the non-organic voice disorders. To test if these assumptions were correct a sample of voice productions recorded from 24 subjects with phonasthenia, rated according to the same diagnosis scheme as previously described, were compared statistically regarding MSD scores with the scores of the organic voice disorders in this material. The listeners were four speech therapists with good rating interreliability (W = .706 significant at the .001 level tested with Kendall's Coefficient of Concordance). This comparison indicates that the MSD of scores on creakiness for the non-organic voice disorders in the larger sample was significantly higher than the scores for voice productions made by the patients with pathological changes of the vocal folds (at the

Table 2. Statistical comparison of summated diagnoses on nine voice quality terms for 5 patients with chronic laryngitis and 4 patients with phonasthenia. Significance tested with Mann - Whitney U Test (p = probability that there is no difference in distribution of judgements for the two samples).

Vocal quality term	Chronic laryng. MSD_{cl}	Phonas-tenia MSD_{fo}	Diff. MSD_{cl}- MSD_{fo}	p. (1-tailed)	Level of signif. .01	.05
Tense	1.53	1.02	.51	.452	NS	NS
Lax	.69	.38	.31	.206	NS	NS
Creaky	.82	.88	-.06	.452	NS	NS
Hard glott.attacks	.74	.73	.01	.452	NS	NS
Breathy	1.74	.54	1.20	.016	NS	x
Gratings	.48	.17	.31	.056	NS	NS
Rough	1.27	.15	1.12	.008	x	
Interm.aphonia	.65	.08	.57	.143	NS	NS
Diplophonia	.39	.04	.35	.095	NS	NS

.005 level tested with the Mann-Whitney U Test). The scores on hard glottal attacks were also higher but not significant above the .5 level. The vocal terms tense and lax were both scored significantly higher for organic voice disorders in this comparison (tense: p = .0446, lax: p = .0091). The earlier differences on all other vocal quality terms were accentuated in this later comparison. In view of these results, the voice diagnosis scheme used in this study is found to be acceptable as an instrument differentiating between organic and non-organic voice disorders.

Referring to section 2, there are findings indicating that persons working in polluted environments run a risk of acquiring chronic laryngitis. The voice productions in this material made by the patients with chronic laryngitis are therefore of special interest. The scores on the voices with this diagnosis were therefore compared separately to the scores on the non-organic voice disorders. This comparison is presented in Table 2.

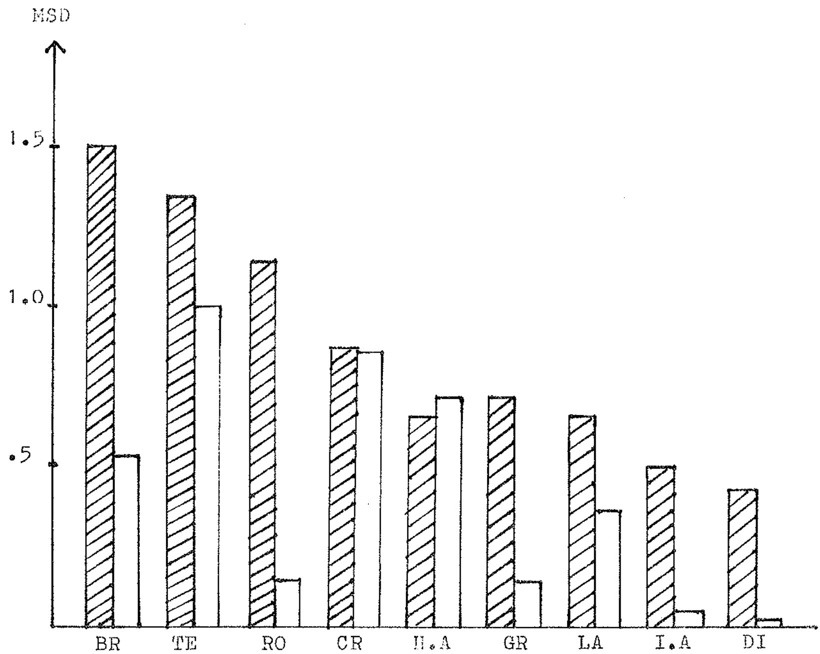


Figure 1. MSD on each of the vocal quality terms for organic voice disorders (marked with lines) and non-organic voice disorders.

Table 2 shows significantly higher scores for voice disorders due to chronic laryngitis on the vocal terms breathy and rough. The score for this diagnosis on gratings is nearly significantly higher than the score for phonasthenia. The scores in Table 2 are in good agreement with the scores in Table 1, and as can be seen the scores on the vocal terms generally associated with organic hoarseness are with one exception higher for chronic laryngitis than for all the organic voice disorders taken together. The exception is the term gratings which probably gets a lower score on chronic laryngitis because of its connection with the diagnosis of vocal nodules. Gratings is described as high-pitched aperiodic noise which is frequently observed in the voices of patients with vocal nodules on their vocal folds.

Mean scalar degrees of the listeners' judgements on the nine vocal quality terms are presented for organic and non-organic voice disorders in Figure 1.

Table 3. Degree of hoarseness (MSD) for the different diagnoses in the two groups of voice disorders.

Diagnosis	Degree of hoarsness (MSD)
Chronic laryngitis	1.93
Paralysis of the recurrent nerve	1.69
Vocal nodules	1.33
Phonastenia	.95

Figure 1 shows the great differences between the two groups of voice disorders with regard to the terms breathy, roughness, gratings, intermittent aphonia and diplophonia. The abbreviations stand for: BR = breathy, TE = tense, RO = rough, CR = creaky, H.A = hard glottal attacks, GR = gratings, LA = Lax, I.A = intermittent aphonia and DI = diplophonia.

To test significance ot using the voice diagnosis scheme as an instrument to detect hoarseness the listener's judgements of degree of hoarseness for organic and non-organic voice disorders were compared. The patients with organic voice disorders are those with chronic laryngitis, paralysis of the recurrent nerve and vocal nodules. The patients with non-organic voice disorders are those with phonasthenia. The mean scalar degrees on hoarseness for the different diagnoses are presented in Table 3.

The difference in degree of hoarseness between organic and non-organic voice disorders was significant at the .25 level (tested with Mann-Whitney U Test). The highest scores of hoarseness were given to the same group of voice disorders that were accorded the highest scores on the vocal quality terms that in general correspond best to organic hoarseness. The voice diagnosis scheme used in this study is therefore suitable as an instrument for detecting hoarseness. The differences in degree of hoarseness between the four groups of diagnoses are presented in Figure 2.

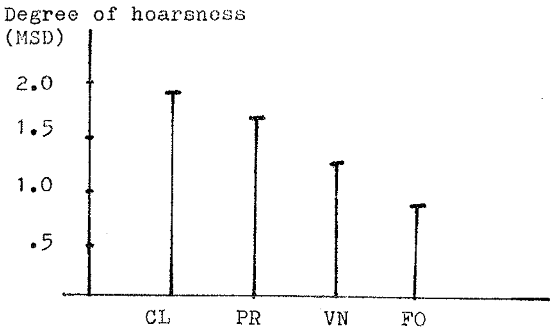


Figure 2. Mean scalar degrees on degree of hoarseness for chronic laryngitis (CL), paralysis of the recurrent nerve (PR), vocal nodules (VN) and phonasthenia (FO).

Figure 2 shows that the greatest difference in hoarseness in this material lies between chronic laryngitis and phonasthenia. This result is in agreement with the high scores for chronic laryngitis on the vocal terms expressing organic hoarseness in Table 2.

5.5 Concluding remarks

The study aimed at testing a voice diagnosis scheme to be used as an instrument for detecting organic hoarseness. The results of the study indicate that the scheme used here is acceptable as such an instrument. There was good agreement among the listeners' ratings and the voice productions with the highest scores on hoarseness were also rated highest on the vocal quality terms usually associated with hoarseness in clinical practice. The organic voice disorders received significantly higher scores on these vocal quality terms than did the non-organic voice disorders.

REFERENCES

Askenfelt, A. & Hammarberg, B. 1981. Speech waveform perturbation analysis revisited, STL- QPSR 4

Bulteau, V. 1976. Nose and throat: occupational hazards, The medical journal of Australia 1

Coleman, R. F. 1971. Effect of waveform changes upon roughness perception, Folia Phoniatrica 23: 314-322

Deal, R. E. & Emanuel F. W. 1978. Some waveform and spectral features of vowel roughness, Journal of Speech and Hearing Research 21: 250-264

Eklund.,E. L, Ludwigsson, N . & Åstrand I . 1977. Occurence of voice disturbances in personel working with preschool children, Examination Paper in Logopedics, Stockholm

Emanuel, F. W. & Sansone, F. E, 1969. Some spectral features of normal and simulated 'rough' vowels, Folia Phoniatrica 21: 401-415

Fritzell, B. 1973. Foniatri för medicinare, Almqvist & Wiksell

Glasenapp, G. B. 1975. Beitrag zur Berufsbedingten Entstehung des Kehlkopfkarzinoms, Laryng. Rhinol. 54: 565-568

Hammarberg, B., Fritzell, B., Gauffin, J., Sundberg, J., & Wedin, L. 1980. Acoustic and perceptual analysis of vocal dysfunction, Phoniatric and logopedic progress report no. 2

Hanson, W. & Emanuel, F. W. 1979. Spectral noise and vocal roughness relationships in adults with laryngeal pathology, Journal of Communication Disorders 12: 113-124

Hecker, M. H. L. & Kreul, E.J. 1970. Description of the speech of patients with cancer of the vocal folds, JASA 1275-1282

Iwata, S. & von Leden, H. 1970. Pitch perturbations in normal and pathological voices. Folia Phoniatrica 22: 413-424

Johansson, J. E. & Södersten, M. 1977. Voice problems in gymnastics masters in Stockholm area, Examination Paper in Logopedics, Stockholm

Kitzing, P. 1979. Glottographic frequency analysis, Malmö

Lieberman, P. 1961. Perturbations in vocal pitch, JASA 35 (5)

Lieberman, P. 1963. Some acoustic measures of the fundamental periodicity of normal and pathological larynges, JASA 35 (3)

Moore, G. P. 1971. Organic voice disorders, Prentice Hall, Inc. Englewood Cliffs, N.J.

Wendahl, R. W. 1966. Laryngeal analog synthesis of jitter and shimmer auditory parameters of harshness, Folia Phoniatrica 18: 98-108

Yanagihara, N. 1967. Significance of harmonic changes and noise components in hoarsness, Journal of Speech and Hearing Research 10: 531-541

Yumoto, E., Gould, W. & Bear, T. 1982. Harmonics-to-noise ratio as an index of the degree of hoarsness, JASA 71 (6)

Zenk, (von) H. 1968. Arbeitsbedingte Auswirkungen auf den Kehlkopf, Z. Ärztl, Fortbildn. H 22

VOWEL REDUCTION IN BULGARIAN

Working Papers 25 1983
Linguistics - Phonetics
Lund University, 151-178

Thore Pettersson and Sidney Wood

1. INTRODUCTION

In this report we present phonological vowel reduction in Bulgarian and review the phonetic data given in the literature, as an introduction to a cinefluorographic and spectrographic study to be published in future reports. While this investigation deals specifically with Bulgarian, it is also of interest for the problem of vowel reduction in general, the phonetic character of schwa-like vowels, and consequently for theories of speech production and motor control and for phonology.

We shall have occasion to refer to both formal and informal speech, dialect and standard forms, since phonological reduction in Bulgarian is subject to both stylistic (formality and situation) and dialect constraints. Contemporary Standard Bulgarian (CSB) is defined by Scatton (1975) as "the contemporary literary norm of the Bulgarian capital, Sofia, as reflected in their <u>formal</u> speech and in <u>normative</u> grammars".

We distinguish between phonological reduction (regular vowel to vowel alternations depending on whether the syllable is lexically stressed or not) and phonetic reduction (a tendency for a vowel quality to become indistinct as it weakens and shifts towards schwa). In many languages these two processes may coincide, the phonological reduction also being towards schwa. In Bulgarian they do not coincide.

A classical account of the Bulgarian vowel system is that given by Trubetskoy (1939). In many languages, says Trubetskoy, the indeterminate vowel only appears in partial systems in those phonic positions where several oppositions based on degree of aperture and oppositions of timbre are neutralized. But Bulgarian is an example of a language where

> an <u>indeterminate</u> vowel in a triangular system can become a <u>specific</u> vowel by entering into a relation of bilateral opposition with a. The Bulgarian indeterminate vowel has approximately the same degree of aperture as o and e, but it is neither rounded nor palatal. It would hardly be possible to assume a pure opposition of timbre between Bulgarian ə and o or between Bulgarian ə and e. But the proportions o:a=u:ə, e:a=i:ə and the proportions u:o=i:e=ə:a deduced therefrom may well be established. The conditions in unstressed syllables (at least in a part of the local types of pronunciation) are proof that this proportion corresponds to reality. For in these syllables o, a and e are not permitted, only u, i and ə are. In other words the oppositions based on degree of aperture u-o, i-e and ə-a are neutralized, while the triangular character of the vowel system is preserved. Graphically, this may be presented as follows:

a

o e ə

stressed ə u i unstressed

u i

The orthographic representation of the "indeterminate vowel" in Bulgarian is ъ, "yer". We prefer to transcribe this vowel, conventionally, as /ă/ rather than /ə/ in order to distinguish it from true weak schwa. This is a typographical measure that enables us to keep an open mind as to the actual phonetic character of /ă/.

We also have a methodological goal. It has long been well-known that the Bell model (presumed high-mid-low and front-central-back tongue positions for vowels) fails to offer a correct or even an adequate description of vowel articulation. This has not impeded work in phonology so long as interest has been concentrated on abstract relations between units, without regard to physical data. Provided units are uniquely classified it does not matter what labels the classifying features bear. The Bell model breaks down when it is held to represent physiological fact (see further Wood 1975a, 1982a). This is a particularly severe drawback when speech production is to be related to phonology, for example when a production model requires phonological directives to control speech directly (as is the case in the model of Chomsky & Halle 1968), or when a continuous link is to be established between phonology, motor control, articulation, sound production and perceptual cues in the speech wave (as in the various models of the late Roman Jakobson).

For this introductory report we shall present Bulgarian vowel reduction in the traditional and familiar terminology of the Bell model, but our analysis of the problem and our subsequent reports on the phonological and phonetic processes involved will be framed in terms that more closely reflect current knowledge of speech production.

Phonetic and phonological descriptions of Bulgarian have been published by Scatton (1975), Stojkov (1966), Tilkov (1970, 1982) and Tilkov & Bojadžiev (1981). Various dialect, morphological, sociolinguistic and stylistic aspects of vowel reduction in Bulgarian have been treated by Bojadžiev (1980), Ivančev (1980), Janakiev (1960), Pašov (1980a, 1980b) and Stojanov (1968). Pašov (1980a) has reviewed earlier grammars on the subject. Lockwood (1972) uses the Bulgarian vowel alternations between stressed and unstressed syllables as a language example in a theoretical discussion of the role of markedness in conventional generative phonology and in stratificational phonology.

2. PHONOLOGICAL VOWEL REDUCTION IN BULGARIAN

The following alternations occur in informal Bulgarian speech between stressed and non-stressed vowels:

	STRESSED	NON-STRESSED
/i/	i	i
/e/	e, ɛ	i
/u/	u	u
/o/	o, ɔ	u
/ă/	ɐ, ɜ, ə	ɐ, ɜ, ə
/a/	a	ɐ, ɜ, ə

The reductions are easily discerned in morphological stress alternations such as the following examples (stressed syllables are indicated by an acute accent, compared vowels are underlined):

	STRESSED	NON-STRESSED	
/í/-/i/	/í m e/	/i m e n á/	name(s)
	[í m i]	[i m i n á]	
/é/-/e/	/s é l o/	/s e l á/	village(s)
	[s é l u]	[s i l á]	

/á/ -/a/	/r á b o t a/	/r a b ó t n i k/	work(er)
	[r á b u t ă]	[r ă b ó t n i k]	
/ó/-/o/	/ó n z i/	/o n á z i/	that: m (f)
	[ó n z i]	[u n á z i]	
/ú/-/u/	/b ú k v a/	/b u k v á r/	letter/ABC
	[b ú k v ă]	[b u k v á r]	
/ắ/-/ă/	/k r ắ č m a/	/k r ă č m á r/	tavern(er)
	[k r ắ č m ă]	[k r ă č m á r]	

The extent of vowel reduction varies considerably, depending on stylistic, dialect and morphological constraints.

Non-stressed /a/ is most likely to be reduced, the reduction of /o/ is quite common, but in CSB /e/ is frequently not reduced.

Vowel reduction is avoided in very formal speech and is not heard, for example, in the speech of radio announcers (Pašov 1980a). This contrasts completely with Russian where the norm requires reduction, non-reduction being looked upon as rustic.

While speakers are subject to social pressures to adapt their speech in this respect towards the norm, it should be noted that the triggering factor is said to be style rather than social class. This is something that Bulgarians are taught at school (cf. Gyllin 1982). The same speaker can vary vowel reduction from occasion to occasion depending on the formality of the situation. Different individuals vary also in how far they succeed in living up to the norm. We believe that this merits a sociolinguistic study comparable to Labov's study of New York speech (Labov 1972). Reading aloud is a formal situation and, typically, one of our informants remarked that while he was reading our word lists he felt the presence of a schoolmaster standing behind him.

Janakiev (1960) points out that Bulgarians cannot spell non-stressed vowels properly unless they know the etymology. Pašov (1980b) has recorded numerous examples of misspelt non-stressed vowels from university entrance examination papers to degree courses in Bulgarian by above average applicants. Pašov underlines that an above average school result is no guarantee that a student can master the spelling of weak vowels. The spelling mistakes occur in both directions (i.e. they include hypercorrect forms) and are more frequent for /i-e/ and /u-o/ than for /ă-a/. He also gives surprising examples of proof-reading errors from official publications and even from the Academy of Sciences spelling dictionary.

Ivančev (1980) has studied rhymes in Bulgarian poetry. Pure rhymes are based on identical vowel sounds and writers who avoid vowel reduction in their own speech should be less likely to rhyme non-stressed /e, o, a/ with non-stressed /i, u, ă/ respectively. It turns out that pure rhymes between reduced /o/ and /u/ and between reduced /e/ and /i/ are very frequent. In contrast, the late 19th century poet Penčo Slavejkov has frequently rhymed non-stressed /o/ with non-stressed /a/. Ivančev records no fewer than 141 examples, e.g:

.....dvata	the two	.....kogato	when
.....zlato	gold	....vratata	the doorway

but Slavejkov has only once rhymed non-stressed /o/ with non-stressed /u/:

....prez ramo	over the shoulder
....čestta mu	the honour to him

Ivančev attributes the numerous reduced /a-o/ rhymes to

Slavejkov's own pronunciation, and takes them as evidence that Slavejkov's /o/ was pronounced with little lip rounding. But they could also be a contrived breach of convention that was intended to shock the reader.

