LUND UNIVERSITY
DEPARTMENT OF LINGUISTICS
General Linguistics
Phonetics

WORKING PAPERS

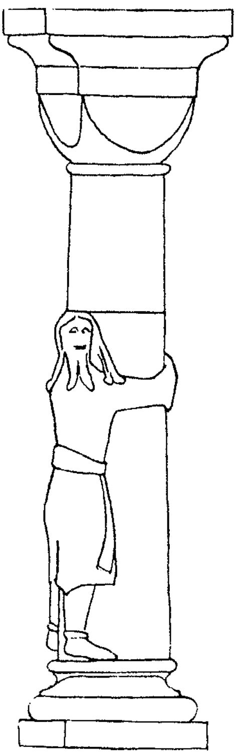
27 • 1984

LUND UNIVERSITY
DEPARTMENT OF LINGUISTICS
General Linguistics
Phonetics

WORKING PAPERS

27 • 1984

ISSN - 0280 - 526X



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

This issue has been edited by Robert Bannert and Eva Gårding

Annual Report

The Tenth International Congress of Phonetic Sciences in Utrecht which was visited by a large group of staff and students from our department created a wave of inspiration for the beginning of the academic year 1983/84.
The research Seminar on Fridays was dominated by the thesis work of our graduate students:

Friday seminars

Sept 23	Summer memories from Umeå, Utrecht and Tällberg	
Sept 30	Vowels and diphthongs in Hausa, Arabic and Chinese	Mona Lindau-Webb, Kjell Norlin, Jan-Olof Svantesson
Oct 14	Perception of interrogative intonation	Anne-Christine Bredvad-Jensen
	Report from the project Speech and Language	Christina Dravins
Oct 28	Report on the symposium Invariance and Variability in Speech processes, MIT Boston	Eva Gårding
Nov 11	The phonology and phonetics of speech rhythm	Gösta Bruce
Nov 18	Computation of the resonances of the vocal tract and some applications	Sidney Wood
Nov 25	Voice and vocation	Ann-Christine Ohlsson
Dec 2	The relationship between prosody and syntax in Greek	Antonis Botinis
Dec 9	Perceptual consequences of changes in Fo range in filtered speech	David House
Dec 16	Some aspects of Finnish sentence intonation	Kjell Weimer

Dec 16	Contrast and focus in French	Paul Touati
Febr 3	Intonation parameters	Eva Gårding
Febr 10	Perceptual compensation strategies in the hard of hearing	David House
March 2	A data base for Scanian place names. Dialect and Place Name Archives, Lund	Stig Isaksson
March 16	Chinese Fo curves extracted by means of the ILS-program	Jan-Olof Svantesson
March 23	Neurolinguistic research paradigms	Dorothea Weniger, Zürich
March 30	Register in Mon Khmer languages	Kenneth Gregerson
April 6	Text prosody in Finnish	Kjell Weimer
April 13	Preliminary results in the project Speech and Brain	Christina Dravins
April 27	Perception of prosody and foreign language learning	Elise André
May 4	Presentation of various computer programs for analysis and synthesis of speech	Sidney Wood
May 11	Experiments with Swedish sentence intonation	Anne-Christine Bredvad-Jensen
	On prosody and syntax: Experiments with perception	David House
	Swedish speech rhythm	Eva Strangert, Umeå
May 25	How Swedes accept Swedish with a foreign accent	Birgitta Kuylenstierna
	An intonation project	Rolf Haberbeck, Berlin

The general seminar held in cooperation with the Department of Logopedics and Phoniatrics and the Child Language Research Center had a more varied program:

Language * Speech * Sound * Hearing

Sept 12	The micro-computer as a pause measurer	Bengt Sigurd
Sept 19	Basic tenets of cyclic phonology/lexical phonology	Jerzy Rubach, Warsaw

Oct 10	Mother-child interaction	Svenka Savit, Novi Sad
Oct 24	Intonation research	David Chrystal, Reading
Nov 7	A model for describing child-adult dialogues verbally, somatically and vocally	Ragnhild Söderbergh
Nov 21	Presentation of the new computer (VAX 730) and demonstration of current research	
Dec 19	Alternative communication	Gisela Håkansson
Jan 30	To speak correctly is good, to speak freely is better	Lennart Norman
Febr 13	Acoustical distinctive features	Olle Engstrand, Uppsala
March 12	New theories about language teaching	Cecilia Thavenius, English Dept. Lund
March 26	Language and the two hemispheres	Dorothea Weniger Neurologische Klinik Zürich
April 9	Linguistic Models and Psycholinguistic Argument	Helen Goodluck
April 16	Strategies in questions and answers	Anna-Brita Stenström
May 7	Speech retarded children and their reading	Kerstin Nauclér, Eva Magnusson, Lilian Rudberg & Gunnel Gahne
May 21	Field work with the Akhas	Inga-Lill Hansson, Dept of East Asiatic Languages, Lund

In addition, two groups of students and teachers meet regularly to discuss their areas of special interest. One area is research in language teaching, Interlanguage, chaired by Robert Bannert, and the other is problems of reading and writing, chaired by Kerstin Nauclér and Eva Magnusson. There were five special courses for graduate students given jointly by the whole department:

Languages in Time and Space	(Bertil Malmberg)
Methods and Models in Linguistics	(Bengt Sigurd)
Psycholinguistics	(Helen Goodluck)
Neurolinguistics	(Dorothea Weniger)
- Introduction to Neurolinguistic	(Christina Dravins)

Language Acquisition

(Ragnhild Söderbergh and Anne-Christine Bredvad-Jensen, Center for Child Language Research)

By the end of this academic year 83/84 27 logopedic students will have completed their two-term phonetics training. In addition we have had a few students with varying interests and backgrounds taking courses of phonetics at different levels. There are at present 8 active graduate students.

Ulrika Nettelbladt defended her PhD thesis, Dysphonology of children, on December 10. Professor Per Linell, Linköping, was the faculty opponent. The dissertation appeared as No 18 of Travaux de l' Institut de Linguistique de Lund. The child Language Research Center held a symposium on May 17-18 visited by guests from Norway and Denmark and sponsored by the Humanistic Research Council and the Einar Hansen Foundation.

Our department was visited by
dr Manfred Pieneman, Sidney, Australia
dr Dorothea Weniger, Zürich
Professor Thorstein Fretheim, Trondheim
Professor Elise André, USA
Ass. Rolf Haberbeck, Berlin

The logopedic students who are now finishing their studies in phonetics visited the Institute of Audiologopedics and the Institute of Phonetics, Copenhagen together with their teachers David House and Sidney Wood. The students presented their work in an informal seminar after attending a presentation of the research facilities and training programs.
The visit was funded by Einar Hansen.

Gösta Bruce spent a sabbatical spring semester at Bell Telephone Laboratories, Murray Hill. Kurt Johansson also had a research sabbatical. Anders Löfquist who now has a part-time research position at the Department af Logopedics and Phoniatrics substituted for him.

The project Phonetic descriptions of some important languages in the third world (Gårding, Lindau, Norlin, Svantesson)[1] is now entering its final year. The projects Vowel Reduction in

Bulgarian (Pettersson and Wood)[1] and Language and the Brain (Gårding and Dravins)[1], and Immigrants Communication: Problems of understanding and being understood (Bannert)[2], are being completed. A new project will start on July 1, Reading and Spelling difficulties. A search for causes. (Nauclér and Magnusson).[1]

The new ILS-program which was purchased at the end of last spring has been installed and local user programs have been developed. Our new digital spectrograph was delivered in July and a new graphic terminal for signal display was acquired. Both were funded by a Wallenberg grant.

Lund in May 1984

Eva Gårding

1) HSFR, Swedish Council for Research in the Humanities and Social Sciences.

2) The Bank of Sweden Tercentenary Foundation.

CONTENTS

Towards a Model for German Prosody

Robert Bannert

ABSTRACT

In this paper, an attempt is made to outline a model for German prosody. The model aims at generating the rhythm and melody of German utterances, i.e. their temporal and tonal structures, by starting from linguistic information in the input and then applying phonological and phonetic rules. The model consists of three main components: the basic temporal and tonal components where the basic temporal and tonal structures of utterances are generated separately from each other and the modification component where the basic structures are modified in different ways according to various demands. Starting out from an inherent duration of the segments spectrally defined in the input, their basic durations are calculated as a consequence of various contextual factors. Operating on the tonal prosodic features of the input, pitch rewriting rules transpose the linguistic features into a notation of tonal points or levels HIGH or LOW which are related to spectral events and of the tonal range WIDE bound to the HIGHs and LOWs. These tonal specifications are transferred in a series of steps by the intonation algorithm into phonetic values. Drawing on acoustic data and on manipulations of the pitch of utterances by means of speech synthesis, some concepts of an intonation theory concerning German are discussed.

1. BACKGROUND AND AIM

In contrast to what is the case for a great number of languages, there does not seem to exist a model for German prosody which is based on acoustic data and which, although qualitatively only, can generate the temporal and tonal structure of utterances by means of rules[1]. Prosody is used here in a somewhat limited sense; the component of voice quality and certain segmentally bound features are excluded. Thus prosody means simply the rhythm and melody of speech.

Rhythm is equivalent to the temporal structure, namely the durations of segments and their interrelationships within larger units such as syllables, phrases, etc. Melody is equivalent to the tonal structure, expressed, for instance, as the Fo-contour which is composed of tonal prosodic features such as accent, intonation type, emphasis, etc. The tonal structure of utterances may be viewed from different angles. From the point of view of the speaker, the tonal contours may be described as either movements or points or levels representing tonal targets. From the point of view of the listener, the tonal structure may be seen in terms of tones or tonal changes associated with certain spectral events such as vowels, consonants, and syllables.

It is the aim of this paper to present an outline of a model for German prosody[2]. I would like to emphasize that this cannot be more than just a first, simple, and incomplete attempt; a great amount of work and research needs to be done to enable us to achieve a more complete and satisfactory understanding of German prosody.

The present outline is based on an acoustic investigation where several important parameters were varied within utterances spoken as one breath group[3]. The second basis for this sketch is to be found in the various models for duration and intonation which have been developed for some other languages (see references). These models served as helpful guide-lines for the more general aspects of the German prosody model.

2. SOME ACOUSTIC DATA OF GERMAN PROSODY

A relatively large material which was read by three university educated female speakers from Northern Germany served as the acoustic basis for the outline of the model for German prosody. The material which is shown in Table 1 consisted of 14 basic utterances[4] the length of which varied as a consequence of the increasing number of accents from 1 to 8. The main syntactic phrase boundaries are indicated.

All utterances were read by each speaker using the following three intonation types: statement (final falling intonation), echo question (same word order, but expressing surprise and

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	utterances													
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.
						#					#				

astonishment, final rising intonation), and information question (inverted word order, final rising intonation). The speakers aimed at giving equal weight to each accent and at producing each utterance as one prosodic phrase without breaking up the longer utterances into several shorter phrases. The utterances containing 2 to 7 accents appeared as pairs differing in the position of the syntactic phrase boundaries. Each utterance was embedded in an appropriate context (cf. Bannert 1983a and for the method Bruce 1977) and read 5 times.

For the sake of convenience, and before outlining the model proper, some basic data concerning certain temporal and tonal aspects of German prosody will be given.

2.1 Sentence and stress group durations

It is well known that segment durations are affected by a number of segmental and prosodic factors. In this respect, the following question is of interest. Are the durations of sentences and stress groups affected by the tonal structures of the different intonation types? This question can only be given a negative answer (for a detailed description cf. Bannert 1983b). The durations of the units of sentence (utterance) and stress group do not vary in a uniform way across the intonation types and the three speakers. No systematic patterns of variation emerge that can be interpreted as a common temporal behaviour reflecting specific linguistic demands shared by all the speakers.

The relationships between the sentence durations of the three intonation types A (statement = reference), E (echo question), and I (information question), according to two methods of evaluation, can be summarized as follows:

$$I \leq A \leq E$$

This means that the three sentence durations are either equal or unequal, i.e. there may or may not be any significant difference between them. If there is a difference, however, the information question has the shortest and the echo question the longest duration. Nevertheless, it should be remembered that the relative differences in sentence duration between the intonation types and across the three speakers are rather small. They vary in a non-systematic way between 0 and 9% of the duration of the statement. As the differences in sentence duration are not systematic and of minor magnitude only, it may be assumed therefore that the differences observed may be perceptually insignificant.

The durations of the stress groups show a similar picture. They do not vary consistently between the three intonation types and across the three speakers. Variation of stress group durations are to be found over the whole utterance. They are not confined to any specific part of the utterance, e.g. the final part where the tonal contour moves in opposite directions (statement vs. questions). Thus it is inferred that the differences in sentence duration observed in the material are the result of a global temporal change which is distributed over the whole utterance.

By means of a graphic representation, the non-systematic pattern of temporal variation at the sentence and stress group level becomes obvious. Speaker B's data are used which are also representative of the two other speakers. Figure 1a shows the relative mean differences of sentence duration between the intonation types statement/information question and echo question/statement of all the 14 utterances. Figure 1b shows the relative mean differences of stress group duration between echo question/statement and their total mean of the longest utterance with 8 accents for all three speakers.

In summary, then, it is concluded that the differences in sentence and stress group duration between the three intonation types are relatively small and by no means systematic or consistent. Therefore, it seems justified, at least as a first approximation, to attribute no obligatory (linguistic) status to the context feature "intonation type". Instead, the duration of intonation types may vary individually. This variation may therefore be considered optional to be incorporated as such into a later stage of the prosody model.

2.2 Fo-contours

As an illustration of the tonal features and their systematic characteristics, Figure 2a shows the normalized and superimposed Fo-contours of the utterance with 6 accents (Der

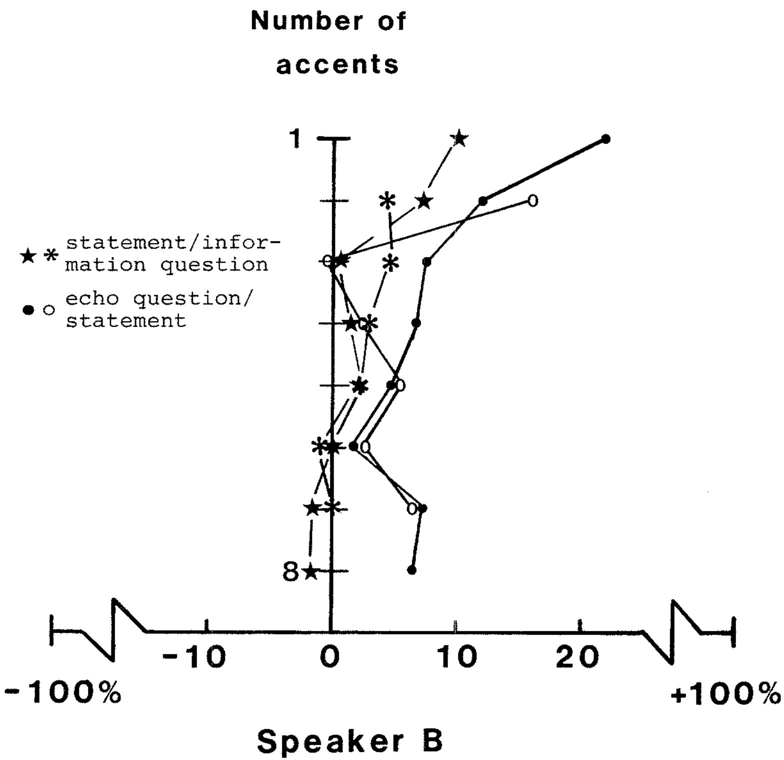


Figure 1a. Relative mean differences of sentence durations between sentence types in the 14 utterances with 1 to 8 accents, statement duration being the reference.

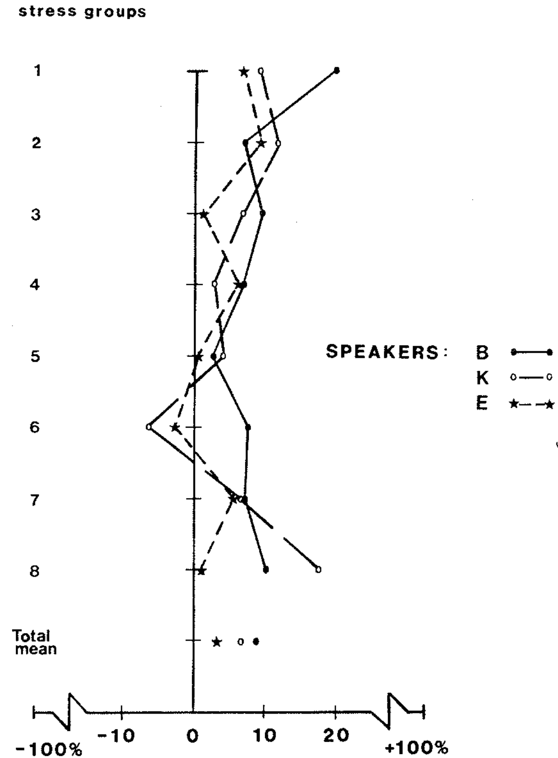


Figure 1b. Relative mean differences of stress group durations in sentence 14. See text.

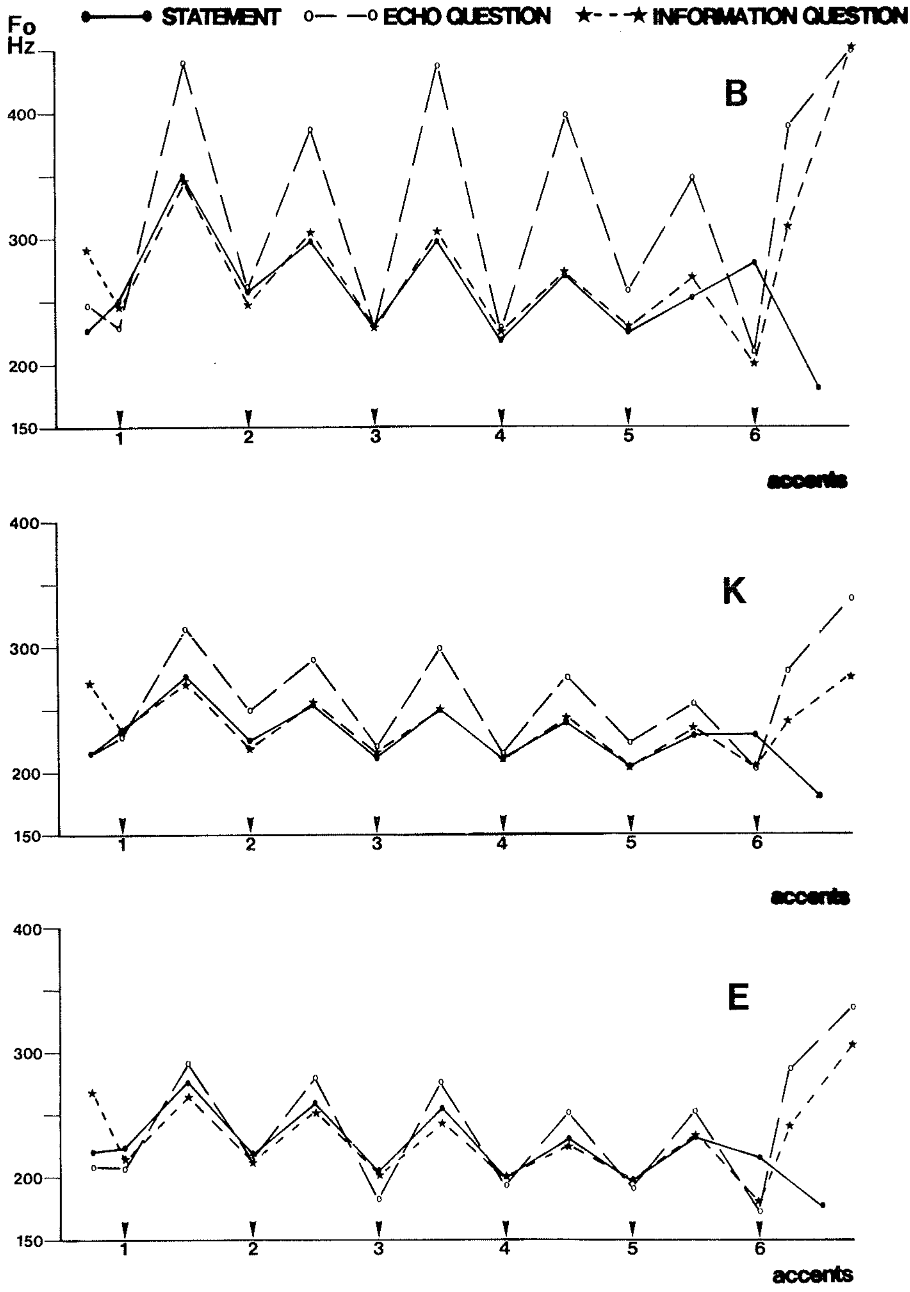


Figure 2a. Normalized and superimposed Fo-contours of the utterance 11 with 6 accents, spoken as a statment, echo question, and information question by three speakers B, K, and E.

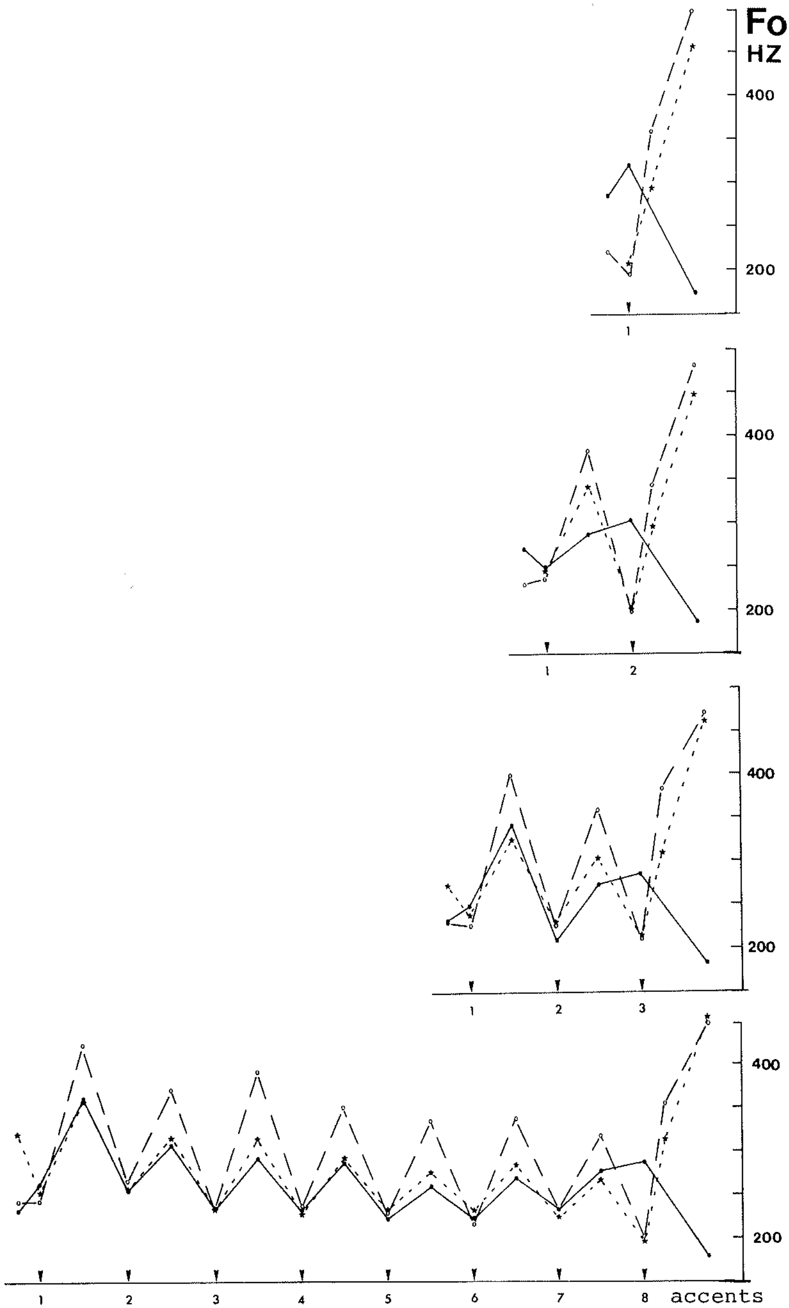


Figure 2b. Normalized and superimposed Fo-contours of four utterances containing 1,2,3,and 8 accents of speaker B. Line-up point is the end of the tonal contour bringing out the systematic structure of the contours.

lullende Müller in Lingen will die längeren Männer immer Lümmel nennen) as the three intonation types for each speaker. Except for one instance, the tonal expression of each accent consists of a tonal rise starting from an Fo-minimum in the consonant preceding the accentuated vowel and ending in an Fo-maximum in the following unstressed syllable. There is only one exception, namely the final accent of statements which is manifested tonally as a fall. It will be recognized that the tonal contour of the two final accents in statements (rising-falling) equals the hat pattern typical for Dutch intonation ('t Hart and Cohen 1973). The tonal differences between the intonation types, above all, are to be found locally at the end of the utterances. The statement is characterized by a rather steep fall in the final accentuated vowel, continuing to the very end of the utterance and ending in a very low tone, the speaker's absolute Fo-minimum. The questions, on the other hand, end in a rising contour reaching a very high tone well above the tonal peaks of the preceding accents. In general, the echo questions show an increased tonal range in the accents, the size of which varies between the speakers. Considering the three superimposed tonal contours representing the three intonation types, one may get the impression that the tonal movements started from a common tonal floor. In most cases, the Fo-minima of all three intonation types differ only slightly. The Fo-maxima, however, may differ greatly, especially for speaker B. In other words, the tonal range between intonation types and speakers varies considerably, the Fo-minima serving as a fixed floor.

It is very easy, starting from Figure 2a, to predict the tonal contours of the other utterances containing 1-5 and 7-8 accents. The Fo-contour of the utterance with 1 accent (Die Männer) by and large corresponds to the contour associated with accent 6, i.e. the final accent, in Figure 2a. The tonal contours of utterances with 2 accents correspond to those of accents 5 and 6; those with 3 accents to those of accents 4, 5, and 6, etc. To be added is the very beginning of the contours, i.e. the initial part preceding the Fo-minimum of the first accent, as shown in Figure 2a (cf. also Bannert 1983a, b).

The systematic nature of the tonal structure when the number of accents is increased, and thus the length of the utterance is also increased, is brought out very clearly in Figure 2b where the normalized and superimposed Fo-contours of utterances with 1, 2, 3, and 8 accents for Speaker B's intonation types are shown. Going from 1 to 8 accents, the final part of the contour, i.e. that contour which is associated with the final accent, is maintained while rising-falling contours, corresponding to each non-final accent, are added before the final accent.

3. THE COMPONENTS OF THE PROSODY MODEL

It is the aim of the prosody model to generate the temporal and tonal structure of a given utterance by means of rules, i.e. by applying phonetic information in a systematic way to a given input structure. In principle, this goal may be achieved in different ways. Two questions, however, should be answered in advance as they have far-reaching implications for the design of such a model.

Firstly, the two-fold aim, namely to generate two structures - one in the dimension of time and the other in the dimension of fundamental frequency, which, in turn, is a function of time -, represents some special difficulties in itself. From the very start, the basic question has to be asked whether the two structures exist each one in its own right, or if one of them can be derived from the other. The answer to this question must have great significance for the structure of the model. If one of the two structures can be derived from the other, then it would not only mean technical savings in the design of the model, but such a relationship, above all, would capture the very nature of the relationship between the temporal and tonal dimensions in prosody.

Not until recently has this two-sided face of prosody attracted the interest of the researchers of prosody. Models for duration and intonation were aimed at generating either the temporal or tonal structure. Over the past few years, however, some alleged connections of tonal movements with segmental lengthening have been reported, and a model which treats vowel duration as a dependent variable of the change of Fo over this segment was presented by Lyberg (1981).

Yet, attractive as this idea might be at first glance and fruitful as it has been for prosodic research during the past few years, it has to be stated that Lyberg's model, in principle and in its absolute form, does not seem to hold against evidence of various kinds presented for Central Swedish (Bruce 1981), Central and Southern Swedish (Bannert 1982a), German (Bannert 1982b), and Danish (Thorsen 1980). It is of course true that nobody will doubt that there are certain dependencies between parts of the temporal and tonal structures. However, their mutual effects are rather small and their interrelationships should be characterized as mutual modification rather than dependence.

The second basic question concerns the input to the model, especially its phonological information. Which phonological features are to be contained in the input which is a linguistically abstract structure showing, among other things, semantic, syntactic, and pragmatic feature specifications? In this instance, too, different alternatives for a solution are possible and plausible, depending on the degree of redundancy that is to be allotted to the phonological structure.

For the present outline I assume that the input shows the following characteristics: under the syntactic category symbols, like N, ADJ, PRON, etc., discrete phonological elements (segments, phonemes) of the canonical lexical units are ordered linearly and specified as to their spectral features. Furthermore, the input contains the prosodic features STRESS and ACCENT of syllables, QUANTITY of vowels, and intonation type COMPLETED of the whole utterance. The input also contains information on various boundaries and the pragmatic features of CONTRAST of words and EMPHASIS of prosodic phrases and PHRASE INTONATION which signals the division of an utterance into minor prosodic units, the prosodic phrases, conditioned by several factors such as tempo, syntactic structure, etc. Thus the input, having the form of an abstract, linguistically defined structure, contains all the necessary phonological features, both spectral and prosodic, and various boundaries. The prosodic features originate from lexical, morphological, syntactic, semantic, and pragmatic features and are inserted into the structure in their appropriate positions on different levels (segment, syllable, word, phrase, utterance, text).

As we can see, the input, as a linguistic and pragmatic structure, is abstract and therefore not specified in the temporal or tonal dimension. It is the task of the prosody model, by applying rules, to generate the concrete temporal and tonal structures in the time-fundamental frequency-field.

Accepting these assumptions, a possible conception of the prosody model is chosen which is shown in Figure 3. The model consists of three main components: the basic temporal and tonal components and the modification component. They may operate in the order mentioned. Thereby two aspects are stressed. Firstly, the temporal and tonal components operate separately, each of them generating a basic temporal and a basic tonal structure. Secondly, tonal features or gestures (tonal changes) like rises and falls or the points or levels HIGH and LOW, and the tonal range WIDE, only exist in reference to events in the time domain, i.e. tonal features are only meaningful if they are associated with temporal units like consonants, vowels, VC-boundaries or syllables.

The basic features of the prosody model outlined here are not unique for German but may also be valid for other languages. Prosodic features that are typical for German, however, are to be found in the phonology of German and in the rules proper which are contained in the temporal and tonal components (the re-writing rules and the intonation algorithm) and in the modification component.

It is the task of the modification component to adjust the basic temporal and tonal structures to their mutual characteristics and demands, and to adjust for assimilations and tempo constraints. Such adjustments are: temporal effects on the tonal contour, for instance the number of unaccentuated syllables between two accentuated ones; conversely, tonal effects on durations, for instance a rising tone may increase segment duration to a certain degree; the so-called micro-

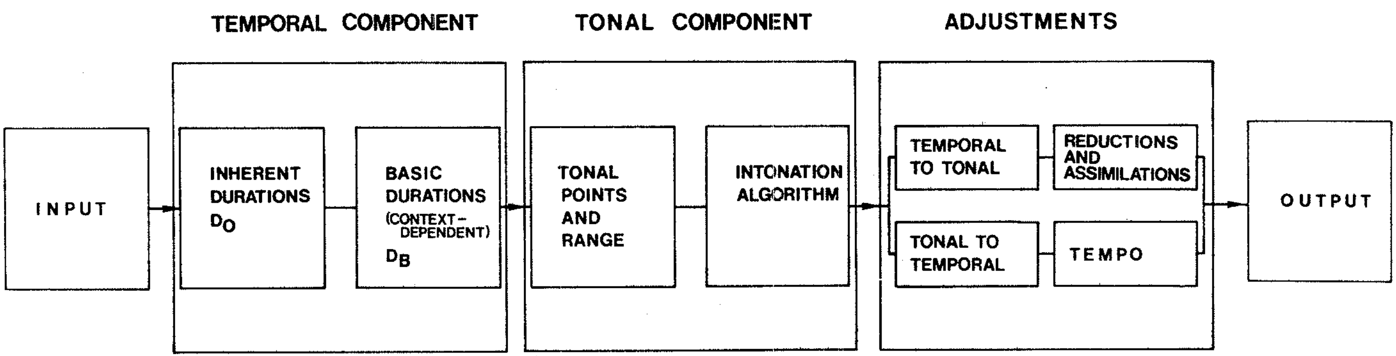


Figure 3. The basic structure of the model for German prosody.

prosody, the effects of the spectral features of the segments on the Fo-contour; speech tempo, reductions and assimilations causing rather drastic alterations in the prosodic structure (and, of course, in the spectral structure, too). It is, however, not self-evident to include tempo and assimilation in the modification component. They could operate separately as well. And, last but not least, a device is needed in order to introduce certain optional (individual) features into the prosodic structure.

As a concluding remark concerning general features of prosody models, I would like to point out another possible conception of a prosody model. Being totally different from the present outline, it does not generate the prosodic structure of an utterance in a linear, step-by-step fashion. Instead, this model derives the target structure in an open procedure, simultaneously on several levels, each level, at any moment, having access to all information on all the other levels. This hint at a dynamic sketch of a prosody model, however, will not be pursued here, although this concept obviously seems to be very attractive.

4. GENERATION OF THE BASIC TEMPORAL STRUCTURE

As I have followed a well-known approach[5], the generation of the basic temporal structure will be sketched here only very briefly. The segment is the basic unit of the temporal structure and its duration is calculated with respect to the different contexts and the domain it appears in.

In the temporal component of the model, the linguistically specified input is used to generate the basic temporal structure in two steps. Firstly, the discrete spectral units are assigned the inherent duration Do which is listed in the catalogue of basic durations. Which inherent durations are contained therein, has yet to be established (but cf. Rietveld 1975). In any case I assume that two temporal vowel categories have to be discerned as a consequence of the prosodic feature of QUANTITY, specified as ±LONG in the lexical units.

In the second step, the inherent duration Do may be modified as a consequence of certain context features. Such contexts are the syllable features STRESS and ACCENT, phrase boundaries, consonant clusters, CONTRAST and EMPHASIS of words and phrases, position within stress groups and phrases. Each context may have the effect of increasing or decreasing the inherent durations, and the size of this change can be expressed by a factor. If several different context conditions are applicable to a given segment or syllable, the final size of the durational change is the sum of all these factors. Thus the basic duration D_B of a given segment is

$$D_B = (a + b + c + d + \ldots + n)\ Do$$

As was shown above (section 2.1), the intonation type has no systematic effect on sentence and stress group durations. Therefore, the context component contains no factor that would apply to a given intonation type.

The concrete basic temporal structure is then added to the original abstract structure of the input. Thus the structure which leaves the temporal component serves as the input to the tonal component. It contains the linguistic features of the input and the phonetic temporal structure in the time dimension.

5. GENERATION OF THE BASIC TONAL STRUCTURE

The generation of the basic tonal structure will be dealt with in more detail. This, in the first place, is due to some novel features in the intonation algorithm. The basic tonal component also operates in two steps[6]. Firstly, the tonal prosodic features, such as accent, intonation type, emphasis, etc., are transposed into a notation of tonal points or levels and of range. Secondly, operating on this information, the intonation algorithm then generates the tonal structure in the time-frequency-field in five steps, thus expanding the amount of concrete information present in the total structure of the given utterance.

5.1 The tonal re-writing rules

Apart from the rhythmic (temporal) prosodic features STRESS and QUANTITY of syllables and words, and different boundaries, at least the following tonal features have to be assumed for

German. From a phonetic point of view, they are described as tonal points or levels HIGH and LOW and as tonal range WIDE:

A. Tonal points or levels

±AKZENT: Accent, manifested in the fundamental frequency dimension, a tonal rise from an Fo-minimum to an Fo-maximum throughout the accentuated vowel. The final accent of a statement, however, appears as a fall, thus reflecting the superiority of the feature of intonation type (sentence intonation), namely a low tonal end, over the syllable or word feature of accent.

±ABGESCHLOSSEN: Completed, feature of the intonation type (sentence intonation), dominating the syllable or word feature of accent (and thus of a smaller domain, lower in the hierarchy), manifested in the fundamental frequency dimension, mainly local manifestation at the end of an utterance[7]: low in completed utterances (often statements), high in non-completed utterances (utterances where questioning is signalled by tonal means).

B. Tonal range

±KONTRAST: Contrast, concerns one word, pointing out one lexical element, manifested tonally by an enlarged range of the Fo-change in the accentuated syllable, accomplished by a heightened Fo-maximum.

±TEILNAHME: Involvement (emphasis), concerns the larger units of phrase and sentence, manifested tonally by enlarged ranges of the Fo-changes of all the accents, accomplished by heightened Fo-maxima.

In the first step of the tonal component, a re-writing algorithm transforms the specification of the tonal prosodic features into a notation consisting of the points HIGH or LOW and the range WIDE. The following rules are to be applied (the arrow means "re-write as"):

A'. Tonal points or levels

+AKZENT → H(IGH) in the pre-accentuated consonant, applies to the final accent in completed phrases (statements) only.

→ L(OW) in the pre-accentuated consonant, otherwise.

+ABGESCHLOSSEN → L(OW) in the final vowel (or syllable)

-ABGESCHLOSSEN → H(IGH) in the final vowel (or syllable)

B'. Tonal range

+KONTRAST → W(IDE) in this LOW or HIGH

+TEILNAHME → W(IDE) in all non-sentence final LOWs and HIGHs.

By applying the tonal re-writing rules, a tonal specification is produced which also reflects the characteristic contrasts between the various prosodic conditions, for example intonation type and involvement. The operation of the tonal re-writing rules is illustrated by the following example taken from the material in the acoustic investigation. The sentence "Der Müller in Lingen will die Männer immer Lümmel nennen" containing four accents is shown as a statement and as an information question (inverted word order):

INPUT

statement

```
[      +ACC         +ACC                +ACC          +ACC              ]
[ Der Müller in Lingen will die Männer immer Lümmel nennen ]   +COMPL
```

information question

```
[           +ACC         +ACC         +ACC          +ACC              ]
[ Will der Müller in Lingen die Männer immer Lümmel nennen ]   -COMPL
```

TONAL RE-WRITING RULES

statement

```
     L         L               L            H         L
Der Müller in Lingen will die Männer immer Lümmel nennen
```

information question

```
          L         L          L            L         H
Will der Müller in Lingen die Männer immer Lümmel nennen
```

The tonal patterns, being the result of the tonal re-writing rules, are now altered into concrete Fo-contours by the intonation algorithm.

5.2 The intonation algorithm

The tonal patterns, specified in terms of L, H, and W and tied to spectral events in the time dimension, are transformed into Fo-values in the fundamental frequency dimension by the intonation algorithm, the second part of the tonal component of the model. This is done in five steps where the basic temporal structure is the necessary prerequisite for the operation of the intonation algorithm. Thus the abstract linguistic and the temporal phonetic structure (information) of the utterance is extended by the addition of the concrete tonal structure.

The intonation algorithm is designed in such a way as to generate the Fo-contour of an utterance from below, starting from the tonal floor of the contour, i.e. its Fo-minima, especially the absolute Fo-minimum.

The application of the intonation algorithm is shown in Figure 4. The same sentence as above "Der Müller in Lingen will die Männer immer Lümmel nennen" containing four accents of equal weight, to be spoken as a one-phrase statement and

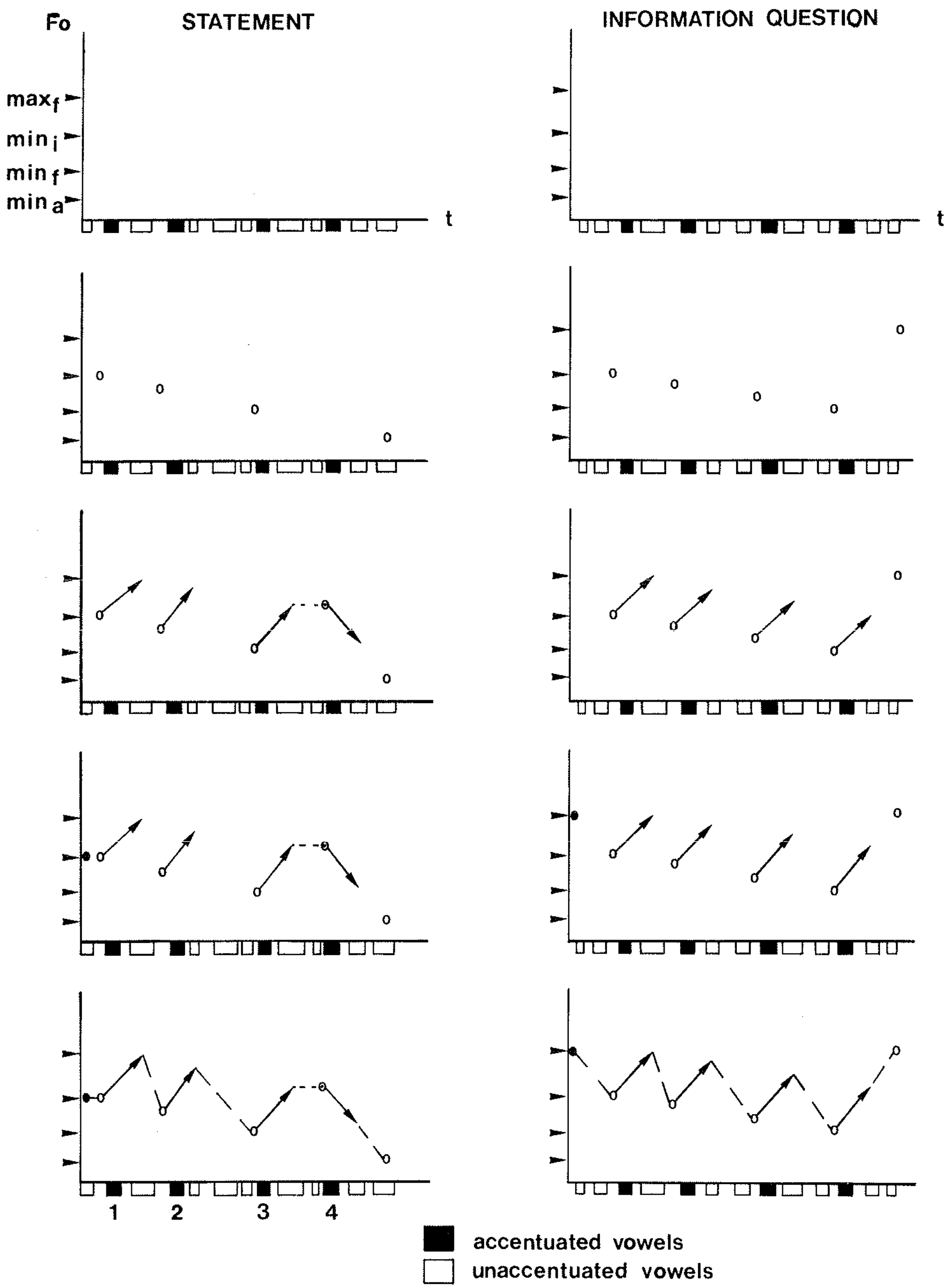


Figure 4. The intonation algorithm illustrating the generation of the tonal structure for a sentence with 4 accents as a statement and an information question.

information question, serves as an example. Where the accent pattern is concerned, this sentence corresponds to other sentences, for instance, "Das Gemälde von Kandinsky ist gestern versteigert worden."