There is a strong dialect component in the tendency to reduce non-stressed vowels. Bojadžiev (1980) among others reports that non-stressed /a/ is neutralized without exception in all dialects whereas the reduction of /e/ and /o/ is limited to eastern dialects. From this fact Bojadžiev draws the conclusion that the reduction of /a/ on the one hand and of /e, o/ on the other consists of two phonologically distinct processes. Similar vowel reductions occur in neighbouring parts of the Balkans. For example they occur in NE but not southern Greek and in E but not central Macedonian.

Pašov (1980a) notes several morphological exceptions. Typical examples are:

tense:	/mólix/	[mólix]	I asked
	/mólex/	[mólex]	I was asking
		(weak /e/	not reduced)
person:	/dovédox/	[duvédox]	I led
		(final weak /o/	not reduced)
vocative:	/sin/	[sin]	son (basic form)
	/síne/	[síne]	son!
		(final weak /e/	not reduced)
	/stánko/	[stánku]	Stanko (basic form)
	/stánko/	[stánko]	Stanko! (vocative)
	(vocative final weak /o/		not reduced)

The unreduced vocative ending is often reinforced by being lengthened:

[síne:] son!
[stánko:] Stanko!

3. PROBLEMS AND RULES

Phonetic data obtained from the literature (Scatton 1975, Stojkov 1966, Tilkov 1970, Tilkov & Bojadžiev 1981, Tilkov 1982) is unfortunately contradictory. This data is reviewed in detail in section 4. In this section we will present various alternative standpoints and see how they affect the possible solutions.

At first sight the task is simple: there is one set of units that is subject to reduction

/e, o, a/

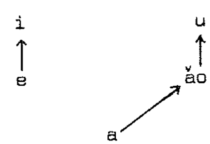
and one set that is not

/i, u, ă/

But what are the defining features for each set and what are the differentiating features?

The ultimate solution is dependent on the classification of /ă/. The following possibilities can all be derived from the published phonetic data reviewed in the next section.

3.1 Is /ă/ a mid back unrounded vowel (i.e. an [ɣ-ʌ]-like spread-lip vowel corresponding to rounded [o-ɔ])? In traditional terms this gives



All non-high vowels except /ă/ shift up one step, i.e. unrounded back has to be excluded (Table I).

Table I. Feature matrix for vowel reduction according to the solution in 3.1. The brackets enclose the features that are affected by reduction. At this stage the specifications are redundant.

	e	i	o	u	a	ǎ
high	(–	+)	(–	+)	–	–
low	–	–	–	–	(+	–)
front	+	+	–	–	–	–
back	–	–	+	+	(–	+)
round	–	–	+	+	–	–

The rule

$$\begin{bmatrix}+\text{low}\end{bmatrix} \longrightarrow \begin{bmatrix}-\text{low}\\ +\text{back}\end{bmatrix} \;\Big/\; \underset{\begin{bmatrix}-\text{stress}\end{bmatrix}}{\underline{\qquad\qquad}} \qquad \text{(I)}$$

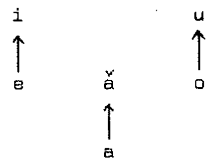
will raise and back a, and

$$\begin{bmatrix}-\text{high}\\ \alpha\text{front}\\ -\alpha\text{back}\\ -\alpha\text{round}\end{bmatrix} \longrightarrow \begin{bmatrix}+\text{high}\end{bmatrix} \;\Big/\; \underset{\begin{bmatrix}-\text{stress}\end{bmatrix}}{\underline{\qquad\qquad}} \qquad \text{(II)}$$

will raise e and o but not ǎ.

The two different rules reflect two different phonetic processes. The complexity of the rules reflects the need to back /a/ and to exclude /ǎ/. (Scatton's classification – /a/ low back and /ǎ/ mid back – is a variant of this solution.)

3.2 Alternatively, is /ǎ/ a mid central vowel, akin to [ɜ] or [ə] (the Trubetskoy solution quoted in the introduction)? This gives



Again, all non-high vowels except /ǎ/ shift up one step, but /a/ does not have to be backed now (Table II). Mid central has to be excluded.

Table II. Feature matrix for vowel reduction according to the solution in 3.2. The brackets enclose the features that are affected by reduction.

	e	i	o	u	a	ǎ
high	(–	+)	(–	+)	–	–
low	–	–	–	–	(+	–)
front	+	+	–	–	–	–
back	–	–	+	+	–	–
round	–	–	+	+	–	–

The rule

$$[+\text{low}] \longrightarrow [-\text{low}] \;/\; \underset{[-\text{stress}]}{\overline{\qquad\quad}} \qquad \text{(III)}$$

will raise a, and

$$\begin{bmatrix} -\text{high} \\ \alpha\text{front} \\ -\alpha\text{back} \end{bmatrix} \longrightarrow [+\text{high}] \;/\; \underset{[-\text{stress}]}{\overline{\qquad\quad}} \qquad \text{(IV)}$$

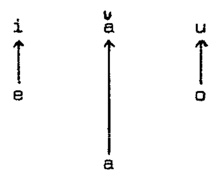
will raise e, o but not mid central ǎ.

There are still two different processes for /e, o/ and /a/ respectively, and /ǎ/ still has to be excluded.

3.3 Is /ǎ/ a high (perhaps central) vowel (Table III)?

Table III. Feature matrix for vowel reduction according to the first solution in 3.3. The brackets enclose the features that are affected by reduction.

	e	i	o	u	a	ǎ
high	(-	+)	(-	+)	(-	+)
low	-	-	-	-	(+	-)
front	+	+	-	-	-	-
back	-	-	+	+	-	-
round	-	-	+	+	-	-



This offers a seductively simple rule: non-high vowels become high (implying that low becomes non-low):

$$[-\text{high}] \longrightarrow [+\text{high}] \ / \ \underset{[-\text{stress}]}{\underline{\qquad}} \qquad \text{(V)}$$

But this solution is the least likely since /ǎ/ is not usually looked upon as a high vowel.

And yet the simple structure of this solution tempts us to ask again: is there one simple feature that differentiates the reducing set from the non-reducing set? The feature that comes to mind is the degree of jaw-opening: is the jaw opening narrower for /i, u, ă/ and more open for /e, o, a/? There is some evidence that it could be.

We shall then need to reintroduce a feature that we can call *open* with the original meaning it once had with reference to the degree of mouth opening depending on the jaw angle (Wood 1982b). With the Bell vowel model generally accepted at the end of the 19th century, the degree of mouth opening was disregarded as a parameter and the terms *close* and *open* were instead associated with the openness of the passage betwen the tongue and the hard palate, thus becoming synonymous with *high* and *low*. This is understandable since the mandible position is a component of tongue height (for palatal vowels at least) and it is virtually impossible to reconcile a mouth opening feature with the Bell tongue features. For example, if [open]⟶[close], then the tongue features have to be respecified too (perhaps [+low]⟶[-low] or [-high]⟶[+high]).

But what we may be faced with in Bulgarian vowel reduction is unmodified lingual activity combined with a narrower jaw opening. This is easier to express (and is physiologically more plausible) in terms of the basic tongue postures (see Fig. 3 and Wood 1979, 1982a):

	palatal	labio-velar	low pharyngeal
close	i	u	ă
	↑	↑	↑
open	e	o	a

If /a/ and /ă/ differ only in mandibular depression, we have the matrix given in Table IV.

Table IV. Feature matrix for vowel reduction according to the second solution in 3.3. The brackets enclose the features that are affected by reduction.

	e	i	o	u	a	ǎ
palatal	+	+	–	+	–	–
velar	–	–	+	+	–	–
pharyngeal	–	–	+	–	+	+
open	(+	–)	(+	–)	(+	–)
round	–	–	+	+	–	–

This gives the following very general rule:

$$\left[\text{+open}\right] \longrightarrow \left[\text{-open}\right] \Big/ \frac{\quad}{\left[\text{-stress}\right]} \qquad \text{(VI)}$$

In the next section the phonetic data published in the literature will be reviewed and interpreted in relation to these possible solutions.

4. PHONETIC DATA

X-ray profiles

Tilkov has published two sets of x-ray profiles (Tilkov 1970, Tilkov & Bojadziev 1981, Tilkov 1982).

The profiles for /ǎ, a, o/ are reproduced in Fig. 1. Tilkov's interpretation of the /ǎ/ profile is that the vocal tract is more or less uniform throughout its length except for a slight narrowing in the pharynx. This narrowing is not so extreme as for /a/ but he notes an evident affinity. He concludes that /ǎ/ is a back vowel (in the sense that it is formed in the

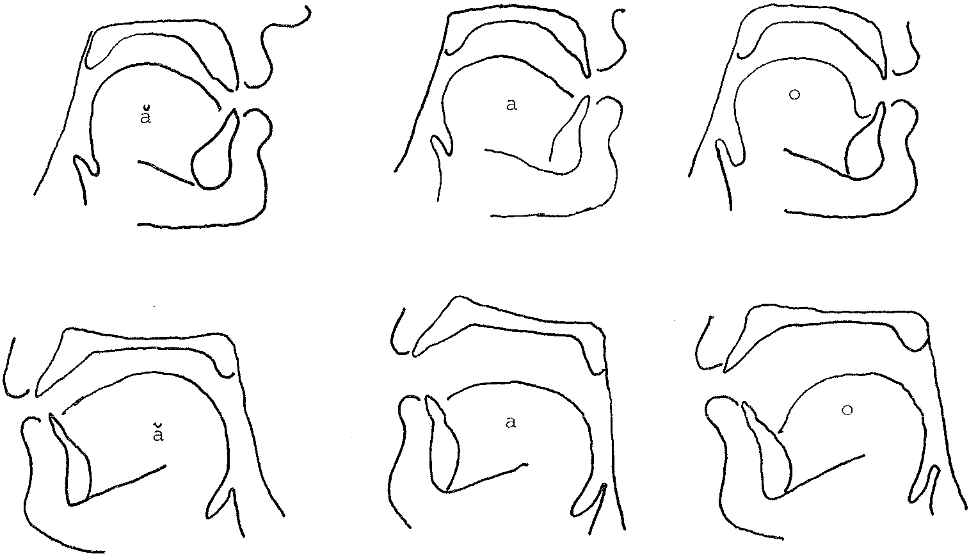


Fig. 1. Profile tracings of /ă, a, o/ after Tilkov & Bojadziev (1981) (above) and Tilkov (1970, 1982) (below).

pharynx, not in the Bell sense).

We have compared the tongue postures relative to the mandible in Tilkov's profiles (Fig. 2). This comparison isolates the lingual manoeuvres the speaker has used for the various vowels. As Figs. 2 and 3 show, the tongue assumes one out of a small set of typical tongue postures relative to the mandible. Each posture can be interpreted in terms of the underlying muscular activity (Fig. 3, for further details see Wood 1979). The tongue forms a major constriction at one of four places in the vocal tract: along the hard palate for [i-ɛ] and [y-œ]-like vowels, along the soft palate for [u-ʊ] and [ɯ]-like vowels, in the upper pharynx for [o-ɔ] and [ɤ-ʌ]-like vowels and in the lower pharynx for [æ-ɑ]-like vowels.

In Fig. 2 we have compared the posture for Bulgarian /ǎ/ with the palatal /i-e/ posture, the low pharyngeal /a/ posture, the upper pharyngeal /o/ posture and the velar /u/ posture taken from the Tilkov profiles. The result of the comparison is similar for both of Tilkov's sets of profiles.

Firstly, Fig. 2a shows that the tongue is less bunched relative to the mandible for the palatal vowels /e/ and /i/. This is typical of the tense-lax palatal [i, e] vs [ɪ, ɛ] contrast (Wood 1975b, 1982b). In Tilkov's profiles, /i/ is close (narrower jaw opening) and tenser (tongue bunched more towards the hard palate) while /e/ is open (larger jaw opening) and laxer (tongue less bunched towards the palate). Compared with these palatal /i, e/ postures, the /ă/ posture is not raised anteriorly towards the hard palate but bulges posteriorly towards the pharynx. This indicates activity in the glosso-pharyngei (superior pharyngeal constrictors) as illustrated in Fig. 3. The Tilkov /ă/ profiles are thus associated with a retracting manoeuvre and not with a palatal manoeuvre. This confirms the usual view that /ă/ is not a palatal vowel.

Figure 2b confirms the similarity of the /ǎ/ and /a/ postures noted by Tilkov. This similarity favours the alternative solution 3.3. But in both examples, the tongue is higher posteriorly for /ă/ than for /a/, suggesting styloglossal or glossopharyngeal activity rather than hyoglossal (i.e. activity directed towards the velum or upper pharynx rather than lower pharynx).

Figure 2c shows that there is a very close similarity between the /ă/ and /o/ postures. The only essential difference is that the tongue blade is depressed for /o/ but not for /a/, which modifies the anterior mouth cavity. This suggests that /ă/ is a spread-lip [ɣ-ʌ]-like vowel corresponding to rounded [o-ɔ], an interpretation that favours solution 3.1 above.

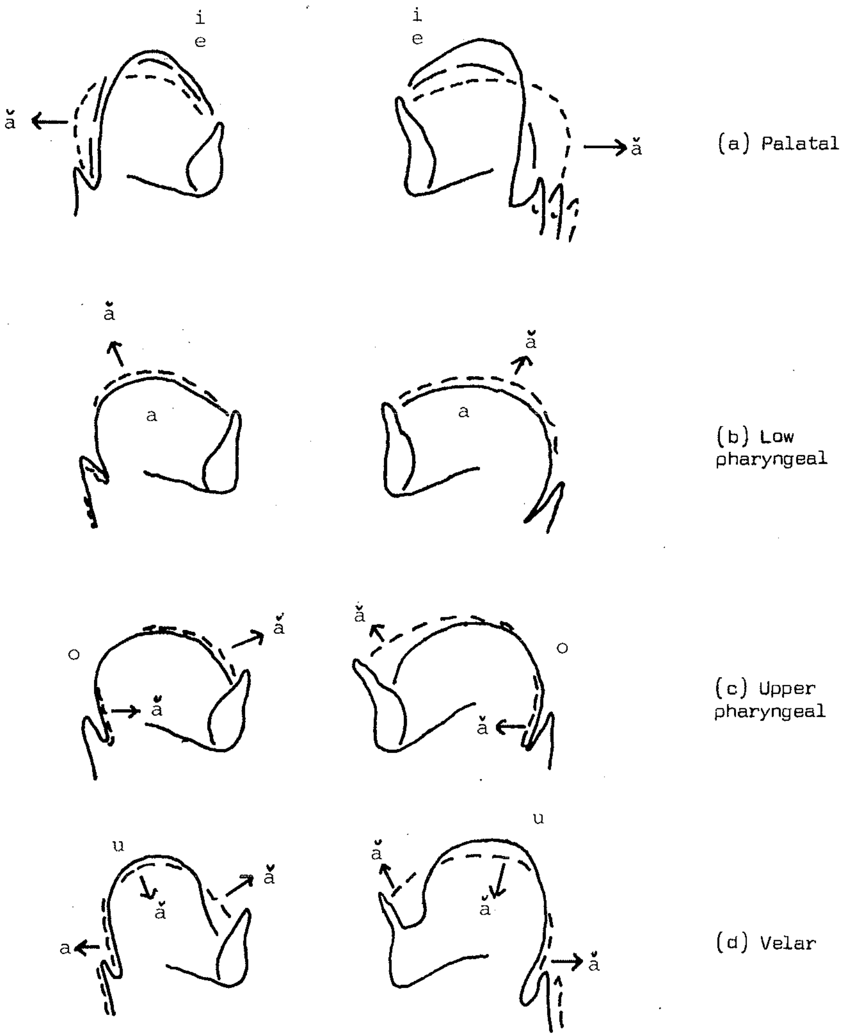
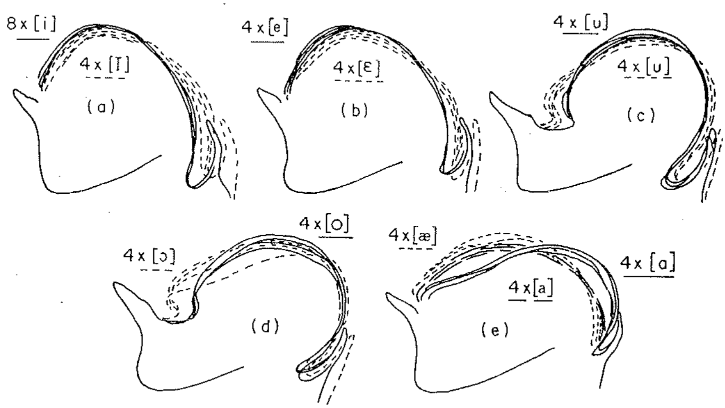
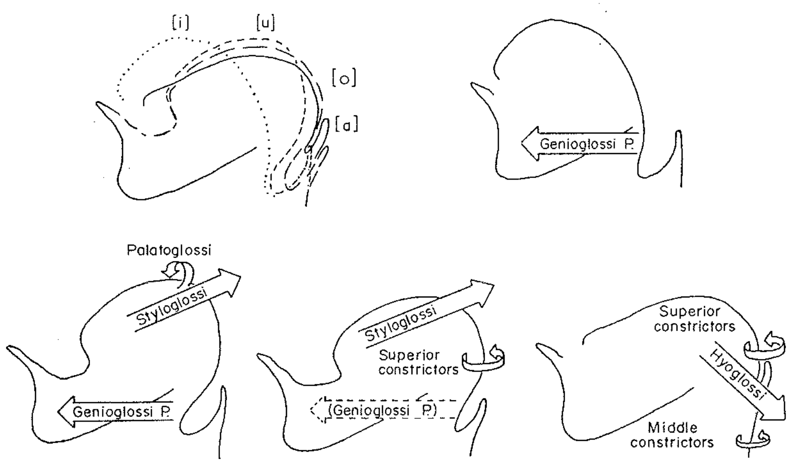


Fig. 2. Comparison of tongue posture of /a/ with the four basic tongue postures relative to the mandible (cf. Fig. 3) after Tilkov & Bojadziev's (left) and Tilkov's (right) profiles.

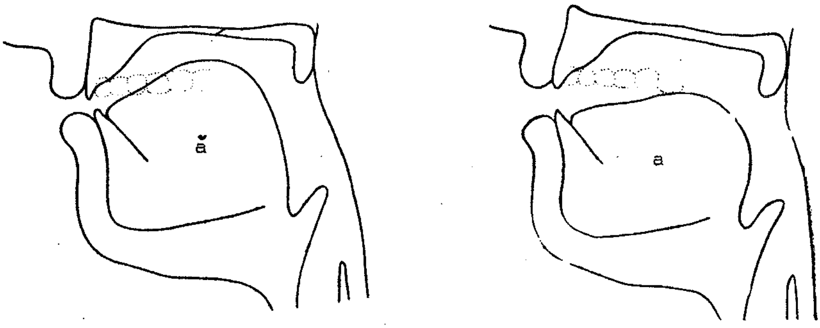


The positions of the tongue relative to the mandible for stressed vowels by the Egyptian Arabic subject.



The directions of contraction of the extrinsic muscles of the tongue and of the pharyngeal constrictors, arranged according to their presumed activity for the formation of the four constriction locations.

Fig. 3. Typical postures of the tongue relative to the mandible for vowels (above) and the associated muscular activity (below). From Wood (1979).



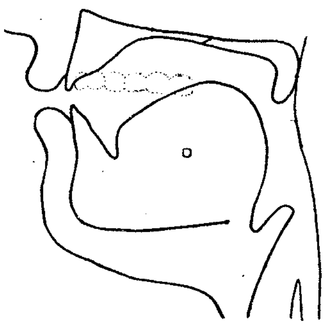


Fig. 4. Profile tracings of /ă, a, o/ after Stojkov (1966).

Finally, Fig. 2d compares /ă/ and /u/. The tongue is less raised relative to the mandible and the tongue root protrudes more into the lower pharynx for /ă/ than for /u/. The tongue blade is also less depressed for /ă/. These lingual differences are typical for the tense-lax [u-ʊ] contrast (see Wood 1975b) and are related to the levels of activity in the styloglossi and posterior fibres of the genioglossi. There is thus a possible lingual affinity between /ă/ and /u/ that would favour a variant of the first solution 3.3 above: /ă/ as a high (possibly back) vowel, corresponding to /u/. This is not an interpretation that native Bulgarian speakers would intuitively accept.

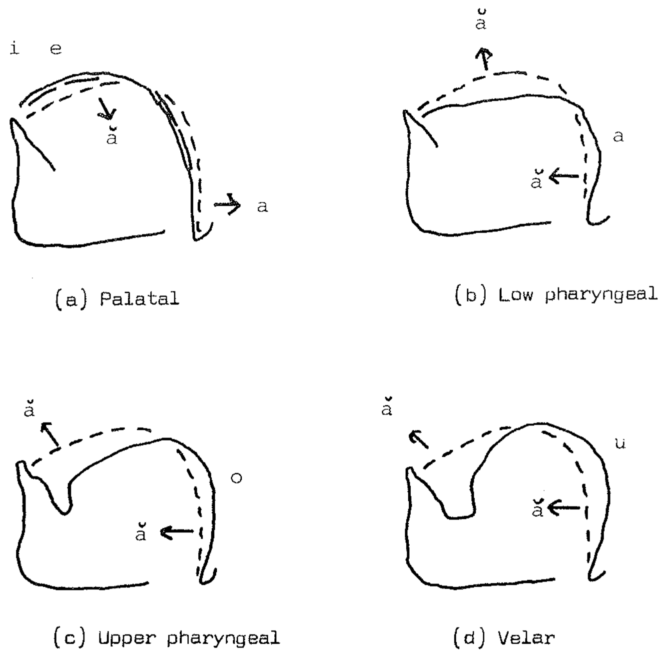


Fig. 5. Comparison of tongue posture of /ă/ and the four basic tongue postures relative to the mandible (cf Figs. 2 and 3) after Stojkov's profiles.

Stojkov (1966) has also published x-ray tracings. The profiles for /ă, a, o/ are reproduced in Fig. 4 and our comparisons of the tongue postures are given in Fig. 5.

The pharyngeal region and the position and attitude of the epiglottis are identical on all of Stojkov's profiles, which indicates that he has only paid attention to the mouth region.

Stojkov's own interpretation of this data is that the tongue is similar for /ă/ and /a/, but somewhat higher and raised anteriorly for /ă/. Figure 4 clearly shows the anterior raising of the tongue for /ă/. Indeed, this profile is more reminiscent of a palatal [ɪ-ɛ] profile rather than a pharyngeal [a] profile. Figure 4 shows a straight back to the tongue for

/a/ with no bulge in the pharynx (typical for palatal vowels).

Firstly, Fig. 5a shows that the Stojkov /ǎ/ posture is very similar to the /i, e/ posture. The tongue is less bunched for /ǎ/ as though it were a lax counterpart to /i, e/ (less activity in the posterior fibres of the genioglossi). This suggests another variant of solution 3.3 above, /ǎ/ as a high (possibly front) vowel similar to lax [ɪ]. This would be a novel interpretation, contrary to the usual view that Bulgarian /ǎ/ is central or back and contrary to the evidence of the Tilkov profiles.

Figure 5b shows no affinity between /ǎ/ and low pharyngeal /a/. The tongue is clearly raised anteriorly for /ǎ/, emphasizing the palatal character just noted.

Similarly, Fig. 5c shows no affinity between /ǎ/ and upper pharyngeal /o/, the tongue being more anterior for /ǎ/. This again points to the palatal character of this particular /ǎ/ profile.

Finally, Fig. 5d also shows an /ǎ/ posture that is more anterior than the velar /u/ posture.

The Stojkov /ǎ/ profile is thus radically different from the Tilkov /ǎ/ profiles. However, it is difficult to know how much confidence to place in Stojkov's profiles in view of his lack of attention to pharyngeal detail.

Acoustical data

Tilkov's (1982 and Tilkov & Bojadziev 1981) and Stojkov's (1966) acoustical F1 and F2 charts are reproduced in Fig. 6.

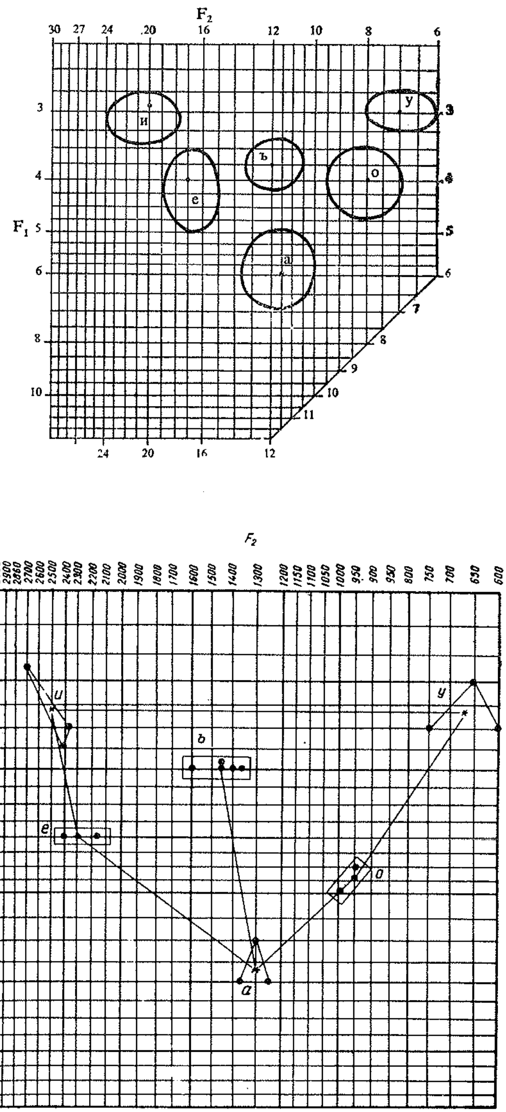


Fig. 6. Acoustical vowel charts from Tilkov & Bojadziev (above) and Stojkov (below).

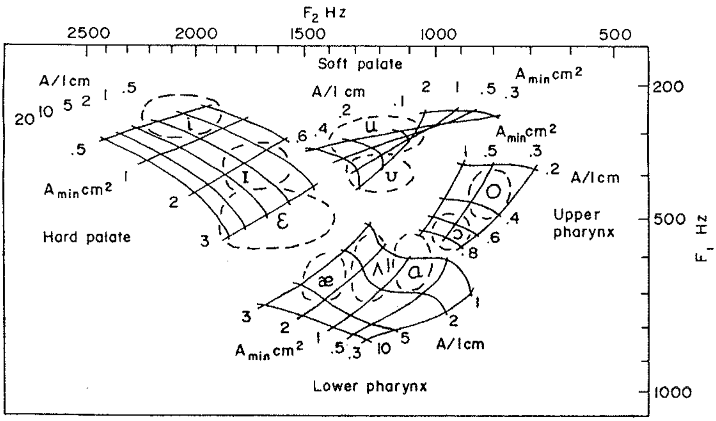
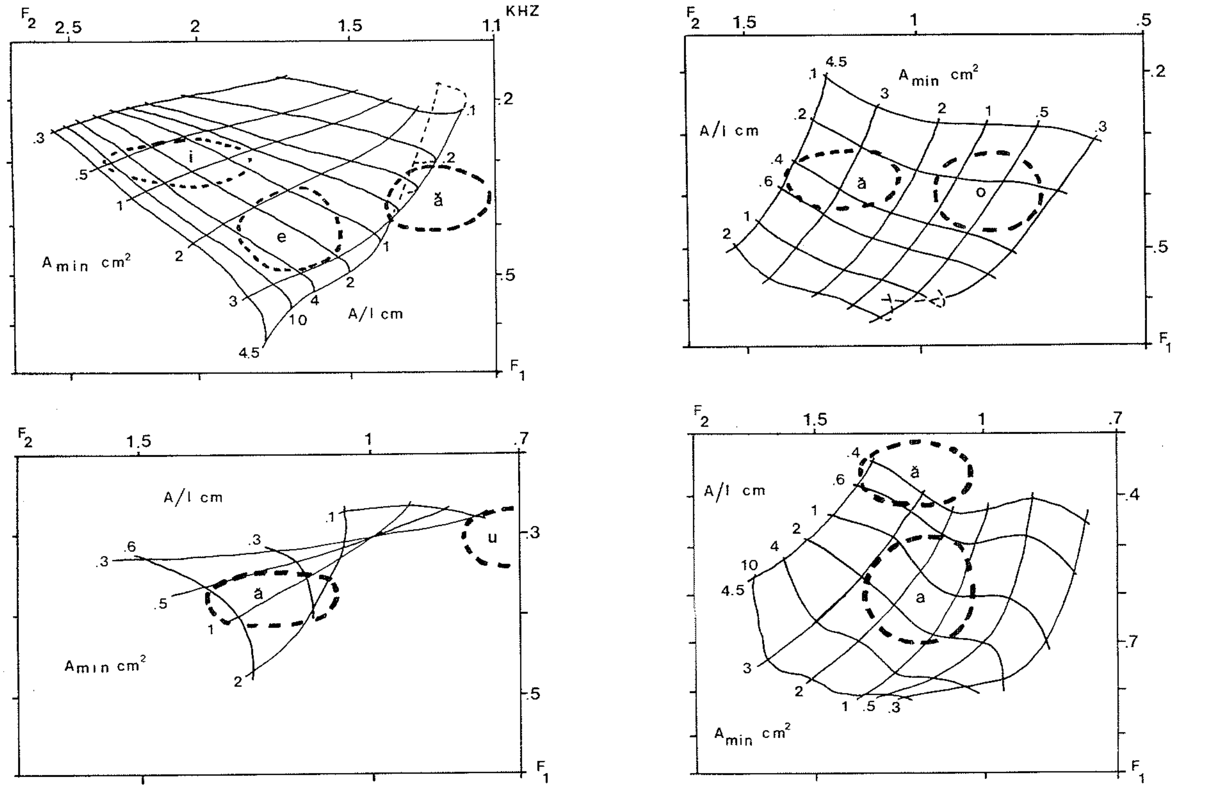


Fig. 7. The frequencies of F1 and F2 generated by the three-parameter model for the four preferred constriction locations, based on nomograms by Stevens & House (1955) (distance from the source to the constriction 12 cm for hard palate, 8.5 cm for soft palate, 6.5 cm for upper pharynx, 4.5 cm for lower pharynx). The superimposed vowel areas are from a sample of Southern British English speech recorded from the radio. From Wood (1979).

It has long been known that judgments of vowel height and backness are more closely correlated with the frequencies of F1 and F2 respectively than with the position of the tongue in the vertical and horizontal planes (see Joos 1948, Lindau 1978), such that high is synonymous with low F1, low with high F1, front with high F2 and back with low F2. The position of /ă/ on the charts reproduced in Fig. 6 is central and midway between /e/ and /o/ (F1 about 350-400 Hz and F2 about 1100-1300 Hz). This would favour solution 3.2 above (/ă/ as a mid central vowel).

A rough articulatory interpretation of these spectra can be obtained by referring them to the Stevens & House (1955) three-parameter model nomograms. Figure 7 shows how the degree of constriction (A_{min} sq cm) and the degree of mouth opening (A/l cm) influence the frequencies of F1 and F2 at the four d_o



(a) Hard palate (above) (b) Soft palate (below) (c) Upper pharynx (above) (d) Lower pharynx (below)

Fig. 8 Comparison of F_1 and F_2 of Bulgarian vowels (Tilkov) and the Stevens and House three-parameter model nomograms for the four basic vowel configurations (cf. Fig. 7) values corresponding to the four relevant constriction locations. In Fig. 8 Tilkov's Bulgarian vowels have been superimposed on the same grid, for comparison.

Firstly, Fig. 8a confirms that the /ǎ/ F1/F2 spectrum is hardly likely to be derived from a palatal configuration. Considerable lip rounding (A/l 0.2-0.5 cm) would be needed to lower F2 of a palatal vowel to below 1500 Hz, whereas /ǎ/ is a spread-lip vowel. This confirms that the Stojkov x-ray profile for /ǎ/ should be taken with caution.

Figure 8b shows that the F1 and F2 of /ǎ/ can be reached from a velar configuration with less rounding than for /u/ (A/l 0.3-0.6 cm against 0.1 cm for /u/) and with a more open velar passage (A_{min} 0.5-2 sq cm). The larger velar opening would be obtained by lowering the tongue body relative to the mandible (which is also a possible interpretation of Tilkov's x-ray data, see Fig. 2d and the discussion above, and represents the first solution 3.3 above). This would make /ǎ/ a spread-lip counterpart to /u/.

Figure 8c shows that the F1 and F2 of /ǎ/ can also be reached by widening a constricted upper pharynx (A/l >2 sq cm would raise F2 beyond 1000 Hz). This is also a possible interpretation of the x-ray data (cf. Tilkov's /ǎ/ and /o/ profiles in Fig. 1 and the discussion above) and represents solution 3.1 above. This would make /ǎ/ a spread-lip counterpart to /o/.

Finally, Fig. 8d shows that the F1 and F2 of /ǎ/ can also be reached from a low pharyngeal configuration by narrowing the mouth opening (A/l <0.6 cm against 0.6-3 cm for /a/), e.g. by not lowering the mandible so far as for /a/, and by widening the constricted lower pharynx (A_{min}>2 sq cm). This is also a possible interpretation of the x-ray data (cf. Tilkov's /ǎ/ and /a/ profiles in Fig. 1) and represents the second solution 3.3 above. This would make /ǎ/ a close (narrower jaw opening) counterpart of open /a/ (larger jaw opening).

5. CONCLUSION

A complete account of phonological vowel reduction is dependent on the analysis of /ǎ/, see for example how the different possible solutions outlined in section 3 affect the formulation of the rules governing the vowel alternations between stressed and non-stressed syllables.

The analysis of the published x-ray data and an articulatory interpretation of the published F1 and F2 frequencies of /ǎ/ yielded several possible solutions: the /ǎ/ configuration can be achieved by modifying any of the three non-palatal configurations and may be related to velar [u], upper pharyngeal [o] or low pharyngeal [a]. Thus, /ǎ/ may be a spread-lip counterpart to rounded [u] or [o] or a close counterpart (narrow jaw opening) to open (large jaw opening) [a]. The solution least favoured by Bulgarians is that based on velar [u] and the solution that is usually preferred is the one related to low pharyngeal [a].

An articulatory interpretation of spectral data based on more than two formants should narrow the choice between possible solutions.

It is our intention to pursue this question further by articulatory analysis of cinefluorographic motion films, by acoustical analysis of spectrographic data and and by computer modelling of individual articulatory manoeuvres.

REFERENCES

Bojadžiev T (1980). Udarenieto i zvukovata struktura na bǎlgarskata duma. In Pǎrvev & Radeva (1980) pp. 200-205. (The stress and sound structure of the Bulgarian word)

Chomsky N & Halle M (1968). The Sound Pattern of English. Harper & Row, New York.

Gyllin R (1982). The traditional normative view of vowel reduction in Bulgarian and its target groups. In Barentsen A A, Sprenger R & Tielemans M G M (eds), South Slavic and Balkan Linguistics, pp. 79-90. Rodopi, Amsterdam

Ivančev S (1980). Rimite na Ivan Vazov i Penčo Slavejkov i redukcijata na glasnite v bălgarskija ezik. In Părvev & Radeva (1980) pp. 249-261. (The rhymes of Ivan Vazov and Penčo Slavejkov and the reduction of vowels in the Bulgarian language)

Janakiev M (1960). Udarnost (akcentuvanost) i bezudarnost (neakcentuvanost). Bălgarsko Stixoznanie, pp. 52-60, Sofia. Also in Părvev & Radeva (1980), pp. 187-194. (Stress (accentuation) and non-stress (non-accentuation); in Bulgarian Metrics)

Joos M (1948). Acoustic Phonetics. Language Monograph No 23. Supplement to Language Vol 24.

Labov W (1972). Language in the inner city: Studies in the Black English Vernacular. University of Pennsylvania Press, Philadelphia.

Lindau M (1978). Vowel features. Language 54:541-562.

Lockwood D G (1972). Markedness in stratificational phonology. In Makkai V B (ed), Phonological Theory, Evolution and Current Practice. Holt, Rhinehart & Winston, New York

Părvev X & Radeva V (1980). Pomagalo po bălgarska fonetika. Nauka i Izkustvo, Sofia. (Manual of Bulgarian phonetics)

Pašov P (1980a). Redukcija na glasnite v săvremennija knižoven bălgarski ezik. In Părvev & Radeva (1980), pp. 230-248. (Vowel reduction in the modern Bulgarian literary language)

Pašov P (1980b). Pravopisni i pravogovorni osobenosti, svărzani s redukcijata na neudarenite glasni. In Părvev & Radeva (1980), pp. 262-268. (Orthographic and orthoepic peculiarities connected with the reduction of the unstressed vowels)

Scatton E A (1975). <u>Bulgarian Phonology</u>. Slavica Publishers, Cambridge, Mass.