STEP 1

In the first step, certain tonal targets (points or levels) have to be determined in order to fix the frame of the Fo-contour of the utterance. These target values are speaker-specific and have to be defined on an appropriate scale[8].

Three values correspond to accent minima (L), the fourth corresponds to the H of the question intonation:

1. $Fo\text{-}Min_a$: The absolute Fo-minimum, the lowest point at the end of a statement. It seems to represent the anchor of the whole tonal structure, the tonal movements (the tonal changes) and the tonal relationships within a given utterance.
2. $Fo\text{-}Min_f$: The final Fo-minimum, the lowest point of the final accent of questions.
3. $Fo\text{-}Min_i$: The initial Fo-minimum, the lowest point of the first accent. This Fo-value is highest for the statement and lowest for the echo question.
4. $Fo\text{-}Max_f$: The final Fo-maximum, the highest point at the end of questions. It may be equal to or higher than the first Fo-maximum.

The tonal points 3 and 1 or 2 respectively define the range of the Fo-minima, the bottom line, the tonal floor, throughout the prosodic phrase. All these points need not be reached in every utterance observed (cf. step 3). They function first of all as auxiliary points. As one global expression of the intonation type and involvement, point 3, the initial Fo-minimum of the first accent varies. Compared to statements, this Fo-minimum is lower in information questions and lowest in echo questions. As a result of this, the range of the bottom line, i.e. the declination, also varies.

STEP 2

In the second step, the fundamental points of the Fo-contour are inserted. These are the Fo-minima of the accents and the end point of the utterance. The Fo-minima are located in the middle of the consonant preceding the accentuated vowel. First the following LOWs are specified from left to right in the utterance: the first LOW is given the value of $Fo\text{-}Min_i$ (point 3), the last LOW is given the value of $Fo\text{-}Min_a$ in statements (point 1) or the value of $Fo\text{-}Min_f$ in questions (point 2).

Second, assigning the intermediate Fo-minima, two cases are to be distinguished depending on the number of accents in the utterance:

1. Up to 5 accents: Divide the tonal interval between $Fo\text{-}Min_i$ and $Fo\text{-}Min_a$ or $Fo\text{-}Min_f$ into equal steps and insert these values into the intermediate pre-accentuated consonants.

2. More than 5 accents: Divide the tonal interval of the bottom line into equal steps. Insert these values, except the last but one, into the intermediate pre-accentuated consonants. Insert a higher value for the last but one accent.

In questions, the value of $Fo\text{-}Max_f$ is inserted for the final HIGH (point 4).

STEP 3

Upon this tonal floor, the tonal movements (rises) associated with the accents are inserted. Thus each Fo-minimum, except the $Fo\text{-}Min_a$ at the end of statements, represents the starting point of the tonal accent movement. Starting from a minimum, the movement can only go upwards, i.e. it must be a tonal rise, as the tonal floor is fixed.

These constant tonal movements are to be seen as tonal differences or tonal steps between syllables. Starting from an Fo-minimum in the consonant preceding the accentuated vowel, the tonal movement ends in the following, unstressed vowel (or in the CV-boundary). This is also true of the final accent of statements, although manifested as a tonal fall. The exact position of the end point of the tonal movement may vary, due to different temporal conditions such as the number of unstressed syllables following the accent or quantity (cf. Bannert 1982b).

There is, however, one accent movement that travels in the opposite direction, namely the final accent fall of statements. This final tonal fall has to be seen as a natural consequence of two factors: the high end point of the preceding accent rise and the low end point of the tonal contour in statements. Due to this tonal conditioning, the final accent fall can be derived easily using the tonal information available at this stage. As is apparent from the Fo-contours in Figure 2, the starting point of the final accent fall shows approximately the same Fo-value as the end point of the preceding accent rise[9]. Therefore this end point value is inserted into the final pre-accentuated consonant and the tonal movement is directed downwards. It terminates in the following unstressed syllable and shows the same speed of Fo-change as the rising movements.

The Fo-change in the accentuated syllable will be expressed by the factor v of a base unit, the absolute Fo-minimum ($Fo\text{-}Min_a$) which seems to be the lowest point or level of the normal speech register of a speaker. The factor v, however, is not to be taken as an exact and fixed value; it is rather to be thought of as a certain Fo-range or interval from which values may be chosen for a given utterance.

If the specification WIDE appears with an accent, the size of the tonal movement v is increased by the factor w which also is speaker-specific and greatly variable. This factor, too, constitutes an Fo-range rather than a fixed value. Thus the total tonal movement V_G in an accent specified for LOW and WIDE is

$$V_G = (v + w)\, Fo_{Min_a}$$

The size of the tonal movements associated with the accents may vary throughout an utterance. This variation, it seems, expresses the concurrence of various features. In order to be heard as an accent, the tonal movement has to exceed a certain minimal value, a tonal threshold, defined from a perceptual point of view and dependent on certain contextual tonal circumstances. Larger tonal excursions of the accent movements are the result of such concomitant tonal features as contrast, emphasis, and phrase intonation. Therefore, the size of the tonal movements may be considered the sum of all contributions of tonal features and, as is the case with the basic durations (cf. section 4), may be defined as

$$V_G = (v + w + k + j + \ldots + n) \, Fo_{Min_a}$$

This means that the total tonal movement V_G associated with an accent is calculated with reference to the absolute Fo-minimum of the speaker. The proportion of the reference value is the sum of contributions of all the tonal features appearing together with a given accent: the minimal movement of the accent v, the range w of the feature WIDE reflecting involvement, etc. However, given equally strong accents in an utterance, I assume the tonal movements to be of equal size on an auditory scale.

It seems obvious that every change in the size of this equal and constant tonal movement signals some prosodic information, namely a deviation from the normal, i.e. expected situation of equal weight from accent to accent. A decrease of the tonal movement means a weakening of the accent decreasing the prominence of the semantic element; an increase, however, expresses a strengthening of the accent thus creating a greater prominence for this semantic element. Data concerning the tonal differences in the accents and the position of the Fo-minima suggest a strong interrelationship between these tonal features and the semantic structure of (parts of) the utterance[10]. Considering the utterances which contain the local adverbial in Lingen, it is found for all three speakers that the accent movement in Lingen is considerably larger than that of the accent in Müller. It is quite obvious that the latter accent appears somewhat weakened. This inequality between accents is shown in Figure 5 for each intonation type and for speaker B and is representative for all speakers. The tonal structures of two sentences are shown which contain seven accents but which differ in the position of the phrase boundaries: "Der lullende Müller # will die längeren Männer

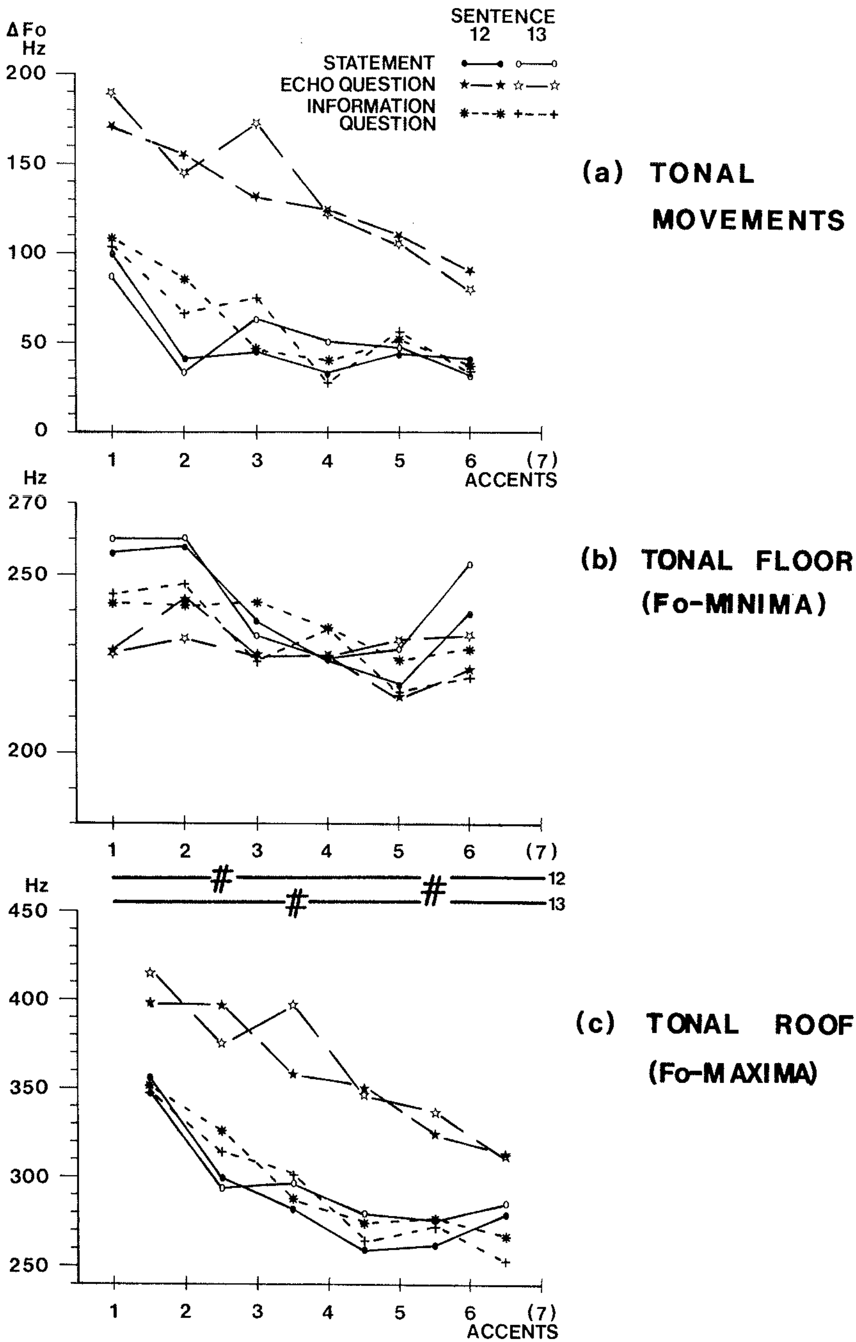


Figure 5. Variation of tonal characteristics due to syntactic and semantic factors. Speaker B. See text.

in der Menge immer lungernde Lümmel nennen" and "Der lullende Müller in Lingen # will die längeren Männer immer lungernde Lümmel nennen". The subject of the first sentence consists of a phrase containing two accents, that of the second sentence consists of a phrase containing three accents (cf. Table 1). In part (a) of Figure 5, the tonal differences (ΔFo) in each accent are shown, in part (b) the bottom line, the tonal floor (Fo-minima), and in part (c) the peak line, the tonal roof (Fo-maxima) are to be seen. It is evident that the tonal differences associated with the second and third accents (<u>Müller</u> and <u>Lingen</u>/<u>längeren</u> respectively) have implications for all three tonal characteristics, namely the Fo-movements, the bottom line, and the peak line. This entails a deviation from the gradually falling or uniform basic tonal structure.

STEP 4

To complete the tonal skeleton, the starting point of the whole Fo-contour is inserted. For statements and echo questions with an unstressed syllable preceding the first stressed one, a value near the initial Fo-minimum is chosen as a first, rough approximation. For the information question showing inverted word order, the starting point is chosen in relation to the initial Fo-maximum in such a way that the starting point equals the first Fo-maximum.

STEP 5

Finally, these fragments or basic elements of the tonal structure have to be connected together in order to arrive at the complete tonal contour of the given utterance. This can be done by straight lines as in Figure 4 which are smoothed out later on or by a cosine function.

The output from the intonation algorithm, the generated basic tonal contours, together with the basic temporal structure, will then be adjusted in various ways in the modification component. However, these adjustments are rather small where size and degree are concerned. The basic features of the Fo-contours, in principle, will not be altered. This means that, in spite of all modifications, the basic character and features of the prosodic structure are maintained.

6. THE MODEL AND INTONATION THEORY

The prosody model presented in this outline, after due amendments and completion, makes the claim of generating the temporal and tonal manifestations of the pure linguistic-prosodic features of a given German sentence, excluding the expression of emotions and other non-linguistic features. Operating on the abstract linguistic information of a sentence, the model components, step-by-step, will produce the concrete phonetic structure in the time and fundamental frequency dimensions.

The linguistic input of this prosody model contains all the prosodic features necessary to generate the prosodic structure of the given utterance. This does not mean, however, that there exists theoretical agreement about which prosodic features are to be assumed for German and how they are distributed in a sentence. Together with research into this issue from a purely general linguistic point of view (cf. Klein and von Stechow 1982, among others), the prosody model sketched here permits controlled experiments to be conducted in order to determine relevant prosodic features in a feed-back fashion. Even at this early stage, however, it seems appropriate to comment on some notions of intonation theory, especially for German as it is described by von Essen (1956).

There are no such independent features/elements like focus, sentence accent, "Schwerpunkt" (centre of gravity) or "progredient" intonation (cf. von Essen 1956) in this model. Rather they are dependent features since they are derived from the model elements[11]. It seems plausible, however, that there exists a difference between focus and contrast in German.

Focus or sentence accent or "Schwerpunkt" is an automatic result of accent and position. It always appears on the final accent of an utterance, be it a completed intonation (statement) or an uncompleted intonation (question). What is crucial, however, is that the following part of the contour remains undisturbed. No tonal movement (protrusion), except the gradually falling or rising declination, may occur after the focus movements. This final part of the utterance without any accent-like tonal movement is called "Nachlauf". From a semantic point of view, it is very often a mere repetition of semantic elements and could readily be deleted. In other words, there are no accents following the sentence accent. Sentence accent is the final accent of a phrase, and together with the rest of the contour, also expressing intonation type, it is heard as "heavier, stronger" than the preceding, non-final accents.

This effect is easily shown by manipulation of utterances by means of speech synthesis. For instance, starting from an utterance containing four accents of equal weight, e.g. "Der französische König war ein launischer Geselle" (The French king was a capricious fellow) and moving the final fall for the statement or the rise for the question in the accent positions to the left, while keeping the Fo-range of this final change constant, the focus (Schwerpunkt) is also moved to the left. The result of this manipulation is a shifting of weight or attention to different parts (semantic elements) of the utterances. Another effect is a shift of context for these utterances. This shifting of focus is illustrated in Figure 6a for statements and in 6b for questions.

In (1) of Figure 6a, the four accents of the statement are of equal weight. This statement may appear as the first sentence of a text, presupposing the general knowledge that there are kings of different countries. Although the accents are equal, each of them reflecting new information, the final accent is heard as the focus. This might be due to two circumstances. First, the final tonal movement (fall or rise) in a tone contour fulfills a particular role in signalling the intonation type and thus the global structuring of discourse. Second, on purely acoustic grounds, the tonal movement (being located in the lowest part of the tonal range of a statement) is perceived as large. The final rise constitutes a relatively large Fo-change from the lowest part of the Fo-range upwards. Both Fo-changes thus give rise to the auditory impression of a strong tonal element. The final tonal movement may be considered the combined result of two factors: the accent movement proper and the movement towards the very end point of the utterance.

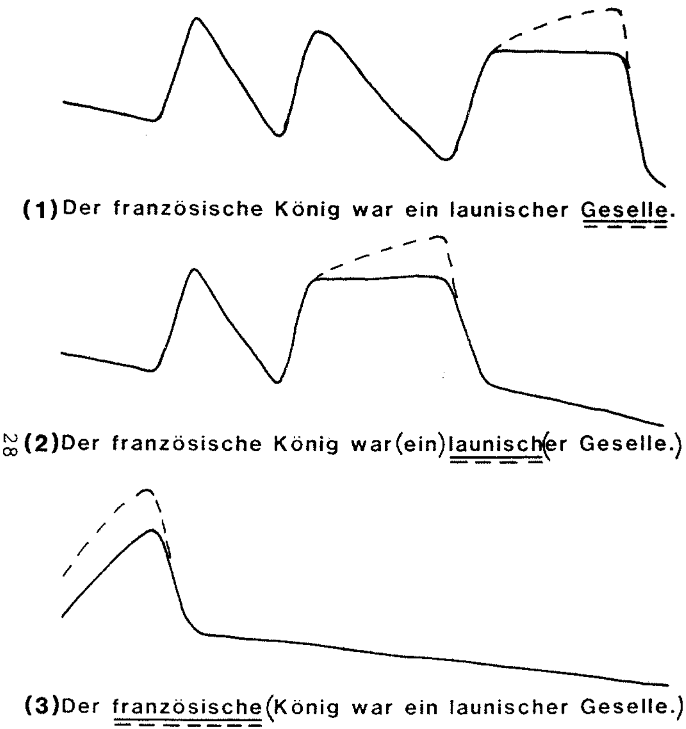


Figure 6a. Shifting of focus (Schwerpunkt) and contrast from right to left in a statement.
Underlined focus word: ———
Underlined contrast word: - - -

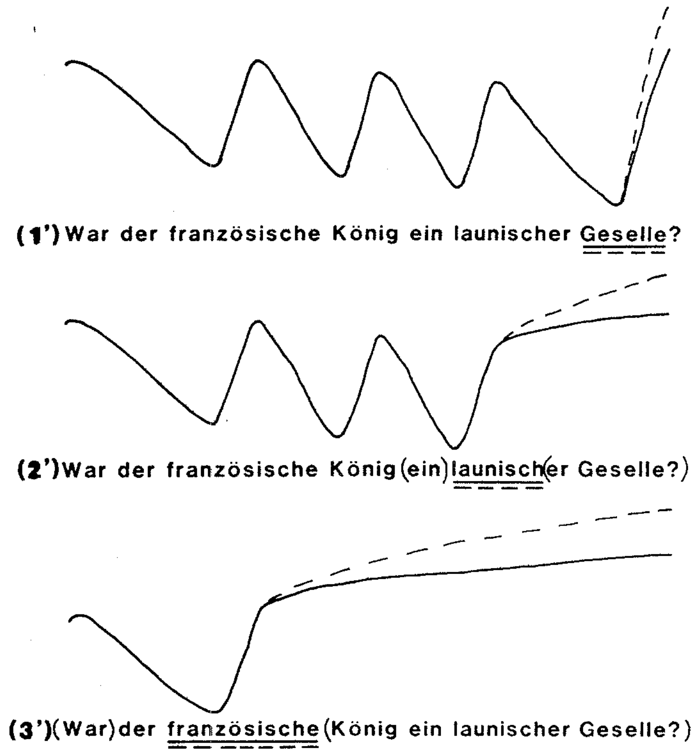


Figure 6b. Shifting of focus (Schwerpunkt) and contrast from right to left in a question.

In (2) the focus is on launisch, in (3), however, on französisch. Shortened (elliptic) but equivalent versions of these two statements would be for (2) "Der französische König war launisch" and for (3) "Der französische". The repetitive parts of the original complete sentences are put into parantheses in Figure 6. The shortened versions could be uttered in the following contexts: "Was für ein Geselle war der französische König?" for (2), where the quality of the French king is in the semantic focus, and "Welcher König war ein launischer Geselle?" for (3), where the nationality of the king who was a capricious fellow is in the centre of interest. Correspondingly, the position of sentence accent is shifted to the left in questions (1'), namely to launisch in (2') and to französisch in (3').

In contrast to the sentence accent, being the utterance-final instance of several equal accents, a semantic element may appear under contrast, which means that this element is particularly pointed out by the speaker either as a paradigmatic device expressing a meaning not expected or unusual in this position or as a correction of a meaning in the answer to a statement previously uttered by the listener. The accent of this emphasized semantic element is stronger than the preceding ones, if there are any. Contrast on a word (a lexical element) is signalled tonally by a larger Fo-change associated with the word accent which is achieved by a higher Fo-peak. This is illustrated by the dashed line in Figure 6a, b (cf. also Bannert 1982b, Figure 1). A contrasting accent always shows up as the last one in a phrase; there can be no pitch protrusion in the Fo-contour following the contrastive accent. In order to bring out the contrast more clearly, the preceding accents, if any, may be weakened.

If the whole utterance is to be stressed, this will be done by means of emphasis, i.e. the whole utterance (its accents actually) will be spoken with high involvement from the part of the speaker. Contrast and emphasis are expressed by the same tonal means, namely increased tonal range, i.e. larger tonal movements in the accents, achieved by raising the Fo-maxima (cf. Figure 2). The only difference between these two features, besides their function, is their domain. Contrast, emphasis, and sentence accent are illustrated in Table 2.

In my intonation model, the so-called "progredient" intonation (cf. von Essen 1956) does not have the same status as the intonation types ±COMPLETED which is the tonal feature of a whole utterance containing one or more prosodic phrases. "Progredient" intonation exists in certain contexts only: it never appears finally in an utterance and it is optional. This intonation seems rather to be a device for structuring an utterance by dividing it into smaller units (prosodic phrases) and thus signalling continuation within an utterance. "Progredient" intonation also seems to be associated with a clause boundary which also may be expressed by temporal and phonatory means. Therefore, in my model, "progredient" intonation corresponds to the feature of phrase intonation which is connected with the phrasing of an utterance. However, the tonal expression of phrase intonation is not yet quite clear, although there exist some data in Delattre et al. (1965).

Different strategies for the tonal expression of phrase intonation can be conceived of, namely a local one and a global one, and of course a combination of both. Locally, there may be a tonal jump from the high Fo-contour to the left of the phrase boundary down to the low Fo-contour to the right of the phrase boundary. In this case the abrupt tonal jump from one phrase to another contrasts with the gradual fall between successive accents within a phrase (cf. Bannert 1983a, Figure 1). Globally, there may be a larger tonal movement in the phrase-final accent, which, compared to all the accents in the whole utterance, expresses the division of the utterance into phrases in a complex way. It seems plausible that phrase intonation marks the subordination of syntactic units within a grammatically complex sentence. These two basic strategies are illustrated in Figure 7[12].

The model also accounts for the devices of accent weakening (backgrounding) and deaccentuation (cf. Bannert 1983a).

It is quite obvious that a considerable amount of work remains to be done in order to develop a complete model for German prosody, including Standard German and the dialects. Nevertheless my outline of a model for German prosody does

Table 2. Illustrations of focus (sentence accent, "Schwerpunkt"), contrast, and emphasis in statements in German. ' = lexical accent (accentuated word), * = accent of involvement (contrast on one word, emphasis on the whole phrase). ( ) = repetitive parts of utterances, usually dropped; / = possible substitution.

Situation:	Zwei Nachbarsfamilien A und B treffen sich an einem Samstag vormittag auf dem Parkplatz eines Einkaufszentrums. Es ist Frühling; die Gartenarbeit wartet. In den Nachbardörfern Schwanfeld und Moosbach ist jeweils ein Gärtner.	
SCHWERPUNKT:	Dialog 1: A: Na, was habt Ihr denn heute vor?/Wo geht's denn hin? B: (Wir wollen) zum 'Gärtner in 'Schwanfeld (fahren).	*Beide Akzente gleichwertig. Schwerpunkt automatisch auf dem letzten Akzent.*
	Dialog 2: A: Margit hat vorgestern gesagt, Ihr wollt heute nach Schwanfeld fahren. B: Ja, (wir wollen) zum 'Gärtner (in Schwanfeld) fahren.	*Schwerpunkt auf dem letzten (einzigen) Akzent.*
KONTRAST:	Dialog 3: A: Fahrt Ihr zum Gärtner in Moosbach? B: Nein, (zum 'Gärtner/zu 'dem) in *Schwanfeld!! Ihr wisst doch, der in Moosbach ist viel zu teuer!!	*Korrektur einer Annahme oder verwunderter Ausdruck, betrifft ein Wort (einer Phrase).*
	Dialog 4: A: Fahrt Ihr zum Schmied in Schwanfeld? B: Nein, zum *Gärtner (in Schwanfeld).	*Schwanfeld deakzentuiert, da es auf das kontrastierende Wort folgt. Identisch mit Dialog 6.*
EMPHASE:	Dialog 5: A: Fahrt Ihr zum Schmied in Moosbach? B: Nein, zum *Gärtner in *Schwanfeld!! Habt Ihr schon wieder vergessen?! Wir haben Euch doch erst gestern gesagt, dass wir heute zum Gärtner in Schwandorf fahren!!	*Ausdruck der Teilnahme (Irritation, Verärgerung), betrifft die gesamte Phrase (mit mehreren Akzenten).*
	Dialog 6: A: Fahrt Ihr zum Schmied? B: Nein, zum *Gärtner!! Habt Ihr denn schon wieder vergessen?! Wir haben Euch doch erst gestern erzählt, dass wir heute zum Gärtner fahren!!	*Ausdruck der Teilnahme (Irritation, Verärgerung), betrifft die Phrase, hier mit nur einem Akzent. Identisch mit Dialog 4.*

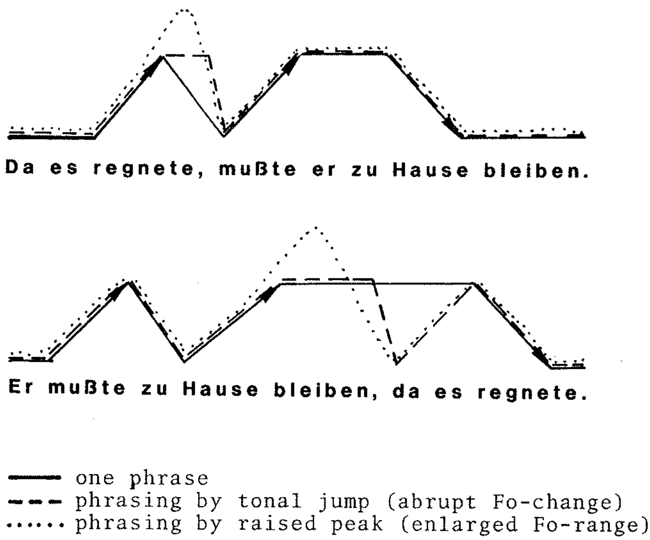


Figure 7. Hypothesized Fo-contours for phrase intonation.

contain some basic aspects relevant to this field of research. Further research work will bring us closer to an understanding of the workings of prosody and thus better qualify us for making intelligible and natural speech synthesis and enable us to more effectively teach German as a foreign language. Future prosodic research will include language-specific German features of Standard German and of the dialects along with features that German shares with other related and non-related languages. As a particular challenge, the close relationship between prosody and syntax and semantics (pragmatics) hopefully will promote the cooperation between phoneticians, general linguists, and psychologists. With these perspectives in mind, we are justified in looking forward to some very exciting and productive research work on German prosody.

FOOTNOTES

1 The book by Isačenko and Schädlich (1970), a translation of their paper (1966), is a pure description of simplified tonal conditions. In spite of its title, it does not contain any rules or algorithms which generate tonal structures.

2 A shortened version of this paper was presented at the Interdisciplinary Symposium "Prosody - Normal and Abnormal" held at the University of Zürich, April 6-8, 1983 and arranged by the Association Européenne Psycholinguistique. It is not the aim of this paper to compare the model proposed here to other models to be found in the literature. For a preparatory work, focusing on the intonation part, see Bannert (1983a).

3 The phonetic aspects, particularly the acoustic prerequisites for the conception of the intonation algorithm and the details of the investigation, are dealt with in Bannert (1983b).

4 Choosing the test sentences, the phonetic demands on a material which is heavily controlled for linguistic and acoustic variables dominated over semantic aspects. Although the test sentences do not constitute the most frequent German utterances, they do form possible linguistic structures. The assumption underlying the strong phonetic choice is represented by the claim that the basic linguistic elements of speech have to be known before it can be meaningful to study the prosody of spontaneous speech with all its numerous variables, the effect of which is not yet known. The longest utterance, containing eight accents, reads in English "The miller in Lingen doing pee-pee will always call the taller men in the crowd loitering louts".

5 The solution adopted here corresponds to the model for segment duration presented by Klatt (1979) and Lindblom et al. (1981), to mention only a few. Exactly the opposite

procedure, working in a top-down fashion, i.e. starting from the largest unit, the sentence, and going down via stress group and syllable to the segments, is to be found in Kohler (1982) and Kohler et al. (1982). Neither will the consequences of the so-called gesture theory suggested by Öhman et al. (1979) be considered here (but cf. Bannert 1982a).

6 The Lund model of intonation (Bruce 1977, Gårding and Bruce 1981) served as a special source of inspiration in outlining the basic tonal component. However, several significant differences, not due to language differences between Swedish and German, were introduced into the model for German prosody. I felt a need to replace the visually-oriented and geometrical conception of lines (base and top line, focal lines) by a dynamic, listener-oriented view of tonal changes and tonal relations (cf. also Bruce 1982).

7 The prosodic primes "utterance" and "phrase" are assumed as given units. Nothing concerning the difficulty of defining them will be discussed here.

8 Mean values for five tonal points in the three intonation types of three speakers are to be found in Bannert (1983b, Table 3).

9 This derivation in statements is only possible if the utterance contains at least two accents. In a statement containing only one accent, the tonal point HIGH of the final and, at the same time, first accent has to be inserted directly by the intonation algorithm.

It cannot be taken for granted that the hat pattern constitutes the only and obligatory tonal contour of the two final accents in statements. Optionally the rise of the last but one accent may be followed by a fall and rise preceding the final accent fall. Thus each accent is characterized by a tonal rise-fall movement, although the tonal gesture is timed differently (earlier) for the final accent. Evidence for this tonal solution is to be found in the Fo-tracings of speaker E.

10 A more detailed account is to be found in Bannert (1983b).

11 This section is based on, among other things, the results of experiments using LPC-synthesis where various utterances of the speakers B and K were manipulated according to the model elements. Thus, by just changing the Fo-contour locally at the end of an utterance, in accordance with the model, a statement is changed into an echo question and vice versa. If the final fall or rise in utterances containing several accents is moved towards the beginning of the utterance, the syllable having received the final tonal movement appears as the "Schwerpunkt" or sentence accent. However, systematic listening tests with naive listeners need to be run.

12 This presentation is based on the Fo-contours of the author. They were observed on a storage oscilloscope connected to a pitch meter. It remains to be investigated, too, how the clause boundary is expressed temporally, for instance by means of final lengthening.

REFERENCES

Bannert R. 1982a. An Fo-dependent model for segment duration? Reports from Uppsala University Department of Linguistics 8, 59-80

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

Bannert R. 1983a. Modellskizze für die deutsche Intonation. Zeitschrift für Literaturwissenschaft und Linguistik 49, 9-34

Bannert R. 1983b. Some phonetic characteristics of a model for German prosody. Working Papers 25, 1-34. Department of Linguistics and Phonetics, Lund University

Bolinger D. 1958. A theory of pitch accent. Word 14, 109-149

Bruce G. 1977. Swedish word accents in sentence perspective. Travaux de l'Institut de Linguistique de Lund XII. Lund

Bruce G. 1981. Tonal and temporal interplay. Working Papers 21, 49-60. Department of Linguistics and Phonetics, Lund University

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

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

Essen O. von 1956. Grundzüge der hochdeutschen Satzintonation. Ratingen

Gårding E. and G. Bruce. 1981. A presentation of the Lund model for Swedish intonation. Working Papers 21, 69-75. Department of Linguistics and Phonetics, Lund University

't Hart J. and A. Cohen. 1973. Intonation by rule: a perceptual quest. Journal of Phonetics 1, 309-327

Isačenko A.V. and H-J. Schädlich. 1966. Untersuchungen über die deutsche Satzintonation. Studia Grammatica VII, 7-67

Isačenko A. and H-J. Schädlich. 1970. A Model of Standard German Intonation. The Hague

Klatt D. 1979. Synthesis by rule of segmental durations in English sentences. Proceedings of the 9th International Congress of Phonetic Sciences. Vol. II, 290-297. Copenhagen

Klein W. & A. von Stechow. 1982. Intonation und Bedeutung von Fokus. Sonderforschungsbereich 99 Linguistik, University of Konstanz

Kohler K.J. 1982. Rhythmus im Deutschen. Arbeitsberichte 19, 89-105. Institute of Phonetics, University of Kiel

Kohler K.J., U. Krützmann, H. Reetz, and G. Timmermann. 1982. Sprachliche Determinanten der signalphonetischen Dauer. Arbeitsberichte 17, 1-48. Institute of Phonetics, University of Kiel

Lindblom B., B. Lyberg, and K. Holmgren. 1981. Durational patterns of Swedish phonology: Do they reflect short-term memory processes? Linguistics Club, Indiana University

Lyberg B. 1981. Temporal properties of spoken Swedish. Monographs from the Institute of Linguistics 6, Stockholm University

Öhman S., S. Zetterlund, L. Nordstrand, and O. Engstrand. 1979. Predicting segment durations in terms of a gesture theory of speech production. Proceedings of the 9th International Congress of Phonetic Sciences. Vol. II, 305-311. Copenhagen

Rietveld A.C.M. 1975. Untersuchung zur Vokaldauer im Deutschen. Phonetica 31, 248-258

Thorsen N. 1980. Neutral stress, emphatic stress, and sentence intonation in Advanced Standard Copenhagen Danish. Annual Report from the Institute of Phonetics, University of Copenhagen 14, 121-205

Thorsen N. 1982. Two issues in the prosody of Standard Danish: The lack of sentence accent and the representation of sentence intonation. Annual Report of the Institute of Phonetics, University of Copenhagen 16, 19-36

Tonal Geography. Geographical Variation in Declarative and Interrogative Intonation along the West Coast of Sweden

Anne-Christine Bredvad-Jensen

INTRODUCTION

Sentence intonation exhibits clear differences between declarative and interrogative intonation in Swedish, see for example Gårding 1979 and Bredvad-Jensen 1980. The dialect in the southern region of Sweden is classified as type 1A in the tonal typology of Gårding (Gårding 1975). The dialect spoken in Gothenburg is a typical representative of western Swedish and is classified as type 2B. For geographical orientation see figure 1 alongside this page. In the area surrounding Steninge (which is situated about halfway between Malmö and Gothenburg) the tonal configuration of declarative intonation is classified as a transitional form, type 2AB (Gårding, Bruce & Willstedt 1981). This classification is based on statement contours. The Gothenburg dialect is traditionally characterized as a dialect with a rising sentence intonation both in statements and in questions. (This is also said about eastern Norwegian which is spoken not far from western Swedish). How are these two rising intonations manifested in statements and in questions? Do they exhibit the same or different kinds of rising characteristics, e.g. local or global effects on the tonal contour?

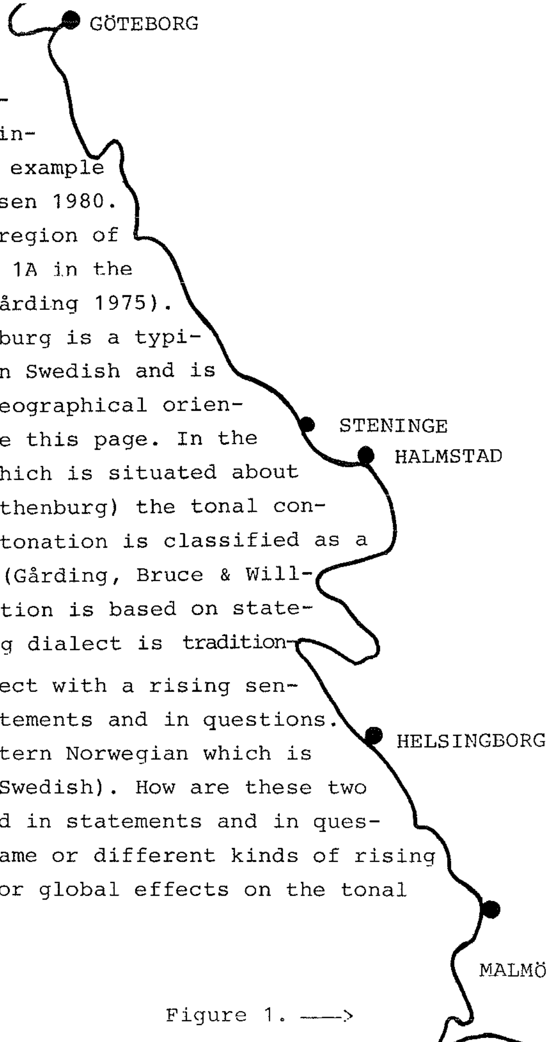


Figure 1. ——>

Are these rising intonations really rising on all occasions, irrespective of e.g. the position of sentence accent? These are the questions I want to study in this paper alongside the following: is it possible to describe the different tonal manifestations of question intonation along the west coast of Sweden as a gradual or continuous tonal development, so that the result of one change at one place is the basis of the next change at another place, or rather will these manifestations be more adequately described as abrupt or discontinuous changes randomly distributed from a geographical point of view?

MATERIAL AND PROCEDURE

The speakers were three women from Malmö, Halmstad (south of Steninge) and Gothenburg, respectively. They were not phonetically trained.

The basis of the speech material that will be discussed here consists of a SVO-sentence consisting of accent 2 words, "Manne lämnar nallarna" ("Manne is leaving the teddy-bears"), which was pronounced both as a question and as a statement. The position of sentence accent, SA, was systematically varied so that either of the three different words in the sentence received SA. Thus three different statements and three different questions were obtained. Each of the sentences was preceded by a context sentence, designed to elicit the appropriate SA in the sentence in question. To obtain continuous Fo-contours which were fairly undisturbed by segmental factors, the test words were composed of sonorant consonants and vowels of the same degree of opening, here non-high vowels.

The sentences were recorded five times each in random order. Tracings of the Fo-contours were made for each sentence and among these one contour could easily be chosen as a typical representative for each of the six sentences.

DISCUSSION OF TONAL CONTOURS

The sentence accented word, focus, is indicated in Figure 2-4, with an arrow in the first CV-boundary of the word. The accent is manifested as a fall beginning in the second consonant in the word in the Malmö and in the Halmstad dialects. In the Gothenburg dialect on the other hand the sentence accent is manifested tonally with a high peak in the first vowel of the postfocal word. The frequency peaks were connected with a straight line, the topline. The direction of the topline indicates the global direction of sentence intonation (1).

MALMÖ

Global characteristics in the tonal contour

As can be seen from Figure 2 there are only small differences between the slope of the toplines (2) before focus in statement and question. Postfocally the topline either falls slightly or is level in the questions, while the corresponding statements exhibit a relatively steeper fall.

All three sentence pairs exhibit an overall frequency level which is higher in the questions than in the corresponding statements.

Local characteristics in the tonal contour

A very pronounced final fall is exhibited in all the three questions, as the starting point of the fall is higher for the questions (due to the higher overall Fo) and as the endpoint is at the same Fo in both question and statement. It seems as if the voice departs from the same resting position in all the sentences regardless of the sentence intonation and that it falls to the same low Fo (final valley) which than could be interpreted as the bottom of the voice.

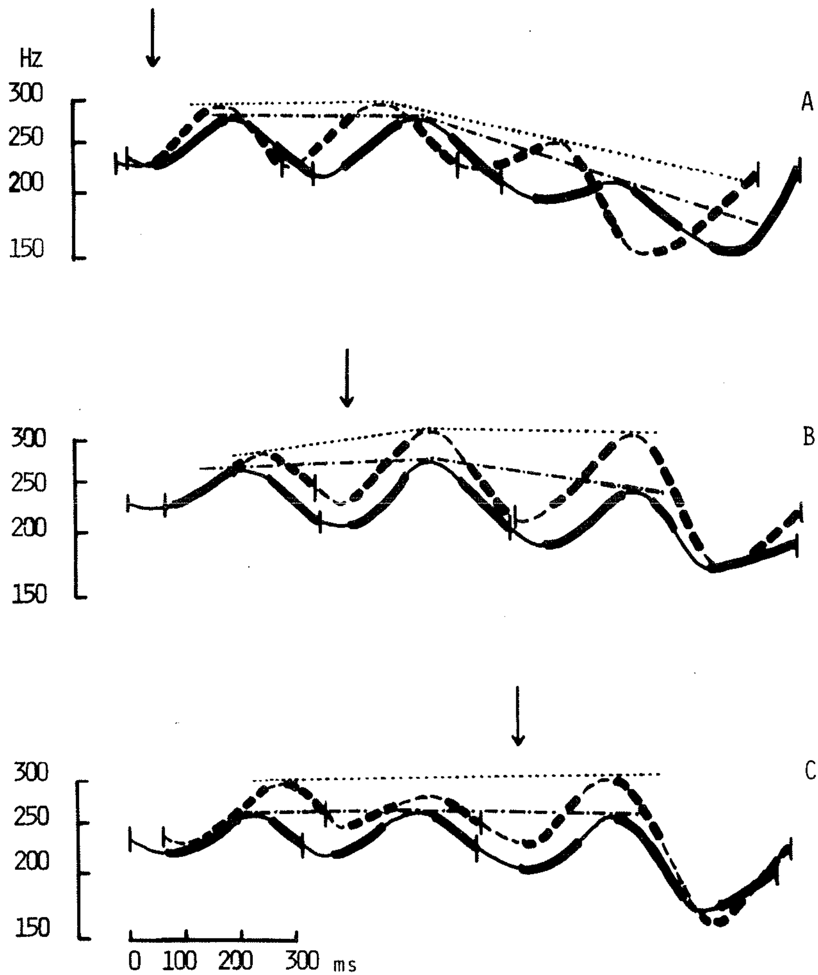


Figure 2. "Manne lämnar nallarna" pronounced as statements (unbroken lines) and as questions (broken lines). Thick contour lines = vowels, thin contour lines = consonants. Vertical bars = word boundaries. A sentence accented word is indicated with an arrow in the first CV-boundary of the word. This boundary is used as a common line-up point in the time domain for each sentence pair. —·—·= topline (statements), ... = topline (questions)(3). MALMÖ

Discussion

Perceptual experiments, using different synthesized versions based on the medially focussed sentence above, showed that the most important cue to the perception of questions in this dialect is the higher overall Fo (Predvad-Jensen 1983). This means that in this case a global intonation characteristic overrides a local one, as it is clear that the higher overall Fo in the qestions is not preserved in order to make possible the pronounced final fall but that the high overall Fo in the questions is preserved because it is important per se.

HALMSTAD

Global characteristics in the tonal contour

The topline ascends postfocally in the questions whereas it descends in the statements thus indicating the difference between a global rising intonation versus a global falling one, as can be seen in Figure 3. This difference is even implied in the final part of the last word in the finally focussed sentence pair (pair C).

The overall Fo is higher in the questions than in the corresponding statements.

A frequency compression of the tonal movements before and after focus can clearly be seen in the questions in comparison with the statements.

Local characteristics in the tonal contour

A wider frequency expansion of the tonal movements in connection with focus is exhibited in the questions as compared to the statements.