Stevens K N & House A S (1955). Development of a quantitative description of vowel articulation and some of its implications. <u>Journal of the Acoustical Society of America</u> 27:484-495.

Stojanov S (1968). Kăm văprosa za zăstava na vokalnata sistema na bălgarskija knižoven ezik. <u>Slavistični izsledvanija</u> pp. 83-89. Nauka i Izkustvo, Sofia. (On the question of the structure of the vowel system of the Bulgarian literary language)

Stojkov S (1966). <u>Uvod văv fonetikata na bălgarskija ezik</u>. Nauka i Izkustvo, Sofia (Introduction to the phonetics of the Bulgarian language)

Tilkov D (1970). <u>Le vocalisme bulgare</u>. Publications de la Société de Linguistique de Paris No. 65. Paris.

Tilkov D (ed) (1982). <u>Gramatika na săvremennija bălgarski knižoven ezik</u>. Tom I: Fonetika. Sofia. (Grammar of the modern Bulgarian literary language, Vol. 1, Phonetics)

Tilkov D & Bojadžiev T (1981). <u>Bălgarska fonetika</u>. Sofia. (Bulgarian phonetics)

Trubetskoy N S (1939). <u>Grundzüge der Phonologie</u>. Travaux du Cercle Linguistique de Prague VII, Prague. English translation by Christiane Baltaxe, University of California Press, 1969.

Wood S (1975a). The weaknesses of the tongue arching model of vowel articulation. <u>Working Papers</u> 11:55-108. Department of Linguistics, Lund.

Wood S (1975b). Tense and lax vowels - degree of constriction or pharyngeal volume? <u>Working Papers</u> 11:110-134. Department of linguistics, Lund. Also in Wood (1982a).

Wood S (1979). A radiographic analysis of constriction locations for vowels. <u>Journal of Phonetics</u> 7:25-43. Also in Wood (1982a).

Wood S (1982a). X-ray and Model Studies of Vowel Articulation. Working Papers 23. Department of Linguistics, Lund.

Wood S (1982b). Radiographic and model studies of the palatal vowels. In Wood (1982a).

This work is sponsored by the Swedish Council for Research in the Humanities and Social Sciences

Working Papers 25 1983
Linguistics-Phonetics
Lund University, 179-194

HOW TO MAKE A TEXT PRODUCTION SYSTEM SPEAK

Bengt Sigurd

Several text production models implemented on computers are able to print grammatical sentences and coherent text, see e.g. Mann & Matthiessen (1983), Allén (1982), Sigurd (1983). It is an interesting task to make such a text production system speak, but few researchers have ventured to solve the additional problems caused by a verbalization system which is intended to speak in a natural way.

Trying to simulate human speech behaviour by computer throws light on the whole production model, particularly if the system is required to have some psychological reality by simulating spontaneous speech with its hesitation pauses, repetitions and errors.

This paper will show how even a simple synthesizer, Votrax, interfaced with the system Commentator, see Sigurd (1983), Fornell (1983) raises interesting questions about the phonetic representation of words, ways of packing sounds into prosodic units, pauses, mistakes in the execution processes and the place of sound laws in a model of a human speaker. The Votrax can pronounce some 60 American English sounds and offers 4 pitch levels; it cannot, of course, be expected to produce high quality Swedish. It is used here for explorative and experimental purposes as is the whole Commentator system.

The problem of producing naturally sounding speech is also attested in experiments with text-to-speech systems, see Carlson & Granström (1978). Such systems, however, have printed text as input which creates additional problems. They have to be able to derive a phonetic transcription from the printed version, which e.g. means pronouncing th, ng and sh, as single sounds, deriving stress in words such as export, which is very difficult and in fact requires comprehension of the text.

COMMENTATOR'S INPUT TO A SPEECH PRODUCTION SYSTEM

The general outline of the Commentator is presented in fig. 1 after Sigurd (1983). The input to this model of verbal production is perceptual data or equivalent coordinate values, e.g.information about persons and objects on a screen. These primary perceptual facts constitute the basis for various calculations in order to derive secondary facts and conclusions about movements and relations such as distances, directions, right/left, over/under, front/back, closeness, goals and intentions of the persons involved. The Commentator produces comments consisting of grammatical sentences making up coherent and well-formed (although often boring) text. Some typical comments are shown below.

A question menu, different for different situations, suggests topics leading to propositions which are considered appropriate under the circumstances and their truth values are tested against the primary and secondary facts of the world known to the system. If a proposition is found to be true, it is accepted as a protosentence and verbalized by various lexical, referential, syntactic and textual subroutines. If e.g. the proposition CLOSE (ADAM, EVE) is verified after measuring the distance between the two referents Adam and Eve and comparing it with the standard for closeness between human beings, the lexical subroutines try to find out how closeness should be expressed in the language (Commentator has mainly produced Swedish), referential subroutines determine whether pronouns could be used instead of the proper names of the persons and textual procedures investigate whether connectives such as however, also or perhaps contrastive stress should be inserted. In the printing version of Commentator all the words in the sentence are packed into a string before they are printed.

The Commentator can deliver words one at a time whose meaning, syntactic and textual functions are well-defined through the verbalization processes. For the printing version of Commentator these words are characterized by whatever markers are needed to get correct printing (spelling, word spaces, punctuation marks). The input to a speaking version of Commentator must be a phonetic transcription, but the details of this and whatever additional information is required in order to produce prosodic units with the proper intonation and stress patterns has to be discovered by empirical and experimental studies and work with models such as the one presented in this paper.

Lines	Component	Task	Result (sample)
10-35	Primary information	Get values of primary dimensions	Localization coordinates
100-140	Secondary information	Derive values of complex dimensions	Distances, right-left, under-over
152-183	Focus and topic planning expert	Determine objects in focus (referents) and topics according to menu	Choice of subject, object and instructions to test abstract predicates with these
210-232	Verification expert	Test whether the conditions for the use of the abstract predicates are met in the situation (on the screen)	Positive or negative protosentences and instructions for how to proceed
500	Sentence structure (syntax) expert	Order the abstract sentence constituents (subject, predicate, object); basic prosody	Sentence structure with further instructions
600-800	Reference expert (subroutine)	Determine whether pronouns, proper nouns, or other expressions could be used	Pronouns, proper nouns, indefinite or definite NPs
700-	Lexical expert (dictionary)	Translate (substitute) abstract predicates, etc.	Surface phrases, words
900	Sentence connection (textual) expert	Insert conjunctions, connective adverbs; prosodic features	Sentences with words such as *ocksa* (too), *dock* (however)
1000	Phonological (pronunciation, printing) expert	Pronounce or print the assembled structure	Uttered or printed sentence (text)

Fig 1 Components of the text production model underlying Commentator

A SIMPLE SPEECH SYNTHESIS DEVICE

The fundamental difference between a printing and a speaking verbal production model must, of course, be that the words of the speaking version have to be coded in some phonetic transcription, which serves as instructions for a speech synthesis device simulating human articulation. The transcription depends on the speech synthesis system to be used.

The experimental system presented in this paper uses a Votrax speech synthesis unit (for a presentation see Giarcia, 1982). Although it is a very simple system designed to enable computers to deliver spoken output such as numbers, short instructions etc, it has some experimental potentials. It forces the researcher to take a stand on a number of interesting issues and make theories about speech production more concrete. The Votrax is an inexpensive and unsophisticated synthesis device and it is not our intention to achieve perfect pronunciation using this circuit, of course. The circuit, rather, provides a simple way of doing research in the field of speech production.

Votrax simulates the human vocal apparatus by means of a harmonic source (for vowels and voiced consonants) and a noise source (for voiceless consonants), supplemented with filters. Information about the sounds to be produced is given according to the LPC technique. Votrax (which is in fact based on a circuit named SC-01 sold under several trade names) offers a choice of some 60 (American) English sounds (allophones) and 4 pitch levels. A sound must be transcribed by its numerical code and a pitch level, represented by one of the figures 0,1,2,3. The pitch figures correspond roughly to the male levels 65,90,110,130 Hz. Votrax offers no way of changing the amplitude or the duration, but choice of pitch level as well as choice between long/short and stressed/unstressed sounds (allophones) can be made with some success.

Votrax is designed for (American) English and if used to synthesize other languages it will, of course, add an English flavour. It can, however, be used at least to produce intelligible words for several other languages. Of course, some sounds may be lacking, e.g. Swedish <u>u</u> and <u>y</u> and some sounds may be slightly different, as e.g. Swedish <u>sh</u>-, <u>ch</u>-, <u>r</u>-, and <u>l</u>-sounds.

Long, short and stressed, unstressed variants are found for some vowels, and if used with some ingenuity the inventory of sounds may serve fairly well for several languages provided the foreign accent and the robot voice are accepted.

English model words containing the sounds and information about the duration of the sounds are given in the manual, but this information is certainly not sufficient for the phonetician. The synthesis device seems to have been constructed for technical purposes without consulting a phonetician or linguist and the presentation of the device is very unsystematic from the point of view of a language professional.

Most Swedish words can be pronounced intelligibly by the Votrax. The pitch levels have been found to be sufficient for the production of the Swedish word tones: accent 1 (acute) as in and-en (the duck) and accent 2 (grave) as in ande-n (the spirit). (For details about the word accents see Gårding, 1977). Accent 1 can be rendered by the pitch sequence 20 and accent 2 by the sequence 22 on the stressed syllable (the beginning) of the words. Stressed syllables have to include at least one 2. Word tones and stress are necessary characteristics of Swedish words and must be given in the lexicon.

Words are transcribed in the Votrax alphabet by series of numbers for the sounds and their pitch levels. The Swedish word höger (right) may be given by the series 27,2,58,0,28,0,35,0, 43,1, where 27,58,28,35,43 are the sounds corresponding to h,ö:,g,e,r, respectively and the figures 2,0, etc after each sound are the pitch levels of each sound. The word höger sounds American because of the ö, which sounds like the (retroflex) vowel in bird, but it is assigned the proper accent 1 by the sequence 20. The 1 on one of the following unstressed syllables is introduced in order to get some variation. No detailed systematic studies of the effects of various combinations of sounds and pitches have been undertaken and the Votrax has not been evaluated properly by phoneticians to my knowledge.

The pronunciation (execution) of the words is handled by instructions in a computer program, which transmits the information to the sound generators and the filters. Some programming details will be given below, but the technical details are outside the scope of this paper.

PAUSES AND PROSODIC UNITS IN SPEECH

The spoken text produced by human beings is normally divided by pauses into parts of several words (prosodic units). There is no generally accepted theory explaining the location and duration of the pauses and the intonation and stress patterns in the prosodic units. Many observations have, however, been made, see Dechert & Raupach (1980).

The printing version of Commentator collects all letters and spaces into a string before they are printed. A speaking version trying to simulate at least some of the production processes cannot, of course, produce words one at a time with pauses corresponding to the word spaces, nor produce all the words of a sentence as one prosodic unit. A speaking version intended to have some psychological reality must be designed to be able to produce prosodic units including 3-5 words, cf Svartvik (1982) and lasting 1-2 seconds, see Jönsson, Mandersson & Sigurd (1983). How this should be achieved may be a called chunking problem. It has been noted that the chunks of spontaneous speech are generally shorter than in text read aloud.

The text chunks have internal intonation and stress patterns often described as superimposed on the words. That is why they may be called prosodic units. Deriving these internal prosodic patterns may be called the intra-chunk problem. We may also talk about the inter-chunk problem having to do with the relations e.g. in pitch, between succesive chunks.

For the purpose of simulating at least some of these features by Votrax we will touch upon these problems and discuss different ways to control chunking and pause placement.

As human beings need to breathe they have to pause in order to inhale at certain intervals. The need for air is generally satisfied without conscious actions. We estimate that chunks of 1-2 seconds and inhalation pauses of about 0.5 seconds allow convenient breathing. Clearly, breathing allows great variation. Everybody has met persons who try to extend the speech chunks and minimize the pauses in order to say as much as possible, or to hold the floor.

It has also been observed that pauses often occur where there is a major syntactic break (corresponding to a deep cut in the syntactic tree), and that, except for so called hesitation pauses, pauses rarely occur between two words which belong closely together (corresponding to a shallow cut in the syntactic tree). There is no support for a simple theory that pauses are introduced between the main constituents of the sentence and that their duration is a function of the depth of the cuts in the syntactic tree. The conclusion to draw seems rather to be that chunk cuts are avoided between words which belong closely together. Syntactic structure does not govern chunking, but puts constraints on it. Click experiments which show that the click is erroneously located at major syntactic cuts rather than between words which are syntacticly coherent seem to point in the same direction. As an illustation of syntactic closeness we will mention the combination of a verb and a following reflexive pronoun as in Adam avlägsnar+sig från Eva. ("Adam distances himself from Eva"), one of the sentences which Commentator often produces. Cutting between avlägsnar and sig would be most unnatural. All other places seem acceptable, although for different reasons. A cut before avlägsnar or Eva would seem to reflect the search for the proper word.

Lexical search, syntactic and textual planning are often mentioned as the reasons for pauses, so called hesitation pauses, filled or unfilled. In the speech production model envisaged in this paper sounds are generally stored in a buffer where they are given the proper intonational contours and stress patterns. The pronunciation is therefore generally delayed to allow context adjustments and various prosodic operations. This delay also offers ways of explaining speech errors. The length of the delay varies.

Hesitation pauses seem, however, to be direct (on-line) reflexes of searching or planning processes and at such moments there is no delay. Whatever has been accumulated in the articulation or execution buffer is pronounced and the system is waiting for the next word. While waiting (idling) some human beings are silent, others prolong the last sounds of the previous word or produce sounds, such as ah, eh, or repeat part of the previous utterence. Hesitation pauses may occur anywhere, but as has been observed they seem to be more frequent before lexical words than function words.

As to the internal prosodic patterns within chunks, it has been observed that sentence final sounds often are prolonged and that the final pitch is low in declarative sentences. Non-sentence final chunks may often also have final lengthening and rising pitch signaling incompleteness. As the Votrax only allows the manipulation of pitch, not duration, we will not go into more detail here (for details of the Nordic Languages see Gårding, Bruce & Bannert, 1978).

In the approach to be presented in detail below we will demonstrate a two buffer model and a chunking mechanism based on the length of the chunk and syntactic structure. It allows natural breathing and avoids unnatural cuts. After the presentation of this approach we will discuss variants and models where other factors may influence the final chunking and pausing in speech.

A TWO BUFFER MODEL OF SPEECH PRODUCTION

The approach proposed in this paper presupposes two buffer memories and trigger mechanisms for filling, emptying and reading these buffers. The buffers work in series and are called the current buffer and the execution buffer. The operation of such a system will be illustrated by using the length of the chunck required as the primary trigger. Words are assumed to be classified as stressed or unstressed and by this difference the text will be divided into chunks generally containing one or several stressed syllables and a greater number of unstressed syllables. These chunks are then given a rising or falling final pitch contour. A computer implementation of this text division method will be discussed in some detail below.

As only pitch, not intensity, is available in Votrax, pitch must be used to signal stress. Unstressed words are assigned pitch level 1 or lower, stressed words get 2 or higher on at least one segment. Words are assumed to be inherently stressed or unstressed as given in the lexicon and the division is assumed to coincide roughly with the division into lexical and grammatical (functional) words often mentioned. In the restricted vocabulary of Commentator the following illustrate lexically stressed words: Adam, Eve, vänster (left), höger (right), porten (the gate), nära (close, near), närma (sig) (approach), också (too), heller (either, neither). The following words are lexically unstressed: han (he), hon (she), honom (him), henne (her), den (it), det (it), om (if), i (in), till (to), och (and), men (but), är (is). Inherently unstressed words may become stressed, e.g. by contrastive stressing, and stressed words may become unstressed.

As described above the verbalization processes of Commentator produces one word at a time and for each word delivered to the first buffer (the current buffer) or the second buffer (the execution buffer) a number of variables may be checked. If a sentence termination is signaled, whatever has been accumulated in the execution buffer is executed (pronounced) and the pitch of the last segment is set at low. If the number of the segments in the chunk being accumulated in the execution buffer does not exceed a certain limit a new word is only stored after the others in the execution buffer. The duration of a sound in Votrax is 0.1 second on the avarage. If the limit is set at 15 the system will deliver chunks about 1.5 seconds, which is what we want. The length of the chunks can be preset in order to simulate different individuals, speech styles or speech disorders.

If the number of segments in the execution buffer exceeds the limit the system proceeds to find out whether there is a tight syntactic link between the last word and the following. Such links (syntactic coherence) is signaled through the process. If not the cut is made after the last word in the buffer and the buffer is pronounced with a rising pitch on the last sounds.

The short sentences produced by Commentator seem to require short chunks. If the limit is set at 15 we would get the following result (L= low pitch, H= high). The number of segments is also given.

1. ADAM Ä(R) TI(LL) VÄNSTER OM (H)(16)	Adam is to the left of
2. EVA.(L) (3)	Eve
3. HON Ä(R) TI(LL) HÖGER OM/	She is to the right of/
PÖGER OM PORTEN/HORTEN.(L)(19)	gight of the gate/the rate.
4. ADAM AVLÄGSNAR SIG (H)(16)	Adam is going away
5. FRÅN EVA.(L)(7)	from Eve.
6. HAN AVLÄGSNAR SIG(H)(15)	He goes away
7. FRÅN PORTEN OCKSÅ.(L)(14)	from the gate too.

The typical speech errors indicated by the slashes will be commented on below.

EXPLAINING SPEECH ERRORS AND SOUND CHANGE

Speech errors may be classed as semantic, lexical, syntactic or phonetic. Semantic errors can be explained in the Commentator model as errors in verifying a proposition, e.g. estimating the closeness when the proposition CLOSE (ADAM, EVE) is to be tested. Lexical errors can be explained as mistakes in picking up the address of a lexical item. Instead of picking up höger (right) the word vänster (left, a semiantonym) stored on an adjacent address is picked up and sent to the current buffer. Syntactic mistakes may occur as a result of mixing the contents of memories used for storing different constituents, forgetting structural conditions etc. Of course various other explanations, including Freudian associations may also have to be evoked.

Phonetic errors, which are of greatest interest in this paper may be context-free (spontaneous) or context-sensitive. Context-free errors consist of substitutions which do not depend on any features in the context. Context-sensitive errors are traditionally divided into 1. progressive, 2. regressive and 3. inversions (metathesis). These cases can be explained in our model if we assume the two buffers and some simple operational mistakes in handling them.