The tonal configuration in the questions displays a final rise ending at a fairly high Fo whereas the tonal configuration in the corresponding statements shows either a pronounced final

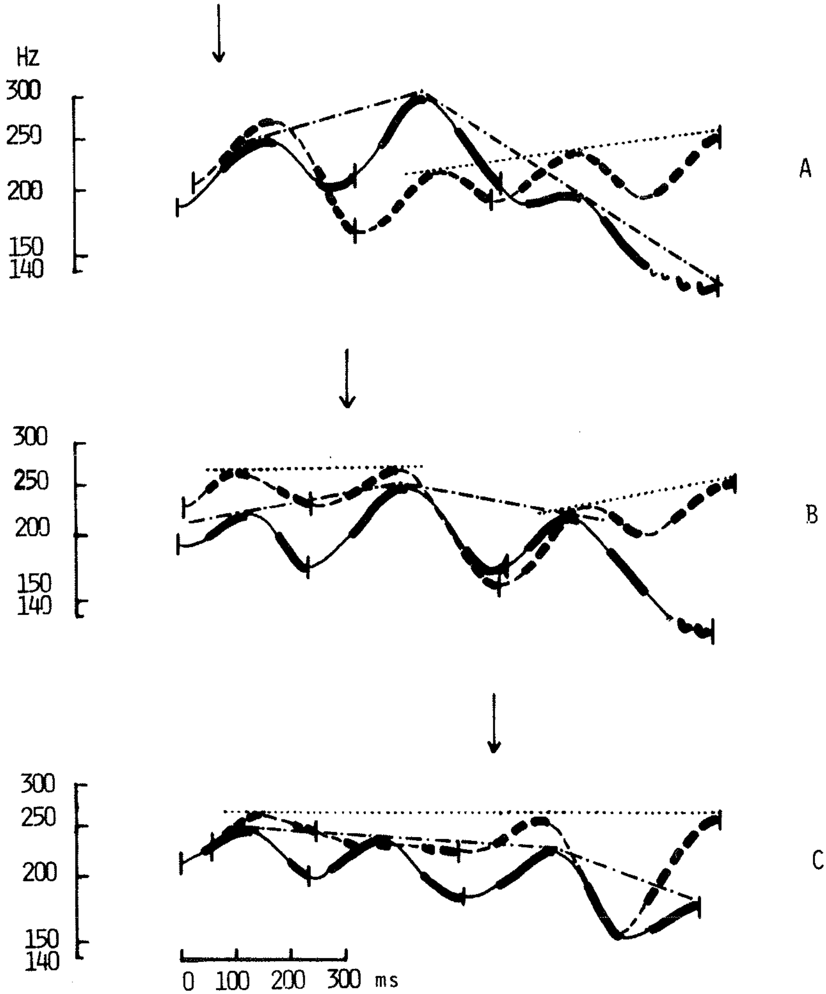


Figure 3. "Manne lämnar nallarna" pronounced as statements (unbroken lines) and as questions (broken lines). Thick contour lines = vowels, thin contour lines = consonants. Vertical bars = word boundaries. A sentence accented word is indicated with an arrow in the first CV-boundary of the word. This boundary is used as a common line-up point in the time domain for each sentence pair. —·—·— = topline (statements), ······· = topline (questions). ꜜꜜꜜ = creaky voice.

fall ending in a creaky voice or a final low Fo as opposed to the very much higher Fo in the corresponding question.

Discussion

The tonal configurations show clear differences between questions and statements in this dialect manifestation concerning both local and global characteristics.

GOTHENBURG

As I am uncertain about the interpretation of the tonal curve of the medially focussed question I will use the corresponding VSO-question for the comparisons. Note that for this sentence only the interrogative function is signalled by both intonation and reversed word order.

Global characteristics in the tonal contour

The topline ascends up to the high peak in the postfocal word in the questions after which it descends, as is shown in Figure 4. This means that there is a global rise in intonation up to the peak mentioned. If the question has final focus as in C in Figure 4, the entire contour can be characterized as a global rise. In the statements the topline starts to level off or to fall slightly at one peak earlier than in the questions; after the next peak the topline descends as in the corresponding questions.

The overall Fo is higher for the questions than for the corresponding statements.

A frequency compression of the tonal movements is exhibited before the highest frequency peak in the questions as compared with the corresponding statement contours.

Local characteristics in the tonal contour

The questions with medially or finally focussed words display a

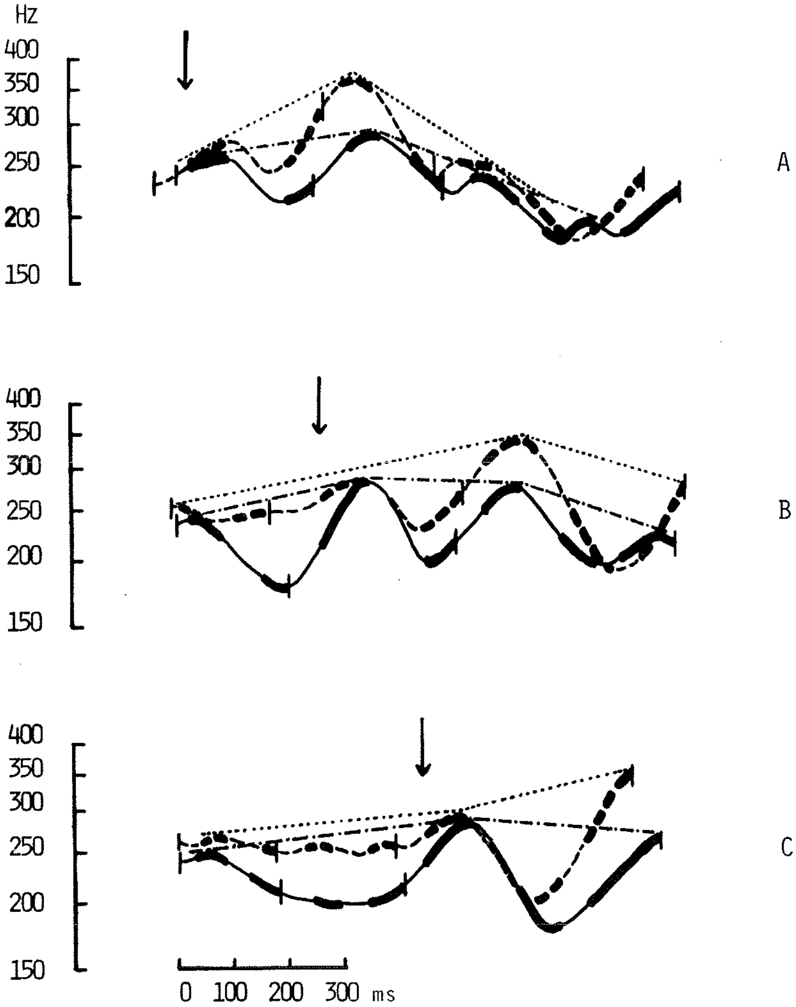


Figure 4. "Manne lämnar nallarna" pronounced as statements (unbroken lines) and as questions (broken lines). Thick contour lines = vowels, thin contour lines = consonants. Vertical bars = word boundaries. A sentence accented word is indicated with an arrow in the first CV-boundary of the word. This boundary is used as a common line-up point in the time domain for each sentence pair. Note that for sentence pair B the question is a VSO-question. -·-·- = topline (statements), ········ = topline (questions).

GOTHENBURG

final rise which is more pronounced and which reaches a higher Fo than the corresponding statement contours.

The postfocal peak mentioned above reaches a higher Fo in the questions than in the corresponding statements for all three sentence pairs.

Discussion

In the introduction of this paper questions about possible configurations of the intonation in statements and questions were proposed. Referring to the discussion about rising intonation it seems clear that the tonal contours of both statements and questions display a global rise, but this rise both lasts for a longer period of time and reaches a higher Fo (the high peak in the postfocal word) in the questions. These differences in timing and Fo result in tonal contours which are predominantly rising for the questions but not so for the statements.

The tonal curves also exhibit local differences, e.g. the size of the final rise which reaches a much higher Fo in the questions than in the statements for sentence pairs B and C. There is also a difference between the high peak in the postfocal word in the statements and in the questions, but this phenomenon is related to the extent of the global rise and will not be discussed further.

Question A, the initially focussed one, is characterized by a global rise to a much lesser extent than the other questions, B and C. This will answer another question asked in the introduction. The position of sentence accent is crucial for the course of the tonal movements yielding a pronounced rising-falling intonation in the initially focussed question whereas the corresponding contour with final focus displays an entirely rising intonation.

To summarize the answers: a) the tonal curves of both questions and statements are similar because they exhibit global rises as well as local rises. b) They are dissimilar with respect to the extent of the rises which are more pronounced in the questions. c) The position of sentence accent determines the extent of the global rise in both sentence types.

CONCLUDING DISCUSSION

In the introduction it was asked whether it was possible to describe the different tonal manifestations of question intonation along the west coast of Sweden as continuous transitions or as abrupt changes. Going from south to north comparisons will be made regarding both global and local tonal qualities.

Global characteristics

The direction of the topline after focus is level or falling in the questions whereas the statements exhibit a fall which is steeper than in the corresponding questions for the Malmö speaker. The tonal contours of the Halmstad speaker show clear rises of the topline postfocally in the questions which contrast well with the falls in the statements. So far the change in the slopes of the topline, going from Malmö to Halmstad, could be attributed to a gradual change. But looking at the postfocal topline slopes for the Gothenburg speaker, displaying either falls (in the statements) or rise-falls (in the questions), the changes would rather be referred to as discontinuous. Prefocally the topline is either level or rising in both questions and statements which can not be interpreted as a continuous change.

As all three dialect samples have a higher overall Fo in the questions there is no change in this respect.

The prefocal and postfocal frequency compression which is absent in the Malmö data, is introduced in the Halmstad data.

It is kept prefocally in the material from the Gothenburg speaker. This change therefor might be interpreted as a gradual development.

The frequency range is wider for the majority of the questions in comparison with the corresponding statements in all three dialects. This characteristic is correlated with others (high overall Fo, high focal or postfocal Fo-peaks), which are discussed above and below. As no change is introduced in this respect, it will not be treated further.

Local characteristics

In the Malmö tonal contours the two sentence types end with final rises but in the Halmstad contours a great difference is exhibited between final rises reaching a very high Fo in the questions and the pronounced fall (or fall-rise) ending at a much lower Fo in the statements. This great difference between the final parts of the tonal contours is not preserved in the Gothenburg dialect, especially not in sentence pair A, which exhibits final rises which are similar to those in the Malmö contours. Therefore these changes could be estimated as discontinuous, going partly back to an earlier condition.

The widening of the frequency range in connection with SA is only vaguely anticipated in the Malmö contours but becomes very apparent in the Halmstad contours. A different tonal pattern elicited by SA is exhibited in the Gothenburg questions (the high postfocal peak). These changes then might be judged as gradual developments the one leading on to another rather than going back to a former state.

Conclusion

Although it did not turn out to be an easy task to categorize these tonal changes as continuous or gradual on the one hand and as discontinuous and randomly distributed on the other hand, the discussion above leads up to the following observations:

Both global and local tonal qualities are represented among those changes which can not favourably be categorized as continuous.

The only changes which might be categorized as gradual changes according to the analysis above are 1) a tonal change in connection with SA (Halmstad, Gothenburg) and 2) a frequency compression outside the domain of SA (Halmstad, Gothenburg). These two changes seem to be correlated with each other, and might be interpreted as only one change resulting in different tonal adjustments in focal and nonfocal positions. Together these adjustments will emphasize the intended tonal change, e.g. the frequency compression outside focus will reinforce the increased Fo-register of the SA-fall in the Halmstad dialect.

For those tonal qualities which show no dialectal changes (higher overall Fo and wider frequency range in the questions) it might be hypothesized that they are the most important ones perceptually.

This is partly confirmed by the perceptual test for the Malmö dialect. Accordingly, perceptual tests are planned for the Halmstad and Gothenburg dialects also.

NOTES

1) The topline is construed using intonational concepts in works of Gårding (e.g. Gårding 1983, Gårding 1984). It seems clear that the second Fo-peak in question C in figure 2 is an example of an undershoot mechanism and that the topline therefore should be drawn as a straight ascending line and not as a fall rise.

2) See note 1 above

3) See note 1 above

BIBLIOGRAPHY

Bredvad-Jensen, A-C. 1980. An intonation-oriented description of questions in Swedish. The Nordic Languages and Modern Linguistics. Hovdhaugen (ed.), 233-241. Oslo: Universitets forlaget

Bredvad-Jensen, A-C. 1983. Perception studies of Swedish interrogative intonation. Paper presented at the Tenth International Congress of Phonetic Sciences, Utrecht University, Aug. 1-6, 1983.

Gårding, E. 1975. Toward a prosodic typology for Swedish dialects. The Nordic Languages and Modern Linguistics 2. Dahlstedt (ed.), 466-474. Lund: Almqvist & Wiksell.

Gårding, E. 1979. Sentence intonation in Swedish. Phonetica 36. 207-215

Gårding, E. 1983. Intonation units and pivots, syntax and semantics. Abstracts of the Tenth International Congress of Phonetic Sciences. Cohen & van den Broecke (eds.), 99-104. Dordrecht: Foris Publications

Gårding, E. 1984. Comparing intonation. Working Papers 27, Department of Linguistics, Lund University

Gårding, E., Bruce, G. & Willstedt, U. 1981. Transitional forms and their position in a prosodic typology for Swedish dialects. Nordic Prosody II. Fretheim (ed.), 40-50. Trondheim: Tapir

Aspects of Fo Declination in Swedish

Gösta Bruce

INTRODUCTION

The topic of the present study is declination (downdrift) - the often observed tendency in many languages for comparable F_0 values to be lower later in an utterance than earlier in the same unit of speech.

Some issues of declination have recently turned out to be controversial (for a review see Cohen et al. 1982).

One such issue is whether declination is to be regarded as a direct consequence of some physiological mechanism, e.g. a property of the respiratory system (Lieberman 1967, Atkinson 1973, Collier 1975, Maeda 1976), or whether it is basically learned behavior and linguistically purposeful (Ohala 1978, Pierrehumbert 1979).

Another issue of debate is how the overall course of a declination is to be interpreted: as a globally specified declining intonation on which local excursions for accentuation are superimposed (see e.g. Öhman 1967, Cohen and 't Hart 1967, Maeda 1976, Fujisaki et al. 1979, Bruce and Gårding 1978, Gårding 1982, Thorsen 1980a and recent contributions in Cutler and Ladd 1983 by Gårding, Thorsen and Vaissière) or as formed by the pitch relations between successive, local excursions for accentuation (Pierrehumbert 1980, Ladd 1983). This issue also involves the question of the possible use of look ahead and look back strategies for the planning and execution of a declination.

The purpose of the present paper is to draw the reader's attention to certain facts about the production of declination (with exemplification from Swedish) that I think are important for a proper understanding of this particular aspect of intonation.

I also intend to let these facts about declination in my interpretation shed light on the debated issues above, in particular the one concerning the implementation of the overall course of a declination.

DATA BASE

The considerations about F_0 declination to be presented in this paper are based on two independent studies of Swedish intonation.

One consists of a series of experiments in Standard Swedish intonation (Bruce 1982a), and the other is a study of text intonation in South Swedish (Bruce 1982b).

In the present paper I will draw upon the results mainly from two of the experiments of the first study. One concerns the relationship between declination and utterance length. In this experiment the number of stress groups (from two to five) is systematically varied as is placement of focus (neutral focus assignment or focus on the first or the last stress groups).

In the other experiment the number of stress groups is varied in the same manner, but at the same time there is also a systematic variation of overall emphasis manifested as two degrees of involvement (detached - involved). The same test material containing meaningful Swedish sentences was used in both experiments, but the recordings took place on separate occasions.

Each test sentence was elicited as an answer to a question. The increase in the number of stress groups was achieved by a syntactic expansion to the right.

The test material was phonetically balanced so as to minimize known microprosodic effects on the F_0 contour. In each test sentence there is an onset of two unstressed syllables and an offset of two (in one case three) unstressed syllables. Between the stresses there are three unstressed syllables. Each stressed syllable carries accent 2, which is analyzed phonologically in Standard Swedish phrased in autosegmental terms as (L) $\overset{*}{H}$ L . Even if no special grouping or focusing is elicited, the first and the last stress group of a sentence will (normally) have a phrase accent in this dialect. It is analyzed phonologically as (L)H immediately following the $\overset{*}{H}$ L of accent 2.

The study of text intonation in South Swedish concerns the relationship between declination and text unit length. In the variation of text unit length three different meaningful Swedish sentences, each containing two stress groups, were used. Each of the sentences can form a complete text unit, but they can also be combined to form two-sentence and three-sentence text units. For these larger units the order of the component sentences has been systematically shifted. The coupling between the sentences in two and three-sentence units is temporal, although this has not been expressed directly by means of temporal adverbs. The recordings of the multi-sentence units contain pauses between the constituent sentences.

The same requirements for the phonetic composition of the test material as in the first study were also complied with in this study. Each constituent sentence has an onset and offset of two unstressed syllables respectively, and there are three unstressed syllables between the stresses.

Each stressed syllable carries accent 2, which can be characterized as $\overset{*}{L}$ H (L) in South Swedish. In each text unit equal prominence of successive accents was elicited. Unlike Standard Swedish there is usually no addition of a phrase accent to the first or last stress group for a neutral focus assignment.

CHARACTERISTICS OF DECLINATION

Facts concerning declination obtained from the two separate studies of Swedish intonation will be summarized below. For a more detailed account see Bruce (1982a, 1982b). The similarities between the two studies seem to be great enough to allow a joint presentation.

It is clear that there are certain linguistic factors that will affect the course of F_0 declination. I have argued elsewhere that placement of focus is one such factor (Bruce 1982a). Up to the focus of an utterance declination appears to be absent or gentle, while after focus there is a conspicuous, stepwise declination. The same declination course can also be found for certain syntactic groupings (Bruce 1982a, section 4). Sentence type (question/statement) will also tend to affect declination. For question intonation (Gårding 1979) the declination is largely suspended. This also holds for a related language, namely Danish (Thorsen 1980a).

In the material chosen for the present paper these linguistic factors have not been varied. Therefore I will concentrate on units of speech that are statements (answers to questions), that have neutral focus assignment, no special syntactic groupings and thus equal prominence of succesive accents.

Declination and unit length

The boundary conditions for F_0 declination are that F_0 starts Low and ends Low, and that the initial Low is higher in frequency than the final Low. The final Low (offset) is usually the lowest F_0 point of an utterance and also the least variable F_0 point. It is constant across variation in unit length, i.e. length of an utterance or a text unit in terms of number of stress groups (see Figs 1 and 2). But interestingly enough the same Low value will also be reached before a pause in non-final

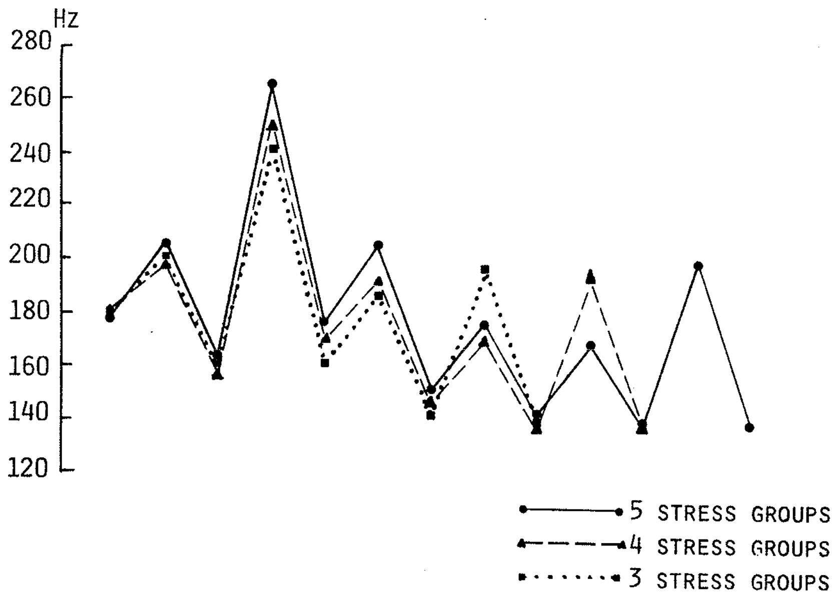


Figure 1. F_0 declination and utterance length. Stylized F_0 contours - means in Hz (9-11 repetitions) of successive F_0 minima and maxima - of utterances of varying length (3-5 stress groups) lined up from the beginning of the utterance. Neutral focus version for a female speaker of Standard Swedish.

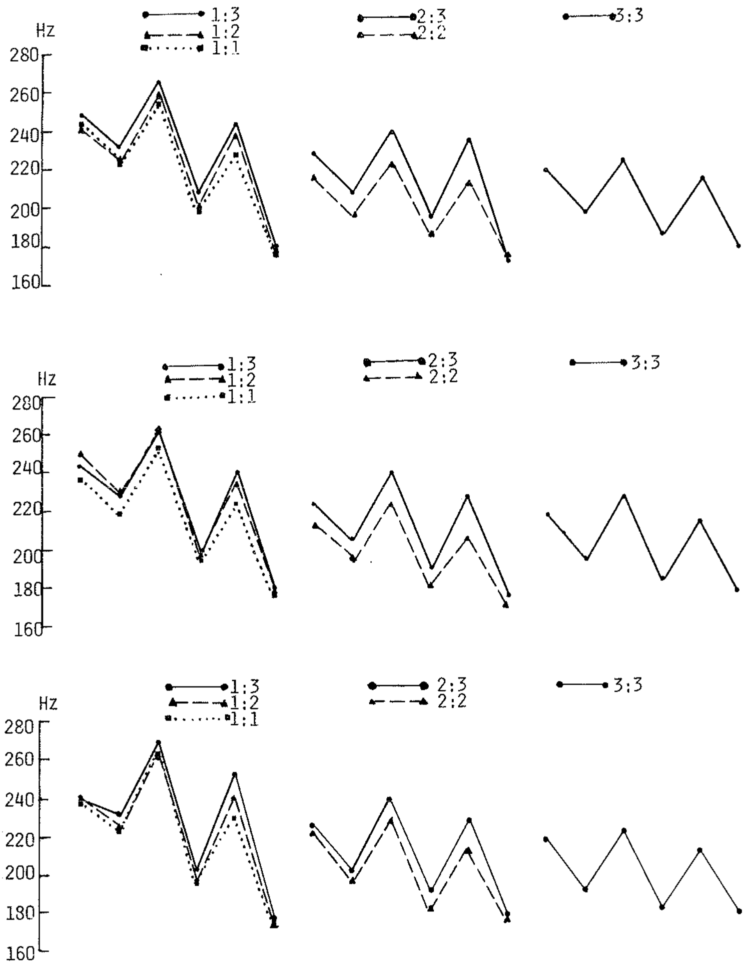


Figure 2. F_0 declination and text unit length. Stylized F_0 contours for three different sentences (upper, middle and lower part) - means in Hz (7 repetitions) of successive F_0 minima and maxima - in different positions of text units consisting of one, two or three sentences (two, four or six accents) lined up from the beginning of the text unit for a female speaker of South Swedish.

position within a text unit (see Fig 2). The final Low is also constant across variation in degree of involvement (see Fig 3). The constancy of this F_0 point is well attested (see e.g. Maeda 1976, Fujisaki et al. 1979, Pierrehumbert 1980).

The initial Low (onset) is constant over variation in unit length (see Figs 1-3), but varies (decreases) with position in a text unit (see Fig 2). It also varies (increases) with degree of involvement (Fig 3), as do all F_0 points except the final Low (see below).

This means that the initial Low of an utterance is a variable distance in frequency above the final Low of the preceding utterance. This is assumed to be the boundary signal, regardless whether or not there is a physical pause in between. The relative degree of frequency shift is an indication of the coupling between the actual utterances.

Another relevant F_0 point in the description of declination is the peak, the highest F_0 point of a unit of speech. It should be noted that in the study of Standard Swedish the peak is equal to the H of the first phrase accent which comes after the first accent, while in the study of South Swedish the peak is the H of the first accent of the unit.

From the peak there is a self-evident, successive lowering of F_0 values of both accent maxima and minima (see Figs 1-3). This is true, if we add the following qualification. In a text unit consisting of two or three sentences the decrease in F_0 values through the unit is not truly successive. The declination is arrested at each new start of a second or third sentence of a unit. This can be regarded as a kind of tonal coupling or adaptation of the earlier part of the F_0 contour of a later sentence to the later part of the F_0 contour of an earlier sentence (Fig 2).

The decrease in F_0 values is steeper in the beginning than later in a declination (see Figs 1-3). For similar results see Fujisaki et al. 1979 for Japanese, Thorsen 1980b, 1981 for Danish, Pierrehumbert 1980 and Sorenson and Cooper 1980 for American English. Ideally the declination appears to be exponential and asymptotic to the F_0 bottom of the speaker's voice range (the final Low)(cf. Pierrehumbert 1980).

There is no apparent difference in slope (rate of declination) between a longer and a shorter unit, lined up from the peak.

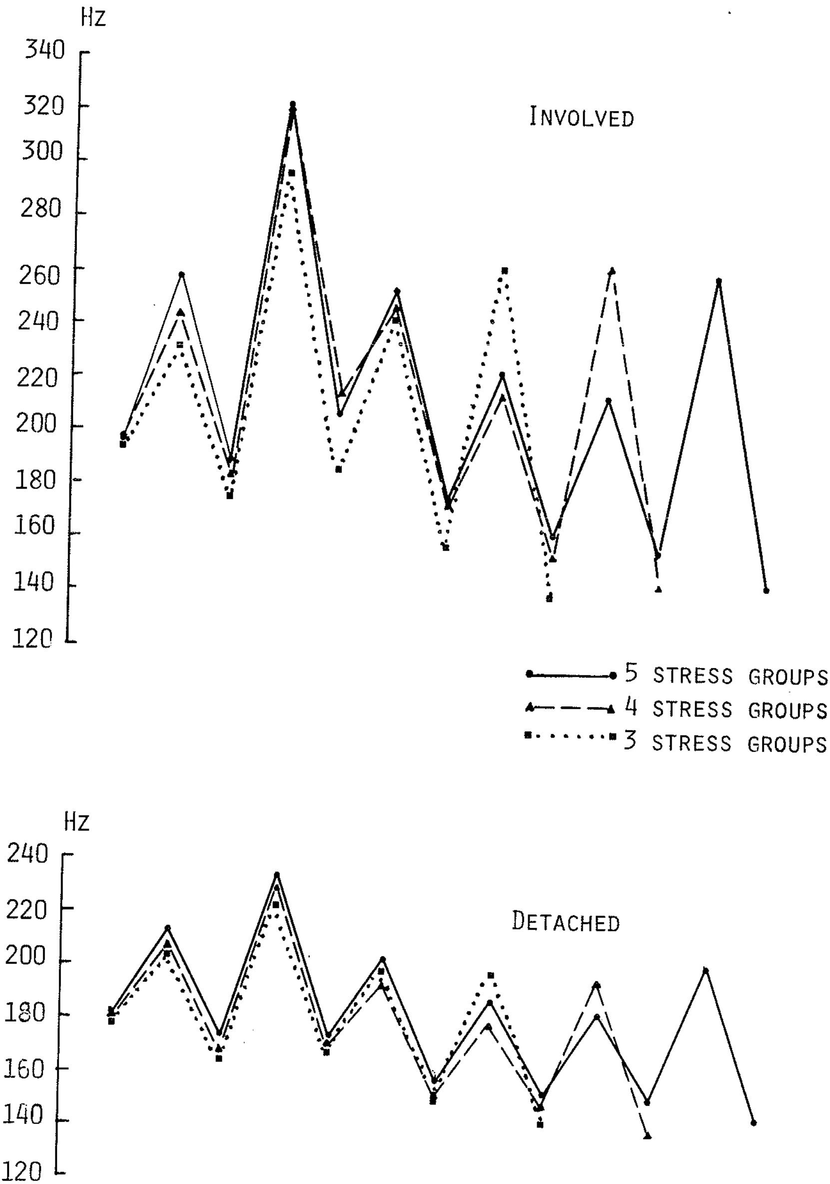


Figure 3. F_0 declination and involvement. Stylized F_0 contours - means in Hz (6 repetitions) of successive F_0 minima and maxima - of utterances of varying length (3-5 stress groups) lined up from the beginning of the utterance in an involved version (above) and a detached version (below) for a female speaker of Standard Swedish.

There is some adjustment of F_0 values to the length of the actual unit expressed as the number of stress groups (accents) contained. This means that lined up from the peak each successive maximum and minimum is usually higher in frequency in a longer utterance except for final values (see Figs 1-3).

I think it is important to note that the decrease in F_0 on a plateau consisting of several unaccented syllables between two accents, if any, is negligible compared with the decrease in F_0 taking place from the minimum preceding the accent excursion to the one following it. According to my data, it is only in utterances containing a single (or two) accent(s) with several unstressed syllables initally or finally that there is an obvious F_0 drop on such a plateau. Therefore, to account for the declination in Swedish described here as primarily one of accentual downstepping it seems possible to assume that the frequency value of each accent L (or H) is a constant ratio of the immediately preceding accent L (or H), as was suggested for American English by Pierrehumbert (1980). The overall declining course of an utterance can be seen as the result of this kind of local rule together with a sensitivity of F_0 values to the length of the unit of speech (For a preliminary test see Bruce 1982a).

I think that a description of the pitch relations between successive, local excursions for accentuation only in terms of F_0 points (minima and maxima) tends to be incomplete and to conceal interesting facts about declination. It seems wise to include also an account of the F_0 changes (rises and falls) contained in these excursions as a complement for an adequate description of declination (For a similar view see Ohala 1982).

There is an apparent asymmetry between rises and falls of the local excursions for accentuation. As a natural component of a declination an F_0 fall generally covers a wider range than the preceding F_0 rise of the same excursion. More interesting is perhaps the fact that the range of successive accent falls decreases gradually within a unit, while the accent rises preceding the falls tend to have a rather constant range independent of their position in a declination (see Figs 4 and 5). This means that the difference in frequency range between fall and rise is greater higher up in a declination. Towards the end of a declination F_0 falls have almost the same narrow range as do

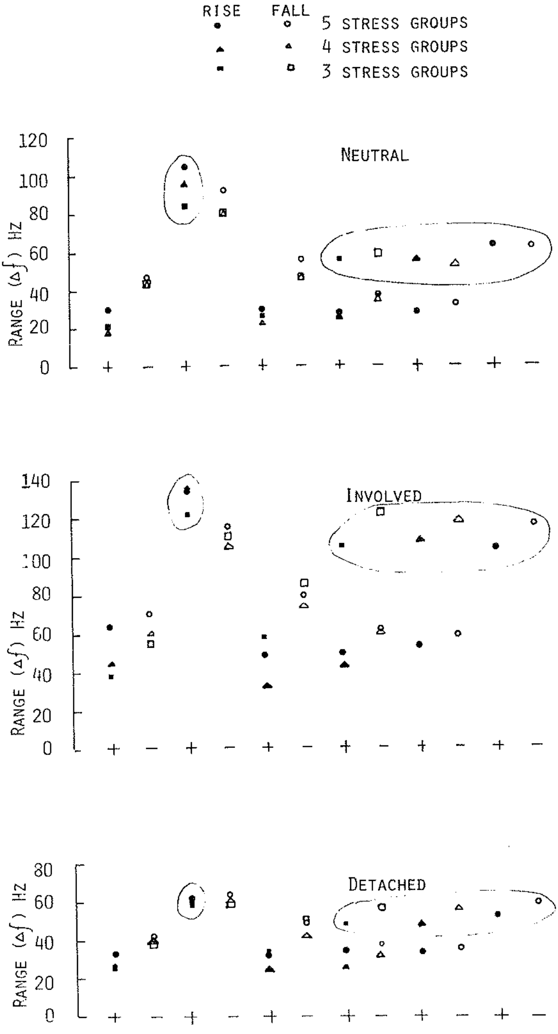


Figure 4. Rise - fall asymmetry. F_0 ranges - means in Hz of successive rises (+) and falls (-) - of utterances of varying length (3-5 stress groups) lined up from the beginning of the utterance. Neutral focus version (9-11 repetitions, upper part), involved version (6 repetitions, middle part), detached version (6 repetitions, lower part) for a female speaker of Standard Swedish. Encircled symbols are manifestations of phrase accent.

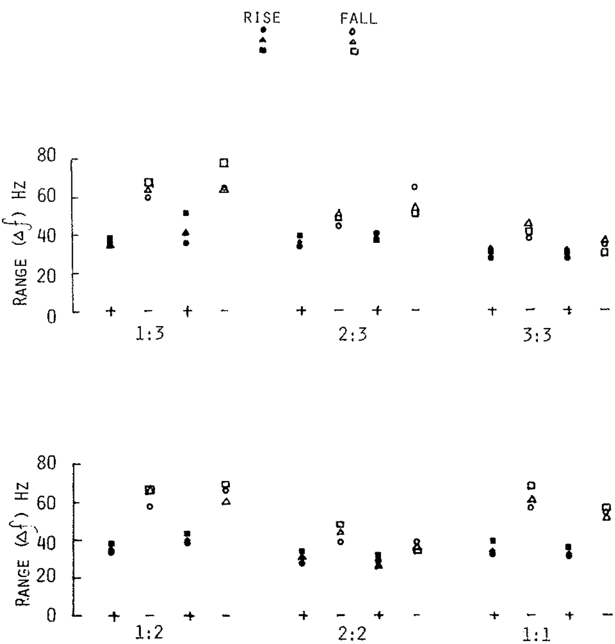


Figure 5. Rise - fall asymmetry. F_0 ranges for three sentences - means in Hz (7 repetitions) of successive rises (+) and falls (-) - in different positions of text units consisting of one, two or three sentences (two, four or six accents) for a female speaker of South Swedish.

corresponding rises. This is true for both studies of Swedish intonation in spite of of the apparent difference in timing of accent rises and falls relative to the stressed syllable between the two dialects described above.

While there is an adjustment to unit length of F_0 points, the range of each accent fall - in the same position lined up from the peak disregarding final characteristics - appears to be more constant (in Hz) across variations in unit length.

As for accent Hs and Ls the relation between successive accent falls of the same declination can profitably be described as one of a constant ratio, so that a particular fall is a fraction of the immediately preceding one.

Finally, in the study of text intonation in South Swedish we observe that both the rise and fall of the second accent tend to be narrower in range than the first accent rise and fall in the same constituent sentence, provided it is text-final. If it is non-final, we find the opposite relation (see Fig 5). This means that there is a concomitant lowering of the text-final accent maximum (see Fig 2). This was not observed for Standard Swedish. Final lowering is a feature typical of American English (Liberman and Pierrehumbert 1982).

Declination and involvement

A change in involvement from detached to involved is clearly expressed as an increase in F_0 range (see Fig 3). Variation in F_0 range is achieved by a frequency expansion upwards; the lower limit of the range (final Low) is fixed, while the upper limit (peak) is highly flexible. F_0 maxima appear to be more affected than F_0 minima. While there is no slope difference between utterances differing in length, there is a superficially obvious slope difference between utterances differing in F_0 range. However, the same F_0 pattern is maintained with a change in F_0 range. Linear regression analysis shows a very high correlation between F_0 values (maxima and minima as well as rises and falls) of a detached and an involved attitude. The relation between successive F_0 values (maxima, minima, falls) is of approximately the same constant ratio in the two different studies. Very similar results are presented in Liberman and Pierrehumbert (1982).

It should be clear by now that a declination is not properly described as a chain of decreasing ranges simulating a successively decreasing degree of involvement. The range decrease in a declination concerns only the falls, while a shift from detached to involved or vice versa affects the range of both rises and falls.

CONCLUSION

I interpret the facts about declination in Swedish presented above as favoring the idea that in a declining intonation the overall course is essentially formed by the pitch relations between successive, local excursions for accentuation. The possible, but not obligatory downtrend in F_0 on unaccented plateaus between accents can be conceived of as an extra means of amplifying a declination. Therefore, what is mainly responsible for the declination effect, in my opinion, is the fact that the fall of a local excursion for accentuation is controlled to cover a wider frequency range than the preceding rise of the same excursion and that this process repeats itself with a successive decrease of a constant ratio in the L value and in the range of the fall for each accent. This is the ideal way of expressing the combination of declination and equal weight of successive accents.

According to this view it is also more adequate to say that there is an adjustment of F_0 values to the number of upcoming accents rather than to utterance length per se.

A drop in F_0 (close) to the bottom of the speaker's voice range somewhere in the course and then a resetting to the previous course of declination will favor the perception of a boundary, e.g. a clause boundary. There is also preliminary evidence showing that a departure in the other direction from the ideal downstepping contour with instead no decrease in successive L values will be perceived as indicating less prominence and thus a grouping of the actual accents.

I think it is natural to assume from the present data on declination in Swedish that both a look ahead and a look back are used in the control of F_0 declination. The look back to the preceding accent will ensure equal prominence of successive accents (or any prominence relations aimed at by the speaker) (cf. Pierrehumbert 1980). The look ahead will be involved in the adjustment of F_0 values to the number of upcoming accents in order to ensure that the F_0 bottom of the speaker's voice will not be reached until the end of the utterance.

I would be inclined to take the facts concerning declination presented here as favoring an interpretation of declination as learned behavior, actively controlled, linguistically purposeful and useful in the textual organization of speech.

The relative weight of what happens at the local excursions for accentuation compared to what happens in between, the rise-fall asymmetry of these excursions in combination with the effect of linguistic factors such as placement of focus, certain syntactic groupings and sentence type (question/statement) on declination are all indicative of declination as a linguistically integrated phenomenon, which is a phonetically motivated and natural process that does not seem to be accounted for in any direct way by any specific physiological constraint.

REFERENCES

Atkinson, J. (1973) Aspects of intonation in speech: implications from an experimental study of fundamental frequency. Unpublished doctoral thesis. University of Connecticut

Bruce, G. (1982a) Developing the Swedish intonation model. Working Papers, Department of Linguistics, Lund University 22: 51-116

Bruce, G. (1982b) Textual aspects of prosody in Swedish. Phonetica 39: 274-287

Bruce, G. and Gårding, E. (1978) A prosodic typology for Swedish dialects. In Gårding, E., Bruce, G. and Bannert, R. (Eds.) Nordic prosody: 219-228. Department of Linguistics, Lund University

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

Cohen, A., Collier, R. and 't Hart J. (1982) Declination: Construct or intrinsic feature of speech pitch? Phonetica 39:254-273

Collier, R. (1975) Physiological correlates of intonation patterns. JASA 58: 249-255

Cutler, A. and Ladd, R. (Eds.) (1983) Prosody: models and measurements. Heidelberg: Springer

Fujisaki, H., Hirose, K. and Ohta, K. (1979) Acoustic features of the fundamental frequency contours of declarative sentences. Annual bulletin of the research institute of logopedics and phoniatrics, University of Tokyo 3: 163-172

Gårding, E. (1979) Sentence intonation in Swedish. Phonetica 36:207-215

Gårding, E. (1982) Contrastive prosody: a model and its applications. Studia Linguistica XXXVI: 146-165

Ladd, R. (1983) Peak features and overall slope. In Cutler, A. and Ladd, R. (Eds.) Prosody: models and measurements: 39-52. Heidelberg: Springer

Liberman, M. and Pierrehumbert, J. (1982) Intonational invariance under changes in pitch range and length. Manuscript, Bell Laboratories

Lieberman, P. (1967) Intonation, perception and language. Cambridge: MIT Press

Maeda, S. (1976) A characterization of American English intonation. Unpublished doctoral thesis. Massachusetts Institute of Technology

Ohala, J. (1978) Production of tone. In Fromkin, V. (Ed.) Tone: A linguistic survey: 5-39. New York: Academic Press

Ohala, J. (1982) Physiological mechanisms underlying tone and intonation. In Fujisaki, H. and Gårding, E. (Eds.) Working group on intonation, Preprints. The XIIIth International Congress of Linguists. Tokyo

Öhman, S. (1967) Word and sentence intonation: A quantitative model. STL-QPSR 2-3: 20-54

Pierrehumbert, J. (1979) The perception of fundamental frequency declination. JASA 66 (2): 363-369

Pierrehumbert, J. (1980) The phonology and phonetics of English intonation. Unpublished doctoral thesis. Massachusetts Institute of Technology

Sorenson, J. and Cooper, W. (1980) Syntactic coding of fundamental frequency in speech production. In Cole, R.A. (Ed.) Perception and production of fluent speech: 399-440. Hillsdale, New Jersey: Lawrence Erlbaum Associates

Thorsen, N. (1980a) A study of the perception of sentence intonation. JASA 67: 1014-1030

Thorsen, N. (1980b) Intonation contours and stress group patterns in declarative sentences of varying length in ASC Danish. ARIPUC 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. ARIPUC 15: 13-47. University of Copenhagen

Prosody and the Hemispheres

Christina Dravins

ABSTRACT

A test was designed in order to asses the ability of brain-damaged subjects to process prosodic information. The test was administered to three groups of subjects; patients who had suffered left or right hemisphere stroke and patients who have had a transitory ischemic attack. The results of the test did not show any significant difference in the performance of the three groups. The purpose of this note is to discuss the test and suggest possible improvements.

1. INTRODUCTION

The role of the right hemisphere in language and communication remains obscure. A widely spread opinion is that the right hemisphere normally has no part in language processing, the exception being persons with reversed or mixed laterality. Some authors claim that the right hemisphere does contribute to language and communication although to a lesser extent than the left (Zaidel 1978, Bradshaw and Nettleton 1981). During the last decade an increasing interest has focused on potential language and communication functions of the right hemisphere. It has been suggested that it plays an important role in the interpretation and handling of paralinguistic aspects of verbal communication, e.g. grasping the abstract content of common words, interpreting figures of speech, solving verbal syllogisms and rembering the general theme of stories (Gardner and Denes 1973, Winner and Gardner 1977, Caramazza, Gordon, Zurif and DeLuca 1976, and Wapner, Hamby and Gardner 1981). Some authors report on disturbances of the ability to evaluate everyday situations and to appreciate humorous material

(Wechsler 1973, Gardner, Ling, Flaum and Silverman 1975, Heilman, Schwartz and Watson 1978).

Several studies have reported the right hemisphere to be involved in the processing of emotional and non-verbal information (Tucker, Watson and Heilman 1977, Ross and Mesulam 1979, Ross 1981). It has also been suggested that the right hemisphere is dominant for the processing of prosodic information (Blumstein and Cooper 1974, Ross and Mesulam 1979, Heilman, Scholes and Watson 1975) but there are authors who reject this hypothesis (Zurif 1972, Schlanger, Schlanger and Gerstman 1976). There are several reports on disturbances of prosodic functions in aphasic subjects (Danly and Shapiro 1982, Danly, Cooper and Shapiro 1983, Ryalls 1982). On the other hand it has been suggested that the cerebral representation of prosodic functions in the right hemisphere mirrors the representation of language in the left hemisphere and that there are clinically distinct syndromes of aprosodia analogous to the syndromes of aphasia (Ross 1981). The author is concerned only with prosody expressing attitudes or emotions.

These results taken together suggest that prosodic functions in speech are affected differently by right and left hemisphere lesions. Right hemisphere lesions might give rise to difficulties in producing and comprehending affectively toned stimuli, whereas left hemisphere lesions might give rise to difficulties in tonal contrasts.

2. PROCEDURE

A test was designed to investigate comprehension, repetition and production of prosodic functions. The test consists of two parts; one exploring verbal contrasts, the other emotional expressions. The linguistic material includes the following contrasts:

- <u>accent 1 versus accent 2</u>

Six minimal word accents pairs were selected. The distinction between members of a pair is mainly marked by pitch variations.

- <u>accent location</u>

Four pairs of words were selected in which the location of accent (stress) is the distinctive feature. The distinction is marked by pitch variation, accentuation (stress pattern) and vowel quality.

- compund noun versus noun-phrase

Four compound nouns were contrasted with four noun phrases, e.g. blåklocka 'hare-bell' and blå klocka 'blue bell'. The distinctions between the members of a pair is carried by variation in pitch and accent pattern.

All target words and phrases were illustrated and recorded by an actress. The targets were printed on cards in phrases with the exception of the compound words and the noun phrases which were printed in isolation.

The emotional stimulus material consisted of six semantically neutral sentences and four sentences verbally expressing different modes. The modes were: happiness, anger, astonishment and sadness. All sentences were recorded by an actress expressing these modes. The sentences were written down on cards, together with a note on the mode of expression in brackets. Five non-verbal, emotional sounds were recordes. The sounds were: crying, laughter, groaning, moaning and screaming. The outline of the test is schematically displayed in table 1.