If a speaker says pöger instead of höger in the text illustrated we would have a regressive error. The p of porten is said to influence its phonotactic equivalent in the preceding stressed word of the text. As Merringer noted already in 1895, errors generally concern corresponding elements in stressed or unstressed syllables: the initial consonantal constituent of a stressed syllable is substituted for the initial consonant constituent of another stressed syllable etc. Substituting the initial parts of stressed syllables seems particularly frequent. If the words are coded accordingly our model, where long strings of sounds are stored before pronunciation, offers simple ways of explanation.

In terms of our model the regressive mistake may be explained as a case where a sound of the buffer is substituted or picked up too soon for pronunciation.

If a speaker says horten instead of porten in the text illustrated, we would say that he has made a progressive error. In terms of our model the speaker has then replaced a later consonant in the buffer.

If a person says pöger om horten, he is said to have produced an inversion of h and p. In terms of our model such a mistake may come about if the two mistakes mentioned are both at work.

Models of speech errors often assume mistakes in reading buffer memories, although they do not generally describe the procedure as detailed as we have done, which in fact allows us to simulate speech errors by a computer program. The serial two buffer model seems to have advantages compared to the competing parallel buffers asumed by Baars (1980). Although several processes may go on at the same time,cf Kempen & Hoenkamp (1982) it seems less natural to assume that the same activity e.g. pronunciation is carried out tentatively with two (or several) processes using several buffer memories, whose contents are then sometimes erroneously mixed.

Most explanations of speech errors assume an unconscious or a conscious monitoring of contents of the buffers used during the speech process. This monitoring may also result in sound changes as the speaker may want to adjust his pronunciation to a new norm. There are several places in the Commentator where sound changes may be introduced. Some changes may correspond to changes in the hardware (Votrax). Changes may be brought about in the execution buffer, deleting unstressed vowels, etc. Changes may be brought about in the execution buffer as a result of monitoring comparing its contents to norms in ways resembling word processing systems which apply automatic spelling correction. Sound changes may be brought about by changing the phonetic representation of one or several words in the lexicon (lexical diffusion).

The chunking shown above often attaches unstressed words to the end of stressed words. It is natural to associate this enclitic buffer procedure with enclitic processes characteristic of language change. Our 2-buffer model attaches unstressed pronouns, auxilliaries and particles to the preceding word (deleting word boundaries). Historical enclitic processes often consist of attaching such function words, reducing their shape and sometimes turning them into inflectional morphemes. The definite articles are well-known examples from the Nordic languages. The development of the passive in the Nordic languages is also a result of enclitic attachment of the reflexive pronoun to the verb: kallas<kallask<kalla-sik<kalla sig. In our sample text nära den may result in the enclitic form nära-n and vänster om in vänstrom. The place and domain of enclitic phonological reduction or deletion rules seems to be the executive buffer.

The idea of packing words in a buffer before they are pronounced is reminiscent of an interesting proposal put forward by Lindblom et al. (1976) in order to explain the inverse relation between segment duration and number of segments in the utterance. Lindblom assumes that segments are packed in a buffer where they are compressed before being uttered. The more segments inserted in the buffer the more they are compressed. This holds for all segments except the last, which keep their inherent duration and therefore seem to be longer than the others. According to Lindblom one should therefore talk about non-final shortening instead of final lengthening.

Lindblom associates his buffer with short term memory, a central concept in psychology. The capacity of his buffer of pronunciation is assumed to be a few syllables which, however, seems to be below the size generally assumed for short term memory.

COMPUTER IMPLEMENTATION

The ideas presented above have been implemented on a microcomputer as an addition to the Commentator program. The Commentator program will not be presented here, as it can be found elsewhere, Sigurd (1983) and Fornell (1983). We will limit ourselves to presentation of the fragment relevant to pronunciation (see Appendix).

The words are represented as data, illustrated in lines 3001-3014. The first figure gives the number of items, and the following pairs of numbers are the code numbers of the speech sounds according to the Votrax manual and the pitch required for that sound (0-3). The pitch (stress) pattern of a word can be identified by reading every other figure starting from the third.

Lines 2005, 2020 illustrate how the lexical numbers given by the Commentator to the variable U make different data lines available for reading. Lines 2050-2070 read the sound and pitch numbers into the current buffer (B1). Line 2064 raises the pitch of a pitch segment, if the contrastive flag (C6) has been set previously by the verbalizing program. Line 2068 sets the flag if a full stop (62) has been entered into B1. Line 2063 shows how a sound change L1>L2 may be introduced in the system.

Line 2075 checks whether the sentence has ended. If not the content of the current buffer is passed on to the executive buffer by the subroutine 2080-2090. If the number of segments in the execution buffer (Ö) exceeds the preset limit (C9) the variable marking syntactic adherence has to be checked (2078-9). If cutting is allowed the execution buffer is sent to pronunciation. Lines 2100-2115 place sound numbers in S and pitch numbers in P. Line 2117 adds a switch-off signal (63). If the last segment number is found to be 62 (sentence termination, full stop) the pitch of the last sound is lowered, otherwise raised (2120). Line 2122 shows another place for sound change. Lines 2200-2228 handle the pronunciation through the interface and the Votrax.

Speech errors may be simulated by copying a sound of one buffer into the other buffer, interchanging within a buffer or reading mistakes as described before.

The resulting pronunciation gives intelligible words and prosodic chunks, but the interchunk and intrachunk prosodic quality is far from perfect. The prosodic chunks and the pauses may be given variable duration simulating more or less hesitant speakers. The pitch levels may be varied and different triggers set during the verbalization process may be used to control pitch. After some experimentation we know that it is easy to simulate many kinds of speech disorders. But it is indead very difficult to simulate normal speech.

ACKNOWLEDGEMENT

I am endebted to Gösta Bruce for discussions of some of the issues of this paper. Gustav Jönsson och Bengt Mandersson have assisted in interfacing the hardware, Jan Fornell in designing the software.

APPENDIX

```
2000 IF U%>20% THEN U%=U%-20% : GOSUB 2020 : RETURN
2001 IF U%>10% THEN U%=U%-10% : GOSUB 2010 : RETURN
2005 ON U% RESTORE 3001,3002,3003,3004,3005,3006,3007,3008,3009,3010 : GOSUB 205
0 : RETURN
2010 ON U% RESTORE 3011,3012,3013,3014,3015,3016,3017,3018,3019,3020 : GOSUB 205
0 : RETURN
2020 ON U% RESTORE 3021,3022,3023,3024,3025,3026,3027,3028,3029,3030,3031 : GOSU
B 2050 : RETURN
2050 READ N% : REM FILL B1(CURRENT BUFFER)
2060 FOR J%=1% TO N% STEP 2%
2062 READ B1%(J%),B1%(J%+1%)
2063 IF B1%(J%)=L1% THEN B1%(J%)=L2%
2064 IF C6%=1% THEN B1%(J%+1%)=B1%(J%+1%)+1% : C6%=0% : REM INCREASE PITCH IF CO
NTRAST VARIABLE SET (C6%)
2068 IF B1%(J%)=62 THEN P1%=1 : REM FLAG IF STOP IN B1
2070 NEXT J% : READ A$ : S5$=S5$+A$
2072 I=RND : IF R5<I THEN B2%(1)=B1%(1) : REM  REGRESSIVE SPEECH ERROR SIMULATIO
N
2074 GOSUB 2080 : REM B2 FILL
2075 IF P1%=1 THEN GOSUB 2100 : P1%=0 : RETURN : REM PRONOUNCE IF STOP IN B1
2078 IF Ö%<C9 THEN RETURN : REM IF NOT TOO LONG CHUNK RETURN
2079 IF A1%=1 THEN A1%=0% : RETURN ELSE GOSUB 2100 : RETURN : REM IF SYNTACTIC A
DHERENCE RETURN ELSE CHUNK
2080 FOR J%=1% TO N% : REM FILLS B2 FROM B1 WHICH IS EMPTIED
2082 Ö%=Ö%+1% : B2%(Ö%)=B1%(J%) : B1%(J%)=0%
2085 NEXT J% : B1%=0%
2090 RETURN
2100 FOR J%=1% TO Ö% STEP 2 : REM READS NUMBERS INTO SOUND(S) AND PITCH(P) REGIS
TERS
2105 V%=V%+1%
2106 S%(V%)=B2%(J%) : P%(V%)=B2%(J%+1) : B2%(J%)=0% : B2%(J%+1%)=0%
2115 NEXT J% : Ö%=0%
2117 I=RND : IF I>F9 THEN S%(V%+1%)=63 : REM SWITCHOFF UNLESS FLOORHOLDER
2120 IF S%(V%)=62 THEN P%(V%-1%)=0 ELSE P%(V%-1%)=2 : REM LOWERING PITCH AT END
OF SENTENCE(62) ELSE RAISING
2122 IF S%(V%)=L1% THEN S%(V%)=L2%
2200 OUT 1,133 : FOR J%=1 TO V%+1% : REM UTTAL
2210 K%=INP(0) : IF K%<>254% THEN 2210
2215 OUT 0%,S%(J%),2%,P%(J%)
2217 IF S9=1 THEN ; S%(J%)":"P%(J%),
2220 OUT 3,1 : OUT 3,0 : OUT 3,1
2225 NEXT J% : B2%=0%
2228 V%=0% : FOR Z=1 TO P5% : NEXT Z : RETURN
3001 DATA 8,14,2,53,2,30,1,49,0 ,    " BÅDA" :    REM LEXIKON
3002 DATA 6,27,0,50,1,13,0," HAN "
3003 DATA 8,21,2,30,3,50,1,12,0  ," ADAM "
3004 DATA 6,27,0,22,1,13,0 ," HON "
3005 DATA 6,6,2,15,3,50,1,"  EVA "
3006 DATA 2,0,0," ÄR "
3007 DATA 4,42,0,9,0," TILL "
3008 DATA 10,27,2,58,0,28,0,35,0,43,1," HÖGER "
3009 DATA 18,21,2,15,2,24,1,47,0,28,0,31,0,13,1,50,0,43,1," AVLÄGSNAR "
3010 DATA 4,31,0,6,1," SIG "
3011 DATA 8,29,0,43,1,52,1,13,0," FRÅN "
3012 DATA 14,13,2,47,2,43,0,12,1,50,0,31,0,42,1," NÄRMAST "
3013 DATA 14,15,2,2,0,13,0,31,1,42,0,35,1,43,1," VÄNSTER "
3014 DATA 12,13,2,47,1,43,0,12,1,50,0,43,1," NÄRMAR "
```

REFERENCES

Allén, S. (ed) 1983. Text processing. Nobel symposium. Stockholm: Almqvist & Wiksell

Baars, B.J. 1980. The competing plans hypothesis. In: Dechert & Raupach (1980), pp 39-49

Carlsson, R. & B. Granström. 1978. Experimental text-to-speech system for the handcapped. JASA 64, p 163

Ciarcia, S. 1982. Build the Microvox Text-to-speech synthesizer. Byte 1982:Oct

Dechert, H.W. & M. Raupach (eds) 1980. Temporal variables in speech. The Hague: Mouton

Enkvist, N-E. (ed) 1982. Impromptu speech: symposium. Åbo: Åbo akademi

Fornell, J. 1983. Commentator, ett mikrodatorbaserat forsknings-redskap för lingvister. Dep of Linguistics. Lund (mimeo-graphed)

Fromkin, V.A. (ed) 1973. Speech errors as linguistic evidence. The Hague: Mouton

Fromkin, V.A. (ed) 1980. Errors in linguistic performance: slips of the tongue, ear and hand. New York: Academic Press

Gårding, E. 1977. The Scandinavian word accents. Travaux de l'institut de linguistique de Lund. Lund: Liber

Gårding, E., G. Bruce & R. Bannert. (eds) 1978. Nordic Prosody. Travaux de l'institut de linguistique de Lund. Lund: Liber

Jönsson, K-G, B., Mandersson & B. Sigurd. 1983. A microcomputer pausemeter for linguists. In: Working Papers 24. Lund. Department of linguistics

Kempen, G & E. Hoenkamp. 1982. Incremental sentence generation. Proc. ninth Intern. Conf, on computational linguist. Prague

Lindblom, B., B. Lyberg & K. Holmberg. 1976. Durational patterns of Swedish phonology. Do they reflect short-term motor memory processes? Department of Linguistics. Stockholm

Mann, W.C. & C. Matthiessen. 1982. Nigel:a systemic grammar for text generation. Information sciences institute. USC. Marina del Ray. ISI/RR-83-105

Meringer, R. 1895. Versprechen und Verlesen: Eine psychologisch-linguistische Studie. Stuttgart: Goeschen

Sigurd, B. 1982. Text representation in a text production model. In: Allén (1982)

Sigurd, B. 1983. Commentator: A computer model of verbal pro-duction. Linguistics 20-9/10 (to appear)

Svartvik, J. 1982. The segmentation of impromptu speech. In Enkvist, N-E. (1982)

Working Papers 25 1983
Linguistics-Phonetics
Lund University, 195-211

ACOUSTIC ANALYSIS OF CHINESE FRICATIVES AND AFFRICATES

Jan-Olof Svantesson

The purpose of this investigation is to find a method of analyzing fricatives (and the fricative phase of affricates) which makes it possible to describe the available "fricative space" of a language, and to apply this method to Standard Chinese.

For this purpose parameters for characterizing fricative spectra are suggested and applied to the Chinese fricatives. Using these parameters the Chinese fricatives (and fricative components of affricates) as said by different speakers can be compared, and also the totality of the Chinese fricatives can be compared to those of other languages, such as Arabic, which has been investigated with the same methods (Norlin 1983). A comparison with the Swedish fricatives is also made, although this is difficult due to the lack of comparable data. Nartey 1982 uses a similar method of analyzing fricative spectra but the Chinese fricatives are not analyzed by him.

The fricatives and affricates in Standard Chinese

Standard Chinese has six fricatives and six affricates:

f	s	sh [ʂ]	x [ɕ]	h [x]
		r [ʐ]		
	z [ts]	zh [tʂ]	j [tɕ]	
	c [tsh]	ch [tʂh]	q [tɕh]	

The official pīnyīn transcription is used (with IPA symbols in brackets if they differ).

There are restrictions on the possible combinations of consonants and a following vowel, so that only these combinations occur:

	___ i [i]	___ i[ɿ] ___ i[ʅ]	___ ü[y]	___ u	___ uV	___ V (other vowel)
s sh r z c zh ch	−	+	−	+	+	+
x j q	+	−	+	−	−	−
h	−	−	−	+	+	+
f	−	−	−	+	−	+

The palatals (x, j and q) are thus in complementary distribution with h and f, and in nearly complementary distribution with s, sh, etc., the only contrasts in the phonemization implied by the pīnyīn transcription being before i. The allophones of i after the fricatives and affricates are:

xi[ɕi]	si[sɿ]	shi [ʂʅ]	ri [ʐʅ]
ji[tɕi]	zi[tsɿ]	zhi [tʂʅ]	
qi[tɕhi]	ci[tshɿ]	chi [tʂhʅ]	

Thus it is possible to regard [ɿ] and [ʅ] as allophones of a phoneme which is distinct from i, and to let the palatals be allophones of one of the other series.

On the other hand, there are pairs of syllables such as:

sa [sa]   xia [ɕia]~[ɕa]
sao [sao]   xiao [ɕiao]~[ɕao]

where the i written in the transcription can be regarded as a vowel onglide conditioned by the palatal, so that these pairs may be seen as minimal pairs for [s] and [ɕ] . Zhāng et al. 1982 have shown that there is less perceptual confusion between s and x than between s and sh although, as shown below, the acoustic difference between s and x is smaller than that between s and sh. Obviously the different environments of s and x make perception easier.

Historically the palatals have developed from alveolars and velars in palatalizing environments:

[*s], [*x]> x [ɕ]
[*ts],[*k]> j [tɕ]
[*tsh], [*kh]> q [tɕh]

Procedure

The acoustic properties of the fricatives and affricates were investigated using four informants, who said each sound twice. The fricatives and affricates were said in word-initial position in a sentence frame, so that they were preceded and followed by the vowel [a] . The palatals (x, j and q), which do not occur before a were said before ia (see above).

Four male native speakers of Standard Chinese were used. Two of them (C and D) were born and grew up in Běijīng, one (B) was born in Jiāngsū province and moved to Běijīng in his sixth year and the fourth (A) was born in Liáonīng in northeast China and moved to Běijīng in his teens. All speak Standard Chinese with Běijīng pronunciation.

The recordings were made in sound-treated rooms in Lund and in Stockholm.

Analysis

Fourier transform (FFT) spectra were made in the middle of each fricative and in the fricative components of the affricates. The sampling frequency was 20 kHz, and the sampling time 26.5 ms. By running the tape at half speed when sampling, spectra in the frequency range up to 10 kHz could be made.

The FFT spectra were converted into critical band spectra according to Schroeder et al. 1979. The critical bands were determined by Schroeder's formula

$$f_n = 650 \sinh(n/7)$$

where f_n is the upper boundary of band n.
The spectrum is thus described as 24 bands, with bandwidths of approximately 100 Hz below 500 Hz, and of about 1/6 of the

center frequency above 1000 Hz. The critical bands computed by this formula are:

n	f_n	n	f_n
1	93	13	2031
2	188	14	2357
3	287	15	2732
4	392	16	3136
5	505	17	3658
6	628	18	4228
7	764	19	4884
8	915	20	5640
9	1086	21	6512
10	1278	22	7516
11	1497	23	8674
12	1746	24	10010

The mean level (in dB) within each critical band was estimated, and the spectra were redrawn as histograms with each critical band represented as a bar with constant breadth, and with the baseline at -30 dB. This amounts to a rescaling of the spectra to an auditively more correct form, since each critical band corresponds to an equal distance (1.5 mm) on the basilar membrane, or to 1200 primary auditory nerv fibres (Schroeder et al. 1979)

For practical reasons, only bands 2-24 were used.

For the characterization of the spectra of the different fricatives, we propose to use the center of gravity and the dispersion of the critical band spectra, and also the mean intensity level (in dB), as computed by the following formulas:

center of gravity $$m = \sum_{n=2}^{24} n \cdot 10^{(x_n/10)}/F$$

dispersion $$s = \left(\sum_{n=2}^{24} (n-m)^2 \cdot 10^{(x_n/10)}/F\right)^{1/2}$$

mean intensity level $$\bar{x} = 10 \log (F/23)$$

where $$F = \sum_{n=2}^{24} 10^{(x_n/10)}$$

and x_n is the mean level (in dB) in band n, as estimated from the FFT spectra.

The center of gravity is a measure of the overall pitch level of the spectrum, and the dispersion can be regarded as a measure of its flatness.

Results

Figure 1 shows oscillograms of the sound waves, Fourier transform spectra in linear scale, the same spectra in logarithmic units (dB), and critical band spectra of each fricative.