To investigate the capacity of the test nine patients with right hemisphere lesions, four with left hemisphere lesions and four patients with the diagnosis transitory ischemic attack (TIA) were tested. All were native speakers of Swedish and right-handed. Patients with severe hearing loss, signs of dementia and age over 70 had been excluded. All brain-damaged patients had suffered ischemic stroke. The patients with left hemisphere lesion all had a history of aphasia but by the time of testing the symptoms were very mild or could not be demonstrated. Yet some patients reported having trouble now and then. The patients were given the test in one or two sessions depending on the endurance of each person. Each testitem was presented once. Since there were many examples of each category this was considered to be sufficient.

3. RESULTS

The test performances were analyzed and the number of errors were tabulated. The average number of errors for each test-group and each task is displayed in table 2, with the exception for the naming of emotionally toned sentences which are displayed in table 3.

Table 2: The average number of errors per test task.

The answers in the naming task were classified into four categories:

a) target: the answer is as intended.

b) synonymous: the answer is synonymous with the intended one.

c) unrelated: the answer has no obvious relation to the intended one.

d) reversed: the answer is reversed to the intended one.

Table 3: The average number of answers in the non-target categories.

		ERRORS	
GROUP	RELATED	UNRELATED	REVERSED
RIGHT (N:9)	1.1	1.0	1.2
LEFT (N:4)	0.75	1.0	0.75
TIA (N:4)	2.25	1.25	0.25

The performance on the repetition and production of emotionally toned utterances were extremely varied and were not subject for analysis. Identification of non-verbal emotional sounds was performed without mistakes.

Tabel 1. Schematical display of test items and test tasks.

Test items	Number of stimuli	Test tasks			
		discrimination	comprehension/ identification	repetition	production
word accents	12		matched to pictures	repeat auditory stimuli	read aloud from card
word stress	8		-"-	-"-	-"-
noun versus noun phrase	8		-"-	-"-	-"-
emotionally toned utterances a) semantically neutral	6	tell whether auditory stimuli are expressed the same way or differently	name auditory stimuli	-"-	
b) semantically loaded	4	-"-		-"-	-"-
non-verbal emotional sounds	5		-"-		

	VERBAL ASPECTS									NON-VERBAL ASPECT
GROUP	COMPREHENSION			REPETITION			PRODUCTION			DISCRIMINATION
	ACCENT 1-2	ACCENT LOC	COMP NP	ACCENT 1-2	ACCENT LOC	COMP NP	ACCENT 1-2	ACCENT LOC	COMP NP	
RIGHT N:9	1.7	0.1	1.2	1.7	0	1.3	0.2	0	1.2	0.25
LEFT N:4	1.0	0.25	0.25	1.25	0	0.25	0	0	0.25	0.25
TIA N:4	0.5	0.25	0.25	0.25	0	0	0	0	0.25	0.44

4. DISCUSSION

No conclusion about the nature of prosodic disturbances can be drawn from this material. But on the basis of the results the tasks can be classified into three groups:

Group I: Tasks that were too easy; i.e. all subjects performed them easily and with few mistakes. This group includes accent location and naming of nonverbal emotional sounds.

Group II: Tasks that had so varied results that the tasks themselves must be regarded as inadequate. This group includes repetition and production of emotionally toned utterances.

Group III: Tasks that were hard enough to create difficulties for some of the subjects. In this group belong contrasts between accent 1 and 2, contrasts between compound nouns and noun-phrases and discrimination and naming of emotionally toned sentences.

It should be pointed out that when the repetition tasks are successful in the linguistic part of the test, comprehension and repetition scores are very similar.

The results of the three groups do not show any difference in performance. However, I cannot avoid the suspicion that this might be due to the small number of subjects. The test will be tried on a larger sample of subjects. For the continuation of this study, tasks classified into group I and II will be eliminated as well as the repetition tasks mentioned above.

A new, more comprehensive test will be developed. The following prosodic functions will be explored in the new test:

- syntactic structure
- syllabic structure
- speech acts
- focusing

An additional requirement is that the new test must be less boring than the present one.

REFERENCES

Blumstein, S.E. and Cooper, W.E. 1974. Hemispheric processing of intonation contours. Cortex 10: 146-158

Bradshaw, J.L. and Nettleton, N.L. 1981. The nature of hemispheric specialization in man. The Behavioral and Brain Sciences 4: 51-91

Caramazza, A., Gordon, J., Zurif, E.B. and DeLuca, D. 1976. Right-hemispheric damage and verbal problem solving behavior. Brain and Language 3: 41-46

Danly, M., Cooper, W.E. and Shapiro, B. 1983. Fundamental frequency, language processing, and linguistic structure in Wernicke's aphasia. Brain and Language 19: 1-24

Danly, M. and Shapiro, B. 1982. Speech prosody in Broca's aphasia. Brain and Language 16: 171-190

Gardner, H. and Denes, G. 1973. Connotative judgements by aphasic patients on a pictorial adaption of the semantic differential. Cortex 9: 183-196

Gardner, H., Ling, P.K., Flamm, L., and Silverman, J. 1975. Comprehension and appreciation of humorous material following brain damage. Brain 98: 399-412

Heilman, K.M., Scholes, R. and Watson, R.T. 1975. Auditory affective agnosia: disturbed comprehension of affective speech. Journal of Neurology, Neurosurgery and Psychiatry 38: 69-72

Heilman, K.M., Schwartz, H.D. and Watson, R.T. 1978. Hypoarousal in patients with the neglect dyndrome and emotional indifference. Neurology 28: 229-232

Ross, E.D. 1981. The aprosodias. Functional-anatomic organization of the affective components of language in the right hemisphere. Archives of Neurology 38: 561-569

Ross, S.D. and Mesulam, M.H. 1979. Dominant language functions of the right hemisphere: prosody and emotional gesturing. Archives of neurology 36: 144-148

Ryalls, J.H. 1982. Intonation in Broca's aphasia. Neuropsychologia 20: 355-360

Schlanger, B.B., Schlanger, P. and Gerstman, L.J. 1976. The perception of emotionally-toned sentences by right hemisphere damaged and aphasic subjects. Brain and Language 3: 396-403

Tucker, D.M., Watson, R.T. and Heilman, K.M. 1977. Discrimination and evocation of affectively intoned speech in patients with right parietal disease. Neurology 27: 947-950

Wapner, W., Hamby, S. and Gardner, H. 1981. The role of the right hemisphere in the apprehension of complex linguistic material. Brain and Language 14: 15-33

Wechsler, A.F. 1973. The effect of organic brain disease on recall of emotionally charged versus neutral narrative texts. Neurology 23: 130-135

Winner, E. and Gardner, H. 1977. Comprehension of metaphor in brain-damaged patients. Brain 100: 719-727

Zaidel, E. 1978. Auditory language comprehension in the right hemisphere following cerebral commissurotomy and hemispherectomy: A comparison with child language and aphasia. In Language Acquisition and Language Breakedown: Parallels and Divergencies. Eds. Caramazza, A. and Zurif, E.B. Baltimore: John Hopkins Press

Zurif, E.B. 1972. Auditory lateralization: Prosodic and syntactical factors. Brain and Language 1: 391-404

Comparing Intonation

Eva Gårding

A comparison of intonation in different prosodic systems, analysed with the same method and described in the same manner, supports the opinion that such systems share some basic features. These common features are present in the underlying structures as well as their surface manifestations.

The goal of my paper is to substantiate this point. I shall draw from analyses of two tone languages, Hausa with two tones and Standard Chinese with four, and three accent languages, Swedish with two accents, English with one movable accent and French with one fixed accent (or none, depending on the analysis). (1)

Intonation without further specification is defined as the sum total of the main factors that shape the FO curve, i.e. lexical and phrasal accents or tones, accentual or tonal expressions of information weight, speech act and boundaries.

In the first part, the intonation parameters used for analysis and description will be presented in schematic form together with their communicative functions. The second part gives real-life examples of these parameters and their correlations to lexical, syntactic and semantic properties. The third part describes a scheme which has been used to generate pitch for comparable sentences in some of the exemplified languages.

Intonational similarities and differences can be described in terms of this model.* Most of the material used in the following summary has been described in more detail elsewhere (see references).

1. Intonation parameters in schematic form

Figure 1 summarizes the main concepts of the model. The drawings, which are meant to have a logarithmic frequency scale, illustrate idealized cases consisting of at least two evenly accented words. There are several reasons for using such uncommon patterns. Focus has been avoided since it causes the F0 curve to deviate from its general direction, not just over the focused part but also over the neighbouring syllables. One-accent phrases have not been included since in such phrases the general direction is not so easy to catch. Phrases with sonorant segments have not been chosen to eliminate articulatory disturbances. An intonation curve of this idealized type has local maxima and minima, i.e. turning points. These turning points are part of a larger pattern, the grid, which is most easily seen if the main maxima are connected by a topline and the main minima by a baseline. When the turning points have been connected, as in the drawings, the grid appears as a sequence of units which are clearly rising, falling or level.

Each unit has a normal, expanded or compressed width, even to the extent of being best represented by one line only. The part of the grid where the direction or width is changed, or where the grid takes a jump, is called a pivot. An intonation unit is

* This last section was the main part of my contribution to the Working group on intonation at The XIIIth International Congress of Linguistics, Tokyo 1982, and has been published elsewhere (Gårding 1983c).

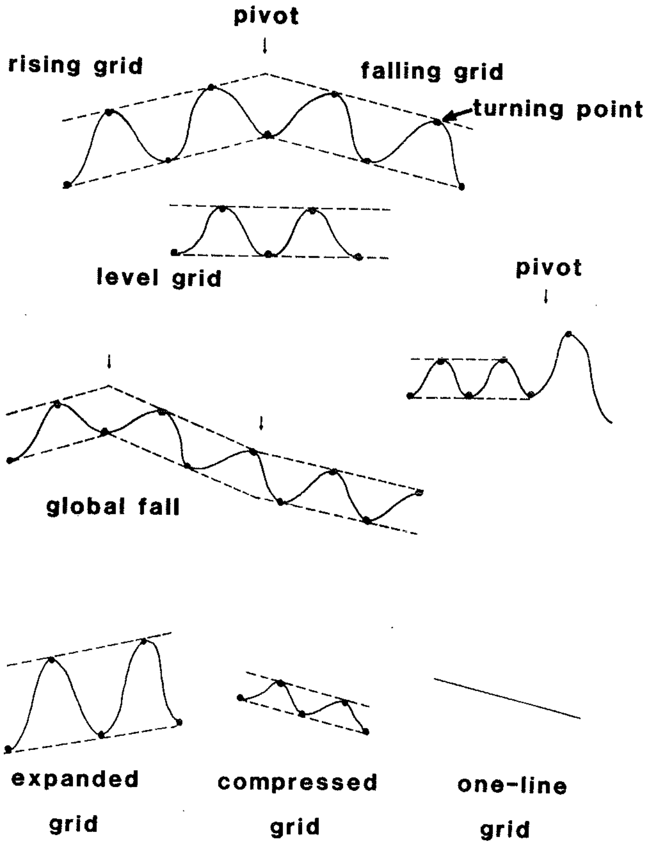


Figure 1. Concepts of the model, illustrated by schematic F0 curves connecting turning points over sonorant segments. Grids are marked by broken lines, pivots by arrows.

Intonation parameters	Function	
	Semantic	Syntactic
turning points	words, morphemes	d:o
pivots	constituents (theme/rheme)	d:o (subject/predicate)
grid: direction	speech act type	sentence type
grid. width, position	information weight (focus)	clause type

Table 1. Communicative function of the intonation parameters.

a piece of an utterance with an unbroken grid. The drawings reflect a superposition principle: lexical intonation is superimposed on a slowly varying phrase or sentence intonation, expressed by the grid. (2)

The concepts illustrated in Figure 1 are associated with the lexicon, syntax and semantics of the utterance, as summarized in Table 1. The local turning points signal words and morphemes. The pivots serve as semantic and syntactic boundaries. The general direction of the grid over the utterance, often in combination with the direction of the last intonation unit, is associated with speech act type and sentence type. These types may be declarative, interrogative and imperative to use the classical terms.

The width and position of the grid for an intonation unit signal its information weight relative to other intonation units. In this way the width and position of the grid may express coordination and subordination of various parts of an utterance. Of course, the prosodic structure, signalled by the intonation parameters, need not coincide with the semantic or syntactic structure of an utterance. It is possible to make a syntactic statement into a prosodic question and syntactic subordination may be completely eliminated by prosodic focus. In fact, there are many examples of prosody overruling syntax, particularly in spontaneous speech where intonation seems to be more closely related to semantics than to syntax. As a consequence, what is theme and rheme from the speaker's point of view can in general be identified unambiguously in such speech. Some examples will be shown in the next section.

2. Intonation parameters in real life

Turning points

An intonation curve can be efficiently described by the position in time and frequency of its end points and turning points. In fact, being reasonably smooth, the curve can be obtained by smooth interpolation between these points over the voiced segments. The turning points are in general connected with accents and tones. The timing relations between segments and turning points are relatively constant in different prosodic contexts and crucial for the identification of words and morphemes in tone languages as well as accent languages. These relations are also important for the identification of boundaries.

In Hausa, a language with two tones, analysed as high and low, the turning points occur at the end of the syllable and are independent of sentence intonation (Fig.2). Note the wider grid of the question and its slightly different direction. In Standard Chinese, which has four tones, the highs and the lows are timed in relation to the syllables which in the majority of cases are also morphemes. Figure 3 shows a sentence consisting of a sequence of alternating falling and rising tones, uttered in different sentence intonations and focus patterns. Note that the timing of the turning points in relation to the segments is nearly constant throughout the prosodic contexts.

The two accents of Swedish have been described in different ways. Schematically both accents have triangular shapes, and in all dialects Accent 2 comes later than Accent 1. Their positions relative to the syllable vary with dialect. In some dialects the accent pair comes early relative to the accented syllable, in others late (Fig.4. Bruce and Gårding 1978).

Seen from the accented syllable only, Accent 1 can be high and Accent 2 low in one group of dialects. In another group of dialects the situation is reversed (e.g. Malmberg 1967). Neither way of analysing the accents seems linguistically interesting. The most attractive analysis is probably one that describes the accents by their highs and lows in the accented syllable and the adjacent ones (Elert 1970, Bruce 1983, Gårding 1981, Rischel 1963, Öhman 1966). This is necessary for a

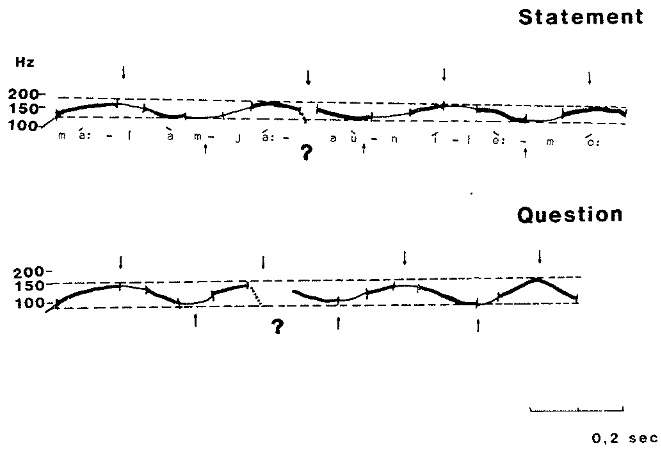


Figure 2. Turning point fixation in Hausa. Máalàm yáa aùní leèmóo. Teacher distributes oranges. The high /´/ and low /`/ tones occur at the end of the syllable marked by hyphen in the transcription. Thick lines denote vocalic segments. Arrows point to the turning points. Data from Lindau-Webb 1983.

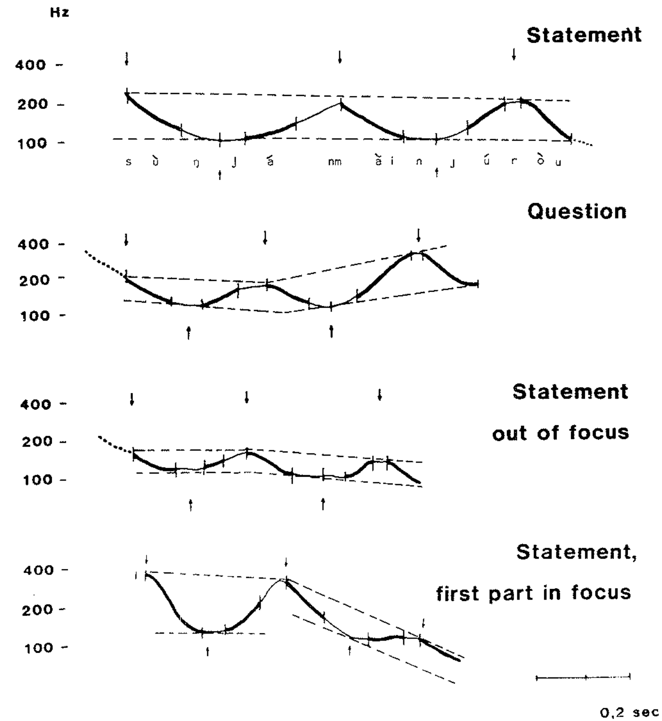


Figure 3. Turning point fixation in Standard Chinese. Sòng Yán mài niúròu. Song Yan buys calf's meat. Thick lines denote vocalic segments. Arrows point to the turning points. From Gårding et al. 1983.

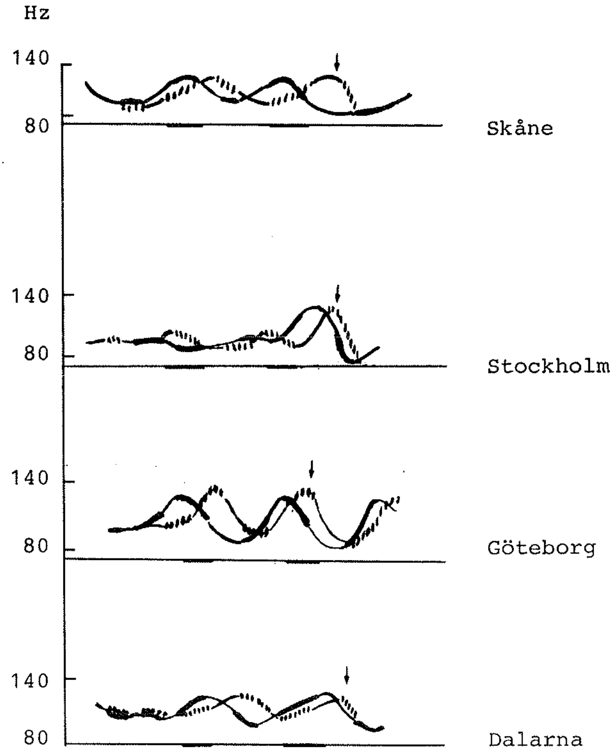


Figure 4. Comparable sentences with Accent 1 and Accent 2 in some Swedish dialects. Note the triangular shape of the accents and that Accent 2 follows Accent 1 in all cases. Note also that the accent pair is earlier in e.g. Skåne than in Dalarna relative to the accented syllable. In the accented syllables, marked by horizontal lines, Accent 1 is high and Accent 2 low in one dialect, e.g. Skåne. The situation is reversed in other dialects. Further discussion in the text.
From Bruce & Gårding 1978.

satisfactory analysis of dialectal variation and brings out the original function of the accents as markers of the word structure.

In Germanic languages outside of Scandinavia the turning points of accents are more or less fixed within the syllable but may alternate between high and low depending on the speech act intonation and the position in the phrase. Figure 5 shows how a focal high in a falling statement intonation corresponds to a low in a question (Bannert 1983, Gårding 1983c).

The grid

The sketchy definition of a grid given earlier needs some additional comments. When this notion is applied in the actual analysis of FO curves, the analyst has to bear in mind that the curve is the sum total of various elements in speech, segmental as well as suprasegmental. We cannot expect toplines and baselines, drawn automatically by connecting local maxima and minima to be a true representation of what is perceived as phrase or sentence intonation. Coarticulation effects, accentual patterns, boundary phenomena and pivots have to be recognized. There is at present no general algorithm for the construction or recognition of a grid.

Nevertheless, when the pivots have been identified, there is in most cases a natural grid, at least over intonation units that are not too short. This grid should reflect speech act, phrasing and accentuation. Connecting the maxima and minima of natural speech does not ordinarily give the idealized shapes of Figure 1 but more often wedge-shaped ones depending on the accent pattern. Figure 6 illustrates the difference between an automatically drawn grid and a grid that reflects the linguistic prosodic elements present in the intonation curve.

The intonation curve, the same in both cases, is derived from the Swedish sentence ´Hon gick inte o(ch) la sej´, She did not go to bed, in an interrogative intonation. With an

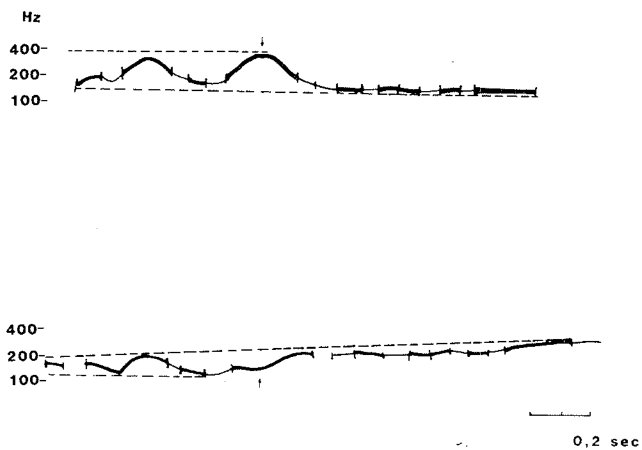


Figure 5a. Turning point fixation in English. I heard the BULLS bellow in the lane above and Did you hear the BULLS bellow in the lane bellow. The arrows point to the high and low focal turning points of bulls in the two cases.

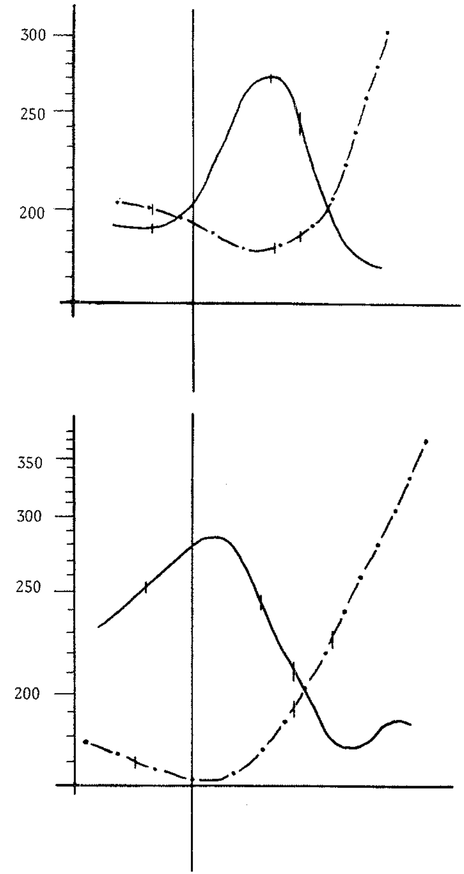


Figure 5b. Turning point fixation in German. Der Mahner. Two speakers of northern German. Solid line for statement and broken line for question. Vertical line indicates the beginning of the accented vowel. Bannert 1982.

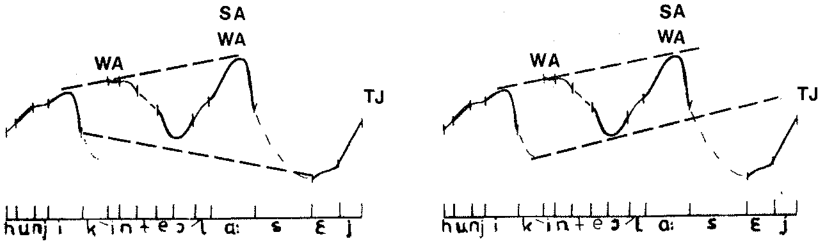


Figure 6. An automatically drawn grid left and a grid where the terminal juncture (pivot) has been recognized to the right. WA=word accent, SA=sentence accent, TJ=terminal juncture

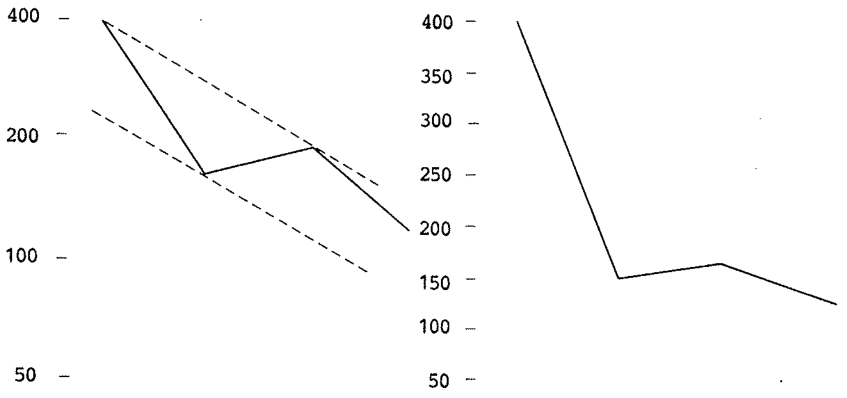


Figure 7. Fourth example of Figure 3 in logarithmic scale to the left and in linear scale to the right. The logarithmic scale lends support to the grid construction.

automatically drawn grid, as in the curve to the left, the baseline would be realized as falling and the topline as rising. This construction does not give a very clear picture of the prosodic events and their contributions to the curve.

If we presuppose the existence of a grid reflecting sentence intonation, as in the curve to the right, we get a rising grid enclosing the accents as a global expression of interrogative intonation and at the end a pivot. This pivot is manifested by the wide F0 range in connection with sentence accent (SA), and the terminal juncture (TJ), a rise from the bottom of the voice, is a local marker of the interrogative speech act.

As in the example above, the grid should as much as possible consist of approximately parallel lines on a logarithmic scale, broken at the pivots. In this way it brings out the direction of intonation while the width of the grid reflects the information weight of the corresponding phrase.(An accentual pattern with alternating strong and weak accents will have to be discussed separately.) To sum up, a single line is not sufficient to express all the information that can be provided by the grid.

For the grid to fit large pitch ranges, it is essential that the scale be logarithmic. (A logarithmic scale has also been used and motivated by Fujisaki, 1969.) Compare the fourth example of Figure 3 with the same phrase drawn to a linear frequency scale (Fig. 7).

With the linear scale the upper part of the curve is blown up out of proportion with the lower part and does not convey the auditory impression of a gradual fall.

As already mentioned the direction of an intonation unit is closely associated with the sentence type and speech act type and with its position in the sentence. The use of a falling grid to express a declarative speech act and a rising or level one to express interrogation are common features for perhaps the great majority of the languages of the world (Bolinger 1977). An important part of the manifestation of speech acts seems to be carried by the last intonation unit. This means that speech acts can have a global and/or local expression. Figure 3 shows a level grid for a statement with a local drop to creaky voice at the end, a rising grid from the subject (theme) in the question and a falling grid from the subject (focused rheme?) in the last utterance. In the Hausa example (Fig.2) the change of direction is smaller, but here the width and the tempo increase in the question.

Focus is in general expressed as an expansion of the grid in connection with the part in focus, and this expansion is often combined with a compression of the part which follows after focus. Examples are shown in Figure 3 and Figure 8. In Hausa the focused word is always initial and a prosodic signal is not necessary. In languages without distinctive accents or tones, the unfocused part is compressed into a one-line grid. Figure 8 shows how the accentual up-and-down movements are almost completely flattened out in French and Greek after focus but not in an accent language such as Swedish, nor in a tone language such as Chinese (Fig.3) A support for a tonal grid encompassing the lexical tonal prosody is given by the strongly compressed accent and tone configurations after focus. Here the lexical points have retained their positions not only in relation to the segments but also to the grid. For a strongly falling grid this may have the effect that a rising tone may have constant pitch (Gårding 1983).

Pivots

There are many kinds of pivots depending on the nature of the two adjoining intonation units and the possible large-scale movements involved at the juncture. Apart from the ones shown in Figure 1, some additional examples are given in Figure 9. It is clearly possible to assign different weights to pivots where greater weight is associated with more important syntactic and

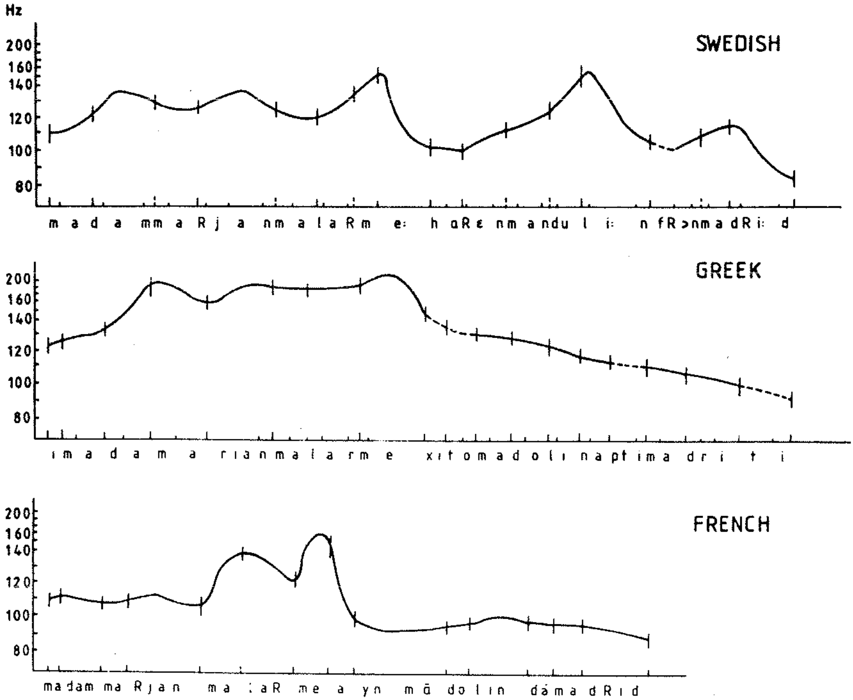


Figure 8. Language-specific ways of manifesting focus. Madame Marianne Mallarmé a une mandoline de Madrid in French and in comparable sentences in Swedish and Greek.

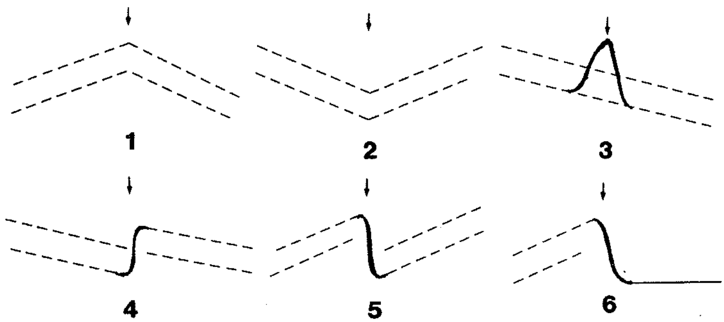


Figure 9. Different kinds of pivots. The direction of the grid changes (1,2). A pivot which does not change the direction of the grid (3). Pivots with resettings of the grid (4,5,6). Flattening of the grid after pivot (6).

semantic-pragmatic boundaries. A change of position as in 4 seems to mark a stronger break in coherence than a change of direction as in 1 or 2.

A strong degree of coherence between two sentences of a text may be shown by a unidirectional falling grid as in Figure 10 with a junctural pivot to mark the sentence boundary. A lesser degree of coherence is produced by a new, upshifted grid as in the same figure. Similarly two rising interrogations are connected by a downshifted grid (Fig.10). Figure 11 shows fundamental frequency curves of "Marie a une mandoline de Panama" and the corresponding sentence in Swedish, both pronounced as the first sentence of a short story. The intonation has certain traces of the syntactic structure of the sentence. In connection with the main syntactic boundary, there is a light pivot with a change in the direction of the grid in both languages.

The introduction of focus in the sentence, as in examples 2A

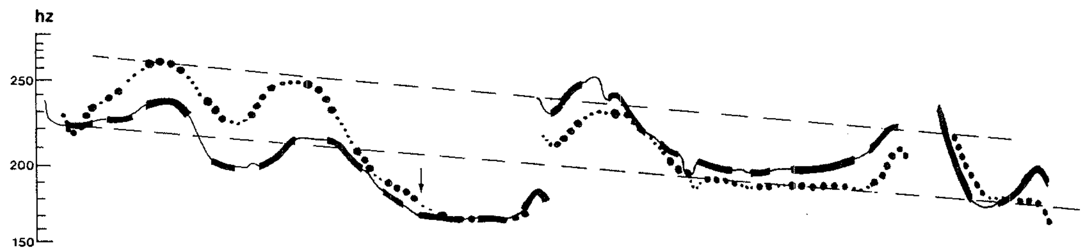


Figure 10. Coherence of phrases.

(a) Man anammar lundamodellerna 'They accept the Lund models', Lundamodellerna är lockande 'The Lund models are appealing'. The dotted lines shows declarative intonation stretching over two sentences with a common grid indicated by broken lines. The pause between the two sentences has been omitted. The solid line shows corresponding sentences each with a declarative intonation of its own and an upshift of the grid (not marked in the figure). This gives the impression of a fresh start and more emphasis to the second sentence. The arrow points to the terminal juncture fall, similar in both situations. From Bruce 1982 apart from the grid lines.

(b) Hon hade hög feber? 'She had a high temperature?', Hon gick inte och la sig? 'She did not go to bed?' Interrogative intonation in two connected sentences with a downshift of the grid. The speaker avoids going beyond her natural range. The arrow points to the terminal juncture rise. From Gårding 1982.

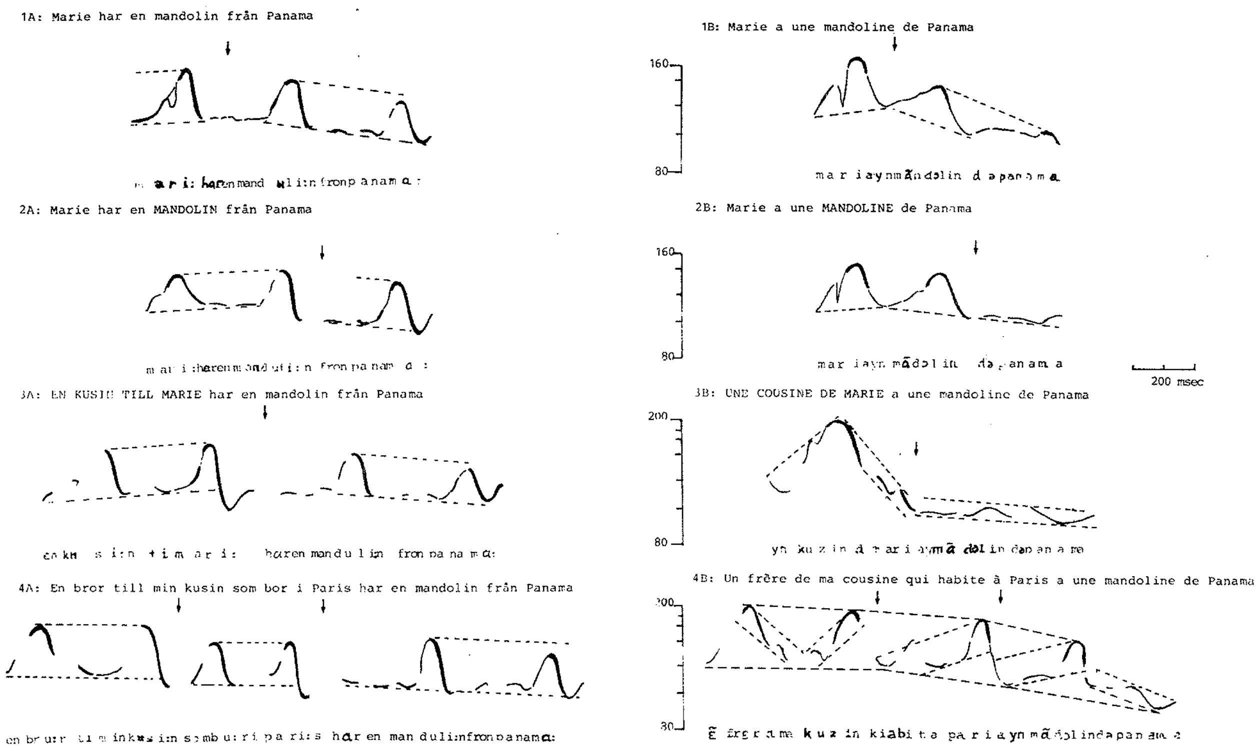


Figure 11. Some functions of intonation in French and Swedish.
From Gårding 1983

and 2B, produces a strong pivot after the focused word, splitting the sentence into a thematic part and a rhematic part. In the Swedish dialect exemplified here, the strong pivot is manifested as a large-range downward movement preceding the directional change of the grid. In the French example the grid is compressed after focus. In both languages the position of focus is free and the focal pivot divides the sentence into parts which are not given by the syntax but by the theme-rheme division chosen by the speaker.

The last pair of examples from French and Swedish show similarities and differences in the use of pivots and grids. In the Swedish example, the non-restrictive relative clause is set off by a compressed grid and the unity of the main clause is signalled by a continuation of the gradually declining topline of the first intonation unit. The intonation units here are well correlated with syntax but they have at the same time a more general semantic function, that of subordinating information parenthetical to the main message. In the corresponding French example the grid is broken more frequently by pivots. In this way the information units are shorter and have steeper slopes. In fact, intonation units in French can be said to play the role of accents in Swedish, namely to signal content words with appended functional items. In the French sentence the relative clause is set off by jump-down pivots. To show the global direction of the sentence intonation, the highs and lows of the intonation units have been connected. This global grid is falling as in Swedish, and both languages have an extra local fall on the last semantic unit, Panama.

3. Comparison in terms of a generative model for intonation

The analytic scheme demonstrated in the preceding sections has been worked out in connection with a generative model for prosody which has been the object of several articles since the mid seventies (see Gårding 1981, 1983a and references there).

Its object is to generate prosody in different languages in a uniform way. Here I shall only give a brief sketch.

The most important feature of the model is that it factors out utterance prosody from lexical prosody by a grid construction. The input to the model is a sentence with phonological markings. In the first stage correct syllable durations are generated from the input markings. In the second stage the phonological markings are subject to intermediate phonological rules which concern context rules at the symbol level. Exemples of such rules are the deletion of accent marks after focus in some languages and Chinese sandhi rules. In the following stage the accents or tones are converted into highs and lows or combinations of these features. Finally pitch is generated. Characteristic of the pitch generating scheme is that the sentence and phrase intonation are generated first from the speech act markings and boundaries in the form of a grid. The lexical highs and lows are then given their proper positions relative to the syllables and their frequency values relative to the grid. The next step modifies these positions by accentual rules and context rules. Finally the curve is obtained by smooth interpolation over the voiced segments through the points generated earlier. Figure 12 shows the model and the principles of the pitch generating scheme.

The generative model is useful for characterizing prosodic systems and comparing them. A summary of the intermediate phonological rules, the intermediate pitch representations and the rules of the pitch algorithm constitutes the main features of the intonational profile of a language or dialect. By adding an account of the function and distribution of the prosodic features needed in the input, we obtain something that could be called a prosodic grammar (Gårding 1983a). The uniform way in which such grammars can be obtained makes prosodic comparisons, also point by point, meaningful. To carry out this program for a sizable corpus of languages is of course a very large undertaking, but even with our small set of prosodically

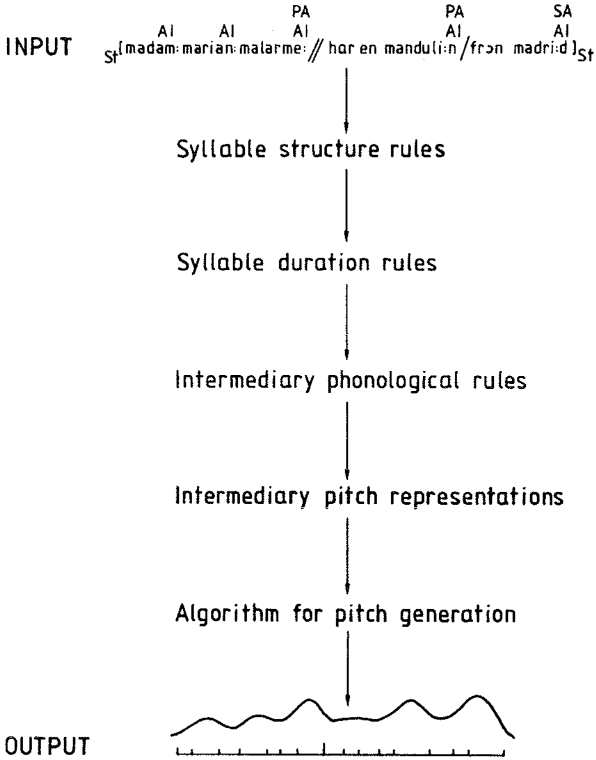


Figure 12a. Model for prosody and principles of the pitch algorithm.

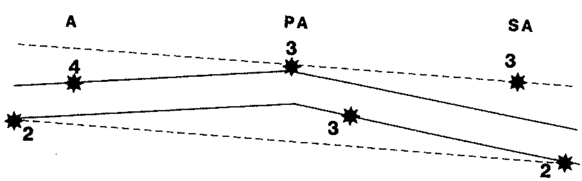


Figure 12b. Priciples of the pitch algorithm. Typical grid with insertion of Highs and Lows.
PA=phrase accent, SA=sentence accent, A=word accent

different systems and limited analyses, it is possible to venture some conclusions.

Grids and pivots

The notions of grid and pivot are applicable in a natural way to the languages analysed so far. Speech acts as manifested by grids seem to exhibit the following general features. They can have a global and/or local expression. In polyphrasal utterances it is the last phrase that carries the main part of the speech-act information. The intonation of the preceding phrase(s), at least in preplanned speech, is adjusted to fit the last phase of the sentence in an optimal way. A rising speech-act contour is preceded by a falling or level one and vice versa.

A falling final intonation unit expresses terminality in a general sense and is strongly connected with the declarative speech act. A rising or level intonation unit expresses non-terminality and connects with an interrogative speech act. The same is true of falling and rising (level) global patterns. This is a general tendency in the languages of the world (Bolinger 1978).

Focus is expressed as an expansion of the grid in connection with the part in focus and this is often combined with a compression of the part which is outside focus. At this point there is a difference between languages with lexical tones and accents and those without lexical prosody. Lexical tones and accents seem to prevent too much compression of the grid.

Accents and tones

The manifestation of accents and tones are local features of an intonation curve where highs or lows or a combination of highs and lows are fixed in time to specific parts of the carrying segments. This turning-point fixation seems to be a general feature. On the other hand, morpheme-bound tones and word-bound accents bring about some essential differences of phonological as well as phonetical nature. Take for instance Chinese and Swedish. Pitch goes up and down more often in Chinese than in Swedish, evidently because of the greater functional load of the Chinese tones. The pitch range is wider in Chinese, and this is at least partly due to the fact that a rising or high tone in focus expands the range upwards towards the ceiling of the voice whereas falling and low tones go in the other direction (Gårding et al. 1983, Gårding 1984). In Swedish, focusing is achieved by the same principle, only here the effect will be dialect dependent. In a dialect with rising accents, focus raises the highs, while in a dialect with falling accents, the lows will lowered.