Table 1 gives the centers of gravity of the critical band spectra (measured in critical band units, and also given in Hz), the dispersion, and the mean intensity level (in dB). The mean intensities are given as deviation from the average for each series of fricatives (read at the same occasion), in order to make them roughly comparable also between different speakers.

Figure 1 shows that s and x in Chinese are characterized by a high peak in the upper part of the spectra (bands 20-23 = 5-9 kHz). They differ by s having a more abrupt fall than x down to the level in bands 17-19 (3-5 kHz). This is reflected in the lower centers of gravity and greater dispersions of x than of s (Table 1). The s of speaker D differs from the other s by having a lower center of gravity and much greater dispersion.

In the spectra of sh, the higher levels of the spectra go further down (to bands 14-15 = 2-3 kHz), and the high level area is relatively large and flat. Thus the center of gravity falls further down, and the dispersion increases compared to s and x.

The intensity of s is somewhat lower than that of sh and x.

The spectrum of f is flatter and has a lower level than these of the sibilants. This is reflected in a greater dispersion, and in a lower mean intensity. Also h usually has greater dispersion, and lower intensity than the sibilants, and lies between them and f in these respects.

The only voiced fricative r differs markedly from the others by having an energy concentration in the low bands (below band 9 = 1 kHz) and also a peak in the area around band 15.

The fricative components of the affricates are generally similar to the corresponding fricatives (i.e. the fricative components of z and c are similar to s, those of j and q are similar to x, and those of zh and ch are similar to sh).Also here speaker D differs: his zh and ch have a lower center of gravity and higher dispersion than his sh.

In Figure 2 the dispersion is plotted against the center of gravity for each fricative. This is one way of representing the fricative space of Chinese in a way which makes comparison with other languages possible.

In Figure 3 an alternative representation of the fricative space is given. Here the mean intensity level over the critical bands is plotted against the center of gravity.

On both diagrams the individual fricatives are fairly well separated, but there is some overlapping, especially between different speakers, which may indicate that the perception of fricatives involves normalizing between different speakers, as is the case for vowels.

Discussion

The method of making critical band spectra and computing the proposed parameters from them makes it possible to characterize the place of the fricatives of a language within the available fricative space, and to compare the fricatives of different languages.

The fricatives space of Chinese, as defined by the center of gravity and dispersion (Fig 2) is rather crowded in the voiceless sibilants area (low dispersion and relatively high center of gravity), and there is less variation between different productions of the same sibilant than with h and especially r, which is the only voiced fricative.

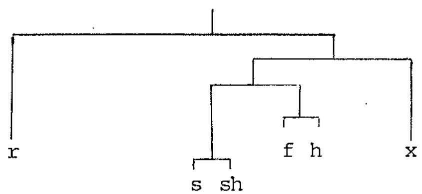
Arabic has also been analyzed with the same methods (Norlin 1983). Comparison of the three voiceless sibilants in Arabic (s, ṣ, and š) with the Chinese ones show that, at least for some speakers, Chinese s is more high-frequent than Arabic s, and that Arabic š and Chinese sh lie in approximately the same area. The high dispersion and low center of gravity area is more utilized by Arabic, which has 7 voiceless and 4 voiced fricatives, than by Chinese.

Swedish also has three sibilants which are roughly comparable to the Chinese ones. Available data on Swedish fricative spectra

(Lindblad 1980) are not directly comparable to our data, but judging from the spectra published by Lindblad it appears that the Swedish sibilants are related to each other in a way similar to Chinese ones, i.e. that the center of gravity decreases and the dispersion increases in the series [s] - [ɕ] - [ʃ] (or [ʂ]; other variants of Swedish /ʃ/ are not comparable). It also seems that the Chinese sibilants have the rise into the high-level area of the spectrum higher up than the Swedish counterparts (which) probably implies that they have higher centers of gravity).

Zhang et al. 1982: 201 give a hierarchical clustering diagram, which shows the perceptual similarity (as measured by the tendency to be confused in a perception test) between Chinese initial consonants.

If the fricatives are isolated from this diagram it becomes:



This can be compared to the dispersion/center of gravity plot (Fig 2). In both diagrams the voiced r is clearly set apart from the others (by its low center of gravity in Fig 2). The pairs s, sh and f, are also well separated in both diagrams (by the dispersion parameter in Fig 2).

The position of x [ɕ] differs markedly in the two diagrams, however. Acoustically it is similar to s and sh , and comes between them in Figure 2 but perceptually x is clearly separated from s and sh. As noted above, this discrepancy can be explained by the nearly complementary distribution of x and s or sh.

The features compact/diffuse and grave/acute are defined by Jakobson, Fant and Halle 1952 in terms of spectral properties, and measures similar to those proposed here are discussed by them.

They define compact phonemes in the following way:

> Compact phonemes are characterized by the relative predominance of one centrally located formant region (or formant).

> They are opposed to diffuse phonemes in which one or more non-central formants or formant regions predominate."

They also suggest that the second moment about the mean (i.e. a measure similar to the dispersion measure used here) should be used as a measure of compactness. From the definition and the examples given by them (/ʃ/ compact, /s/ diffuse for instance), it appears that compactness is related to a non-extreme value of the center of gravity (and perhaps a not too high dispersion).

The feature grave/acute is defined as:

> "... the predominance of one side of the significant part of the spectrum over the other. When the lower side of the spectrum predominates, the phoneme is labeled grave; when the upper side predominates, we term the phoneme acute."

As a measure of this, the third moment about the center of area is suggested (and also the center of area, a measure similar to the center of gravity as used here). It seems that this feature cannot be directly related to any of the parameters used here. The term "significant part of the spectrum" used in the definition seems to imply that a local measure of the center of gravity could be used to distinguish pairs of phonemes which differ in this feature, rather than the "global" center of gravity used here.

Acknowledgements

The work reported here is supported by the Swedish Council for Research in the Humanities and Social Sciences.

References

Jakobson, R., Fant, G., and Halle, M. 1952. Preliminaries to speech analysis. 2nd pr. Cambridge, Mass.: MIT Press

Lindblad, P. 1980. Svenskans sje- och tje-ljud i ett allmänfonetiskt perspektiv. Lund: Gleerup

Nartey, J. 1982. On fricative phones and phonemes. UCLA Working Papers in Phonetics 55, Los Angeles

Norlin, K. 1983. Acoustic analysis of the fricatives in Cairo Arabic. To be published in Working Papers, Phonetics Laboratory, Lund

Schroeder, M.R., Atal, B.S., and Hall, J.L 1979. Objective measure of certain speech signal degredations based on masking properties of human audititory perception. In Björn Lindblom and Sven Öhman, (Ed.) Frontiers of speech communication research, London: Academic Press 1979, pp 217-229

Zhāng, J., Lǔ, S. and Qí, S. 1982. A cluster analysis of the perceptual features of Chinese speech sounds. Journal of Chinese Linguistics, Berkeley, 10, pp 189-206

Table 1. Center of gravity, dispersion and mean intensity level of the critical band spectra of fricatives and fricative components of affricates in Standard Chinese.

	Speaker	Center of gravity crit. band units	Hz	Dispersion crit. band units	Mean intensity dB
s	A	22.31	7314	1.24	3.74
		22.12	7118	1.06	5.84
	B	20.66	5771	1.26	3.83
		20.84	5923	1.44	1.57
	C	21.48	6493	2.34	-2.79
	D	19.90	5174	2.82	4.41
		18.53	4246	4.57	-0.89
x	A	19.84	5129	1.95	6.28
		21.38	6400	1.69	4.22
	B	19.33	4766	1.61	7.22
		19.73	5049	1.65	9.10
	C	19.82	5114	2.18	3.10
		20.93	6000	1.74	3.04
	D	20.29	5472	1.97	4.41
		19.82	5114	2.70	0.16
sh	A	17.11	3456	2.02	8.76
		16.54	3181	2.35	5.59
	B	17.96	3910	2.69	6.81
		19.52	4898	2.32	6.08
	C	17.41	3610	2.71	4.97
		16.17	3014	1.98	7.35
	D	16.05	2961	2.95	3.24
		17.49	3652	2.50	3.60
h	A	19.45	4849	2.98	-2.18
		17.67	3749	2.70	2.23
	B	14.43	2333	4.45	-2.63
		17.35	3579	3.50	-1.93
	C	19.79	5092	2.22	3.13
		15.64	2789	3.48	2.21
	D	12.78	1822	2.57	3.73
		13.13	1921	2.53	2.89
f	A	20.25	5441	3.57	-7.83
		20.10	5325	3.14	-10.23
	B	18.03	3950	3.98	-6.43
		19.51	4891	3.48	-6.85
	C	17.97	3916	3.38	-6.58
		19.25	4711	3.70	-9.18
	D	15.84	2872	3.77	-11.36
		15.40	2692	3.05	-5.87
r	A	11.45	1485	3.55	-8.80
		12.56	1762	2.36	-7.66
	B	13.86	2143	4.52	-8.79
		11.46	1488	5.39	-7.98
	C	4.83	428	1.56	-1.83
		6.35	609	3.94	2.02
	D	6.08	575	4.56	-4.40
		5.29	480	2.88	0.14

Table 1 (cont.)

	Speaker	Center of gravity		Dispersion
z	A	22.52	7538	0.84
	B	21.15	6192	1.88
	C	21.19	6228	2.46
	D	19.70	5027	3.69
c	A	21.80	6798	1.11
	B	20.55	5681	2.55
	C	21.44	6456	2.44
	D	20.32	5496	4.36
j	A	20.84	5923	1.63
	B	20.33	5504	1.34
	C	20.92	5991	1.41
	D	20.41	5568	2.29
q	A	20.18	5386	2.03
	B	19.84	5129	1.30
	C	20.43	5584	1.81
	D	20.27	5457	1.76
zh	A	15.54	2748	2.59
	B	20.41	5568	1.77
	C	15.45	2712	3.36
	D	13.47	2022	4.63
ch	A	17.96	3910	2.78
	B	19.55	4919	2.43
	C	16.48	3153	3.10
	D	15.69	2809	4.96

Figure la. Wave-forms of Standard Chinese fricatives.

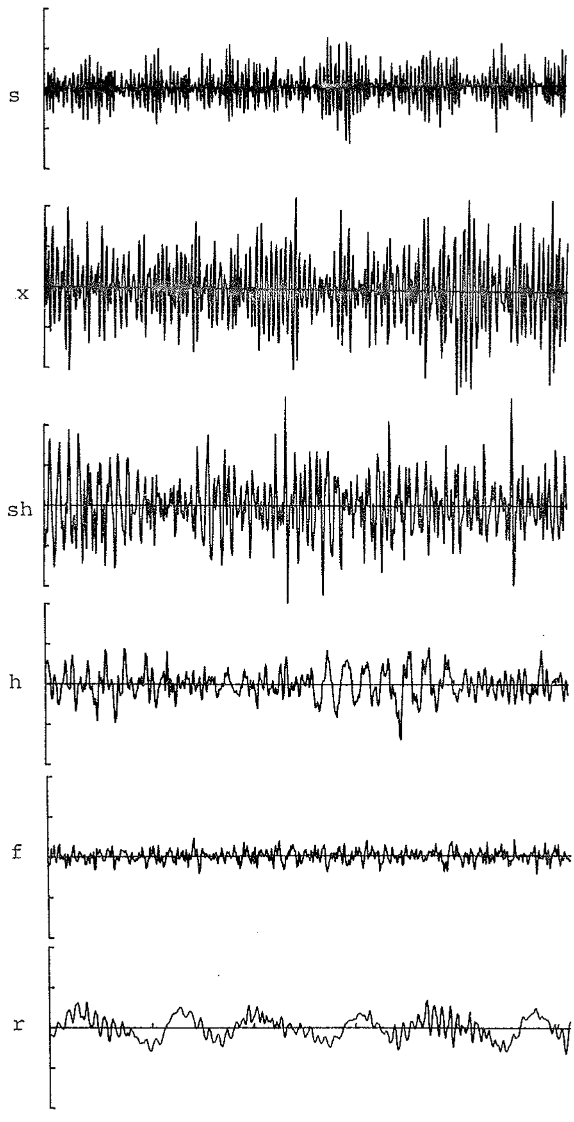


Figure 1b. FFT spectra in linear scale

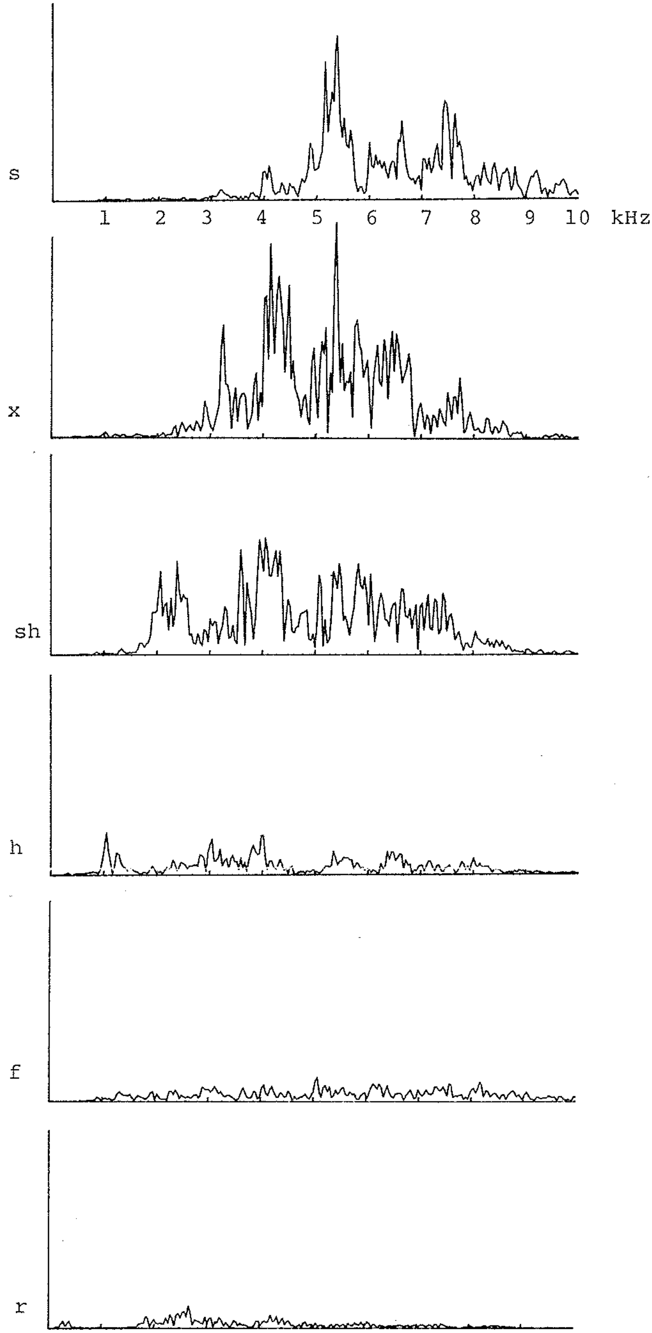


Figure 1c. FFT spectra in logarithmic scale (dB)

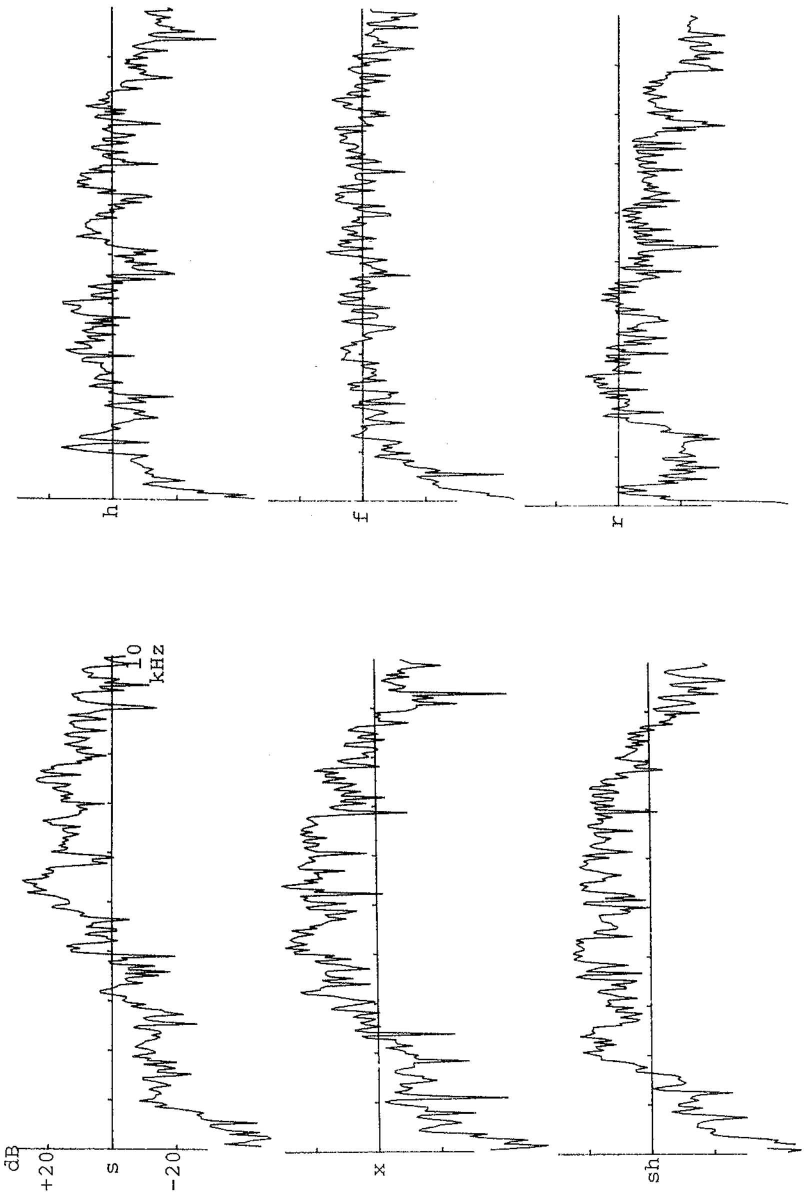
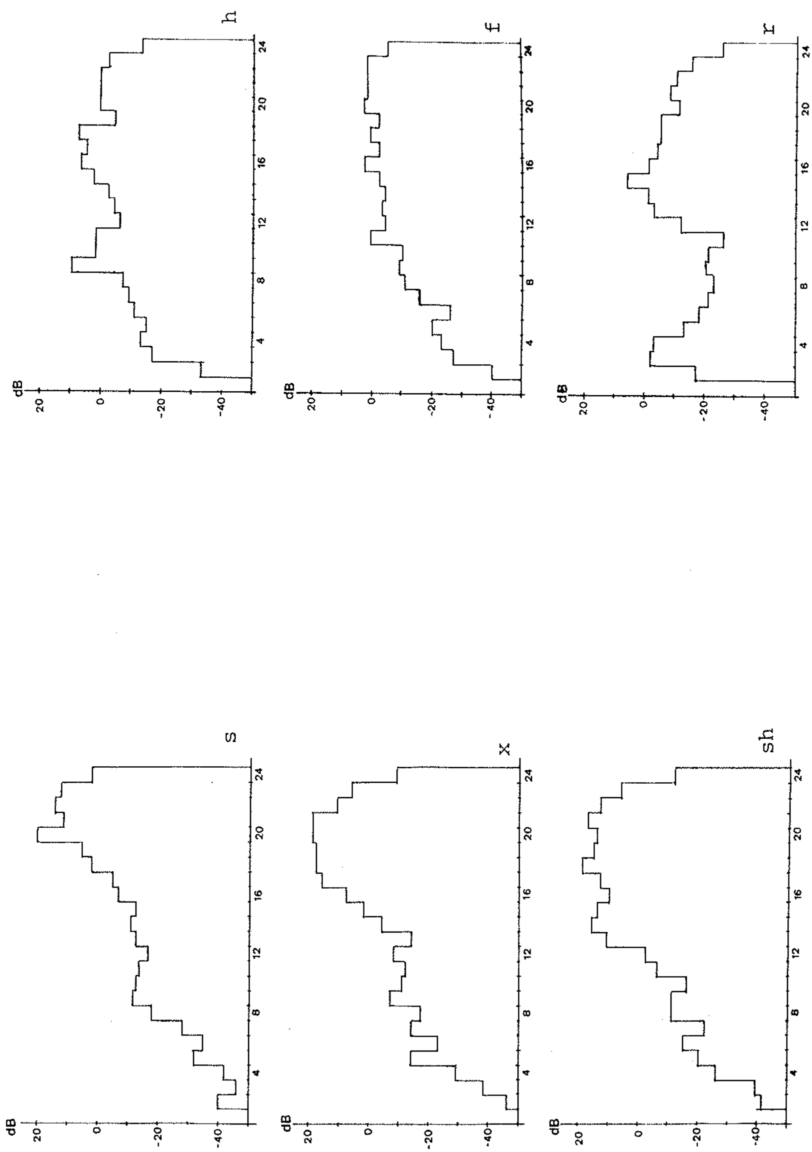