Prescriptions for the location of tones and accents in relation to the syllables show some important differences in tone and accent languages, even if the phonological representation is the same. The turning point of a Swedish accent, it may be high or low, tends to be located in the middle of the vocalic segment which is closely associated with the intensity maximum. This position is slightly shifted in focal or emphatic context (Gårding 1983c). This does not seem to be the case in Chinese. Here the turning point is connected with the syllable-initial consonant. We are reminded of the classical discussion of tones and accents as having different physiological mechanisms.

Superposition

The principle of regarding intonation as the sum of a phrase intonation component and a lexical component is suitable for all the languages that have been analysed here. It finds its expression in the grid as a frame for the accentual and tonal pitch movements. More precisely, the accents or tones are superimposed on a slowly varying intonation.

Conclusion

To conclude this paper let me try to express an overall view of the manifestation of intonation which has emerged from the comparison of some different prosodic systems. Intonation curves in general can be said to have a threefold structure, each part encompassing the next, namely

(1) a global pattern of linear elements, related to sentences and speech acts
(2) a short-term linear structure, related to phrases and
(3) a local wave structure, related to lexical accents and tones.

Notes

1) Intonational similarities and differences between languages have also been studied by Jacqueline Vaissière, see e.g. Vaissière 1983, with overlapping results.

2) Superposition is implied in the metaphors used by Bolinger for accents as 'ripples on waves' (1964) and similary by Chao for tones (1968). Superposition was used explicitly in a quantitative model of intonation by Öhman (1967), Fujisaki (1969) and in my own qualitative model (1970 and subsequent papers), Mc Allister (1971), Thorsen (1978) and several others. The superposition principle has been considered contrary to the tonal sequence principle advocated by Pierrehumbert (1981). See Thorsen (1983).

REFERENCES

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

Bolinger, D. 1964. Intonation as a universal. Proceeding of the International Congress of Linguistics 1962, ed. Lunt. The

Hague: Mouton

Bolinger, D. 1977. Intonation across languages. In Universals of human language, ed. Greenberg, Vol.2: Phonology. Stanford University Press: 471-524

Bruce, G. 1977. Swedish word accent in sentence perspective. Travaux de l'Institut de Linguistique de Lund XII. Lund: Gleerups

Bruce, G. 1982. Experiments with the Swedish intonation model. In Preprints of Working group on intonation, eds. Fujisaki and Gårding, Tokyo: 35-45 and Working Papers 22, Department of Linguistics: 51-116

Bruce, G. and Gårding, E. 1978. A prosodic typology for Swedish dialects. In Nordic Prosody, eds. Gårding, Bruce & Bannert, Travaux de l'Institut de Linguistique de Lund XIII: 219-228

Chao, Y.R. 1968. A grammar of spoken Chinese. Berkely and Los Angeles

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

Fujisaki, H. and Nagashima, S. 1969. A moel for synthesis of pitch contours of connected speech. Annual Report Eng. Res. Inst., University of Tokyo, vol. 28: 53-60

Gårding, E. 1970. Word tones and larynx muscles. Working Papers 3, Phonetics Laboratory Lund University: 20-46

Gårding, E. 1981. Contrastive prosody: A model and its application. Special lecture to the AILA congress 1981. Proceedings II, eds. Sigurd & Svartvik, Studia Linguistica 35, 1-2, 81, Gleerups: 146-166

Gårding, E. 1982. A comparative study of intonation. In Preprints of Working group on intonation, eds. Fujisaki & Gårding, Tokyo: 85-94

Gårding, E. 1982. Prosodic expression and pragmatic categories. Working Papers 22, Department of Linguistics, Lund University: 117-136. Also in Textstrategier i tal och skrift. Almqvist & Wiksell International, 1982

Gårding, E. 1983a. A generative model of intonation. In Prosody: Models and measurements, eds. Cutler & Ladd, Springer: 11-25

Gårding, E. 1983b. Intonation units and pivots, syntax and semantics. In Abstracts of the tenth international congress of phonetic sciences, eds. Cohen & v.d. Broecke, Foris: 99-103

Gårding, E. 1983c. Superposition as an invariant feature of intonation. Contribution to the MIT conference of Invariance and variability of speech signals. In press

Gårding, E. 1984. Chinese and Swedish in a generative model of intonation. In Nordic Prosody III, ed. Elert, University of Umeå

Gårding, E. and Lindblad, P. 1973. Constancy and variation in Swedish word accent patterns. Working Papers 7, Phonetics Laboratory Lund University: 36-110

Gårding, E., Botinis, A. and Touati, P. 1982. A comparative study of Swedish, Greek and French intonation. Working Papers 22, Department of Linguistics, Lund University: 137-153

Gårding, E., Zhang, J. and Svantesson J-O. 1983. A generative model for tone and intonation in Standard Chinese. Working

Papers 25, Department of Linguistics, Lund University: 53-65

Lindau Webb, M. 1983. Tone and intonation in Hausa. Paper presented to the tenth international congress of phonetic sciences.

Malmberg, B. 1959. Bemerkung zum Schwedischen Wortaccent. Zeitschrift fur Phonetik und allgemenie Sprachwissenschaft 12, 193-207

Mc Allister, R. 1971. Predicting physical aspects of English stress. Speech Transmission Laboratory, QPSR, Royal Institut of Technology, Stockholm: 20-29

Öhman, S. 1966. Generativa regler för det svenska verbets fonologi och prosodi. Förhandlingar vid sammankomst för att dryfta frågor rörande svenskans beskrivning III: 71-87

Öhman, S. 1967. Word and sentence intonation: A quantitative model. QPSR, Speech Transmission Laboratory Royal Institute of Technology, Stockholm, No 2/3: 20-54

Pierrehumbert, J. 1980. The phonology and phonetics of English intonation. MIT doctoral dissertation

Rischel, J. 1963. Morphemic tone and word tone in Eastern Norwegian. Phonetica 10: 154-164

Thorsen, N. 1978. An acoustical analysis of Danish intonation. Journal of Phonetics 6: 151-175

Thorsen, N. 1983. Two issues in the prosody of standard Danish. In Prosody: Models and measurements, eds. Cutler & Ladd, Springer: 27-38

Vaissière, J. 1983. Language-independent prosodic features. In Prosody: Models and measurements, eds. Cutler & Ladd, Springer: 53-66

Experiments with Filtered Speech and Hearing-Impaired Listeners

David House

ABSTRACT

Speech perception experiments using LPC pitch-edited low-pass filtered speech stimuli were presented to normal and hearing-impaired listeners in an attempt to study the interplay of two kinds of frequency movement (Fo and formant transitions) in the perception of stops and CVC words in a speech context, and to test the hypothesis that increased movement of Fo can aid stop identification. The results indicated little difference in stop identification relating to Fo movement. There were, however, certain considerable differences in results between listeners with normal hearing and listeners with a noise-induced hearing loss indicating the possibility of an acquired perception stratgegy used by the hearing-impaired. Correlation between filter frequency, hearing loss and test performance was also observed.

1. INTRODUCTION

Intonation has been shown to be a major factor in determining the perception of syllable stress (Fry, 1958). Furthermore, it appears that fundamental frequency movement in sentence intonation provides an overriding cue to stress, and that this movement may produce a perceptually all-or-none effect, the important factor being movement per se and not magnitude of movement (Lehiste, 1970). Focus of a sentence, however, as defined by Jackendoff (1972) as being the "new" information not shared by the speaker and the hearer, can be signaled by introducing emphatic sentence stress and thereby significantly increasing the movement and range of Fo (House, 1983).

Although intensity and duration are correlates of stress, fundamental frequency movement seems to provide relatively stronger perceptual cues to the location of stress (Lehiste, op.cit.). Fo movement then provides, from the point of view of dynamic perception (Johansson, 1975), a change in frequency over time which could be registered as an event by the perceptual mechanism. This event would in turn sharpen attention and aid in short-term memory retrieval of spectral cues. Since resolution of spectral cues can be seen as more crucial in the "bottom-up" processing of new information than in presupposive "top-down" speech processing (Marslen-Wilson & Tyler, 1980), it would be interesting to investigate to what extent segmental resolution could be facilitated by varying degrees of fundamental frequency movement where semantic focus is constant in "bottom-up" processing using semantically non-redundant speech stimuli.

Cutler (1976), using phoneme-monitoring reaction time experiments, concluded that a significant difference in reaction times can be attributed to the prediction of upcoming stress locations even when the target-bearing word in stress position does not contain a high-stress intonation contour. Reaction times were, however, substantially shorter when the target-bearing words contained high-stress contours.

Intonation, then, points toward the focussed word where segmental resolution is facilitated by virtue of sharpened attention on the part of the listener coupled with changes in pitch, vowel duration and intensity. The question raised here is one concerning the role of frequency movement in segmental resolution. Can the same movement in frequency used to signal location of sentence stress be used by the perceptual mechanism in interaction with other cues to aid in segmental resolution?

One possible interaction could be between pitch movement realized as frequency movement of harmonics of the fundamental in the vowel and vowel formant transition movement realized as resonance induced amplitude shifts between successive harmonics.

Since formant transitions are important cues for stop identification, such interaction might facilitate perception of transitions and aid in stop identification. On the other hand, interaction might not necessarily result in an amplification of formant transitions and would therefore not facilitate stop identification.

A further question that arises is that if these two kinds of movement in frequency interact to facilitate segmental resolution, what are the possible implications for listeners with hearing disabilities? It is well documented that individuals with moderate sloping sensorineural hearing losses have difficulty in identifying place of articulation especially in voiceless stops. However, subjects having similar audiometric configurations can differ radically in their performance in both synthetic and natural speech tests (Van de Grift Turek, et al. 1980; Risberg & Agelfors, 1978; Picket, Revoile, & Danaher, 1983). Could Fo movement as a correlate of stress aid hearing-impaired listeners in identifying stop consonants?

In attempting to answer these questions, word identification tasks could be presented through a filter roughly corresponding in frequency to a typical noise-induced audiometric configuration for hearing-impaired listeners. By presenting the stimuli both to listeners with normal hearing and to hearing-impaired listeners, the experiment would have two goals: 1.) to test frequency movement interaction as an aid to stop consonant identification in filtered speech, and 2.) to compare performance of listeners whose audiometric configurations correspond to the filter frequencies used in the presentations. The latter goal might also help in exploring the relationships between perception of the speech wave filtered before reaching the auditory perifery and perception of the speech wave altered by an impairment of the auditory perifery.

2. METHOD

A. Linguistic material and stimuli

There is much debate concerning differences in perception of sense vs. non-sense speech utterances and utterances in and outside of a sentence frame in listening experiments (Pastore, 1981; Johnson & Strange, 1982). It seems, however, that when dealing with stimuli involving both sentence intonation and local segmental cues and their interaction, the closer the stimuli can be to real-life speech, provided of course that variables can be sufficiently controlled, the more we can learn about the communicative aspects of speech perception. (See also Gårding, 1967 for differences between juncture perception in sense and non-sense words.)

The carrier sentence "de´ va´ _ ja´ sa´" (It was _ I said.) was selected such that a vowel would immediately precede the target word. 26 single-syllable CVC words were chosen having an initial voiced or voiceless stop (Table 1). Target word initial stop was considered as the perceptual target phoneme and Fo movement in the preceding and following vowel was to be altered to test interaction between stop identification and frequency movement in the two adjoining vowels.

Five of the test sentences were recorded by a male speaker of Southern Swedish in two versions, first with emphatic high stress given to the target word, then with indifferent low stress. The 26 test sentences were then recorded by the same speaker using neutral intonation and stress.

The five low-stress and five high-stress tokens were digitized

	TAL (speech)	KAL (bare)	BAL (dance)	DAL (valley)	GAL (to crow)
PAR (pair)	TAR (take)	KAR (tub)	BAR (bare)		
	TUR (luck/turn)	KUR (cure)	BUR (cage)	DUR (major key)	
	TÅR (tear/toes)	KÅR (corps)	BÅR (stretcher)		GÅR (go/walk)
PÅG (boy)	TÅG (train)		BÅG (cheat)		
	TAM (tame)			DAM (lady)	GAM (vulture)
	TOK (fool)	KOK (potful)	BOK (book)		

Table 1. Target words used in the stimulus sentences.

using a VAX computer at a sample rate of 10,000Hz. The tokens were then analyzed with a linear prediction analysis method (ILS program package). The five contours for the two intonation types were averaged to serve as a natural model for pitch editing of the sentences with neutral intonation. The 26 neutral intonation sentences were digitized and analyzed in the same manner.

Two stimulus versions of each sentence were then synthesized from the linear prediction coefficients with the pitch contour selected to conform to the low and high-stress intonation models respectively (Fig. 1). The resulting 52 stimuli were randomized in the computer. A pilot test using four listeners was run to determine a satisfactory low-pass filter cut-off frequency which would allow correct initial stop identification of around 50% in the target word. An eight order Butterworth low-pass filter with the cut-off frequency set at 900Hz, -48dB/octave allowed the 50% correct identification in the pilot study.

The stimuli were recorded on tape through the filter in thirteen blocks each containing four stimuli with a 10-second interval between stimuli and a 20-second interval between blocks. An additional block was placed at the beginning as a practice buffer.

B. Subjects

25 beginning speech pathology students at Lund University with normal hearing and unfamiliar with the test material participated in the experiment as part of a course requirement. 13 patients at the Department of Otorhinolaryngology, University Hospital, Lund, with noise-induced hearing losses of roughly similar audiometric configuration (beginning of slope at around 1000Hz, approximately -40dB at 2000Hz and -60dB at 3000Hz) voluntarily participated in the experiment as an extended part of routine audiological examinations. Patients were selected on the basis of previous pure-tone audiograms

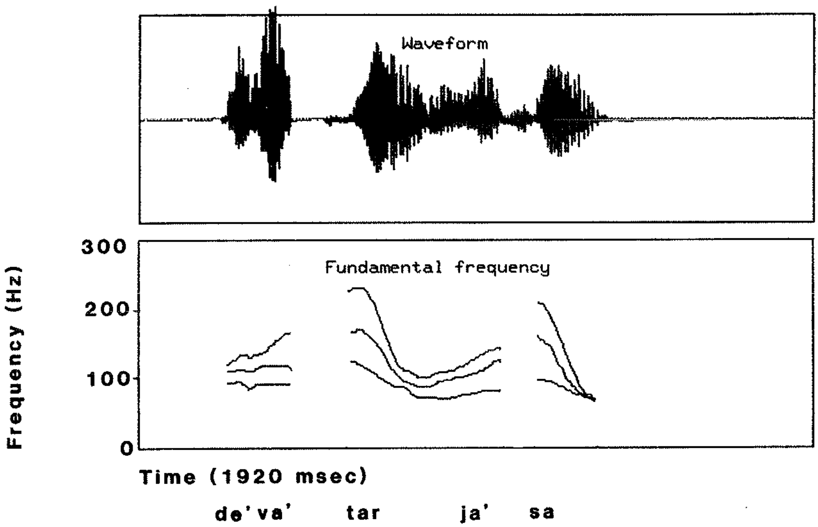


Figure 1. Waveform and fundamental frequency contours of one stimulus sentence. The middle contour represents the original neutral stress, the upper contour is the high-stress edited version and the lower contour is the low-stress edited version. The waveform represents the neutral version. Only the two edited versions were used in the test.

where relative symmetry of hearing loss was a criterion as well as break frequency and degree of slope.

C. Procedure

The normal-hearing subjects were tested in three groups on three different occasions. Written instructions and answer sheets were handed out and the instructions were read aloud by the experimenter. The subjects were informed that they would hear the carrier sentence "Det var _ jag sa." presented through a filter and that they should try to write the word following "var" in each sentence. If they were unsure they were requested to guess the word closest to the sound they heard. After the practice block was run subjects were allowed to ask questions. Testing took place binaurally in a sound-treated perception laboratory using a Revox A77 tape recorder and Burwen PMB6 Orthodynamic headphones. Sound level was checked as comfortable during the practice block. The test took 15 minutes.

The hearing-impaired persons were tested individually while sitting in a sound-insulated room. Routine pure-tone and speech audiograms were first made after which Békésy sweep-audiograms with pulsed stimuli were performed for each ear. This was done to more closely define the steepness and frequency location of the hearing loss. The same instructions were presented to the listeners except that they were asked to repeat the words orally instead of in writing. The responses were monitored outside the sound-insulated room over a loudspeaker and recorded on tape.

The stimuli were presented monaurally through a Revox A77 tape recorder, via the speech channel of a Madsen Clinical Audiometer Model OB70 and matched TDH-39 headphones with MX-41/AR cushions. The stimuli were presented at most comfortable level established during the practice block and generally corresponding to 30dB over speech threshold and to the level for maximum speech discrimination established during speech audiometry.

Monaural presentation was deemed advisable since hearing loss was not completely symmetrical. A second randomized version of the tape was made, and learning effects were minimized by presenting half the first version to the left ear, the entire second version to the right ear and then the remaining half of the first version to the left ear. After hearing the filtered stimuli the subjects were given a break and then the test was repeated using non-filtered stimuli. The test, including the Békésy audiograms but excluding the routine audiograms, took approximately an hour and a half. The patients were extremely cooperative especially considering the length of the test.

3. RESULTS

A. Normal-hearing listeners

All three groups showed similar patterns of initial stop identification for the filtered stimuli. Roughly one-half of the stops were correctly identified, and a general bias toward labials was observed. The voiced-voiceless distinction was perceived by all subjects in nearly all target words with place-of-articulation for voiced stops being somewhat easier to identify than for voiceless stops. The mean number of correct stop identifications for the normal-hearing subjects as a group (Group 1, Fig. 2) was higher (14.5 of 26) when the sentence and target word carried the indifferent, low-stress intonation contour than when the sentence and word carried emphatic, high stress (12.7 of 26). The difference was significant, $p<0.05$, running contrary to the hypothesis. Correct word identification, however, did not reveal a significant difference between low and high stress, although more low-stress words were correctly identified. Labials /p,b/ and the voiced velar /g/ were favored by the normal-hearing listeners (Fig. 3). In only one phoneme /d/ were substantial differences observed relating to stress contours. The low-stress versions received more than twice as many correct responses compared to their high-stress counterparts. The

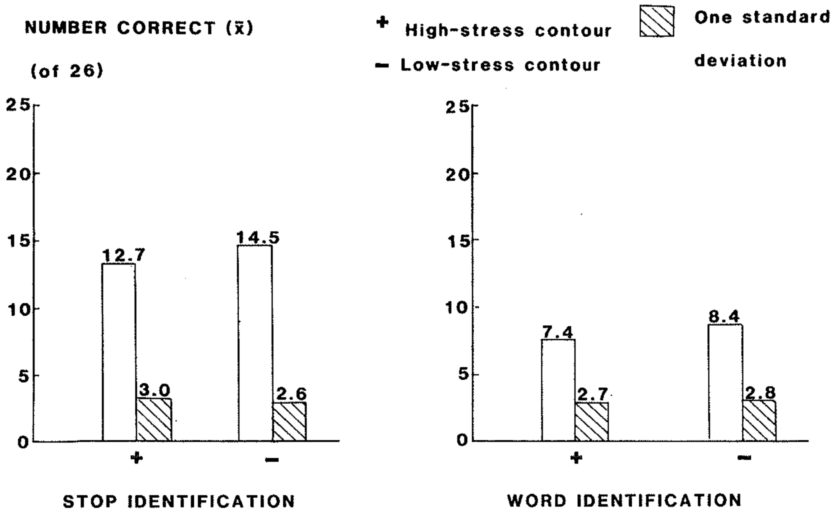


Figure 2. Mean number of correct identifications for stops and words in low-pass filtered sentences with high and low-stress fundamental frequency contours. (Group 1, listeners with normal hearing.)

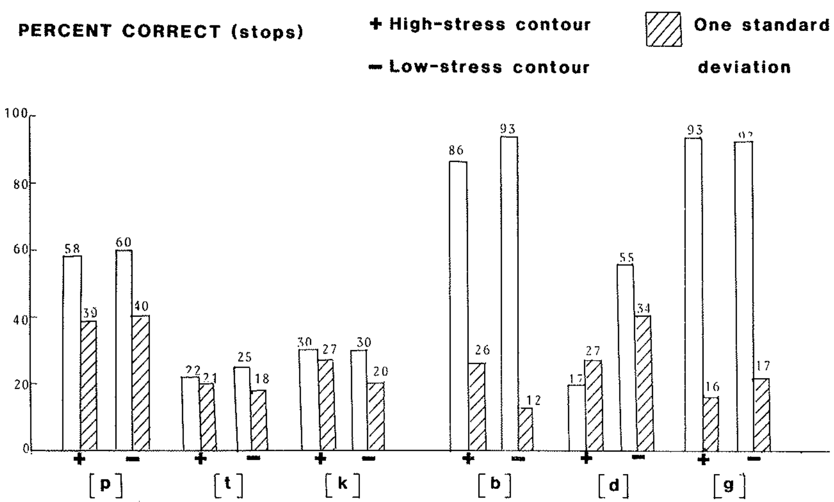


Figure 3. Correct initial phoneme identification in percent for Group 1 (Normal hearing)

vowels were nearly always correctly identified.

B. Hearing-impaired listeners

The number of stops and words correctly identified by the hearing-impaired listeners as a group was about the same as for the normal group: about one in two for stops and one in three for words (Fig. 4). Again, for both words and stops, identification was slightly better for sentences having the low-stress Fo contour, although here this difference was not significant, $p>0.05$. Labials and velars were again favored in the voiced stimuli results, but the dental phoneme /t/ was favored in the voiceless results (Fig. 5). Standard deviation for both stops and words in both stress categories was greater for the hearing-impaired group than for normals. As with the normal group, the vowels were nearly always correctly identified.

C. Differences between the two groups

As previously mentioned, the hearing-impaired group as a whole did better on both stop and word identification in the filtered speech, although the difference was not significant, $p>0,05$. There was, however, a striking difference between the two groups manifested by the preference for the voiceless dental /t/ among the hearing-impaired group which contrasted to the labial preference /p/ among the normal group.

Reactions to the test by subjects of the two groups also differed. Members of the normal-hearing group felt that the test was extremely difficult and frustrating. There were, however, substantial differences among members of the hearing-impaired group in both performance and reactions to the test. These listeners basically fell into two categories. Either they reacted much like the normal group complaining about the difficulty of the task and obtaining results similar to the normal group or they made many correct responses from the very beginning of the test, performed better throughout the test than the other groups, and did not feel that the

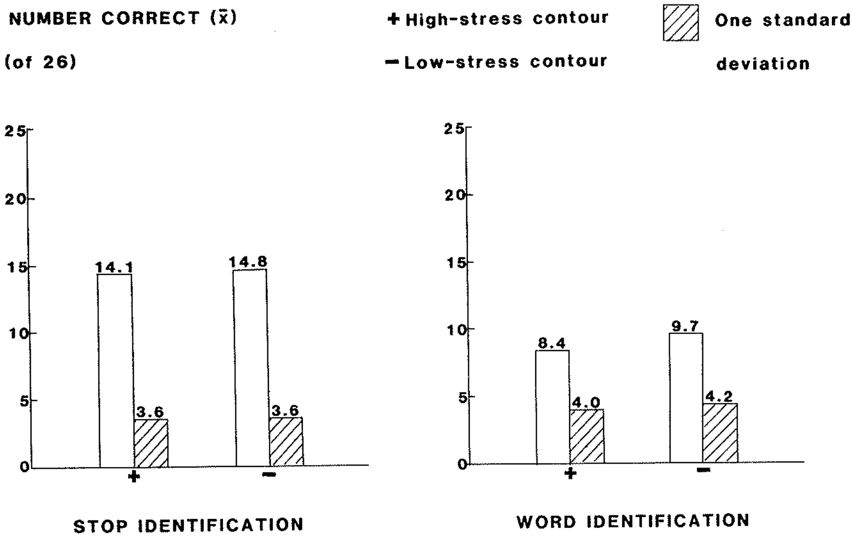


Figure 4. Mean number of correct identifications for stops and words in low-pass filtered sentences with high and low-stress fundamental frequency contours. (Group 2, hearing-impaired listeners.)

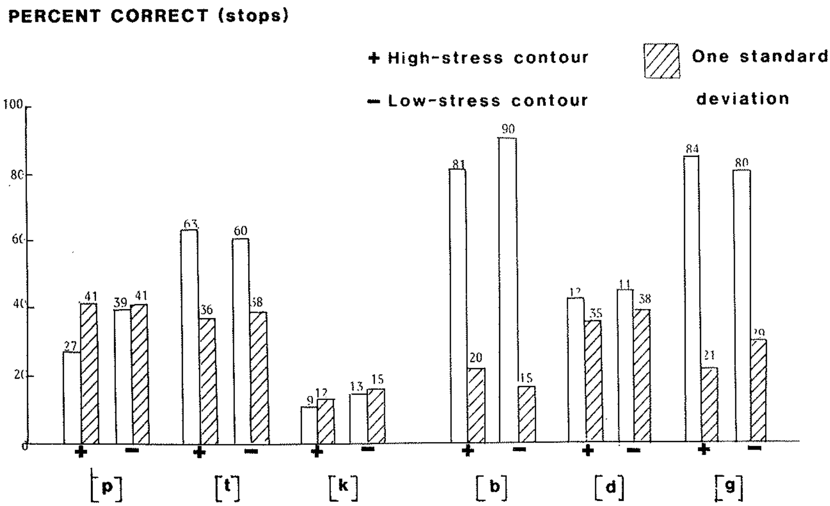


Figure 5. Correct initial phoneme identification in percent for Group 2 (Hearing-impaired).

presentation was particularly difficult or unusual.

In order to interpret these differences, subcategories of the normal-hearing group (Group 1) and the hearing-impaired group (Group 2) were made. The results were combined for low vs. high stress since those differences were generally not significant. The two groups were then divided up on the basis of best results (28 or more correct identifications for stops, Groups 1A and 2A) and worst results (Groups 1B and 2B). Figure 6 shows correct identification for these subcategories. The difference in stop responses between the best hearing-impaired group (2A) and the normal group as a whole (1) was highly significant $p<0.003$. The difference in word identification was also significant, $p<0.05$. This difference is, however, less convincing when the best hearing-impaired group is compared to the best normal group. The hearing-impaired group still performed better (34 correct vs. 30 correct for stops, 24 vs. 18 for words), but the differences were not significant, $p>0.05$.

Since any group can be subcategorized using a best-results criterion, the hearing-impaired group was divided into two new categories using pure-tone audiogram configuration criteria. The categories were (Group 2C) those ears most closely resembling the filter function used in the test, i.e. severity of hearing loss increasing sharply at a drop-off frequency lower than 1500Hz and at least a -35dB threshold drop between 1000Hz and 2000Hz and a threshold of less than -50dB(HL) at 2000Hz; and (Group 2D) those ears which least resembled the filter function, i.e. either the slope was too flat or the drop-off frequency was greater than 1000Hz (see Figures 7 and 8 for example audiograms). Correlation between the best-ear group (2A) and the most-like-filter group (2C) was high with all the ears occuring in 2C also being represented in 2A. Identification results for these two groups, for both stops and words, were also very similar, as can be seen in Figure 6. Differences in identification for both stops and words between

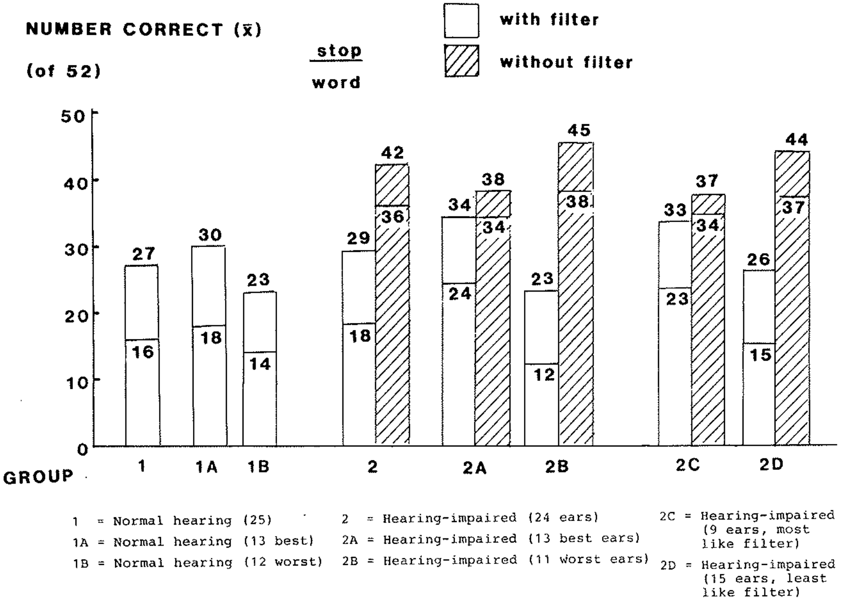


Figure 6. Mean number of correct identifications for stops and words for various group subcategories. (See text for subcategory criteria.)

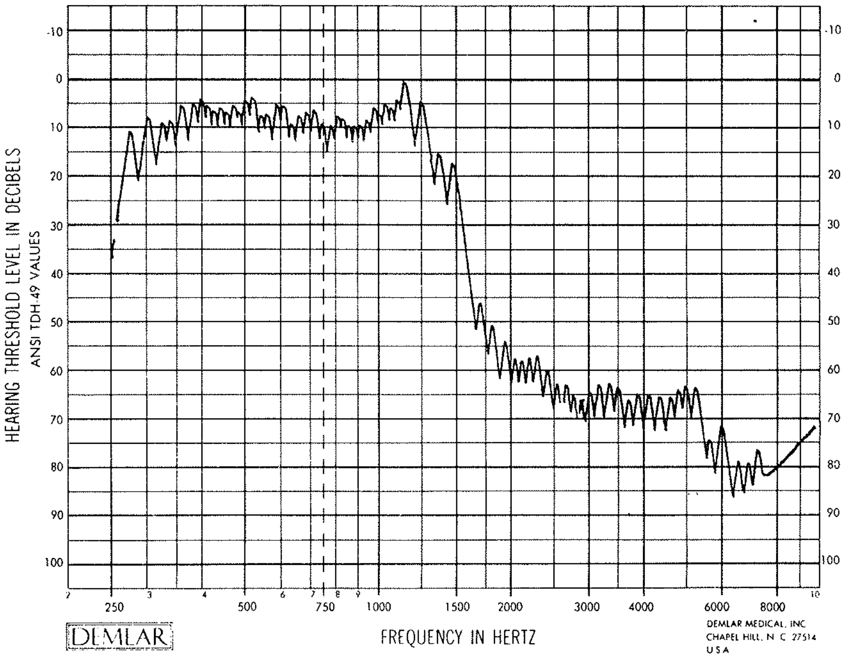


Figure 7. Example audiogram (Békésy) of a "most-like-filter" ear.

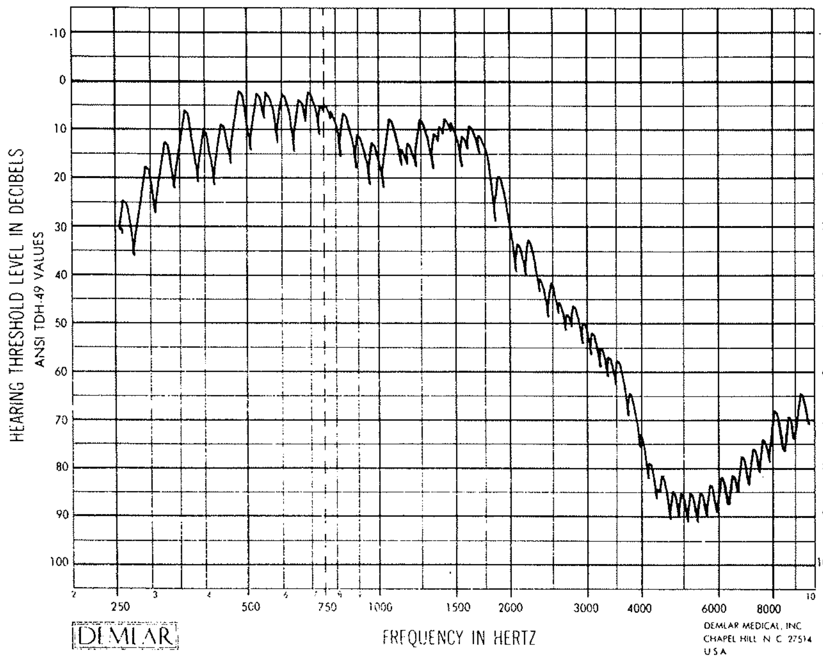


Figure 8. Example audiogram (Békésy) of a "least-like-filter" ear.

Group 2C (most-like-filter) and Group 1 (normals) were significant, $p<0.05$.

A final point of interest when comparing the hearing-impaired subgroups concerns listener reactions to the non-filtered test version. In general, those listeners who performed best on the filtered version noted little or no difference between the filtered and non-filtered versions, while those listeners who performed worst on the filtered version performed best on the non-filtered version (Fig. 6) and noted a considerable improvement in "clarity". There were, however, no significant differences which could be attributed to low or high stress. The normal listeners were not tested on the non-filtered version as an informal test indicated 100% correct identification.

4. DISCUSSION

A. Fo movement and stop identification

The results of these tests provide a negative answer to the question of whether or not increased movement of Fo in the vowels adjoining a stop can aid hearing-impaired listeners in identifying stops. The same results apply to normal listeners using filtered speech. In fact, in all groups and on all tests, identification of both stops and words was slightly better when the vowels carried low-stress pitch movement, i.e. lower Fo and little absolute movement. An explanation for this could lie in the fact that a lower fundamental produces a tighter series of harmonics which could in turn supply more energy to the critical formant amplitude shifts in the transitions thereby enabling better identification. The improvement, however, can only be seen as highly marginal as the differences were not statistically significant.

It seems then that at least regarding filtered speech in Swedish and for hearing-impaired listeners, heightened Fo movement related to focus and sentence stress serves as a marker to direct the attention of the listener to the focussed word but does not intrinsically aid the perceptual mechanism in segmental resolution. It could be that the perception of Fo movement is integrated over a longer time interval than the perception of segments and, at least where global Fo movement is concerned, Fo movement perception is related to the pragmatic intentions of the speaker rather than to the phonemic content of the word in focus. If this is the case, then the perceptual mechanism could rely on increased intensity in the vowel and increased vowel duration during stress to aid in segmental resolution. During an informal listening session it was felt by several listeners that the high-stress stimuli sounded "thinner and weaker" than their low-stress counterparts. This could be due to the fact that the increased intensity normally associated with high stress was missing.

An aspect of production which could also contribute to a separation in perception of the two kinds of frequency movement dealt with here, i.e. Fo and formant frequency, is that of source. If perception of movement in the speech wave can be coupled to articulator movements as described by Fowler, et al. (1980), then the separate nature of the sources, i.e. tongue and jaw movement to alter resonance and laryngeal movement to alter fundamental frequency, could be perceived as relating to these separate sources from a production standpoint and therefore be processed separately. Clearly, these kinds of production-perception interactions and the processing of different kinds of movement need to be investigated further.

B. Compensation strategies and hearing-impaired listeners

Perhaps the most interesting result of the experiment pertains to the difference in performance between the normal-hearing listeners and the hearing-impaired listeners, especially those whose audiograms best matched the filter. The greatest difference in performance can be attributed to the tendency for hearing-impaired listeners to choose /t/ instead of /p/ or /k/

while the normal listeners tended to choose /p/. As the frequency of /t/-words dominated in the test material, the hearing-impaired group naturally came out ahead. A possible explanation for this could lie in the fact that listeners with a hearing loss, being accustomed to hearing speech resembling the filtered stimuli, were able to distinguish between the presence or absence of /p/ (low-frequency burst and low-frequency F2 transitions) while the low frequency nature of the filtered stimuli sounded labial to those unaccustomed to speech sounding similar to the stimuli. This would also account for the difference among members of the hearing-impaired group. Basically, the closer the correspondence between filter and hearing, the better able the listener is to comprehend the sentence.

While a hearing loss cannot be described as a filter function in absolute terms, the correlation in this experiment between performance results, filter frequency and audiometric configuration seems to indicate a certain performance predictability. Thus, if audiometric similarity is narrowly defined in terms of break frequency and slope, listeners tend to perform similarly when the filter function resembles the audiometric configuration. They even tend to perform better than listeners with normal hearing.

If we assume that those listeners with a hearing loss corresponding to the filter were able to discriminate between presence vs. absence of /p/, why then did they nearly always pick /t/ where the actual stimulus contained either /t/ or /k/? Were the better results simply a matter of chance, there being more /t/-words than /k/ or /p/-words in the test material? In a computer analysis of word frequency in Swedish newspaper material (Allén, 1972) the frequency of /t/-initial words used in the test was greater than /p,k/ in all but one case (Table 2). In the same material, t was also the most frequent letter of the six representing stops (Table 3). On the basis of this material, it could be tentatively conjectured that certain

	TAL 200	KAL	BAL	DAL 15	GAL
PAR 427	TAR 1964	KAR	BAR		
	TUR 137	KUR	BUR	DUR	
	TÅR 10	KÅR 12	BÅR		GÅR
PÅG	TÅG 55		BÅG		
	TAM			DAM 137	GAM
	TOK 8	KOK	BOK 876		

Table 2. Frequency of target words in a study of newspaper material (Allén, 1972).

	ABSOLUTE	RELATIVE
t	456035	8.238
d	225548	4.084
k	170580	3.089
g	166999	3.024
p	86131	1.560
b	64902	1.175

Table 3. **Frequency of letters representing target-word initial stops used in the test. (From Allén, 1972)**

hearing-impaired listeners, when presented with semantically non-redundant speech, choose the most frequent or probable word from the lexicon which matches the incomplete phonetic signal. It is possible that listeners can build up a frequency or probability strategy during the long period of time often associated with the progression of a noise-induced hearing loss. All listeners but one in the "most-like-filter" group were over 58 years of age, had very similar audiograms and performance, and had long histories of working in noisy environments. The one exception, 39 years of age, had of course a similar audiogram but did not perform nearly as well as the older listeners on the test.

An additional point of interest was that one listener with an asymmetrical loss performed better on the filtered stimuli test with his worse ear. This was noticed by the listener himself who expressed considerable surprise over it. His worse ear audiogram, however, fit the filter function almost exactly. His better ear loss began at around 2200Hz.

Finally the fact that persons with impaired hearing tended to choose words beginning with /t/ could have certain clinical implications. A greater awareness of compensation strategies both on the part of those using the strategies and those who are often in contact with persons suffering from a hearing loss could be instrumental in improving the communication ability of a relatively large group of people.

Further work with perception, hearing loss, and compensation strategies could also provide us with interesting insights into speech perception as a whole. It might be possible that the compensation strategies used constantly by the hearing impaired are also available to and used to a lesser degree by normal listeners when engaged in everyday speech in non-optimal, noisy environments.

ACKNOWLEDGEMENTS

I would like to express my sincere thanks to Jarle Aursnes and the Department of Otorhinolaryngology, University Hospital, Lund, for making available both patients and equipment for the clinical part of the experiment. Thanks is also due to Bengt Mandersson for writing the pitch editing program, helping in constructing the stimuli and in interpreting the results of the experiment; and to Gösta Bruce and Eva Gårding for valuable discussion and comments.

REFERENCES

Allén, S. 1972. Tiotusen i topp, Almqvist & Wiksell, Stockholm

Cutler, A. 1976. Phoneme-monitoring reaction time as a function of preceding intonation contour. Perception & Psychophysics 20:1, 55-60

Fowler, C.A., Rubin, P., Remez, R., & Turvey, M. 1980. Implications for speech production of a general theory of action. In B. Butterworth (Ed.) Language Production, 373-420, London, Academic Press

Fry, D.B. 1958. Experiments in the perception of stress, Language and Speech 1:126-152

Gårding, E. 1967. Internal Juncture in Swedish, Lund, C.W.K. Gleerup

House, D. 1983. Perceptual Interaction between Fo Excursions And Spectral Cues, Working Papers 25:67-74, Department of Linguistics and Phonetics, Lund University

Jackendoff, R.S. 1972. Semantic Interpretation in Generative Grammar. Cambridge, Ma.: MIT Press

Johansson, G. 1975. Visual Motion Perception. Scientific American 232:6, 76-88

Johnson, T.L. & Strange, W. 1982. Perceptual constancy of vowels in rapid speech, Journal of the Acoustical Society of America 72: 1761-1770

Lehiste, I. 1970. Suprasegmentals. Cambridge, Ma.: MIT Press

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

Pastore, R.E. 1981. Possible psychoacoustic factors in speech perception. In Eimas & Miller (Eds.) Perspectives on the Study of Speech, Hillsdale, N.J. Erlbaum

Picket, Revoile, and Danaher 1983. Speech-Cue Measures of Impaired Hearing. In Tobias and Schubert (Eds.) Hearing Research and Theory v.2, Academic Press

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

Perceptual Experiments with Duration versus Spectrum in Swedish Vowels

Kurt Johansson

1. INTRODUCTION

It is well known that the vowels in Swedish word pairs like vit:vitt, kal:kall, rot:rott generally differ both with regard to duration and to spectrum, i.e. formant frequencies (perceptually: vowel length and vowel quality, respectively). There are also differences, particulary durational, between the final consonants, short consonants after long vowels and long after short. Consistent durational differences, between vowels as well as consonants, only appear in stressed positions, while some vowel quality differences may be upheld also in unstressed positions where all vowels are short. (1)

Different phonological interpretations are discussed for instance by Elert (1964, 39 ff. and 1970, 54 ff.) and Hadding-Koch - Abramson (1966). In this paper I will not enter this discussion.

Long and short vowels are manifested rather differently in various Swedish dialects. Long vowels may be pronounced as monophthongs, but they are generally more or less diphthongized (Elert 1981). Short vowels may also be diphthongized but ordinarily they are more monophthongal. (2)

The perceptual investigation carried out by Hadding-Koch -

Abramson (op.cit.) dealt, as does this paper, with the question of whether vowel duration or vowel spectrum is the primary cue for distinguishing minimal pairs of the above type. They prepared their stimuli from 3 Swedish word pairs by using a tape cutting and splicing technique. (3) The male speaker and the listeners came from Southern Sweden. For practical reasons only non-diphthongized vowels were included. The main results were that vowel duration was shown to be the primary cue for [ɛ:]:[ɛ] and [ø:] : [œ] and formant frequencies, i.e. quality, for [ʉ:] : [ɵ](4).

PURPOSE OF THE PRESENT INVESTIGATION

One of the reasons for carrying out the experiments presented below was that I wanted to cover all 9 vowel pairs of Central Swedish:
([i:]:[I],[e:]:[ɛ],[ɛ:]:[ɛ],[y:]:[Y],[ø:]:[œ],[ʉ:]:[ɵ],
[u:]:[U],[o:]:[ɔ], and [ɑ:]:[a]). (5)

My hypothesis was that vowel duration would be the main perceptual cue for some of the vowel pairs and vowel spectrum for others. What is actually tested, then, is which manner of transcribing, for example /vi:t/:/vit/ or /vit/:/vIt/, is nearest to the perceptual reality. (6)

The background for this hypothesis was of course, beside the findings of Hadding-Koch - Abramson, the observations of many investigators that the magnitude of vowel spectrum differences, and sometimes also of vowel duration differences, may vary for different vowel pairs. My own informal tests with electronically gated speech also contributed to the assumption.

Another reason for my experiments was in fact to give a background for spelling methodology in Swedish schools. The use of single or double consonants in Swedish spelling is largely dependent on the previous vowel, and there has over the years been much discussion on the issue of whether vowel length or vowel quality should be considered the appropriate methodological starting-point. One reason why I chose to work with Central Swedish was that dialects from Central Sweden have been normative for Swedish spelling.

In this paper, however, I do not intend to go into this discussion, as I have done this elsewhere (Johansson 1981) and my standpoint may be inferred from the results presented below.

3. PREPARATION OF THE TEST MATERIAL

In order to be able to control as many variables as possible I chose to work with synthetic speech. The stimuli were created with an OVE III speech synthesizer, controlled by an ALPHA LSI computer.

I preferred, as did Hadding-Koch - Abramson, to use non-diphthongized vowels. Diphthongized vowels will be dealt with elsewhere.

The test vowels appeared in 9 monosyllabic word pairs, always preceded by [h] and followed by [s] .(7)

The [h] was constructed as a voiceless counterpart to the adjoining vowel.

The [s] was characterized by two formant bands centered around 4 800 and 7 600 cps. It should be mentioned that the reason for choosing [s] was that it does not have any voiced counterpart in Swedish. Not much experimenting is needed to show that listeners tend to interpret, where this is possible, a consonant as voiced or unvoiced, not only in accordance with voicing or aspiration but also depending on the duration of a preceding vowel. In the experiments reported below this fact would without doubt have created an unnecessary complication.

In spite of the fact that the duration of a postvocal consonant is complementary in stressed words in most Swedish dialects, it has generally been assumed that durational differences between consonants are not distinctive, and this was also supported by the Hadding-Koch - Abramson experiments. I had planned to investigate this further, but an unfortunate location of some test stimuli made the results less reliable, and I have preferred to return to this question on another occasion. I am, however, apt to believe that differences in the durations of the consonants are of no consequence in the present investigation. Anyhow I choose a value for the [s] duration averaged between long and short according to Elert's measurements (1964, p 143), in this case 210 milliseconds.

The formant frequencies used were the ones reported by Fant for Central Swedish speakers (appendix 1). This was the only investigation available giving formant frequencies for both long and short vowels. In spite of the fact that the measurements had been made on isolated vowels the listeners did not react or comment on the resulting stimuli as being unnatural.

Steady state vowels, i.e. vowels without formant transitions, were chosen, which is of little perceptual consequence before [s] with its high intensity and strong intrinsic cues.

As fundamental frequency variations are known to influence the perception of length the frequency was held constant (at 100 cps).

In the test the vowel qualities were kept unchanged and the only thing that was varied was the duration of the vowels. The variations were made in 20 millisecond steps from 80 milliseconds to 200. (8)

In this way 126 different stimuli were created. On the test tape a buffer of 10 stimuli was recorded in order to let the listeners become acquainted with the test procedure and as these stimuli had rather normal durations they may also have served to give the listeners a tempo reference. The rest of the stimuli were randomized and each stimulus recorded exactly in the same way three times in a row, with one-second pauses. Between different stimuli there was a pause of 2.5 seconds. After every tenth stimulus a longer pause of about 5 seconds was made. After 50 and 100 stimuli there was a longer pause. Each stimulus only appeared once in the test.

The whole test lasted for about 20-25 minutes.

4. LISTENERS

In all, 50 listeners took part in the test, 42 women and 8 men with an average age of 40. 35 of the participants were school teachers specializing in speech, reading, and spelling therapy, 10 were students or teachers of speech pathology, and the others were specialists in other fields.

5. LISTENING TEST

The test stimuli were presented to the listeners from a Revox A 77 tape recorder via Burwen PMB 6 head-phones.

The listeners were given a test form having two choises for each stimulus, and they were asked to underline the word they thought had been presented (forced choise).

6. RESULTS

As was mentioned earlier the formant frequencies of the vowels were kept constant and only the durations changed, the purpose being to investigate whether vowel duration or vowel spectrum constitutes the major perceptual cue for distinguishing within the 9 word pairs.

If for instance a word his [hi:s] with a gradually shortened vowel eventually will be judged as hiss [hIs], in spite of the fact that vowel quality has not been changed, duration must be considered the essential cue. If, on the other hand, the identy of the word does not change, quality must be the important thing.

Figure 1, treating all listeners as one group and all vowels, reveals that on the whole duration must be considered the major cue. When the vowels are short, both vowels with qualities appropriate for long and for short vowels are judged as short. In the middle of the figure a change of identity takes place, and both types of vowels are judged as long when they have been lengthened enough.

If, however, we look at each vowel pair separately, it becomes clear that all pairs do not behave in the same manner. Most pairs change from short to long responses within the range of 120-180 milliseconds. (9) This is the case for [i:]:[I],[e:]:[ε],[ε:]:[ε],[y:]:[Y],[ø:]:[œ], and [u:] [U], i.e. 6 out of 9 pairs. The listeners tend to judge a vowel as short up to a duration of about 120 milliseconds and as long from about 160 (with 7 out of 12 vowels exactly at these values). These values may be compared with the average values reported by Elert for Stockholm Swedish (1964, p 109), 108 milliseconds for short vowels in his one-word list and 163 for long vowels. (The values in his sentence list are somewhat lower, but not very much.) The "short-point" above is 75% of the "long-point", a percentage that may be compared with what is generally reported for short vowels in relation to long (65-70%). (10)

The distance between the points where listeners judge the above vowels as short and long, respectively, can be described by positive values, i.e. a vowel judged as long is always longer than a vowel judged as short, and within these vowel pairs it does not matter that vowel formant frequencies are different. The remaining 3 pairs, [ʉ:]:[ɵ],[ɑ:] :[a], and [o:]:[ɔ], give negative values, i.e. there is considerable overlap so that a long vowel may be shorter than the "short" vowel, and the short vowel longer than the "long" vowel, without making the listeners judge them as short and long, respectively. Vowel quality is obviously the important thing here.

In order to get a better picture of the contribution of duration versus spectrum differences within the word pairs, the data will be presented in another way.

Figure 2 shows what could be expected under ideal conditions, if duration alone were the distinctive cue. Along the x-axis we have gradually increased the duration of the vowel, along the y-axis we have the percentage of "long"-responses both for "long" and "short" vowels. If duration were the only cue, the results should be as indicated in this figure. Both "long" and "short" vowels should be judged as short or long in accordance with the duration of the vowel, and there should be an uncertainty region in between.

The expected results, if vowel quality were the only distinctive cue, should be in conformity with figure 3. A "short" vowel should be judged as such and give 0% "long"-responses, whatever the duration of the vowel, and in the same way a "long" vowel should give 100%.

In figure 4 all listeners are treated as one group, without any consideration of dialect, and all vowels are included . It is quite clear, as it was from figure 1, that duration is the primary cue. The curves for "long" and "short" vowels are very much the same as the one presented in figure 2, but the curve for the "short" vowel is consistently below the other curve, which indicates that quality is not unimportant.

It can be seen from figures 5-13 that the vowel pairs may be divided into 3 different groups:

<u>The first group</u> consists of the 6 pairs commented on earlier, [i:]:[ɪ], [e:]:[ɛ], [ɛ:]:[ɛ], [y:]:[Y], [ø:]:[œ], and [u:]:[U] (figures 5-10). It might have been expected that [ø:] [œ] would have shown a greater dependence on the quality cue, but the results here are quite in agreement with the findings of Hadding-Koch - Abramson. Within the group the pair [e:]:[ɛ] displays the greatest distance between the curves, but there can be no doubt that duration is the essential cue here, too.

Only one vowel pair, [o:]:[ɔ], belongs to <u>group 2</u> (figure 11). The curves are here considerably further apart than for the above group, and although they look very much like the curves for group 1, I am, considering what I said earlier regarding this pair, apt to consider quality as the distinctive cue. But duration is certainly not unimportant.

<u>The third group</u>, consisting of [ʉ:]:[ᵾ] and [ɑ:]:[a], is of a different kind (figures 12-13). The curves are more of the type presented in figure 3, although they have the appearance of being turned on end as a result of the fact that duration even here is not completely unimportant. One reason for this behaviour of the curves might be that some of the listeners come from regions that do not have the usual quality differences between "long" and "short" vowels. In fact some of the dialect information given on the test forms indicates this, but the information is not detailed enough to allow a definite conclusion. However, the behaviour of the curves in figures 12-13, together with what was said about overlapping earlier, are to my mind indication enough that quality is the distinctive cue. Another thing that supports this conclusion is that the "long" vowels with their shortest durations, and the "short" vowels with their longest, do not change the responses further than to values indicating guessing.

7. DO LISTENERS FROM DIFFERENT DIALECTS BEHAVE DIFFERENTLY?

A question that has sometimes been discussed by Swedish phoneticians is if for instance people from Scania in the south of Sweden, with their tendency to diphthongize, particularly the "long" vowels, use these quality differences distinctively and not duration differences.

By coincidence there were among the listeners 15 Scanians and 15 people speaking various dialects of Central Swedish. The groups are of course too small to enable any definite answer, and as the listeners belong to different dialectal subgroups and are rather sophisticated they can not be regarded as representatives of genuine dialects.

If we, bearing this in mind, take a look at figure 14 we can see that the people from Central Sweden give more "short"-responses when the vowels are short and more "long"-responses when they are long. This is valid for "long" as well as "short" vowels. To put it another way, the Scanians are not quite as apt as the people from Central Sweden to judge according to duration.

The difference between the two groups is not very great, but together with what we know from other investigations, above all that Scanians tend to diphthongize more than most Swedes and that duration differences both between "long" and "short" vowels and consonants are less than for Central Swedish, we have indications that the role of quality variations should be investigated further. I am currently preparing some experiments that I hope will serve to elucidate this particular question.

8. SUMMARY

A listening test with synthetic stimuli, based on Central Swedish , was carried out using 50 listeners, the purpose being to investigate whether duration or formant frequency is the primary cue for distinguishing between [i:]:[I], [e:]:[ε], [ε:]:[ε], [y:]:[Y], [ø:]:[œ], [ɯ:]:[ɤ], [u:]:[U], [o:]:[ɔ] [ɑ:]:[a] respectively.

The following conclusions may be drawn:

a) If we look at the whole material, duration must be considered the major cue.

b) The vowel pairs may be divided into three groups, depending on which is the essential cue:

   1 The duration group containing [i:]:[I], [e:]:[ε], [ε:]:[ε], [y:]:[Y], [ø:]:[œ], and [u:]:[U].

   2 The quality group containing [ɯ:]:[ɤ] and [ɑ:]:[a].

   3 The intermediate group, consisting only of [o:]:[ɔ]. For this pair both duration and quality are of importance. I am, however, inclined to consider this group as a subgroup of the quality group.

c) A comparison of a group of Central Swedish speakers and Scanians indicates, together with other things, that dialectal differences should be investigated further, particularly dialects with strong diphthongization.

NOTES

1) I will in this paper only deal with vowels in stressed positions.

2) As for instance in the Malmö dialect (Bruce 1970).

3) The word pairs used were: väg 'road':vägg 'wall', stöta 'push': stötta 'prop up', and ful 'ugly': full 'full'.

4) The symbols taken from the Swedish dialectal alphabet. The corresponding IPA symbols ordinarily used are [ʉ] and [ɵ], respectively.

5) As mentioned earlier Hadding-Koch - Abramson only covered three out of nine pairs, and they used a speaker and listeners from Southern Sweden. Furthermore they seem to mean (p 106) that vowel duration is the essential cue for all vowel pairs exept one, viz. [ɯ:] : [ɤ] , a view that was not compatible with my own experiences.

6) In spite of the fact that I have used a colon as a marker of length, I do not, so far, wish to take sides concerning the question of whether length should be considered a separate phoneme or a distinctive feature of the vowel.

7) Some of the resulting monosyllables are nonsensical, but they are all possible Swedish words.

8) Elert (1964, p 109) reports 108 milliseconds for short vowels and 163 for long as average duration values in his one-word list. Before [s] which was used in my experiments, he reports exactly the same durations in his sentence list (p 115 and p 118).

9) I have made no statistical treatment of the data in this paper but instead used a rough estimation and considered the listeners convinced that a certain word has been presented, when the percentage falls within the upper or lower third.

10) See for instance Gårding _et_ _al_. (1974, p 108).

11) Occasional irregularities in the curves generally depend on preceding stimuli.

FREQUENCIES FOR THE FIRST THREE FORMANTS (after Fant 1957)

	F_1	F_2	F_3
[i:]	240	2050	3000
[I]	300	2050	2700
[e:]	340	2100	2500
[ε]	385	2000	2450
[ε:]	440	1800	2400
[ε]	385	2000	2450
[y:]	255	1950	2400
[Y]	300	2000	2400
[ø:]	350	1700	2200
[œ]	400	1550	2300
[ɯ:]	260	1600	2150
[[illegible]]	450	1050	2300
[u:]	280	700	2250
[U]	340	700	2600
[o:]	410	750	2450
[ɔ]	500	850	2550
[ɑ:]	650	1000	2500
[a]	750	1250	2500

REFERENCES

Bruce, G. 1970. Diphthongization in the Malmö dialect. Working papers 3. Phonetics Laboratory, Lund University.

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

Elert, C.-C. 1981. Ljud och ord i svenskan 2. Umeå: Almqvist & Wiksell.

Elert, C.-C. 1964. Phonologic studies of quantity in Swedish. Uppsala: Almqvist & Wiksell.

Fant, G. 1957. Den akustiska fonetikens grunder. Kungliga Tekniska Högskolan, Stockholm, Avd för Telegrafi-Telefoni, Rapport No 7.

Gårding, E., R. Bannert, A.-C. Bredvad-Jensen, G. Bruce and K. Nauclér. 1974. Talar skåningarna svenska? Svenskans beskrivning 8:107-117. Lund.

Hadding-Koch, K. and A.S. Abramson. 1966. Duration versus spectrum in Swedish vowels: some perceptual experiments. Studia Linguistica XVIII:94-107. Lund.

Johansson, K. 1981. Bör dubbelteckningsmetodiken bygga på längd- eller klangfärgsskillnader? Utvecklingsarbete och fältförsök, Rapport 2/81. Lärarhögskolan, Malmö.

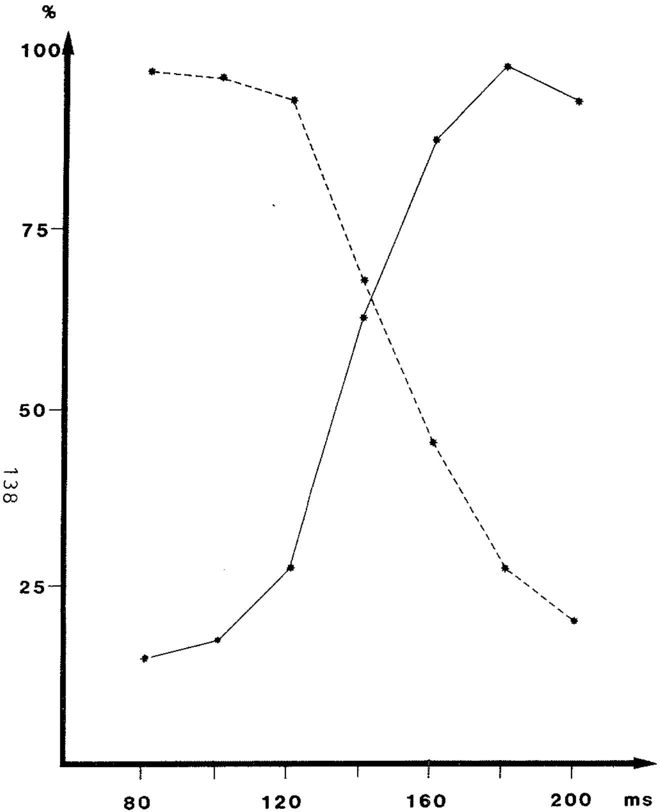


Figure 1 Percentage of correct responses for all "long" vowels (continuous line) and all "short" (dashed).

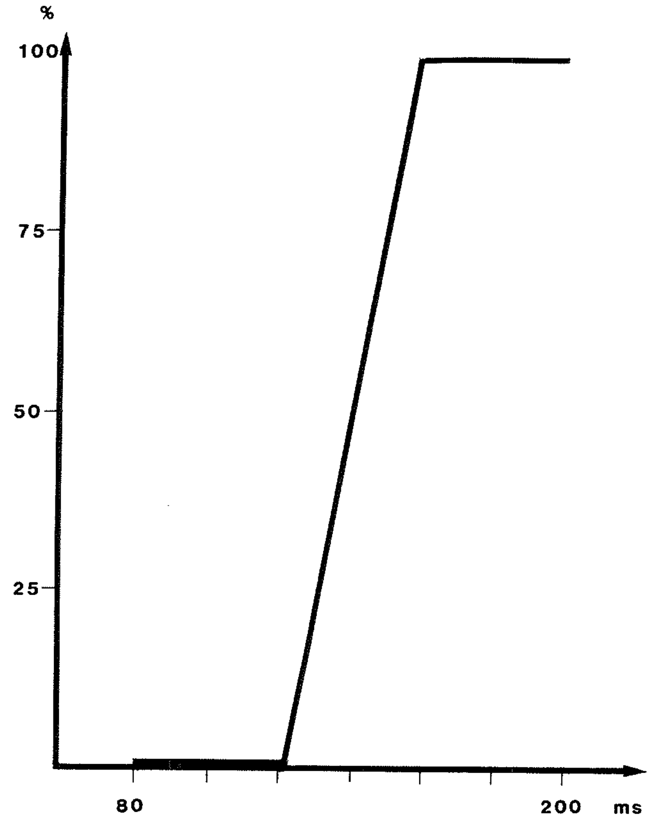


Figure 2 Expected results, if vowel duration alone were the distinctive cue. Duration along the x-axis, percentage of "long"-responses along the y-axis.

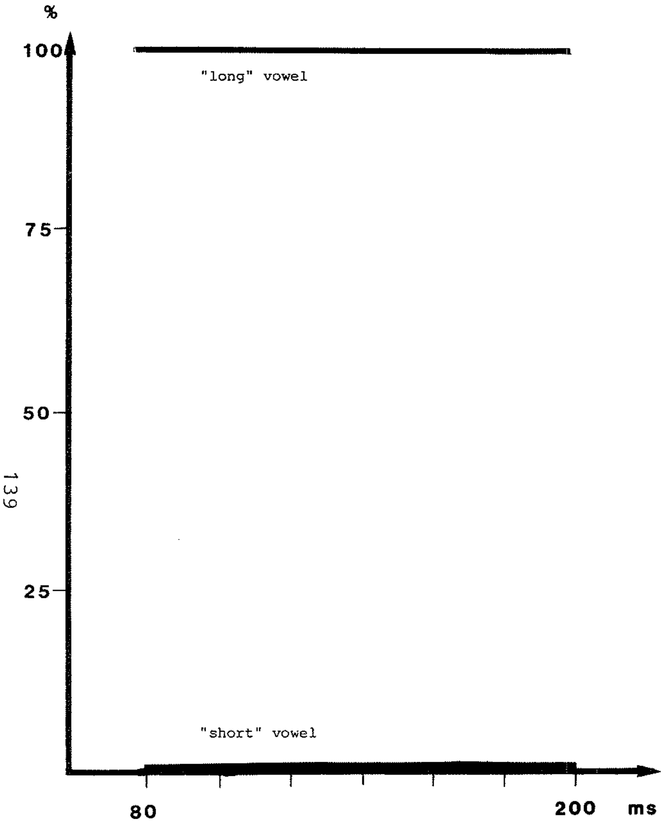


Figure 3 Expected results, if vowel quality alone were the distinctive cue.

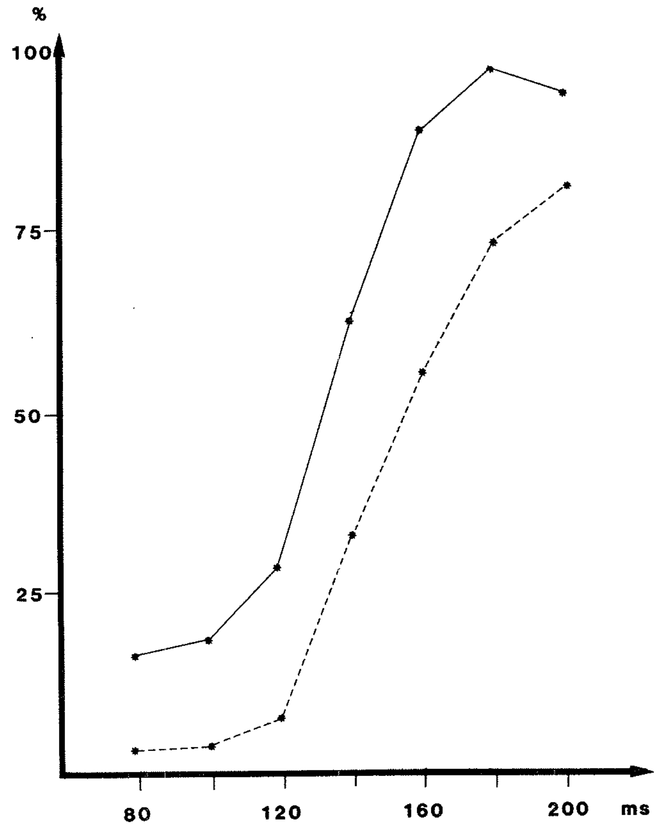


Figure 4 Percentage of "long"-responses to stimuli with gradually increased vowel duration. Continuous line = "long" vowels, dashed line = "short" vowels.

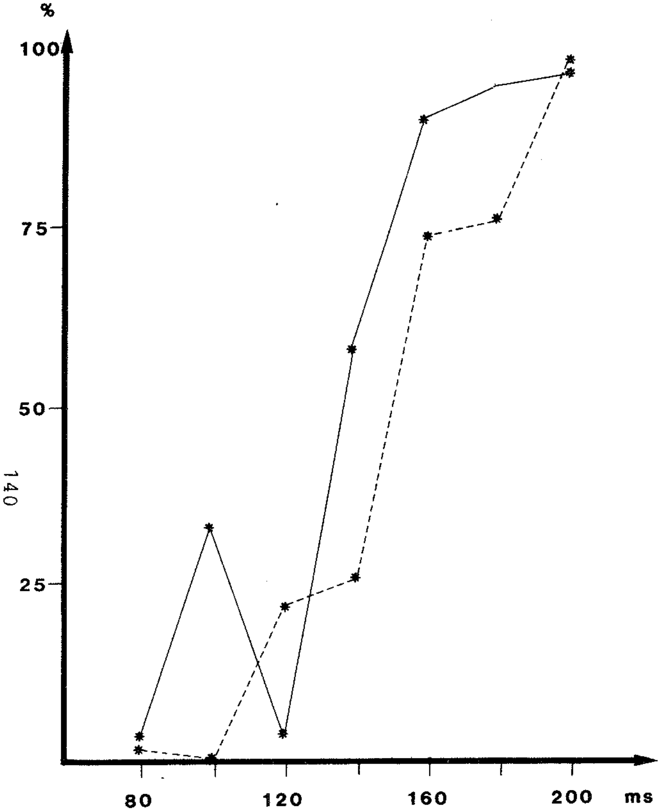


Figure 5 Percentage of "long"-responses for [i:] (continuous line) and [I] (dashed).

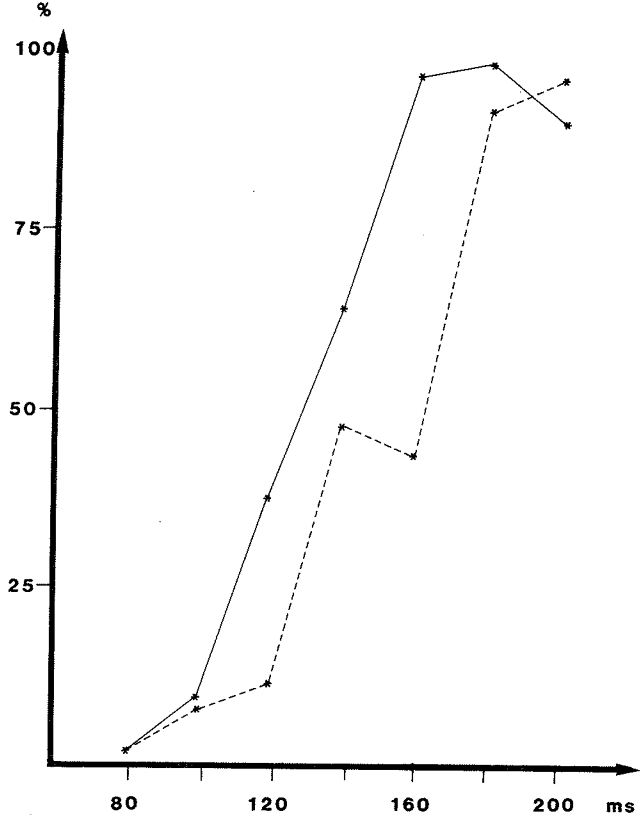


Figure 6 Percentage of "long"-responses for [e:] (continuous line) and [ε] (dashed).

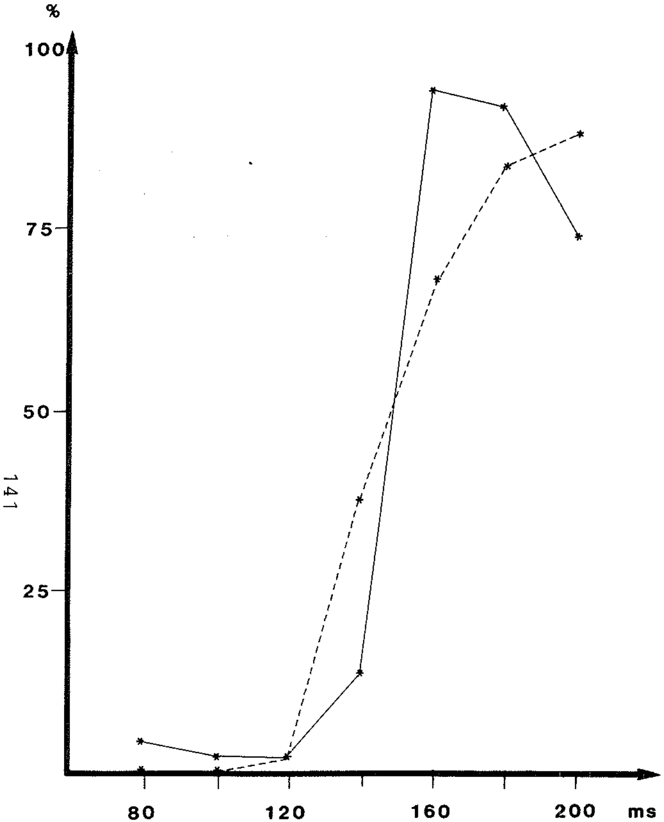


Figure 7 Percentage of "long"-responses for [ɛ:] (continuous line) and [ɛ] (dashed).

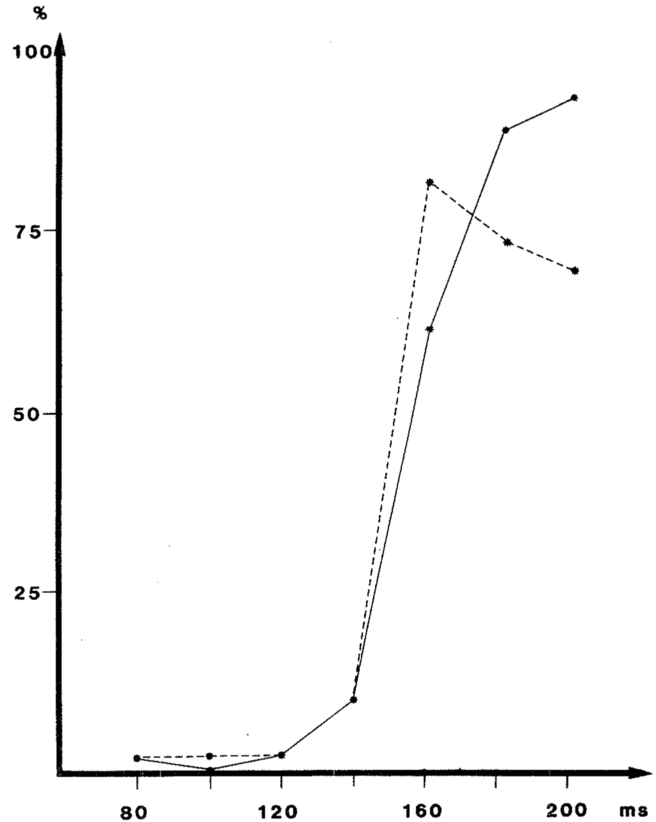


Figure 8 Percentage of "long"-responses for [y:] (continuous line) and [Y] (dashed).

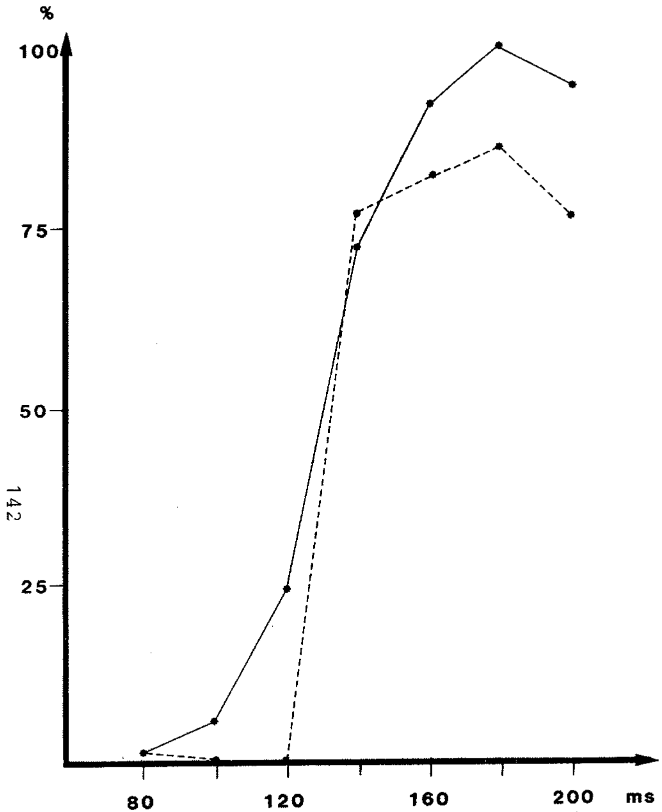


Figure 9 Percentage of "long"-responses for [ø:] (continuous line) and [œ] (dashed).

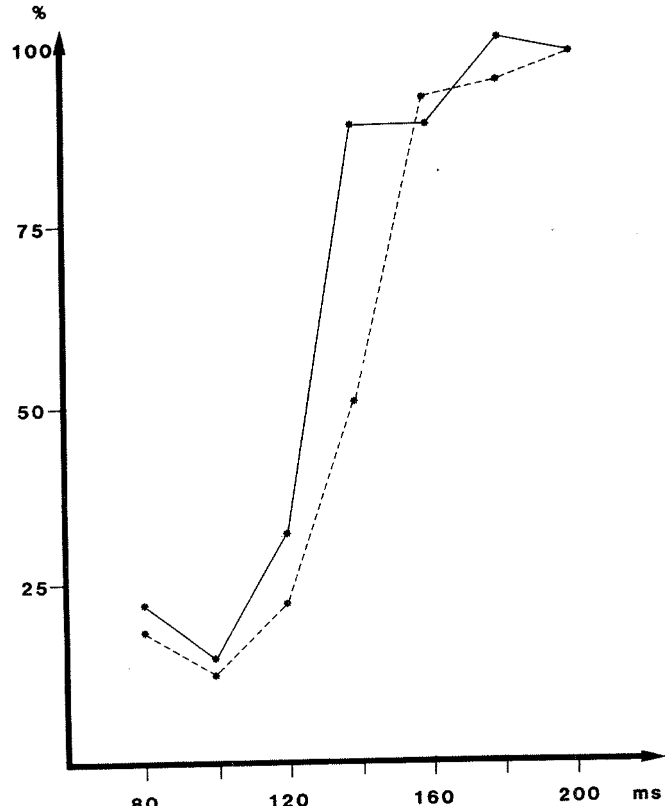


Figure 10 Percentage of "long"-responses for [u:] (continuous line) and [U] (dashed).

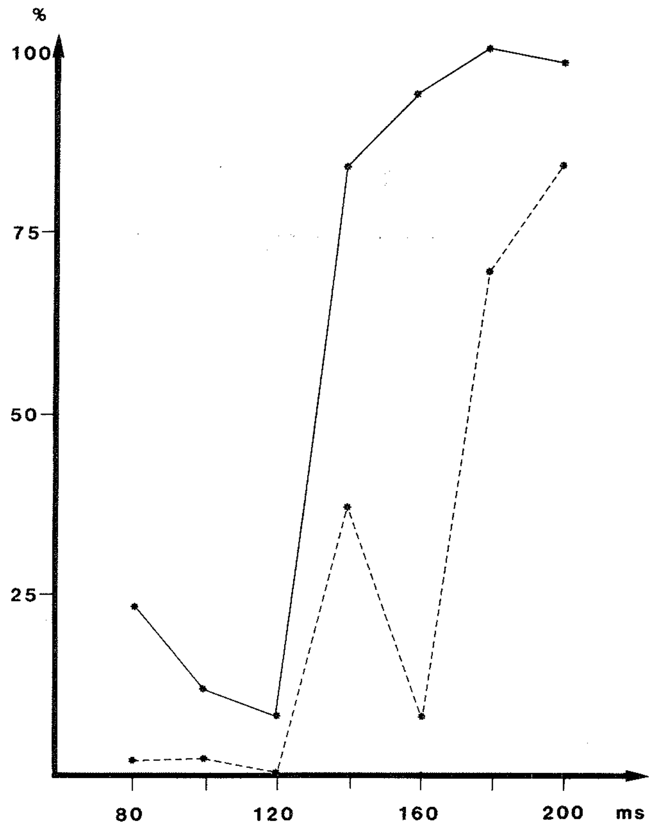


Figure 11 Percentage of "long"-responses for [o:] (continuous line) and [ɔ] (dashed).

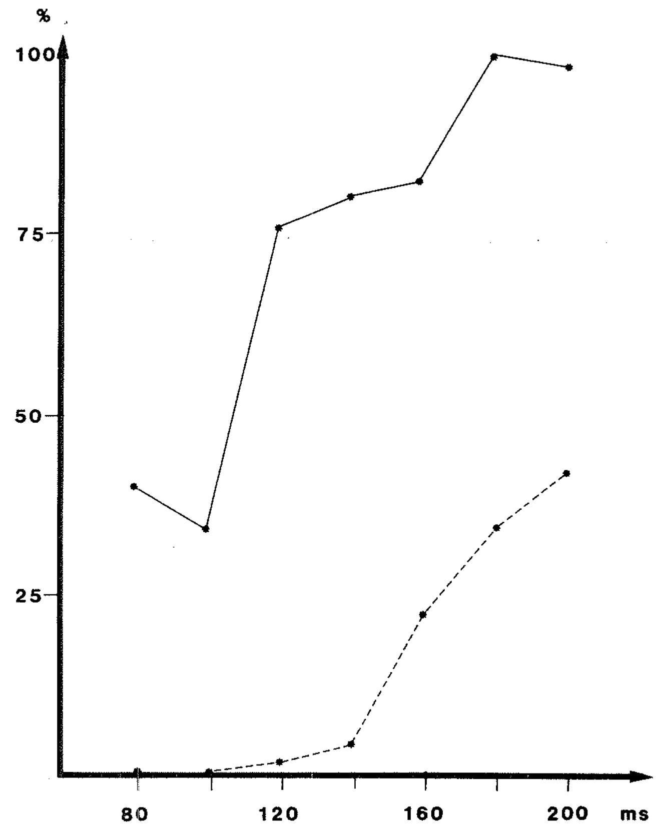


Figure 12 Percentage of "long"-responses for [ɯ:] (continuous line) and [ɯ̆] (dashed).

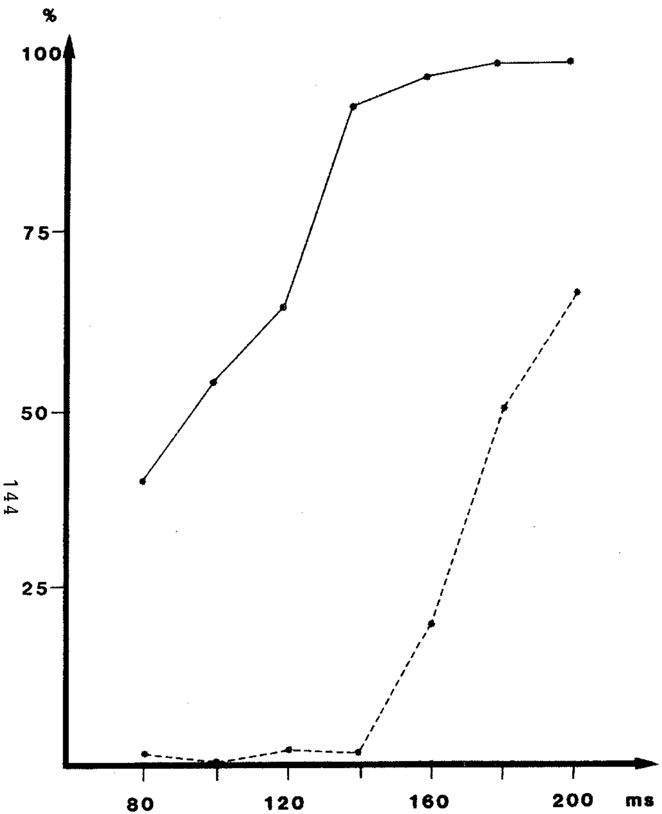


Figure 13 Percentage of "long"-responses for [ɑː] (continuous line) and [a] (dashed).

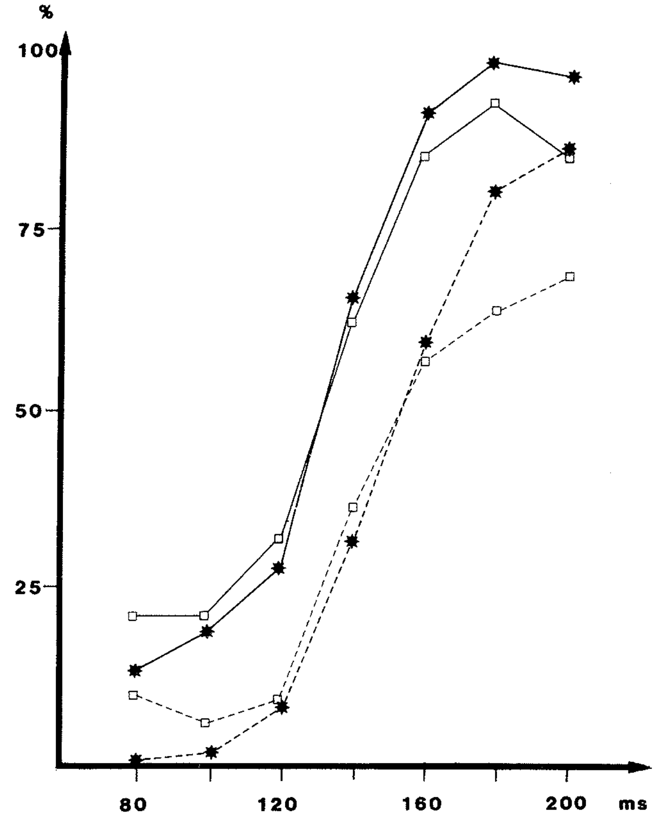


Figure 14 Percentage of "long"-responses for "long" vowels (continuous line) and "short" (dashed) for two groups of listeners, one from Central Sweden (stars) and one from Scania (squares).

Testing a Model for Hausa Intonation

Mona Lindau

This study reports a test of a formal model of the relationship between tone and intonation, using Hausa, a Chadic language spoken in Nigeria. This language has a relatively simple tone system of basically two tones, High and Low. The language also has a falling tone that can be analyzed as a combination of a High and a Low, and will not be considered further here (Meyers 1975, Cowan and Schuh 1976). The goal of this study is to test a model of intonation by matching actual fundamental frequency curves with curves generated by rule for Hausa statements and questions on two different tone patterns. The model is based on that developed at Lund in Sweden (Gårding and Bruce 1981, Gårding 1981). The results will also be useful in speech synthesis of Hausa utterances.

Figure 1 shows a model Hausa sentence. The input is a phonetic transcription with boundaries and tones marked. The model describes intonation as grids of baselines and toplines. In this description the grids are stored lexically as a basic slope for a particular number of syllables for each type of utterance, and are generated by rule. The boundary marks, and the Highs and Lows are inserted onto the grid lines. Local rules may modify the underlying patterns. The last step is to concatenate the maxima and minima into a smooth curve.

The model can be tested by the degree to which it will account for both tonal and intonational patterns used in common by different speakers. An analysis was made of Hausa utterances. The data consist of tape recordings from nine male speakers of Kano Hausa, recorded in Kano. The speech material was designed to illustrate intonation in statements and questions in appropriate contexts, and consisted of sentences on alternating High and Low tones, and on High tones only. Each tone pattern occurred in a short and a long version. There were two short basic sentences on High tones only, "Muudii yaa zoo gidaa" (Muudii came home), and "Muudii yaa ga zoomoo" (Muudii saw a hare), and one long sentence, "Uwargidan Muudii taa ga zoomoo bayan gidaanaa" (Muudii's senior wife saw a hare behind my house). The basic short sentence on strictly alternating High and Low tones was "Maalam yaa auni leemoo" (The teacher weighed the oranges), and the long sentence on this tone pattern was "Maalam yaa auni leemoo gaban garinmuu" (The teacher weighed the oranges in front of our town). In addition there was one sentence of Highs and Lows with two adjacent High tones, "Maalam yaa raba naamaa", HLHHLHL (The teacher divided the meat). Basic sentences on Low tones only were excluded by the fact that Hausa does not have any low tone verbs. The sentences were organized in blocks of one theme. Each block began with an introductory statement about a given topic; there then followed various question versions of the statement, each followed by a proper reply. This study examines a subset of the data, consisting of the introductory statements, the yes/no questions, and a question beginning with the question-word "waa...?" (who?).

The sentences were read by the speakers for the recording. The style of speech is fairly formal, but a non-colloquial style was intended, so as to maximize contrasts at this exploratory stage of the investigation. The speakers did not always produce the intended prosodic patterns on each utterance, so the results may reflect less than nine speakers. The number of speakers analyzed is specified in each case in describing the results.

Particularly the long sentences were affected in that some speakers seemed to prefer breaking them up into smaller phrases with pauses. Only the long sentences without such pauses were included in the analysis.

The recorded material was analyzed from narrow band spectrograms made on a digital spectrograph. Wide band spectrograms were used to aid in segmentation and for durational measurements. The fundamental frequency curves were traced from the 10th harmonic. One advantage of tracing pitch on spectrograms is that a pitch curve is visible even when the intensity drops at the end of an utterance, or at creaky phonation. Several of the speakers had quite creaky phonation, particularly at the end of utterances.

In order to test the model, measurements for all speakers were made of the fundamental frequency at the beginning points (BEG), the end points (END), and at all turning points (H,L) in statements and questions (figure 1). The model predicts that it should be possible to specify tones in terms of an intonational grid. Accordingly, straight lines were fitted by eye through the High and Low turning points to form a grid. Each line was drawn so as to fit as many maxima and minima as possible in the utterance. Occasional turning points may fall outside the grid, but they can usually be accounted for by local rules. For example, in utterances with alternating Highs and Lows the last High is often lowered, and in the long utterances on alternating Highs and Lows the penultimate High is often raised. In yes/no questions the last High is always considerably raised and is considered to be outside the grid. The rates of slope of these grid lines were measured and calculated in % per second. In addition the total duration for each statement and corresponding yes/no question was measured.