Figure 1d. Critical band spectra



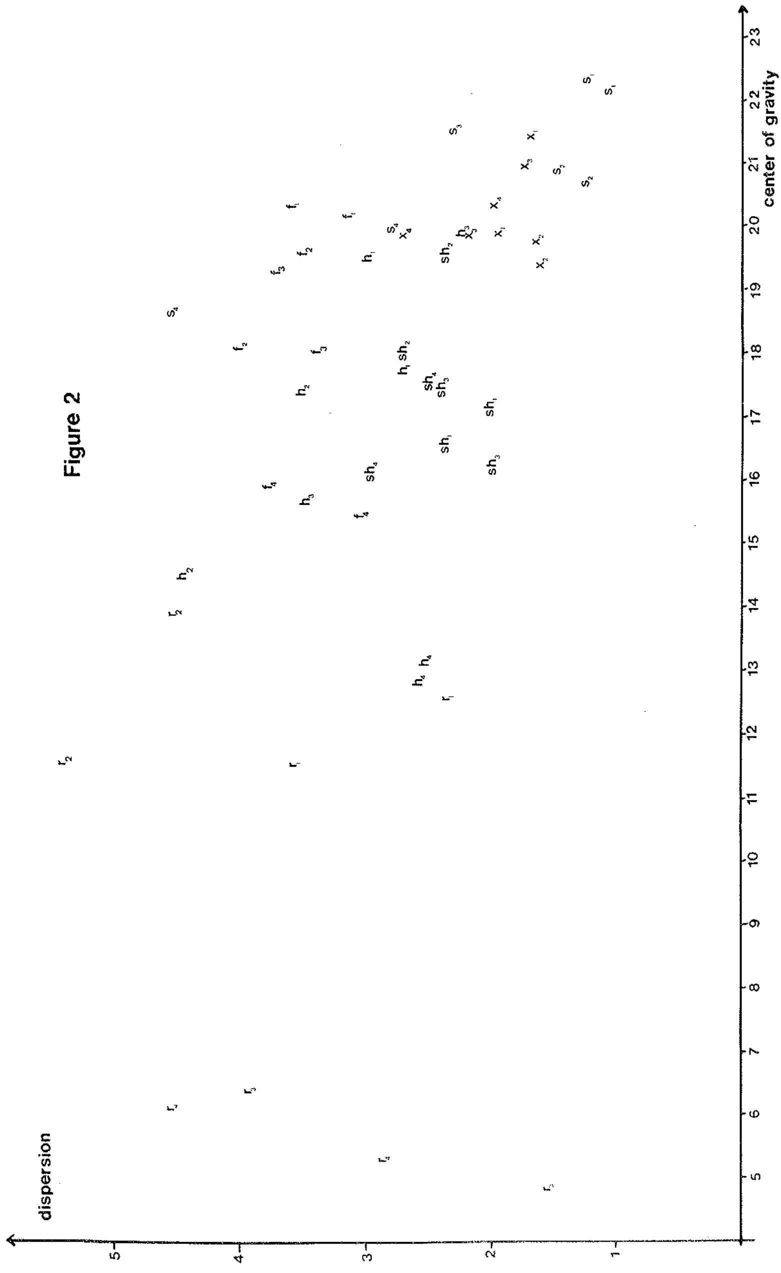

dispersion
center of gravity
1 2 3 4 5
5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

Figure 2

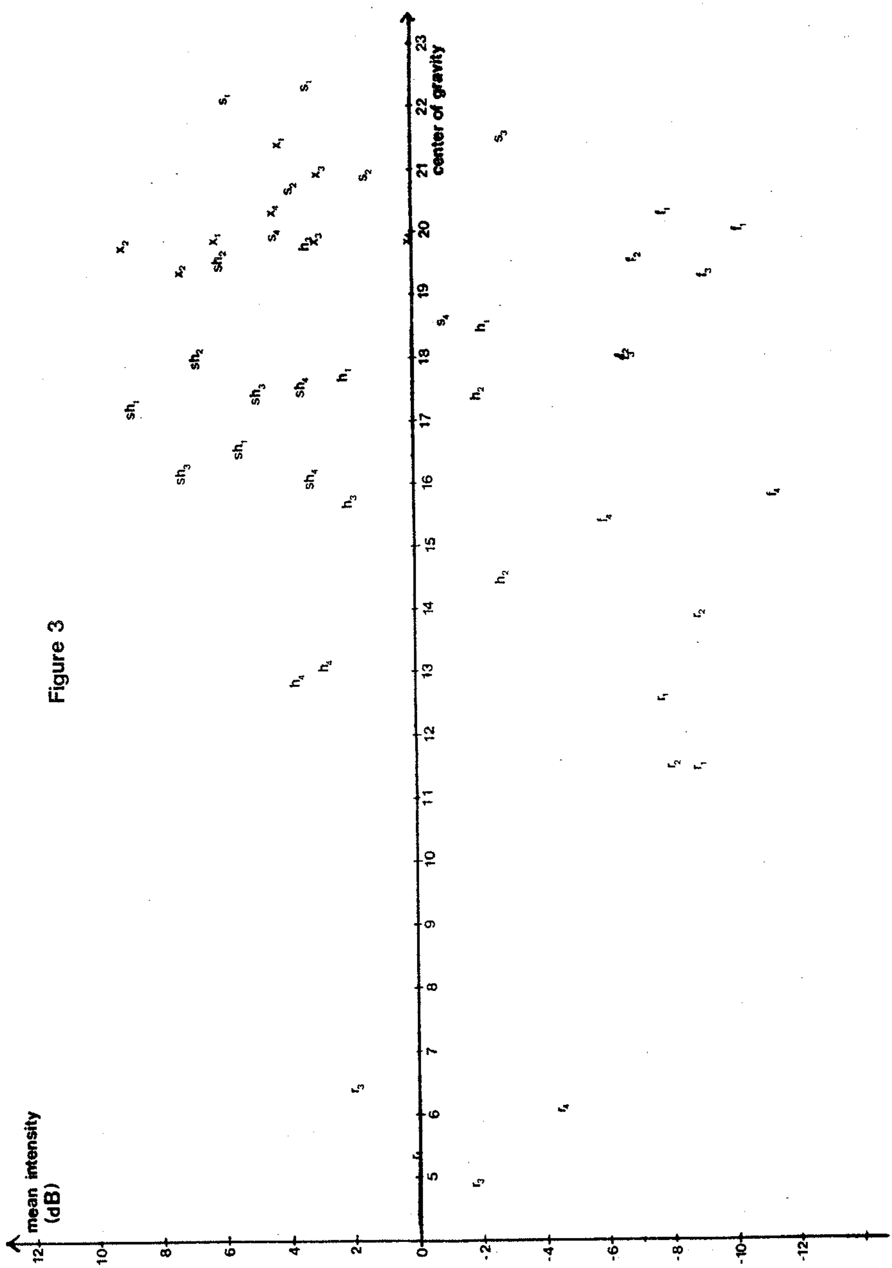

mean intensity
(dB)
center of gravity
12
10
8
6
4
2
0
-2
-4
-6
-8
-10
-12
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23

Figure 3

Working Papers 25 1983
Linguistics-Phonetics
Lund University, 213-218

PERCEPTUAL CRITERIA FOR DIFFERENTIATING BETWEEN DIALECT TYPES

Ursula Willstedt

INTRODUCTION

The Swedish Prosody Project is continuosly engaged in appraising and revising the model that was orginally developed and described by Gårding (1973), Bruce (1977) and Gårding and Bruce (1979, 1980). This model, which is mainly production oriented, derives and generates dialect prosodic variants in Swedish from a small number of grammatical parameters. Four dialect types are recognised, denoted 1A, 1B, 2A and 2B, represented by Malmö, Dalarna, Stockholm and Göteborg respectively. These are differentiated by the timing of word accents and sentence accent (focus) in each type. For exemple, type 2A has the earliest word accent timing and type 1B the latest, types 2A and 2B have a high focus position which gives two-peaked accent 2 on words in focus position while in type 1A and 1B the corresponding words have one peak.

Recent work has been devoted to the problem of whether intermediate types can be described and generated with the aid of the model. This report queries whether the number of types corresponds to perceptual impressions or whether more types are needed. At the Trondheim prosody symposium we assumed from an analysis of intermediate forms that the Småland dialects, with Växjö as prototype, could also be considered a separate Swedish dialect type. We denoted it 2AB. The 2 refers to the two-peaked accent 2 in focus, A to the sentence accent manifestation which is similar to Stockholm (2A) and B to the word accent manifestation which is similar to Göteborg (2B). This is illustrated in fig. 1 wich shows observed tonal patterns for the various dialects.

In Växjö speech the word accent has a relatively late timing, which may explain the obvious F_0 rise during the post-accent vowel when the sentence accent falls on sentence-final accent 2 words (solid line in Fig. 1). This also occurs in Göteborg speech, where the sentence accent manifestation is distinct from the word accent just as in Växjö. When the focus is on the final word, there is presumably insufficient time left for signalling the sentence accent by any other means than by a rise. Figure 1 also shows that the fundamental frequency curve for Växjö is similar to Stockholm, but the 70 ms timing shift means that the word accents are manifested differently in the two dialects. Figure 2 clearly shows how this difference of timing affects in both dialects.

A1
Man vill
•••ANAMMA några längre nummer
—anamma några längre NUMMER

A2
Man vill
•••LÄMNA några långa nunnor
—lämna några långa NUNNOR

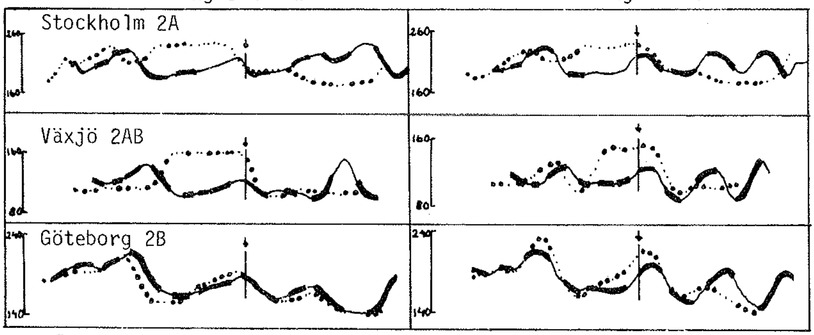


Fig. 1. Representative fundamental frequency curves for sentences with accent 1 and accent 2 words respectively. Examples of initial and final sentence accent location are given for both sentences. The thicker lines indicate vowels, thin lines consonants.

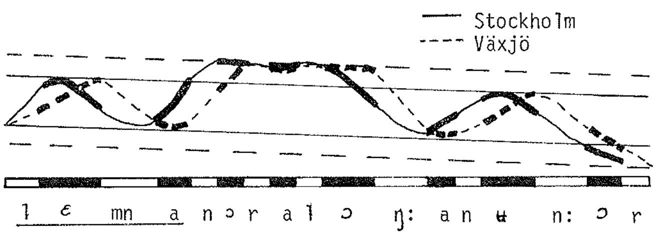


Fig. 2. Normalized fundamental frequency and duration patterns for a sentence with accent 2 words. Time scale: 1 cm = 0,1 s

Despite obvious similarities in the outer configuration, Fig. 2 shows how the focused accent 2 word lämna is located in the rising-falling tonal movement in the Växjö pattern whereas the corresponding word in the Stockholm example is falling-rising. The difference should have some perceptual significance. The observation leads to the question as to whether it is so important perceptually that it is justifiable to consider these two variants to be members of quite different dialect categories.

MATERIALS AND METHODS

To obtain an answer to this question ten subjects were recorded via a laryngograph, which enables pitch information from the voice source to be heard prior to the supraglottal filtering. The informants represent dialectvariants from Växjö (2AB), Kalmar (2A), Stockholm (2A), Göteborg (2B), Malmö (1A) and Krisitanstad (1A). The utterances that were compared are the proverb när katten är borta dansar råttorna på bordet (" when the cat´s away the mice will play") and contrived sentence man anammar lundamodellen ('one absorbs the lund model") consisting virtually of sonorants only to show up the timing of word and sentence accents. Similar conditions (only sonorants and comparable vowel qualities) also pertain in the utterances used for the tone curves of Figs. 1 and 2.

The proverb, pronounced with even intonation was used in a listening test. One or two representative renderings from each dialect were selected and paired at random to provide 25 stimuli. Eleven listeners representing various dialects have judged whether the renderings in each pair can be classed as belonging to some dialect area or not.

RESULT AND DISCUSSION

The results of the listener test are given in Fig . 3. The horizontal axis contains the stimuli comparing paired dialect variants. The vertical axis records the number of listeners who made "same dialect" judgment for each pair.

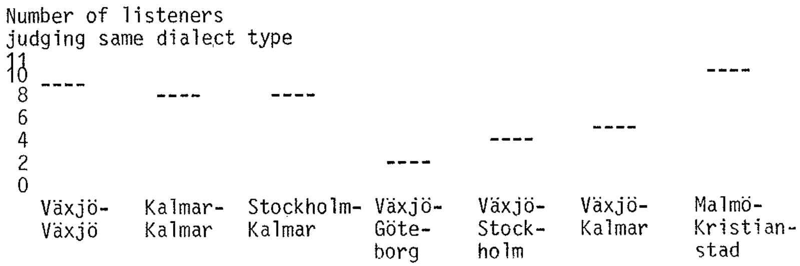


Fig. 3. Result of the listener test. Pairs of dialects are given on the horizontal axis. The vertical axis represents the number of listeners who judged the dialects in each pair as belonging to the same dialect type.

For stimuli comparing two speakers known to be of the same dialect, the "same dialect" judgement are high. For example, the two Växjö informants were judged to have the same dialect by 10 of the 11 listeners.

The Växjö and Göteborg renderings were judged to be the same dialect by only 1 listener. The similar word accent do not appear to be sufficient perceptual ground for assigning two variants to the same dialect type if the sentence accent are different. Växjö and Göteborg have different sentence accent patterns when the sentence accent is located finally or medially in the sentence. Gårding and Bruce (1978) have also concluded that the sentence accent pattern is the primary criterion for differentiating dialects.

The Växjö renderings were also compared with the two variants of the 2A dialects type, Stockholm and Kalmar, where the prosodic pattern is similar. These were judged to belong to the same dialect as Växjö by 3 and 4 listeners respectively. It thus seems to be a more difficult task to differentiate Växjö from Stockholm or Kalmar than Växjö from Göteborg. Nevertheless, Växjö tends to be classed as a separate dialect. This merits further investigation with more informants and more listeners. However, the present report does indicate perceptual justification for considering that the Småland dialects, with Växjö as prototype, constitute a seperate dialect type.

The Malmö and Krisitanstad dialects (tonal patterns in Fig. 4) were also compared in the listener test. According to the model, they are both classed as type 1A. The result of the listener test shows that they are also perceptual similar, 9 of the 11 listeners judging them to be the same dialect. A closer look at the tonal pattern reveals a slight difference of timing, just as there was between Växjö and Stockholm. This is illustrated in Fig. 5.

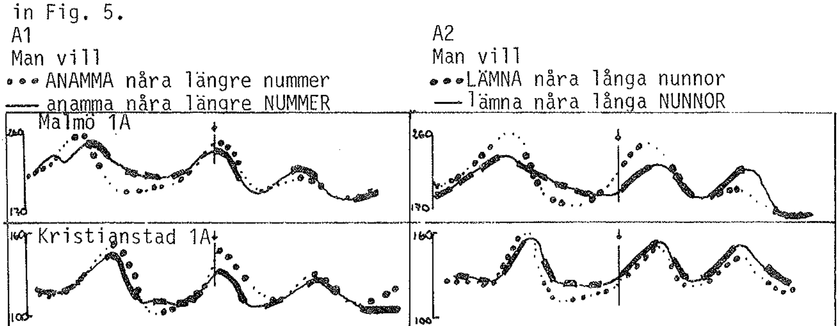


Fig. 4. Representative fundamental frequency curves for sentences with accent 1 and accent 2 words respectively. Examples of final and initial sentence accent location are given. Thicker lines indicate vowels, thin lines consonants

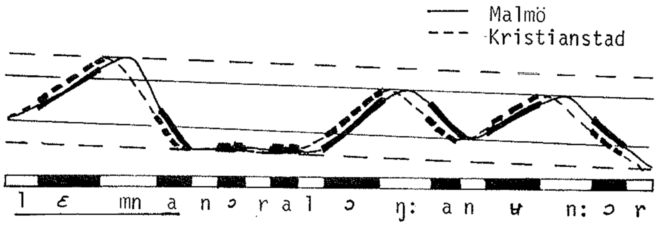


Fig. 5. Normalized fundamental frequency and duration patterns for a sentence with accent 2 words and initial focus. Time scale: 1 cm = 0,1 s.

The timing difference is a mere 30 ms and has no effect on the tonal pattern of the word accents. There was, however, a clear effect for Växjö and Stockholm, where the timing differences was more than twice as large. Since the Malmö and Kristianstad dialects are classed as the same dialect type by the listener test, there is presumable a critical limit for whether or not timing shifts differentiate between dialect types.

The timing of the F_0 fall in accent 1 and accent 2 in one and the same dialect was studied by Bruce (1977). The timing of the high turning point for both accents was varied continously in a synthetic rendering of inga malmer. This sentences can be interpreted as accent 1 ("no ores") or accent 2 (forname and surname, "Inga Malmer"). The ensuing listener test revealed a critical region with a timing difference of 20-30 ms where it was difficult to differentiate between the accents. In view of this result, it is not unlikely that a timing shift of about 70 ms between two otherwise similar tonal patterns is perceptually relevant and the variants would be assigned to different categories.

The result of the present investigation shows the perceptual importance of the tonal movement of the accented syllable. The considerable timing difference between Växjö and Stockholm yields quite different word accent patterns although the sentence accents are very similar. In Växjö the pitch of the accented syllable is rising to a high whereas in Storkholm it is falling towards a low (cf. Fig. 2). This is presumably why the two dialects were judged as belonging to different dialect types. It is interesting to compare this with the Edinburgh and Glasgow accents which were described as high and low respectively by Gilian Brown et al. (1980). There may be perceptual justification for discribing the Växjö and Stockholm dialects respectively as high and low.

REFERENCES

Brown, G. & Currie, K & Kenworthy J. 1980. Questions of Intonation. London Croom Helm

Bruce, G. 1977. Swedish word accents in sentence perspective. Gleerup Lund.

Bruce, G. & Gårding E. 1979. A prosodic typology for Swedish dialects. Nordic Prosody, Travaux XIII.

Bruce, G. & Gårding, E. 1980. A presentation of the Lund Model for Swedish intonation. Nordic Prosody II ed by Torsten Fretheim

Gårding, E. 1973. Toward a prosodic typology for Swedish dialects. Proc. 2nd Int Conf. of Nordic and General Linguistics. Ed by Dahlstedt

Gårding, E. & Bruce, G. & Willstedt, U. 1980. Transitional forms and their position in a prosodic typology for Swedish dialects. Nordic Prosody II, ed by Torsten Fretheim

Willstedt, U. 1980. An investigation of Prosody in some dialects of Småland. Working Papers 19. Phonetic laboratory, Lund University

RECENT NUMBERS AND CONTENTS

LECTURES AND PUBLICATIONS ELSEWHERE

1982

Bannert, R. An Fo-dependent model for duration? RUUL 8: 59-80 (Reports from the Institute of Linguistics, Uppsala University)

Bannert, R., O. Engstrand, H. Eriksson & L. Nordstrand. Phonetics and identity. RUUL 8: 22-58 (Reports from the Institute of Linguistics, Uppsala University)

Bílý, M., L-Å. Henningsson & B. Sigurd. Från taggning till förståelse. PILUS 47

Bílý, M. & T. Pettersson. A typology of case. 7th Scandinavian Conference of Linguistics, Hanasaari (Helsinki), Finland, 17-19 December 1982

Bruce, G. Reglerna för slutledsbetoning i sammansatta ord i nordsvenskan. In, Nordsvenska: språkdrag i övre Norrlands tätorter. C.-C. Elert and S. Fries (eds.):123-148. Universitetet i Umeå. Almqvist & Wiksell International, Stockholm

Bruce, G. Textual aspects of prosody in Swedish. Phonetica 39: 274-287

Bruce, G. Experiments with the Swedish intonation model. XIIIth International Congress of Linguists , Tokyo, 29 August-4 September 1982

Dahlbäck, H. Fonologisk målspråksvariation och förståelighet. The 7th Scandinavian Conference of Linguistics, Hanasaari, Helsinki, 17-19 December 1982

Engdahl, E. A note on the use of lambda conversion in generalized phrase structure grammars. Linguistics and Philosophy 4(4): 505-515

Engdahl, E. Constituent questions, topicalization, and surface structure interpretation. In D. Flickinger, M. Macken and N. Wiegand (eds.) Proceedings from the First West Coast Conference on Formal Linguistics, Stanford

Engdahl, E. Restrictions on unbounded dependencies in Swedish. In E. Engdahl and E. Ejerhed (eds.) Readings on unbounded dependencies in Scandinavian languages, Almqvist & Wiksell International, Stockholm

Engdahl, E. Relational questions. Workshop on Scandinavian Syntax and Theory of Grammar, Trondheim, June 1982

Engdahl, E. Interpreting questions. Workshop on Syntactic Theory and How People Parse Sentences, The Ohio State University, Columbus Ohio, May 13-15, 1982

Fujisaki, H. & E. Gårding (eds.). Preprints of Papers for the Working Group on Intonation. The XIIIth International Congress of Linguists, Tokyo, 1982

Fujisaki, H. & E. Gårding. Summary of Working Group on Intonation. In Proceedings of XIIIth International Congress of Linguists

Gårding, E. A model for comparison of intonation. XIIIth International Congress of Linguists , Tokyo, 29 August-4 September 1982

Gårding, E. Aspects of intonation. MIT, September 1982

Gårding, E. Prosodic expressions and pragmatic categories. In Acta Universitatis Lundensis Band 38, Textstrategier i tal och skrift . Almqvist & Wiksell International

Gårding, E. Fonetikens villkor och mina. Forskning på osäkra villkor. Humanistisk-samhällvetenskapliga forskningsrådet, Stockholm

Gårding, E. & M. Lindau-Webb. Intonation models for Hausa. Los Angeles, September 1982

Gårding, E. A comparative study of intonation. Workshop on Prosody, Association Européenne de Psycholinguistique, Paris, Maison des sciences de l'homme 21-23 April 1982

Lindell, K., H. Lundström, J-H. Svantesson & D. Tayanin. The Kammu year, its lore and music.

Malmberg, B. Les dimensions du langage vues à travers une analyse de l'expression. Travaux de l'Institut de Phonétique d'Aix vol 8: 7-24

McGarr, N.S., T. Baer, A. Löfqvist, R. Whitehead & D. Metz. High-speed filming and acoustic analysis of deaf glottal pulses. Journal of the Acoustical Society of America 71: S55 (A). 103rd Meeting of the Acoustical Society of America Chicago 26-30 April 1982

Nettelbladt, U. Phonotactic and prosodic development in Swedish children. Symposium on Child Prosody at Child Language Research Forum Stanford 27-28 March 1982

Pettersson, T. In favour of the archiphoneme. In D.L. Goyvaerts (ed.) Phonology in the 1980's. Ghent: E. Story-Scientia, pp. 39-54

Pettersson, T. Prose and poetry. Studia Linguistica 36:1, 1982, pp. 64-87

Sigurd, B. Att ordidrotta - till palindromernas lingvistik. Språkvård 1982:1

Thunberg, A-M., K. Nowak, K-E. Rosengren & B. Sigurd. Communication and equality. A Swedish perspective. Stockholm: Almqvist & Wiksell (1982)

1983

Bannert, R. Phonological variation and language teaching. In: S.W. Felix / H. Wode (eds.), Language Development at the Crossroads, 151-171. Tübingen

Bannert, R. Att testa förståelighet i bruten svenska. FRN-symposium, Stockholm 10-11 March 1983

Bannert, R. Towards a model for German prosody. Interdisciplinary Symposium. "Prosody - Normal and Abnormal", Zürich 6-8 April 1983

Bannert, R. Modellskizze für die deutsche Intonation. Zeitschrift für Literaturwissenschaft und Lingvistik 49

Bannert, R. Problems in learning Swedish pronunciation and in understanding foreign accent. Folia Linguistica XVIII, 1-2

Bruce, G. Accentuation and timing in Swedish. Folia Linguistica, vol, XVII

Bruce, G. Om rytmisk alternation. De nordiska språkens prosodi 3, Umeå 3-4 juni 1983

Bruce, G. Textual aspects of prosody in Swedish. 10th international congress of phonetic sciences, Utrecht 1-6 augusti, 1983

Engdahl, E. Parasitic gaps. Linguistics and Philosophy 6(1): 5-34

Engdahl, E. Review of P. Jacobsson and G. Pullum (eds.) The nature of syntactic representation. Studia Linguistica 37

Engdahl, E. Introduction to the papers in the workshop on Scandinavian Syntax. In F. Karlsson (ed.) Proceedings from the Seventh Scandinavian Conference of Linguistics, Helsinki

Engdahl, E. Parasitic gaps and resumptive pronouns. GLOW (Generative Linguists of the Old World) Colloquium, York, March 28-31, 1983

Engdahl, E. Current approaches to wh-phenomena and binding. Second Workshop on Scandinavian Syntax and Theory of Grammar, Biskops-Arnö, June 4-6, 1983

Gårding, E. & J. Zhang. Tone and intonation in Standard Chinese. Zürich, April 1983. Interdisciplinary Symposium. Prosody - Normal and Abnormal

Gårding, E. På jakt efter en allmän intonationsmodell . De nordiska språkens prosodi 3. Umeå, juni 1983.

House, D. Transitioner i talperception: implikationer för perceptionsteori och applikationer i hörselvården. TLH-symposium, Stockholm, mars 1983

Pettersson, T. & S. Wood. Vokalreduktion i bulgariskan. TLH-symposium, Stockholm, mars 1983

Pettersson, T. Fria adverbial. Divertimento con introduzione. Nydanske Studier & Almen Kommunikationsteori 13, pp. 41-74. København: Akademisk Forlag

Sigurd, B. Artificiell intelligens på mikrodator. Mikrodatorn 1983:3

Sigurd, B. Morse med dator. Automatisk morseläsare. Mikrodatorn 1983:3

Sigurd, B. Möt språkforskningen. Lund: Liber

Sigurd, B. Text representation in a text production model. In: S. Allén (ed.) Text processing. Nobel symposium. Stockholm: Almqvist & Wiksell

Sigurd, B. Commentator: A computer model of verbal production. Linguistics; 20-9/10

Sigurd, B. Om jätte-, kalas- och andra förstärkningsprefix. Språkvård 1983:2

Svantesson, J-O. fc. 'Personal pronouns and referent dynamics in Kammu'. In Svantesson, ed. fc.b.

Svantesson, J-O. fc.a. The phonetic correlates of register in Va. ms.

Svantesson, J-O., ed. fc.b. Papers from the Helsingør symposium on Austroasiatic linguistics and literature (preliminary title). To be published by the Scandinavian Institute of Asian Studies, København

CURRENT RESEARCH PROJECTS

The immigrant and communication technology - understanding and being understood (The Bank of Sweden Tercentenary Fund)

Robert Bannert, Hans Dahlbäck, Gisela Håkansson, Kerstin Tevajärvi

Verbal reforms (The Bank of Sweden Tercentenary Fund)

Bengt Sigurd, Christer Platzack, Anna Orrling

Commentator, a computer system simulating verbal behavior (The Council for Research in the Humanities and Social Sciences)

Bengt Sigurd, Jan Fornell

Language and brain (The Council for Research in the Humanities and Social Sciences)

Eva Gårding, Christina Draviņš

Central Features in Scandinavian Syntax (Nordic Committee for Coordinating Research in the Humanities)

Elisabet Engdahl

Phonological and Phonetic studies of vowel articulation in Bulgarian (The Council for Research in the Humanities and Social Science)

Thore Pettersson, Sidney Wood

Phonetic descriptions of some important languages in the third world (The Council for Research in the Humanities and Social Sciences)

Eva Gårding, Mona Lindau, Kjell Norlin, Jan-Olof Svantesson

PRAKTISK LINGVISTIK

Nr 1 - 1979	Eva Gårding & Robert Bannert Optimering av svenskt uttal 10:- (incl. cassette recording)
Nr 2 - 1979	Kenneth Hyltenstam (Ed) Invandrarundervisning - aspekter på alfabetisering och den första nivån 10:-
Nr 3 - 1979	Robert Bannert Ordprosodi i invandrarundervisningen 20:- (incl. cassette recording)
Nr 4 - 1979	Eva Wigforss Svensk pluralbildning hos finska invandrar-barn - en psykolingvistisk utvecklingsstudie 20:-
Nr 5 - 1981	Robert Bannert Svårigheter med svenskt uttal: inventering och prioritering 30:- (incl. cassette recording)
Nr 6 - 1981	Gösta Bruce (Ed) Tema prosodi - seminarieuppsatser i fonetik 20:-
Nr 7 - 1982	Don Kulick, Karina Vamling, Lars-Åke Henningsson Om tolkning 20:-
Inquiries to:	Praktisk lingvistik Institutionen för lingvistik Lunds Universitet Helgonabacken 12 S-223 62 LUND Sweden

(Prices in Swedish Crowns)

Swedish Post Office giro account 439 34 92 - 6

STUDIA LINGUISTICA

Eds Bengt Sigurd and Jan Svartvik

One volume of Studia Linguistica is published each year in two parts. The subscription price is 60 SEK (approximately $ 13). Orders may be sent to any international bookseller or to the publishers, CWK Gleerup, Box 1205, S-221 05 Lund, Sweden

TRAVAUX DE L'INSTITUT DE LINGUISTIQUE DE LUND
FONDÉS PAR BERTIL MALMBERG
PUBLIÉS PAR EVA GÅRDING ET BENGT SIGURD

I	*Carl-Gustav Söderberg.* A Typological Study on the Phonetic Structure of English Words with an Instrumental-Phonetic Excursus on English Stress. 1959.
II	*Peter S. Green.* Consonant-Vowel Transitions. A Spectrographic Study. 1959.
III	*Kerstin Hadding-Koch.* Acoustico-Phonetic Studies in the Intonation of Southern Swedish. 1969.
IV	*Börje Segerbäck.* La réalisation d'une opposition de tonèmes dans des dissyllables chuchotés. Étude de phonétique expérimentale. 1966.
V	*Velta Rūķe-Draviņa.* Mehrsprachigkeit im Vorschulalter. 1967.
VI	*Eva Gårding.* Internal Juncture in Swedish. 1967.
VII	*Folke Strenger.* Les voyelles nasales françaises. 1969.
VIII	*Edward Carney.* Hiss Transitions and their Perception. 1970.
IX	*Faith Ann Johansson.* Immigrant Swedish Phonology. 1973.
X	*Robert Bannert.* Mittelbairische Phonologie auf akustischer und perzeptorischer Grundlage. 1976.
XI	*Eva Gårding.* The Scandinavian Word Accents. 1977.
XII	*Gösta Bruce.* Swedish Word Accents in Sentence Perspective. 1977.
XIII	*Eva Gårding, Gösta Bruce, Robert Bannert* (Eds). Nordic Prosody. 1978.
XIV	*Ewa Söderpalm.* Speech Errors in Normal and Pathological Speech. 1979.
XV	*Kerstin Nauclér.* Perspectives on Misspellings. 1980.
XVI	*Per Lindblad.* Svenskans sje- och tje-ljud (Some Swedish Sibilants). 1980.
XVII	*Eva Magnusson.* The Phonology of Language Disordered Children. 1983.
XVIII	*Jan-Olof Svantesson.* Kammu Phonology and Morphology. 1983.

Numbers V, VII-XII, XIV-XVIII are available from the publisher, LiberLäromedel, Box 1205, S-221 05 LUND.

Numbers III, VI, XIII are available from the Department of Linguistics and Phonetics, Helgonabacken 12, S-223 62 LUND.

LINGUISTICS DEPARTMENT

STAFF

Chairman	Eva Gårding
Director of Studies	Kurt Johansson
Library	Jana Geggus, Sidney Wood
Office	Eva Körner, Britt Nordbeck Susanne Törnqvist

General Linguistics

Dahlbäck, Hans	Fil Kand; Research Assistant, extra teacher
Engdahl, Elisabet	PhD; Research Fellow
Fornell, Jan	Research Assistant
Geggus, Jana	Fil Kand; Assistant; extra teacher
Hyltenstam, Kenneth	Fil Dr; Research Assistant
Håkansson, Gisela	Fil Mag; Research Assistant
Malmberg, Bertil	Fil Dr; Professor Emeritus
Nordbeck, Britt	Secretary
Orrling, Anna	Fil Kand; Research Assistant
Pettersson, Thore	Fil Dr; Docent; Lecturer
Prohovnik, Barbara	BA; extra teacher
Sigurd, Bengt	Fil Dr; Professor, Head of Division
Viberg, Åke	Fil Dr; Research Fellow
Research Students	Jean-Jacques Berthout, Hans Dahlbäck, Jana Geggus, Lars-Åke Henningsson, Merle Horne, Gisela Håkansson, Christian Matthiesen, Barbara Prohovnik, Leila Ranta, Christopher Stroud, Jan-Olof Svantesson, Karina Vamling

Phonetics

Bannert, Robert	Fil Dr; Docent; Research Assistant
Bredvad-Jensen, Anne-Chr.	Fil Kand; Assistant; extra teacher

Bruce, Gösta	Fil Dr; Docent
Dravinš, Christina	Logoped; Research Assistant
Gårding, Eva	Fil Dr; Professor; Head of Division
Hadding, Kerstin	Fil Dr; Professor Emeritus
Johansson, Kurt	Fil Lic; Lecturer
Jönsson, Karl-Gustav	Technician
Körner, Eva	Secretary
Lindau-Webb, Mona	PhD; extra teacher; Research Assist; Phonetics Dept, UCLA
Löfqvist, Anders	Fil Dr; Docent; Research Fellow
Mandersson, Bengt	Techn Dr; Civ Ing; Computer Engineer
Nauclér, Kerstin	Fil Dr; Logoped; extra teacher
Nettelbladt, Ulrika	Logoped; extra teacher
Norlin, Kjell	Fil Kand; Research Assistant
Svantesson, Jan-Olof	Fil Kand; Research Assistant
Tevajärvi, Kerstin	Hum Kand; Research Assistant
Wood, Sidney	Fil Dr; extra teacher; Research Assistant
Research Students	Catharina Andersson, Antonis Botinis, Anne-Christine Bredvad-Jensen, Christina Dravinš, David House, Boris Larnert, Elisabeth Lindström-Sederholm, Eva Magnusson, Ulrika Nettelbladt, Ann-Christine Ohlsson, Kerstin Tevajärvi, Paul Touati, Kjell Weimer, Ursula Willstedt