RESULTS AND DISCUSSION

Consider the statements first. Figure 2 shows one speaker's pitch curves of three statements with different tone patterns and lengths. In the utterances with alternating Highs and Lows the two grid lines are be in a fairly constant relationship to each other throughout the utterances, so it is possible to derive one line from the other. As there are no sentences on Low tones only, the topline is taken as basic, and the bottom line as derived.

	short	long
H only	13.89 (4.48) n=18 v=32.3 %	7.86 (2.91) n=7 v=37 %
HL	32.71 (13.28) n=7 v=40.6 %	16.29 (2.98) n=7 v=18.3 %
HLHHLHL	28.75 (6.69) n=8 v=23.3 %	

Table I

Mean slopes in % per second with standard deviations.
v = coefficient of variation (SD/x).

The statement intonation describes a downward slope, even in utterances on High tones only. The parameters that need to be set in order to generate pitch curves for statements include:

1. the slope of the topline for short and long sentences.

2. the frequency of the first High.

Table I shows the mean % per second rate of slope with standard deviations of the top line slopes. The different means show that the slope is affected by both sentence length and tone patterns. The longer sentences have a slower rate of descent of the topline, and paired t-tests show that the differences between short and long sentences are significant ($p<0.001$). As the data only include two sentence lengths for each tone pattern, the precise relationship between sentence length and topline slope cannot be determined here. Very approximately for these data, doubling the number of syllables will halve the rate of the slope. Sentence length will thus have to be part of the specification of the topline. The rate of slope also varies with different tone patterns. The utterances with alternating Highs and Lows have about twice as steep a slope as those with Highs only in utterances of comparable length. Paired t-tests show that the differences in slope between utterances with different tone patterns are significant ($p<0.001$). The slope of the utterance on the HLHHLHL pattern with two adjacent Highs is also significantly different from that of the strictly alternating sentence ($p<0.05$). The downward slope in statements with alternating Highs and Lows is wellknown as Downdrift, a phenomenon where Highs and Lows are lower than in other positions. Downdrift has sometimes been regarded as an intonational phenomenon (e.g. Hombert 1974, Ladefoged 1975, Meyers 1976), sometimes as an effect of local tone assimilations of Highs to preceding Lows (Abraham 1941, Hyman and Schuh 1974). Both views are partly correct. The downward slope in the sentences on High tones only can of course not be an effect of local tone assimilations, but must be considered to be a global statement intonation. The much steeper slope in the sentences with alternating High and Low tones is here interpreted as an effect of combining the global statement intonation from Highs only with a fairly slow rate of slope with local tone assimilation rules that lower Highs following Lows substantially. This view is supported by the fact that in the sentence with two adjacent Highs in the middle ("Maalam yaa raba naamaa"), where only two Highs follow Lows, the slope is less than in the strictly alternating "Maalam yaa auni leemoo", where three Highs follow Lows. The more Low-High sequences there are, the steeper the slope becomes. In this view then, Hausa has both a global statement intonation with the slope of that on High tones only, and local tone assimilations that will change the slope depending on the number of Low-High sequences in the derivation of the final slope. The final slope can be described as an effect of both an almost universal downward slope in statement intonations, and more language specific modification of that declination.

The topline thus varies with sentence length and tone patterns. It does not seem to be predictable from any other parameter. It is possible that the slope could depend on the frequency of some point at the beginning or end of the utterance. But in these data there is no significant correlation between the slope, on the one hand, and the beginning or endpoints, or the initial or final Highs, on the other. The topline slope in statement intonation is a parameter that functions like a primitive. Barring a more precise knowledge of the relationship between sentence length and slope, it has to be set for each length in generating pitch curves.

The topline has to be anchored at some point, presumably at the beginning or end of a statement. Paired t-tests were used to determine which point differed least between the various types of statements. The first High was the only point that showed nonsignificant differences between most sentences of different tone patterns and different lengths. In the relatively few cases where the first High did differ between sentences, the difference was in the direction of this point being higher for longer sentences than for shorter sentences. The first High is also the point with the smallest coefficient of variation for different types of sentences within speakers. The first High will thus be used as a further set value to anchor the topline.

This is a different situation than in Swedish and in English, where the first High is higher the longer the utterances are (Bruce 1982, Pierrehumbert 1980). The end point tends to be invariant, as well, so the slope in these languages is in a sense a function of these two points. In Hausa, however, the first High, as well as the starting point (BEG), are points that, within a speaker, show very little variation. It remains an open question how typical the differing results for Hausa are in this respect. It is possible that the differences in the behaviour of the first High could be an effect of different type of speakers rather than a language difference. Many studies of intonation have used speakers that are highly trained in the task of reading in front of a microphone, while the speakers used in this study were not used to this.

Within each type of statement there is a great deal of variation between speakers in the topline slope, as evidenced by the fairly large standard deviations in table I. Although a speaker will vary the topline slope with sentence length and tone patterns, the precise amount of slope appears to be idiosyncratic. The variation between speakers is not correlated with the speakers' mean fundamental frequency or range. In generating the grid for a statement, the mean slope will be taken to represent a typical speaker.

In the sentence with alternating Highs and Lows speakers also vary to some extent in their range. This variation probably reflects a non-linguistic factor of the attitude of the speaker. An involved speaker uses a larger range than a detached speaker (Hadding-Koch 1961, Bruce 1982). The range can be expressed as the mean ratio between Highs and following Lows throughout the utterance. Speakers tend to keep this ratio fairly constant throughout an utterance. The mean ratio between Highs and following Lows for the short "Maalam yaa auni leemoo." is 1.25 (SD=0.05, n=7), and for the longer "Maalam yaa auni leemoo gaban garinmuu." this mean ratio is 1.2 (SD=0.06, n=5). The mean ratio will be used in deriving the bottom line from the topline in the grid.

The timing of turning points in the fundamental frequency curve in relation to the segmental structure of the utterance is also a necessary part of the specification. In sentences with alternating Highs and Lows there is a strong tendency for the turning points to occur right at the syllable boundaries. In sequences with open syllables, like CVV-CVV, and CV-CVV the turning point coincides with the V-C boundary, but in sequences with closed syllables, CVC-CVV, the turning point occurs in the middle of the double consonant. The fundamental frequency curve associated with a High tone syllable will thus describe a movement towards a maximum during the syllable, and reach its maximum at the border to the next syllable. Similarly, the fundamental frequency curve associated with a Low tone syllable will reach its minimum at the very end of the syllable. The turning points in the fundamental frequency curve thus serve as a phonetic correlate to syllable boundaries. This pattern in the timing of the turning points is not much influenced by position in the sentence, but it is relatively stable throughout an utterance.

Figure 3 shows how a pitch curve for the High/Low sentence "Maalam yaa auni leemoo" can be generated.

<u>First</u>, set the duration for the given number of syllables (here, the seven syllables take an average of 130 centiseconds). Set the value for the first High (the average speaker's value is 160 Herz). Construct the intonational grid, anchored on an arbitrary (speaker-specific) value for the first High. The basic topline slope for a statement of this length is taken to be the slope for High tones only, about 14% per second.

Construct the bottom line as a ratio (here 1.25) of the topline. Secondly, insert start and end points. As a convention, every statement begins and ends with the pitch for a Low, even if the first (or last) tone is a High to (or from) which the pitch must immediately rise (or fall). Thirdly, insert High and Low tones on the grid lines. The Highs and Lows will constitute turning points on the grid lines. Align the sentence so that the Highs and Lows occur at the syllable boundaries. The fourth step involves applying the tone assimilation rules and this will result in a steeper slope, the more Low-High sequences the sentence has. For this sentence the rate increases from about 14% per second to 33% per second. At this stage optional rules may apply that move specific tones away from the grid lines. These are processes that are found for some speakers. They include lowering the final High tone when it occurs after a Low, raising a Low after the first High, and raising a penultimate High in longer sentences. Lastly, concatenate all the turning points into a smooth curve. For these simple statements the resulting pitch curve approximates the curves in the data well.

Now the intonation in questions can be considered. In Hausa, yes/no questions are signalled by intonational means, not by syntactic, or morphological means. Cowan and Schuh (1976), and Miller and Tench (1980) characterize this type of question intonation as similar to intonation in statements but with less of a slope, and a local rise of the final High. Newman and Newman (1981) posit a separate question morpheme at the end of all questions. This morpheme consists of a Low tone with length. Short final vowels lengthen in questions. The low tone is justified by a claim that in questions the final raised High is followed by a fall.

Typical fundamental frequency curves of yes/no questions are illustrated in figure 4. The salient features of yes/no questions are a suspension of the statement downward slope to zero slope, and a considerably raised pitch of the last High tone. Yes/no questions on High tones never exhibit any downward slope, and in the questions with alternating High and Low tones only two out of nine speakers follow the pattern described in the references with somewhat of a downward slope of the grid before the final raised High. The rest have zero slope. The width of the grid is not significantly different from that in statements. The frequency of the first High tone does not differ significantly between statements and corresponding yes/no questions, so Hausa questions are not characterized by a raised register, as questions may be in other languages, for example in Swedish (Hadding-Koch 1961, Gårding 1979, Bredvad-Jensen 1983).

The raised last High is sometimes, but not always followed by a fall. Figure 5 shows the fundamental frequency curves of the last two syllables of all the speakers superimposed for each yes/no question. Unbroken lines indicate a final rise that is not followed by a fall, and dashed lines indicate a rise that is followed by a fall. Many speakers do not end a yes/no questions with a fall. In fact, even when the last syllable is on a Low tone, as in the utterances ending with "...naamaa?", three of the nine speakers do not end in a fall. A final Low tone question morpheme is not part of a general pattern in the yes/no questions. The final falls that do occur can be viewed as instances of the general convention of starting and ending low, rather than as a manifestation of an abstract low tone question morpheme.

In addition, statements and corresponding yes/no questions differ significantly in total duration, The questions are about 10% shorter than the corresponding statements. The difference is due to an overall shortening of the question utterance as compared to the statement, not to shortening of any particular part of the utterance. This shortening is not found in all languages, for example not in German (Bannert 1983). If it is language specific, this type of overall shortening will have to be part of the rules for question formation in Hausa.

Figure 6 shows how a yes/no question "Maalam yaa auni leemoo?" can be generated. First, set the duration of the question at 10% less than the corresponding statement (here = 117 csec.). Construct the grid with zero slope, anchored on the first High as for statements. Secondly, insert a low starting point, and optionally a low endpoint. Thirdly, insert the tones, and fourthly, raise the last High tone. The amount of increase of the last High varies considerably, but is typically around 5-10%. Lastly, concatenate the turning points into a smooth curve.

Typical pitch patterns of question word questions are illustrated in figure 7. They display characteristics that place them as something between the pitch patterns in statements and yes/no questions. A question-word question of a sentence with alternating High and Low tones will have a downward slope, but it is significantly less than that in statements. The mean slope of "Waa ya auni leemoo?" is 23% per second (SD=3.1, n=8). But "waa"-questions of the basically High tone sentences have a zero slope, just like the yes/no questions. Also like the yes/no questions the last High tone may be raised, but only for two out of the nine speakers. Mostly the last High tone is not raised. There is however always a fall after the final High tone. This fall is most simply accounted for by assuming it to be part of the general convention of starting and ending Low.

SUMMARY

The intonation of simple sentences in Hausa can be represented as grids of (near)parallel lines. The rate of slope of the grid is at least related to sentence type (statement and question), sentence length, and tone pattern. Pitch curves are generated by rules from underlying grids. The base form of a grid is a topline from which the bottom line is derived as a ratio. The topline is anchored on the first High tone. The slope of the base intonational grid is that which appears from High tones only, when there is no influence from tonal assimilations. All questions are specified with zero slope, and statements with a slope that depends on the length of the sentence. In this study only two slopes appear as base toplines, one for short, and one for long sentences, but obviously a function can be worked out from more data on different sentence lengths. Lexical tones and boundary marks are mapped onto the grid lines. In sentences with alternating High and Low tones assimilatory processes apply that lower Highs following Lows, so that the slope of the grid becomes steeper, thus claiming that downdrift is an effect of both intonation and tonal assimilations. These assimilations apply in statements and question-word questions, but not usually in yes/no questions. Local tone rules may move tones away from the grid lines. Lastly, a smooth curve is generated through the boundaries and tones. At least for the simple sentences dealt with above, the resulting pitch curves approximate the data well.

Acknowledgments: Many thanks to Will Leben, Nicolas Faraclas, and Brian McHugh for recording the speech material in Kano. I am much indebted to Peter Ladefoged, Ian Maddieson, and the UCLA Phonetics Lab for comments on earlier versions. Eva Gårding provided much inspiration. Financial support came from the National Science Foundation and the Swedish Council for Research in the Humanities and Social Sciences.

REFERENCES

Abraham, R. 1941. Modern Grammar of Spoken Hausa. Kaduna: Government Printer.

Bannert, R. 1983. Some phonetic characteristics of a model for German prosody. Working Papers. 25:1-34. Department of Linguistics, Lund University.

Bredvad-Jensen, A.-C. 1983. Perception studies of Swedish interrogative intonation. Abstracts of the Tenth International Congress of Phonetic Sciences. 578.

Bruce, G. 1982. Experiments with the Swedish intonational model. In H. Fujisaki and E. Gårding (eds.) Preprints of Papers for the Working Group on Intonation. The XIIth International Congress of Linguists,Tokyo 1982.

Cowan, R. and R.Schuh. 1976. Spoken Hausa. Ithaca: Language Services, Inc.

Gårding, E. 1979. Sentence intonation in Swedish. Phonetica 36.207-215.

Gårding, E. 1981. Contrastive prosody: a model and its application. Studia Linguistica 35:146-165.

Gårding, E. and G. Bruce. 1981. A presentation of the Lund model for Swedish intonation. In T.Fretheim (ed.) Nordic Prosody II: 33-39. Trondheim: Tapir.

Hadding-Koch, K. 1961. Acoustico-phonetic studies in the intonation of Southern Swedish. Lund: Gleerups.

Hombert, J.-M. 1974. Universals of downdrift: their phonetic basis and significance for a theory of tone. In W. Leben (ed.) Studies in African Linguistics Supplement 5:169-183.

Hyman, L. and R.Schuh. 1974. Universals of tone rules: evidence from West Africa. _Linguistic Inquiry_ 5.81:115.

Ladefoged, P. 1975. _A Course in Phonetics_. New York: Harcourt, Brace, Jovanovich.

Meyers, L. 1976. Aspects of Hausa tone. _UCLA Working Papers in Phonetics_ 32.

Miller, J. and P.Tench.1980. Aspects of Hausa intonation: utterances in isolation. _Journal of the International Phonetic Association_ 10,1-2: 45-63.

Newman, P. and R. Ma Newman. 1981. The question morpheme q in Hausa. _Afrika und Ubersee_ 64,1:35-46.

Pierrehumbert, J. 1980. _The Phonology and Phonetics of English intonation._ PhD thesis, MIT.

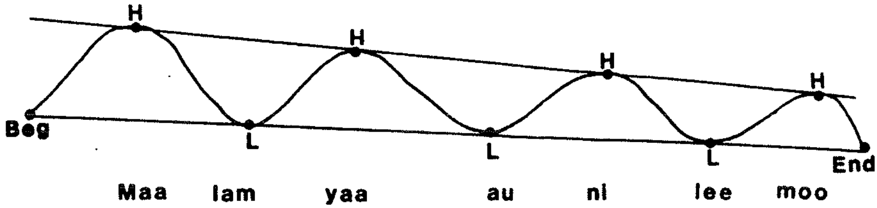


Figure 1
A model Hausa utterance.

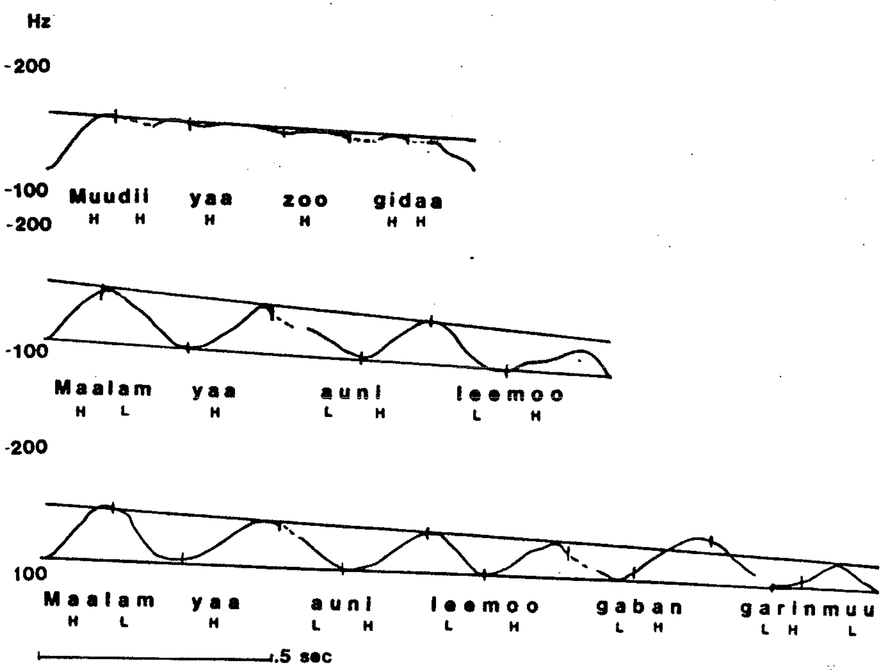


Figure 2
Pitch curves of statements on different tone patterns and different lengths of one speaker.

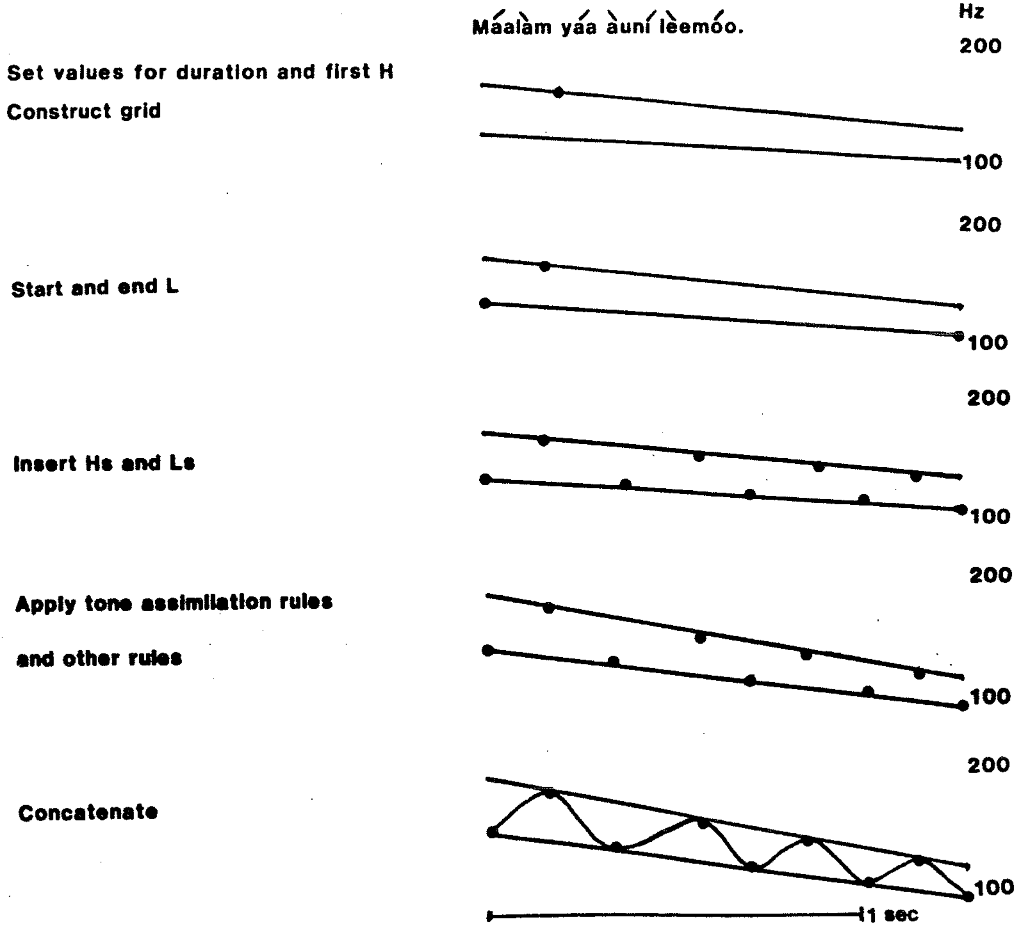


Figure 3

Rules for generating the pitch curve associated with a sentence on alternating High and Low tones.

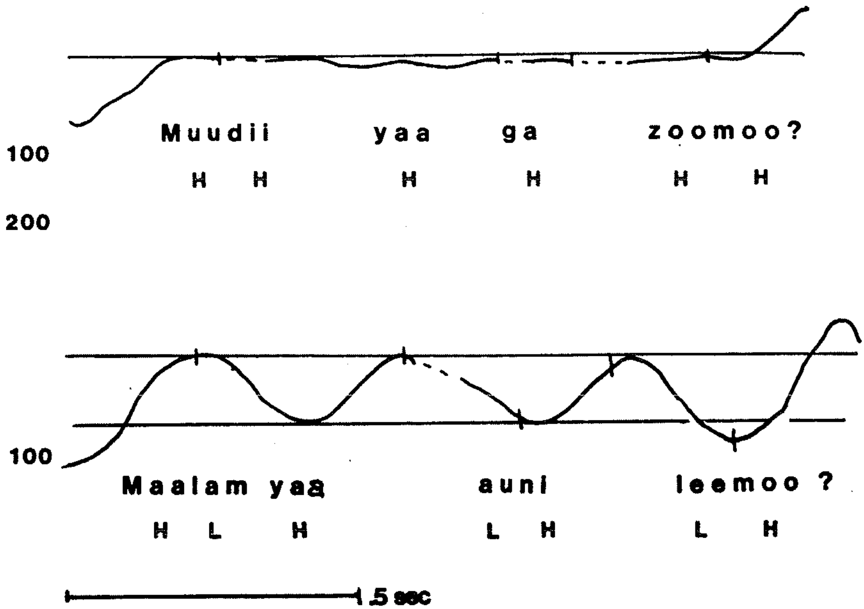


Figure 4
Pitch curves of yes/no questions on different tone patterns of one speaker.

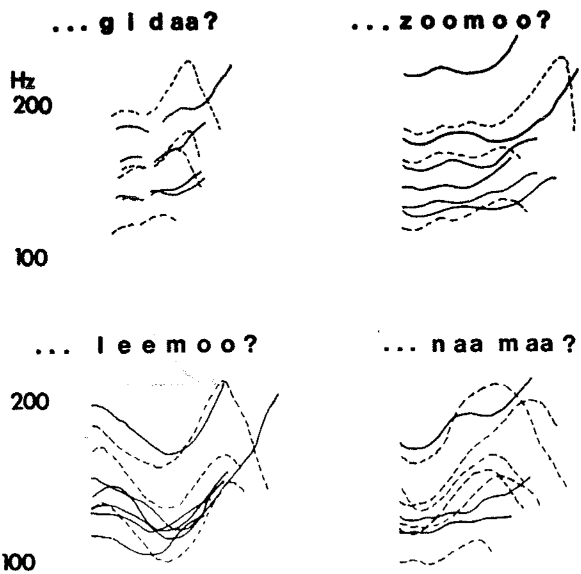


Figure 5
Pitch curves of the last two syllables in four questions for 8 or nine speakers.

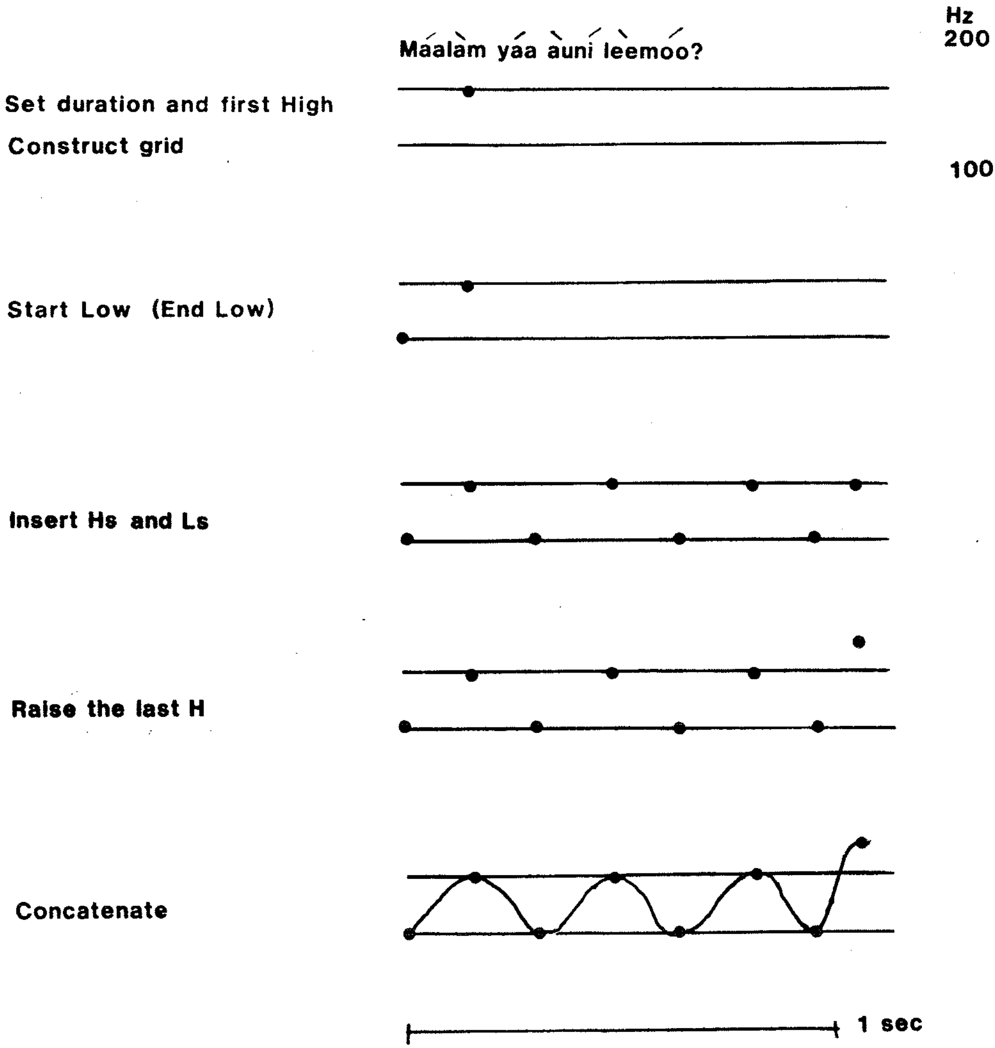


Figure 6

Rules for generating the pitch curve associated with a yes-no question on alternating High and Low tones.

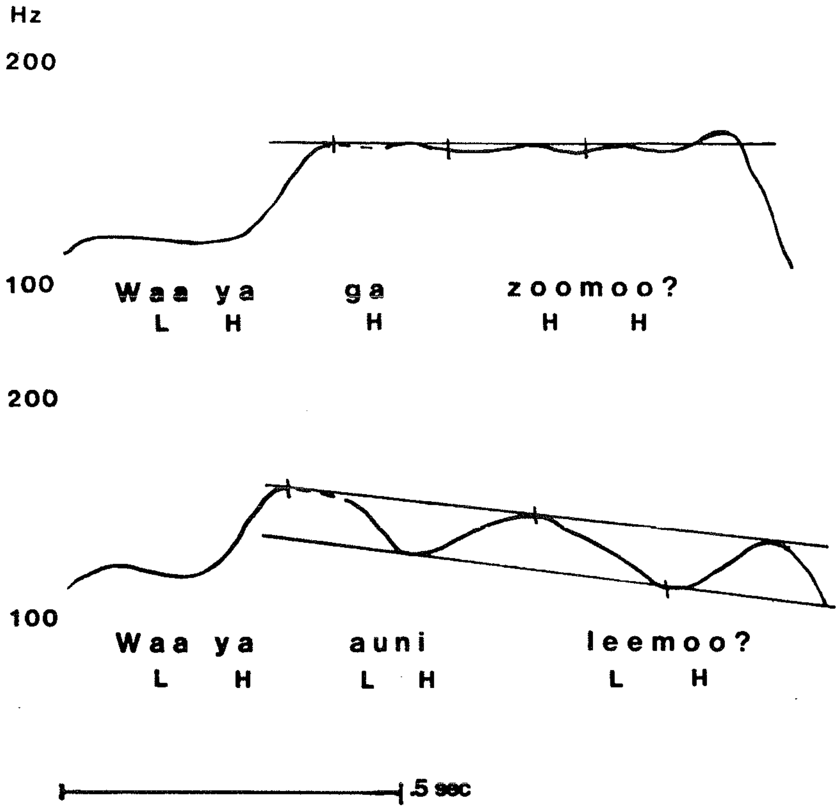


Figure 7
Pitch curves of question-word questions on different tone patterns of one speaker.

Form or Substance? The Linguistic Awareness of Pre-school Children and School Children Investigated by Means of a Rhyming Test

Eva Magnusson, Kerstin Nauclér and Ewa Söderpalm

Revised version of a paper presented at the Xth International Congress of Phonetic Sciences, Utrecht, the Netherlands, 1 - 6 August, 1983

The development of language in children does not only include the production and perception of language but does also entail a certain amount of linguistic awareness. Linguistic awareness is the term used by Mattingly (1972) to describe the speaker's/listener's ability to focus on the linguistic expression rather than on the linguistic content. On the phonological level linguistic awareness can be defined as the ability to disregard the meaning of words and to concentrate on their sound structure, e.g. the ability to realize that "train" is a short word although it refers to a long object and to be able to segment the sound sequence /tre:n/ into four segments although a train may consist of more than four parts.

Several questions concerning the development of linguistic awareness merit a further investigation, i.e.

- -the relation between linguistic awareness and language development
- -the relation between linguistic awareness and cognitive development

-the relation between linguistic awareness and reading and writing acquisition.

There are different opinions about the relationship between the development of linguistic awareness and language development on one hand and cognitive development on the other. Some researchers (e.g. Mattingly 1972, Marshall and Morton 1978) regard language development as a prerequisite for linguistic awareness while others (e.g. Hakes 1980) regard cognitive developmental level as more important for the development of linguistic awareness. Intermediate position is taken by researchers who assign importance to both linguistic and cognitive factors (e.g. Tornéus 1983). It has been shown by e.g. Bruce (1964), Calfee et al. (1973), that younger children do not show the same degree of linguistic awareness as older children do. Liberman et al. (1977) have reported that very few five-year-old children (17%) can indicate the correct number of sound segments in a word while most of the six-year-olds (70%) manage such a task. Syllables, on the other hand, are easily mastered by five-year-old children.

Since linguistic awareness develops gradually it is, however, hard to decide whether the increased awareness in normally developing children is a consequence of the cognitive or of the linguistic development. One way of clarifying this issue is to study the linguistic awareness of children whose language development is not as advanced as that of their peers. If two children of the same age who differ in language development show the same degree of linguistic awareness, it can be assumed that the linguistic awareness is a consequence of the cognitive development. If on the other hand the linguistic awareness of the two children differs, it can be concluded that language development is more important than cognitive level. A suitable test design would thus be to compare the linguistic awareness of groups of children differing in language development but matched for age/cognitive level, that is to compare groups of language disordered children with groups of normally developing children.

The relation between linguistic awareness and reading-spelling acquisition is also subject to different opinions. Some reserchers like Valtin (in press) and Ehri (1979) consider linguistic awareness as the result of reading acquisition whereas others like Calfee et al. (1972), Liberman (1973), and Tornéus (1983) regard linguistic awareness as the prerequisite for reading aquisition.

The most relevant aspect of linguistic awareness in relation to reading and spelling is the awareness of phonological structure. The ability to segment within the syllable is of particular importance. It has been shown in a longitudinal study conducted by Lundberg and his co-workers in Sweden (1980) that linguistic awareness as measured by segmentation tasks is the best predictor of pre-school children's future reading and spelling success.

One such task which requires segmentation ability is rhyming. The manipulations performed in rhyming are to separate prevocalic element(s) of the stressed syllable from the rest of the syllable and to use what is left of the syllable or the word as a model when producing new rhymes. A rhyming test would thus be an appropriate instrument to study linguistic awareness as regards its relation to language and cognitive developmental level as well as to literacy.

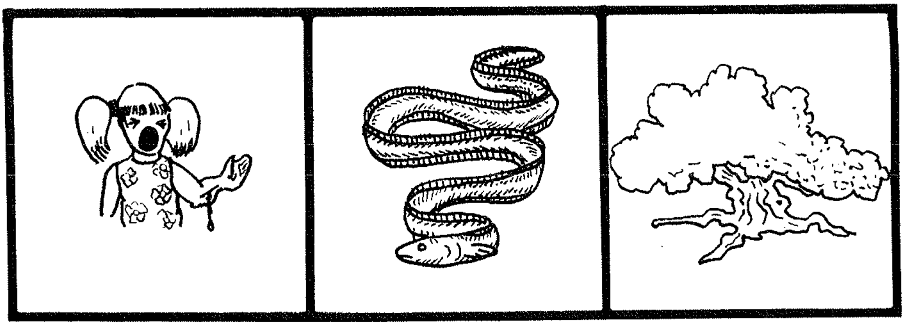
PROCEDURE

A rhyming test was therefore constructed and administered to 10 pre-school and 10 school children. Half of the children in each group had developed language normally and the other half had been diagnosed as language disordered. The ages of the pre-school children were between 3;11 - 5;8 years for the normally developed ones and between 5;10 - 7;2 years for the disordered children. The ages of the school children varied between 7;1 and 7;10 for both the subgroups (first grade).

	Language development normal	disordered
Pre-school children	3;11 - 5;8	5;10 - 7;2
School children	7;1 - 7;10	7;1 - 7;10

Table 1. Age of the twenty subjects

The test consisted of two parts. In part I, the child was presented with 3 pictures and the corresponding words (model words) (see figure 1). The child was instructed to point to the picture which represented the word that sounded most similar to a new word pronounced by the experimenter (test words).



MODEL WORDS [aj] [o:l] [e:k]

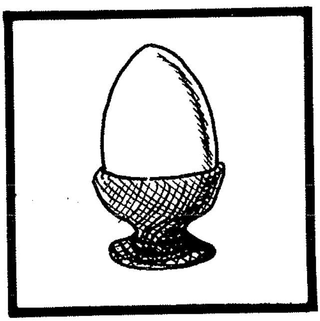
TEST WORDS [bo:l], [ste:k], [maj] etc

Fig.1 Rhyming test. Part I.

This procedure is modeled on a procedure used by Calfee et al. (1980) for testing and training phonetic segmentation. It has, however, been modified to fit the special demands of language disordered children.

Since the children answered by pointing and not by speaking, the deviant speech of the language disordered children did not influence the result. Because of the often limited memory capacity of the language disordered group (Gahne et al. 1983) the demands on memory had to be kept low. Therefore, the words were selected in such a way that only the test words (CVC structure) and not the models (VC structure) had to be segmented. The vowels were phonetically well seperated.

In part I of the test, however, it would be possible to find the appropriate rhyming words by comparing the vowels only. Therefore, part II of the test consisted of tasks of a somewhat different type (see figure 2). Only one model word was presented at a time, illustrated as before by a picture.



MODEL WORD [ɛg] egg

TEST WORDS [hɛg], [nɛb], [tag], [sɛk] etc

Fig.2 Rhyming test. Part II.

The child's task now was to decide whether a number of test words rhymed with the model word or not. The test words were varied in such a way that in the non-rhyming words either the vowel or the post-vocalic consonant differed from that of the model word, some by several distinctive features and others by one feature, e.g. VOICE or ROUNDING. The complete list of test words can be found in appendix 1.

RESULTS

From the error scores presented in table 2 it can be seen that the normally developed children made fewer errors than the language disordered children and school children made fewer errors than pre-school children. However, as a group, the language disordered school children made fewer errors than the normally developed pre-school children. It should be noted that there are some children around the age of four in the pre-school group and they seem to have difficulties in understanding the instructions. Children as young as four have probably not reached the cognitive level necessary for this type of task. The group which made the highest number of errors was the language disordered pre-school children. There seemed to be no correlation between age and linguistic awareness as measured by rhyming in this group.

	Normal			Disordered		
	I	II	Total	I	II	Total
Pre-school children	19	48	67	38	71	109
School children	4	5	9	16	37	53
Total	23	53	76	54	108	162

Table 2. Number of errors in the rhyming test, part I and II, made by the twenty subjects.

All subjects made more mistakes in part II of the test than in part I. This is as could be expected since part I does not require segmentation of the post-vocalic consonant but can be managed by comparing vowels that are phonetically well separated. Part II, thus, puts higher demands on both segmentation ability and ability to discriminate between phonetically close segments.

DISCUSSION

Language development seems to play an important role for the development of linguistic awareness as the normally speaking school children perfomed better than the language disordered school children. The same holds for the pre-school groups, even though the normally speaking pre-school children were considerably younger than the language disordered ones (cf. table 1).

Thus, age and cognitive level can not be the only determinants contributing to the development of linguistic awareness. The fact that school children score higher than pre-school children shows that age/cognitive level are not totally irrelevant. It could even be assumed that reading and spelling instruction to a certain extent promotes the growth of linguistic awareness as the language disordered school-children perform better than the normally speaking pre-school children. It is however obvious that reading and spelling ability is not a prerequisite for the development of linguistic awareness as illustrated by the illiterate pre-school children's rhyming performance.

When examining the results more closely we find that in part I, the most common difficulty among the youngest and/or the language disordered children was their inability to disregard content. Instead of concentrating on sound structure their choices implicated that they focused on content, basing their rhyming choices on semantic associations instead of on the similarity of sounds. For instance, when the test word was skrek (screamed) they pointed to aj (ow) instead of ek (oak), thus demonstrating an inability to disregard substance and to concentrate on form.

In part II, the errors can be of two types: false acceptances and false rejections. A comparison between the distribution of false acceptances and false rejections in part II and the errors in part I suggests that if a child makes more errors in part II than in part I, this can be explained as a lacking of ability to discriminate rather than as an inability to segment. In view of the fact that most of the children who showed this pattern were language disordered the explanation seems reasonable.

The numerous errors in part II as compared to the less frequent ones in part I were earlier attributed to the fact that the words in part II required segmentation while part I might be managed by comparing the vowels in the test words and in the models. However, an alternative explanation for the differential level of difficulty in the two parts of the test can be suggested if we discuss our results in terms of syllable structure. It has been suggested by e.g. Fudge (1969) that the structure of the syllable is hierarchical and divisible into onset and rhyme and that the rhyme is further divisible into peak and coda. Empirical data as e.g. in Treiman (1983) show that it is much easier to segment between the onset and the rhyme of the syllable than between the peak and the coda of the rhyme as a consequence of the hierarchical structure of the syllable. In our test the segmentation required in part I is between the onset and the rhyme while in part II a segmentation within the rhyme, i.e. between the peak and the coda, is required. Thus, our results can be interpreted as reflecting this hierarchical structure. Empirical data of other types such as slips of the tongue also reveal the internal structure of the syllable as pre-vocalic consonants are more often involved in speech errors than post-vocalic ones (Hockett 1967, MacKay 1972, Söderpalm 1979).

Children who are in the process of becoming aware of the segmentability of sequences seem to be dependent on what kind of segments they are supposed to handle within the rhyme. It is obvious from our data that certain consonant types such as nasals are much more difficult to segment from the preceeding vowel than stops or fricatives. It has been shown by House (1982) that non-coronal nasals have special acoustic characteristics depending on the preceeding vowel and thus require special identification strategies. Therefore, initially, it does not seem to be a question of form _or_ substance but rather an interaction between form _and_ phonetic substance contributing to an increasing of linguistic awareness.

TO SUM UP:

Since the normally developed children performed better than the language disordered children on the rhyming test it can be argued that language development is more important than cognitive development for linguistic awareness. Cognitive factors may also be of some importance since the language disordered school children performed somewhat better than the normally developed pre-school children. For this result, however, reading and spelling acquisition could be an influencing factor.

ACKNOWLEDGEMENTS

This research was partly supported by a grant from HSFR, Swedish council for Research in the Humanities and Social Sciences.

REFERENCES

Bruce, D.J. 1964. The analysis of word sounds by young children. British Journal of Educational Psychology 34:158-170

Calfee, R., Chapman, R. & Venezky, R. 1972. How a child needs to think to learn to read. In Gregg (ed.) Cognition in learning and memory. John Wiley & Sons, Inc.

Calfee, R.C., Lindamood, P. & Lindamood, C. 1973. Acoustic-phonetic skills and reading - Kindergarten through twelfth grade. Journal of Educational Psychology 64:293-298

Calfee, R.C. et al. 1980. Stanford Foundation Skills Test. Book B. Phonetic segmentation. Stanford University (unpublished)

Ehri, L.E. 1979. Linguistic insight: Threshold of reading aquisition. In Waller & MacKinnon (eds.) Reading research: Advances in theory and research, Volume 1, Academic Press, New York

Fudge, E.C. 1969. Syllables. Journal of Linguistics 5:253-286

Gahne, G., Rudberg, L-A. & Åström, E. 1983. Undersökning av korttidsminnets betydelse för läsförmågan. 60-poängsuppsats i fonetik. Stencil. Lunds universitet, Institutionen för lingvistik

Hakes, D.T. 1980. The development of metalinguistic abilities in children. Springer-Verlag, Berlin

Hockett, C.F. 1967. Where the tongue slips, there slip I. In To Honor Roman Jakobson. Mouton, The Hague. Also in Fromkin, V. (ed.) 1973. Speech errors as linguistic evidence. Mouton, The Hague

House, D. 1982. Identification of nasals: An acoustical and perceptual analysis of nasal consonants. Working Papers 22, Phonetic Laboratory, Department of Linguistics, Lund University

Liberman, I.Y. 1973. Segmentation of the spoken word and reading acquisition. Bulletin of the Orton Society 23:65-77

Liberman, I.Y., Shankweiler, C., Liberman, A.M., Fowler, C. & Fischer, F.W. 1977. Phonetic segmentation and recoding in the beginning reader. In Reber & Scarborough (eds.), Towards a psychology of reading: Proceedings of the CUNY conference, Hillsdale, New Jersey

Lundberg, I., Olofsson, Å. & Wall, S. 1980. Reading and spelling skills in first school years predicted from phonemic awareness in kindergarten. Scandinavian Journal of Psychology 21:159-173

MacKay, D.G. 1972. The structure of words and syllables: Evidence from errors in speech. Cognitive Psychology 3:210-227

Marshall, J. & Morton, J. 1978. On the mechanics of Emma. In Sinclair, Jarvella & Levelt (eds.) The child's conception of language. Springer-Verlag, Berlin

Mattingly, I.G. 1972. Reading, the linguistic process, and linguistic awareness. In Kavanagh & Mattingly (eds.), Language by ear and by eye. The MIT Press, Cambridge

Söderpalm, E. 1979. Speech errors in normal and pathological speech. Travaux de l'institut de linguistique de Lund XIV. Gleerups, Lund

Treiman, R. 1983. The structure of spoken syllables: Evidence from novel word games. Cognition 15:49-74

Tornéus, M. 1983. Rim eller reson? Språklig medvetenhet och läsning. Psykologiska institutionen, Umeå Universitet

Valtin, R. (in press) The development of metalinguistic abilities. In Downing & Valtin (eds.) Language awareness and learning to read. Springer-Verlag, Berlin

APPENDIX

Model words and test words in part I and part II of the test

Part I

Model words

[aj], [o:l], [e:k]

Test words

[bo:l], [ste:k], [maj], [haj], [gre:k], [mo:l], [blaj], [ho:l], [le:k], [sto:l], [skre:k], [ble:k], [skro:l], [svaj], [ve:k] [kaj], [vro:l], [skraj]

Model words

[i:s], [ʉl], [ɛg]

Test words

[krʉl], [lɛg], [mʉl], [skʉl], [ri:s], [ʃɛg], [di:s], [pri:s], [kri:s], [gʉl], [vi:s], [hʉl], [gnɛg], [vɛg], [hɛg], [tʉl], [gri:s], [klɛg],

Part II

Model word

[i:s]

Test words

[kri:s], [pri:s], [vi:s], [li:v], [gri:s], [ny:s], [di:s], [hi:t], [hɯ:s], [li:k], [he:s], [ri:s]

Model word

[ɛg]

Test words

[hɛg], [tag], [vɛg], [nɛb], [vɛ:g], [lɛg], [sɛŋ], [klɛg], [sɛk], [lʉg], [ʃɛg], [gnɛg]

Model word

[ʉŋ]

Test word

[tʉn], [nʉb], [kʉŋ], [ʃʉŋ], [tɔŋ], [ʃoeŋ], [svʉŋ], [mʉg], [paŋ], [tʉŋ], [pʉŋ], [plʉŋ]

PhD Thesis Summary: Developmental Studies of Dysphonology in Children

Ulrika Nettelbladt

INTRODUCTION

The current investigation focused on dysphonology (delayed or deviant phonological development) in children with language disorders, i.e. children who do not develop language normally. In this thesis, only children with specific language disorders, were considered, i.e. children with no other gross handicaps. The thesis had two main objectives. The first was to describe subgroups of dysphonology in children with language disorders. The description was based primarily on different phonological characteristics of the children's productions. The second purpose, related to the first, was to study the development of children with different types of dysphonology and to find out whether the children follow similar lines of development. One normal subject was included in the study, as a preliminary basis of comparison for phonological patterns in language disordered children and in a young child with normal language development. To achieve these goals, it was necessary to develop a method for describing severe dysphonology.

SUMMARY

The thesis includes two different studies, focusing on production. The first, the exploratory study, was a survey of dysphonology. Thirty-one subjects were included, from four different regions in Sweden: 17 boys and 14 girls, ranging in age from 4;0 to 7;11. The second, or the main study, was a longitudinal study of 10 subjects (5 girls and 5 boys) selected from the exploratory study. These children were chosen from a single dialect region: South West Skåne (Scania). The 10 subjects were studied over a period of from 1 to 4 1/2 years. Longitudinal data from one young normal subject (age 1;8 to 2;5) was included in the main study for reference purposes. The material for the language disordered children in both studies was elicited by means of a picture naming test, balanced with respect to Swedish phonemic, phonotactic and prosodic characteristics. It also included a number of children's tonguetwisters. In the main study, some supplementary testing was carried out, including a sentence-eliciting test, auditory phoneme and accent discrimination and, finally a test of articulatory praxis. All testing sessions were tape recorded, transcribed during the sessions and later retranscribed. The taped material from the normal subject consisted of conversational speech.

The material of the exploratory study was subjected to a broad phonological analysis, mainly in terms of simplification processes, concerning syllables, vowels and consonants. Dysprosody, homonymy, segment variability, relative degree of unintelligibility and resistance to therapy, were also noted.

The major result of the exploratory study was the identification of two main types of dysphonology. The first and most severe group is referred to as the syntagmatic group. It is characterized by strong syntagmatic restrictions in combination with paradigmatic restrictions. Syllable, vowel and consonant processes are all found in this group, but in particular the syllable processes, such as deletions of final consonants and unstressed syllables, conspire to restrict the phonotactic structure and thus the phonological diversity of possible words. The second group, regarded as moderate, is referred to as the paradigmatic group. Here paradigmatic restrictions predominate, i.e. on the consonant phonemic system.

In order to describe the changing phonologies of the children with severe dysphonology in particular, it was necessary to revise the descriptive framework. The new framework was divided into two parts, an autonomous and a process description.

In the autonomous description, the child's words are analyzed independently of the adult target, and it is thus possible to include unintelligible utterances (frequent in severely dysphonological children) in this part of the description. It is divided into three levels: word, syllable and segmental. At the word level, different kinds of syntagmatic restrictions are stated, and stress and accent patterns are described. Phonotactic and prosodic characteristics of words are summarized in a number of typical word patterns. At the syllable level, syllable types and restrictions within them are noted. At the segment level, vowel and consonant phonemes are identified but the appropriateness of defining phonemic contrasts in the earliest stages of development and in severely dysphonological children is questioned. Vowel length contrast and occurrence of diphthongs are noted.

In the process description, the child's productions are related to the adult target words in terms of (simplification) processes of two main types: syntagmatic and paradigmatic. Syntagmatic processes change the phonotactic structure of the target words. Child productions in which syntagmatic processes have applied are usually shorter and less complex than their target equivalents. Extraction processes are introduced as a complement to deletions, to describe child productions where only a small fraction of the target word is left. Other typical syntagmatic processes are, for example, reduplication, additions of different kinds and assimilation. Paradigmatic processes are context-free processes and apply to classes of segments. Two main kinds of paradigmatic processes are considered: Vowel and consonant substitutions.

The phonological development of the language disordered children in the main study and the normal subject were analyzed according to the descriptive framework briefly outlined above. Only selected recordings were used, chosen because noticeable changes occur in comparison to earlier recordings. The analyses are presented as individual case studies: these are relatively detailed for two of the children with the most severe dysphonology and for the normal child. Complete word lists are given for these three children. For the remaining children a selected, illustrative sample is included. Brief case histories are given for the language disordered children, as well as an informal follow-up of their school performance after the study was completed.

The results indicate that it is indeed possible to unravel systematic phonological patterns, even in severe cases of dysphonology, despite the fact that they often have idiosyncratic solutions for their phonological problems. Similar but more varied kinds of patterns were also found in the young normal child. Thus, there appears to be a basic similarity between the primitive phonology of the young normally developing child and the phonology of the children with severe language disorders.

Organizing the data into autonomous descriptions made it possible to develop a preliminary stage model for the earliest stages of phonological development, based primarily on phonotactic and prosodic characteristics. The model is divided into four subsequent stages. The first two stages are characterized by strong phonotactic constraints on words (syllable or segment harmony) and on syllables (predominance of simple open and closed syllables) and absence of prosodic contrasts. The model specifies how children gradually loosen the constraints which their developing language and speech processing systems impose on them. Within the model, a number of implicational relationships appear, e.g. between the relaxation of phonotactic constraints and the development of prosodic contrasts.

There is a sharp break between the second and the third stages of the model. A number of achievements take place between these two stages, and it is suggested that this is the point at which a syntagmatic-paradigmatic shift in phonological development occurs. It is suggested that the child's phonology is reorganized: i.e. there is a shift in attention from word patterns to syllable shapes and later to segment contrasts. From the process point of view, extraction processes, reduplication and assimilation gradually give way to consonant substitutions. More speculatively, it is further suggested that the ability to handle increased phonotactic complexity, as described in the model, may have a parallel in the development of syntax, which appears to start at the time of the syntagmatic-paradigmatic shift. Whether these different achievements are based on a more general ability to handle complex, hierarchically organized language material is considered.

The model has important clinical implications. It is suggested that language disordered children can be classified according to types of dysphonology in terms of severity and how well they fit to the stages of the model. Severe cases of dysphonology exhibit characteristics of the first two stages. Symptoms typical of these stages may thus be considered as risk symptoms of a more severe type of language disorder.

Children in the syntagmatic group, i.e. with characteristics of stages I and II, are much more rigid in their use of word patterns and syllable shapes and also in their use of processes: Syntagmatic processes predominate in this group, and the children tend to stick to a few primitive processes, such as reduplication or assimilation. In this respect, they deviate from the normal subject at a comparable stage of development. These children also tend to relapse easily into earlier types of simplification, and some of them exhibit residual syntagmatic restrictions. On the basis of these considerations, specific suggestions for therapy are also given.

Finally, the present investigation stresses the importance of a differentiated linguistic analysis (in this thesis restricted to phonological aspects). This analysis should be regarded as a vital complement to a thorough neurological and neuropsychological examination of children with language disorders.

Acoustic Analysis of Vowels and Diphtongs in Cairo Arabic

Kjell Norlin

1. INTRODUCTION

This paper reports an investigation on acoustic properties of vowels and diphtongs in Cairo Arabic. The two classes of plain and pharyngalized consonants in Cairo Arabic are also discussed, using the alveolar fricatives /s/ and /ṣ/ as an illustration (Norlin 1983). The effects on vowels in the environment of these consonant classes are demonstrated.

Arabic belongs to the Semitic branch of the Afroasiatic language family. Like all varieties of Arabic, Cairo Arabic has few contrasting vowel phonemes, giving place to a number of vowel allophones. Cairo Arabic has five contrasting long vowels: /ii, ee, aa, oo, uu/, and three short vowels: /i, a, u/. The long mid vowels are derived from the diphtongs /ai/ and /au/. The diphtongal qualities are still preserved in Standard Arabic. In Cairo Arabic the long mid vowels /ee/ and /oo/ are shortened under certain morphological conditions and merge with short /i/ and /u/ phonetically. In addition, Cairo Arabic does have three phonetic diphtongs, [iu, au, ai] in a subset of the vocabulary. These are usually analyzed as /iw, aw, ay / (Harrell, 1957).

Syllable structure is rather simple. The following syllables occur: CV, CVC, CVV, CVVC, CVCC. The last two syllables can only occur in word final position and, of course, form monosyllabic words.

2. PROCEDURE

Most of the data illustrating vowels are taken from real monosyllabic words of the types CVVC and CVCC, set in a sentence frame ʔulna ... kamaan (we said ... again), preceded and followed by dental consonants. All contrasting vowel phonemes and the three diphtongs occur in this position. To get examples of short [e] and [o], some disyllabic words were included, since they cannot occur in monosyllables. Long /ii, aa, uu/ also occur in a stressed syllable in disyllabic words, since vowels are not found between these consonants in monosyllables.

Six speakers of Cairo Arabic recorded the sentences to give five tokens of each utterance. The recordings were made in the studio of the Phonetic Department in Lund. The recordings of vowels and diphtongs were analyzed from broad-band spectrograms from a Kay Digital Sonagraph 7800. For the vowels the first two formants of the five tokens were traced and superimposed on each other to get an idea of variation within each speaker. The variation proved to be very small and can in all cases be considered to be within the measurement error. Therefore three tokens were selected for analysis. The long and short vowels were all monophtongs. Formant frequencies were measured from a steady state portion of the formant. Mean formant values representing each vowel of each speaker were calculated, tables 1 and 2. Using a lab computer the formant values in Hz were converted to mel and plotted on an acoustic chart with F_1 and F_2 in the usual way. Vowel duration was measured from five tokens and mean values calculated, table 3. For the diphthongs the first two formants were measured together with the duration of each steady state and the duration of the transition. The formant frequencies were measured in the middle of each steady state.

3.1. RESULTS: VOWEL DURATION

Vowel durations were first considered separately in plain and pharyngalized environment. The results show that vowel duration is not significantly different between plain and pharyngalized long vowels, nor between plain and pharyngalized short vowels in these environments. Therefore plain and pharyngalized vowels are considered together. Table 3 shows the mean duration for all the vowels. The difference in length between long and short vowels is rather large. Short vowels are about half the duration of long ones.

3.2. RESULTS: VOWEL QUALITY

Figure 1 is a formant chart of plain, long vowels. Plain, long vowels are well separated with no overlapping, except for long /ii/ and /ee/ touching each other. Figure 2 shows a formant chart of the three plain, short vowels. These vowels are also

well separated clusters. Figure 3 also shows plain, short vowels, but it includes the non-phonemic [e] and [o] . Here short [e] and [o] show nearly complete overlapping with short /i/ and /u/. Their non-phonemic status is stated in literature, but phonetic data supporting the linguistic analysis is generally not presented. In some text-books there is even a claim that there is a phonetic difference between [i] and [e] on the one hand, and [o] and [u] on the other (Abdel-Massih 1975).
Figures 4 and 5 show formant charts of the pharyngalized long and short vowels.

4. DISCUSSION

4.1. Vowel length and vowel quality

Difference in vowel length influence vowel quality. For both plain and pharyngalized environments the long vowels are more peripheral, whereas the short vowels are inside the space of the long vowels with the exception of long /aa/ and short /a/ where the quality difference is small. Both plain and pharyngalized long /ii/ differ significantly ($p<0.001$) from short /i/ along both F_1 and F_2. The short /i/ vowels are lower and further back than the long /ii/ vowels. Short /u/ vowels are lower and more front than long /uu/ vowels. Both sets of long /aa/ differ to some extent from their short counterparts along F_2. The short vowels tend to be further back. The differences along F_1 between long and short /a/ vowels are non-significant: long and short vowels have the same vowel height. It seems as if vowels in Cairo Arabic are anchored on the low vowels and the short vowels /i/ and /u/ do not reach the vowel quality of the long ones, figure 6.

4.2 Vowel quality in plain and pharyngalized environment.

An earlier study (Norlin 1983) analyzed all the fricatives in Cairo Arabic. These include the pharyngalized fricatives /ṣ/ and /ẓ/. In this study FFT spectra of the fricatives were converted to critical band spectra. The center of gravity of the critical band spectra was plotted against dispersion, figure 9. The center of gravity in the spectra was also plotted against the mean intensity level of the spectra in dB, figure 10.
The results from this earlier study showed that the plain and pharyngalized consonants are different, even if the difference

is small.
Considering the effects of plain and pharyngalized environments on vowel quality, a comparison between long vowels shows that there is a complete overlapping for for both sets of long /uu/. Pharyngalized long /ii/ shows a small difference from its plain counterpart in that it is slightly lower and further back. Pharyngalized long /aa/, however, is greatly affected, showing a considerable difference in the F_2 dimension from its plain counterpart: the pharyngalized /aa/ is much further back than the plain /aa/, figure 7.

A comparison between the plain and pharyngalized short vowels shows that the pharyngalized short /u/ does not overlap with plain short /u/, as is the case with the long counterparts. It is further back. The short pharyngalized /i/ is also further back than plain /i/. The two sets of short /a/ differ in the same manner: pharyngalized short /a/ is further back. Thus all short vowels are further back in pharyngalized environment.Fig.8.

In conclusion, it is evident that the pharyngalization process affects the whole syllable. On the one hand, plain and pharyngalized consonants differ consistently. On the other hand, vowels in plain and pharyngalized environment differ in more complex ways. Plain and pharyngalized low vowels always show a considerable difference in the F_2dimension, regardless of length, the pharyngalized vowels being more back. High long vowels show small or no difference, whereas short high vowels always are further back than plain ones.

5. DIPHTHONGS

Most Standard Arabic diphtongs /ai/ and /au/ have in Cairo Arabic developed into long /ee/ and /oo/. They also exist in Cairo Arabic, however, in a number of purely dialectal words and some Standard Arabic words commonly used in daily speech. In addition, there exists a third diphtong [iu] due to morphophonematic rules in verb conjugation. A comparison between the short vowels and the corresponding segments in the diphthongs shows some differences in vowel quality. In the diphtongs, F_1 frequencies are always identical with F_1 in short vowels, but F_2 is always lower, making diphthong segments more back.

The rate of transition between the two components in diphthongs is fast, around 30-35 milliseconds in all the diphtongs. It seems as if the diphthongs in Cairo Arabic are made by stringing the short vowels together with a fast transition, figures 11, 12, 13.

ACKNOWLEDGEMENTS

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.

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

Norlin, K. 1983.Acoustic analysis of fricatives in Cairo Arabic. Working Papers 25:113-137. Phonetics Laboratory, Lund University.

Table 1. Formant frequencies of long and short plain vowels in Cairo Arabic.

Vowel	Speaker	F_1	F_2	F_3	Vowel	Speaker	F_1	F_2	F_3
/ii/	1	225	2600	3050	/aa/	1	625	1700	2700
		275	2575	3150			600	1650	2650
		250	2650	3100			625	1675	2675
	2	300	2275	3150		2	550	1750	2475
		300	2325	3000			550	1825	2500
		300	2325	3150			575	1800	2500
	3	275	2250	2900		3	525	1700	2600
		250	2275	2850			575	1750	2675
		275	2250	2800			500	1700	2650
	4	300	2050	2775		4	525	1600	2500
		300	2100	2800			500	1625	2500
		300	2000	2800			500	1525	2500
	5	300	2300	2925		5	575	1750	2600
		300	2325	3025			575	1775	2550
		275	2300	3000			575	1825	2525
	6	350	2200	3150		6	550	1675	2700
		300	2200	3125			550	1750	2700
		300	2200	3200			550	1700	2700
	mean	288	2289	2997		mean	557	1710	2594
/ee/	1	300	2250	2800	/oo/	1	400	750	2650
		300	2400	2800			350	750	2500
		400	2375	2825			400	800	2500
	2	425	2100	2700		2	400	850	2225
		400	2125	2650			400	900	2225
		450	2150	2650			425	875	2250
	3	375	2075	2700		3	400	950	2350
		400	2025	2700			450	1000	2375
		375	2050	2700			400	1000	2400
	4	400	1825	2500		4	350	900	2250
		350	1875	2600			350	850	2200
		400	1900	2500			400	975	2400
	5	350	2275	2700		5	400	825	-
		325	2225	2650			400	850	2400
		350	2225	2700			400	775	2400
	6	350	2200	3100		6	400	775	2400
		375	2200	3100			325	775	2500
		325	2150	3050			350	800	2450
	mean	369	2135	2746		mean	389	856	2381

Table 1 (cont.)

Vowel	Speaker	F_1	F_2	F_3	Vowel	Speaker	F_1	F_2	F_3
/uu/	1	275	775	2675	/a/	1	550	1600	2725
		300	700	2650			575	1600	2700
		275	700	2500			625	1600	2700
	2	250	825	2300		2	650	1500	2450
		300	850	2275			625	1475	2500
		300	875	2325			625	1525	2500
	3	275	750	2475		3	550	1575	2525
		300	825	2500			550	1600	2600
		300	800	2350			550	1600	2525
	4	300	800	2275		4	550	1500	2575
		300	850	2200			525	1600	2550
		300	775	2225			550	1525	2600
	5	325	700	-		5	600	1500	2625
		300	650	-			575	1575	2600
		300	700	-			600	1525	2500
	6	300	725	2350		6	525	1650	2800
		300	725	-			600	1700	2750
		300	775	-			525	1700	2825
	mean	294	767	2392		mean	575	1575	2614
/i/	1	400	1875	2750	/u/	1	400	1150	2750
		375	1950	2725			400	1100	2725
		375	1900	2750			475	1075	2900
	2	475	1850	2425		2	450	975	2250
		475	1825	2550			450	1050	2350
		450	1850	2500			475	1050	2250
	3	425	1875	2550		3	400	1175	2250
		425	1825	2600			450	1175	2375
		400	1825	2600			425	1200	2400
	4	400	1750	2500		4	400	1250	2250
		475	1750	2500			375	1200	2300
		450	1800	2425			425	1250	2225
	5	400	1925	2625		5	350	1100	2500
		375	1925	2625			375	1250	2400
		400	1925	2600			400	1150	2400
	6	450	1950	2825		6	425	1050	2550
		450	1900	2800			400	1100	2400
		450	2000	2800			400	1050	2600
	mean	425	1872	2619		mean	415	1131	2438

Table 2. Formant frequencies of long and short pharyngalized vowels in Cairo Arabic.

Vowel	Speaker	F_1	F_2	F_3	Vowel	Speaker	F_1	F_2	F_3
/ii/	1	250	2400	2800	/uu/	1	325	725	3000
		275	2400	2800			250	675	2975
		275	2325	2825			250	675	-
	2	300	2150	2625		2	300	875	2375
		300	2200	2725			300	775	2450
		300	2200	2700			325	800	2500
	3	300	2200	2600		3	250	825	2500
		350	2125	2575			325	850	2400
		325	2100	2600			300	800	2500
	4	325	2000	2425		4	325	825	2500
		350	2000	2475			300	800	2500
		325	1925	2425			325	800	2575
	5	350	2200	2700		5	300	650	2700
		350	2175	2750			325	700	2650
		300	2200	2700			325	625	2750
	6	350	2175	3150		6	350	800	-
		300	2150	3100			300	750	-
		350	2150	3100			300	750	-
	mean	315	2171	2726		mean	304	761	2598
/aa/	1	600	1125	2900	/i/	1	400	1375	2800
		600	1100	2900			450	1250	2800
		675	1050	2975			450	1300	2900
	2	550	1000	2250		2	475	1525	2500
		600	1000	2375			450	1550	2450
		600	1000	2350			475	1525	2400
	3	600	1100	2300		3	475	1475	2500
		550	1100	2300			475	1475	2475
		550	1050	2300			475	1475	2500
	4	550	1050	2400		4	450	1325	2400
		500	1025	2350			450	1300	2375
		550	1125	2300			425	1300	2300
	5	550	1050	2750		5	350	1850	2550
		525	1025	2675			400	1800	2625
		525	1025	2750			400	1800	2675
	6	575	1225	2775		6	400	1900	2725
		650	1300	2500			400	1800	2725
		550	1200	2725			425	1900	2650
	mean	572	1086	2549		mean	435	1551	2575

Table 2 (cont.)

Vowel	Speaker	F_1	F_2	F_3	Vowel	Speaker	F_1	F_2	F_3
/a/	1	650	1200	2800	/u/	1	500	900	2300
		625	1200	2850			500	1000	-
		600	1225	2750			400	975	2300
	2	650	1100	2500		2	400	800	2350
		650	1100	2525			450	750	2500
		625	1075	2500			450	825	2350
	3	600	1175	2350		3	500	1100	2400
		600	1150	2275			450	1100	2225
		550	1100	2400			475	1100	-
	4	600	1050	2300		4	450	1050	2500
		600	1150	2400			450	1100	2475
		575	1125	2300			400	975	2350
	5	550	1050	2875		5	375	925	2600
		550	1100	2850			400	900	2700
		575	1125	2875			400	975	2725
	6	600	1250	2800		6	400	1000	-
		550	1300	2850			400	1000	-
		575	1325	2800			400	925	-
	mean	596	1156	2611		mean	433	967	2444

Table 3. Vowel duration, plain and pharyngalized vowels, mean values in milliseconds.

/ii/	131	/i/	67
/ee/	153	/a/	84
/aa/	158	/u/	75
/oo/	185		
/uu/	139		

Table 4. Formant frequencies and durations of diphthongs in Cairo Arabic.[1]

Diphtong	Speaker	F_1	F_2	F_1	F_2	t_1	t_2	t_3	tot.
iu	1	300	1850	300	800	60	50	50	190
		450	1850	425	850	40	50	50	175
		425	1800	375	775	30	55	35	170
	2	450	1750	350	1000	30	25	40	125
		450	1700	325	850	60	30	35	155
		475	1750	450	1025	45	25	40	125
	3	450	1800	375	950	50	35	40	150
		500	1800	325	900	65	40	25	160
		450	1850	350	925	50	40	40	150
	4	475	1750	375	800	60	25	45	145
		400	1675	325	800	40	25	40	135
		475	1725	350	850	50	25	45	140
	5	400	1900	300	800	55	40	40	160
		375	1850	300	750	60	40	60	190
		400	1825	325	750	50	40	50	165
	6	425	1700	325	775	50	30	50	150
		400	1650	350	800	50	30	50	155
		450	1625	375	800	50	30	50	170
	mean	430	1769	350	844	49	35	44	156
ai	1	650	1600	400	2450	60	25	85	170
		675	1675	375	2450	60	20	90	170
		700	1700	400	2500	75	25	90	185
	2	650	1700	375	2000	75	30	60	165
		650	1600	400	2000	80	35	65	185
		675	1675	425	2100	70	40	70	180
	3	575	1750	400	2200	70	20	80	165
		600	1650	400	2100	50	30	85	170
		550	1650	450	2025	70	30	75	180
	4	600	1775	325	2075	70	30	55	175
		575	1725	350	2025	70	30	60	160
		575	1675	350	2050	70	30	70	170
	5	525	1700	325	2250	70	30	80	180
		650	1700	375	2250	60	30	80	170
		625	1625	350	2250	60	35	95	190
	6	600	1800	350	2225	75	30	95	195
		600	1750	350	2225	70	30	100	200
		575	1800	300	2200	70	30	85	185
	mean	614	1697	372	2187	68	29	79	177

1. t_1= duration of first steady state, t_2= duration of transition between the steady states, t_3= duration of second steady state.

Table 4 (cont.)

Diphtong	Speaker	F_1	F_2	F_1	F_2	t_1	t_2	t_3	tot.
au	1	575	1375	325	750	70	30	65	190
		625	1400	400	775	65	45	55	180
		625	1450	400	800	70	40	55	180
	2	550	1425	400	1075	70	25	40	140
		575	1450	425	1100	70	25	40	135
		600	1375	425	1100	70	30	35	140
	3	550	1500	425	1025	70	40	35	160
		550	1500	425	1000	80	35	40	150
		550	1500	425	1075	65	40	55	180
	4	525	1425	375	1050	50	25	40	125
		500	1400	400	1100	50	25	30	120
		550	1400	400	950	60	20	30	125
	5	700	1500	450	825	60	30	40	145
		650	1525	375	850	60	40	30	150
		600	1575	350	700	60	35	35	145
	6	600	1625	425	850	50	40	50	180
		575	1575	400	850	50	40	60	180
		600	1575	425	825	55	40	55	190
	mean	583	1476	403	928	63	34	44	156

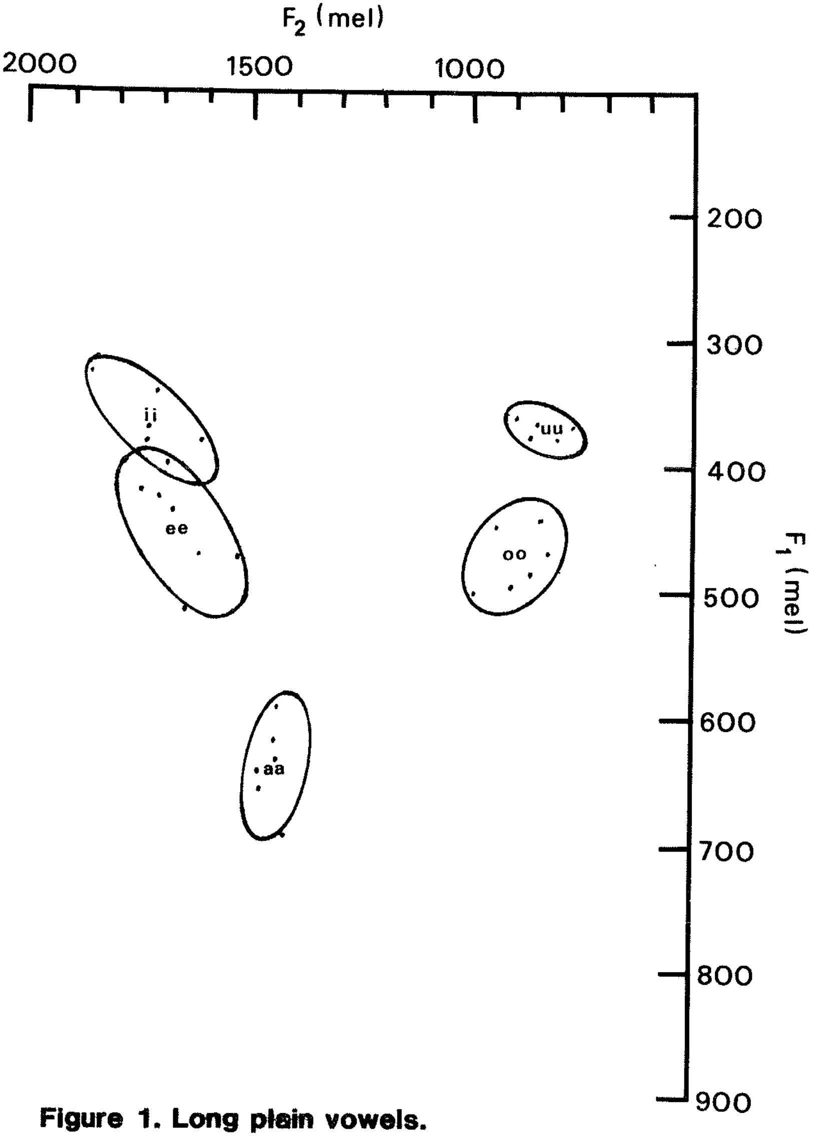


Figure 1. Long plain vowels.

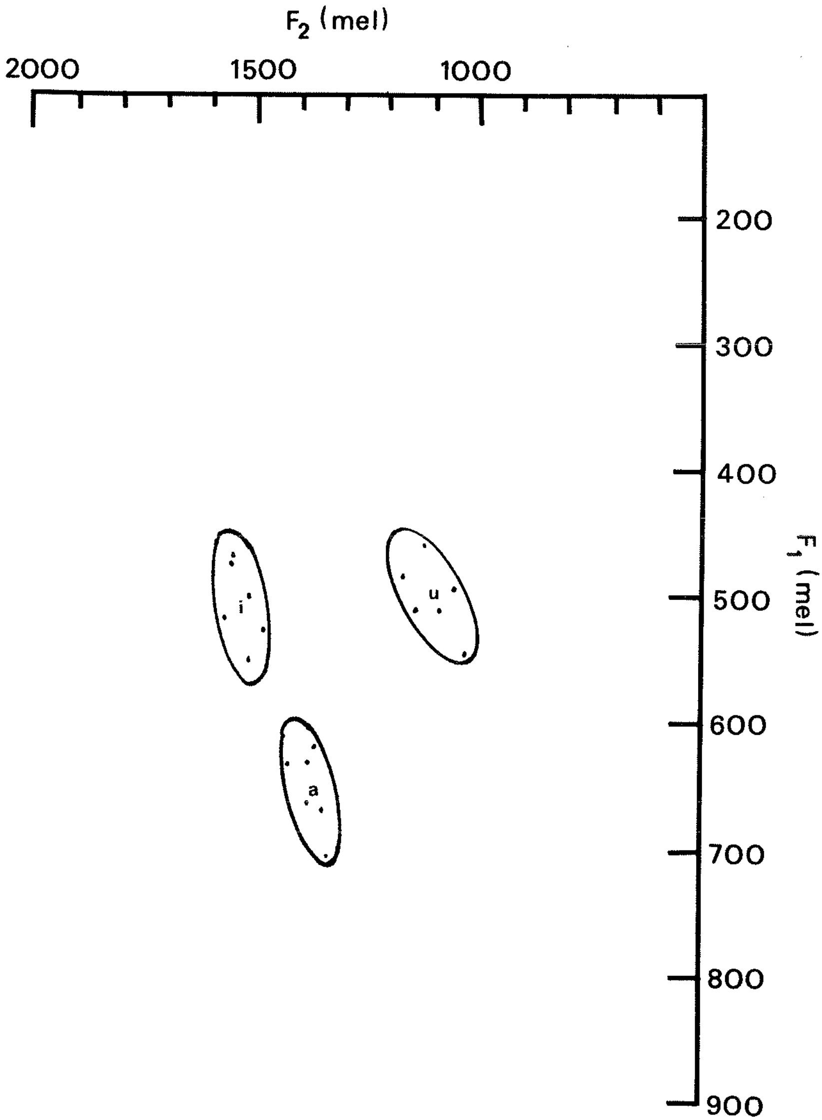


Figure 2. Short plain vowels.

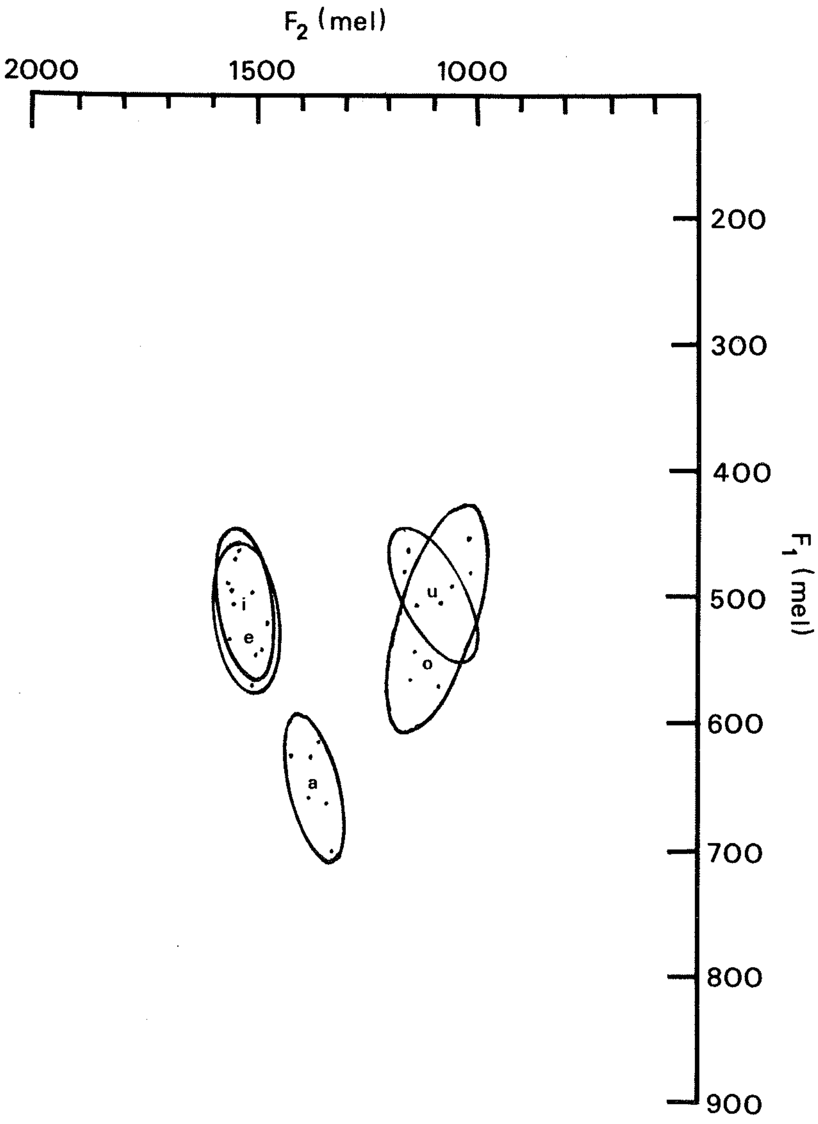


Figure 3. Short plain vowels.

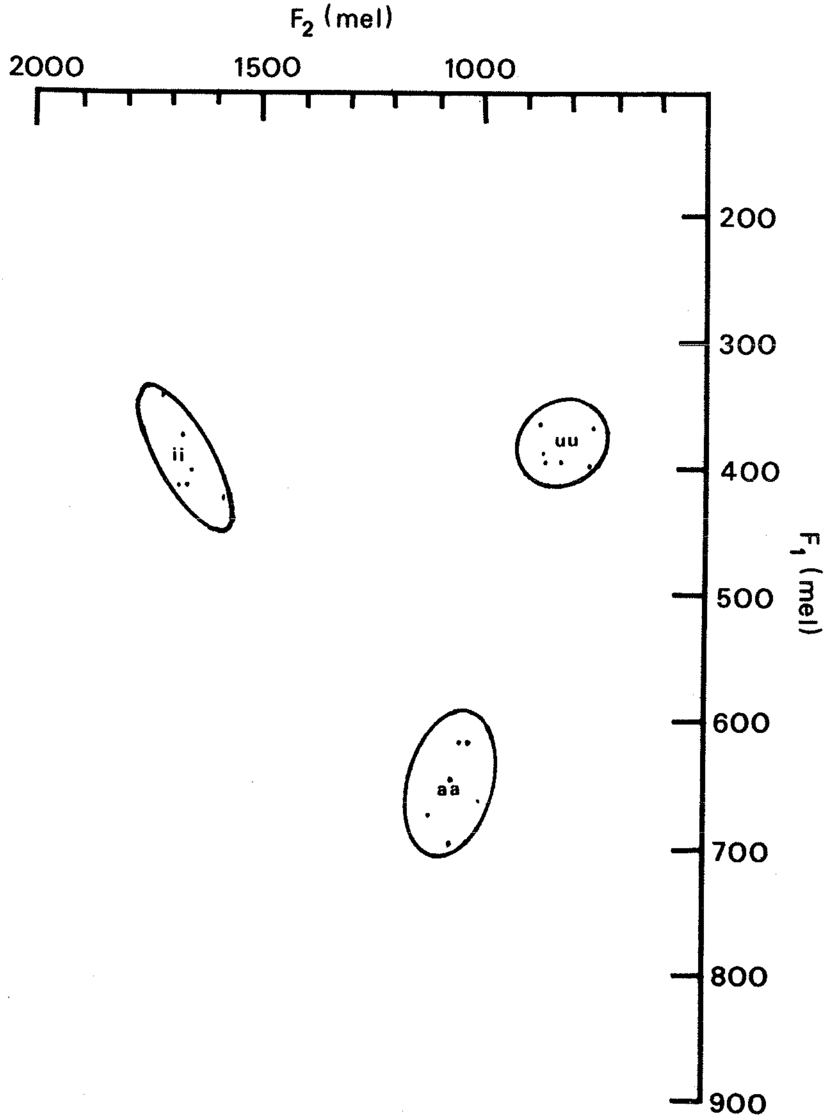


Figure 4. Long pharyngalized vowels.

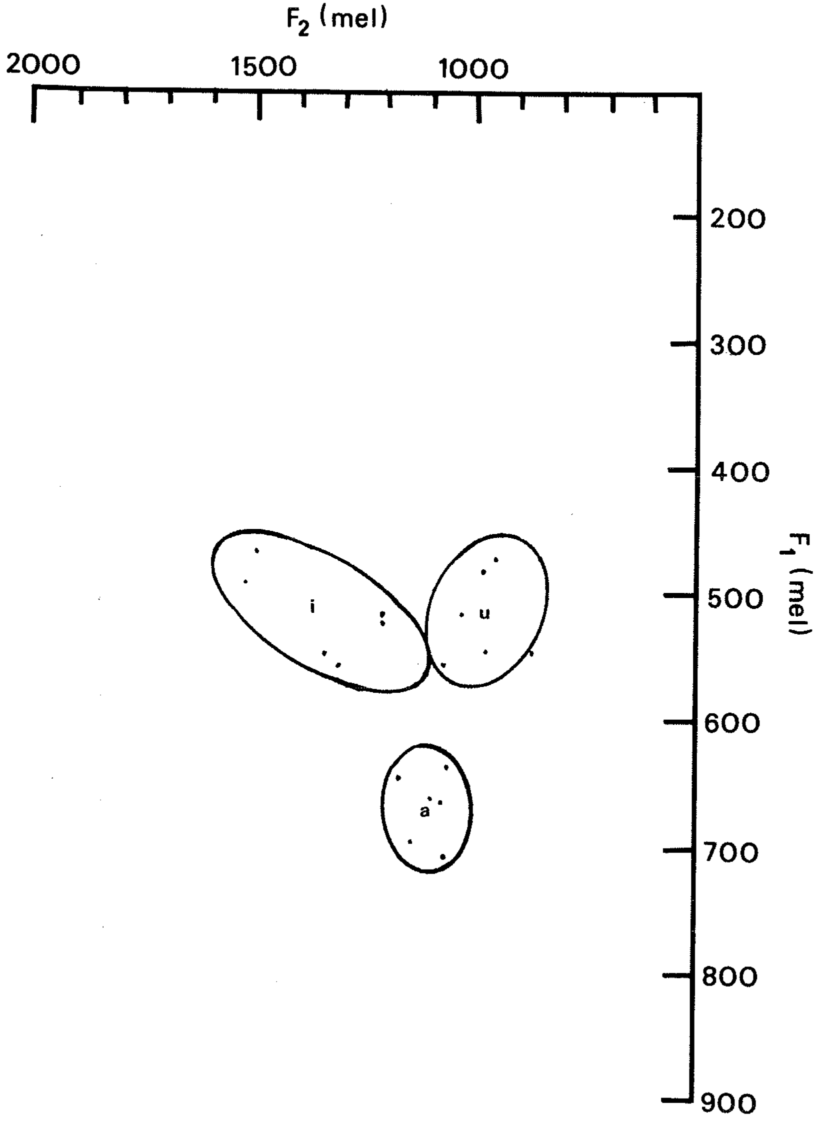


Figure 5. Short pharyngalized vowels.

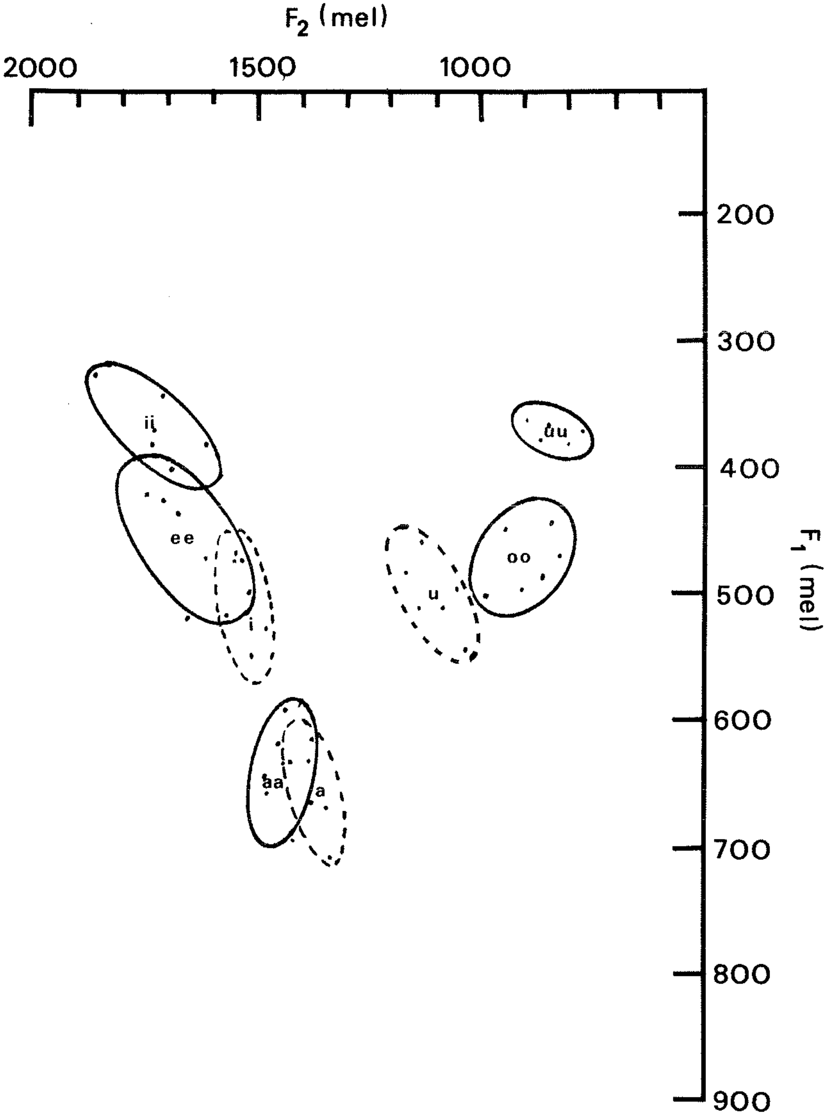


Figure 6. Long and short plain vowels.

—— long vowels

- - - short vowels

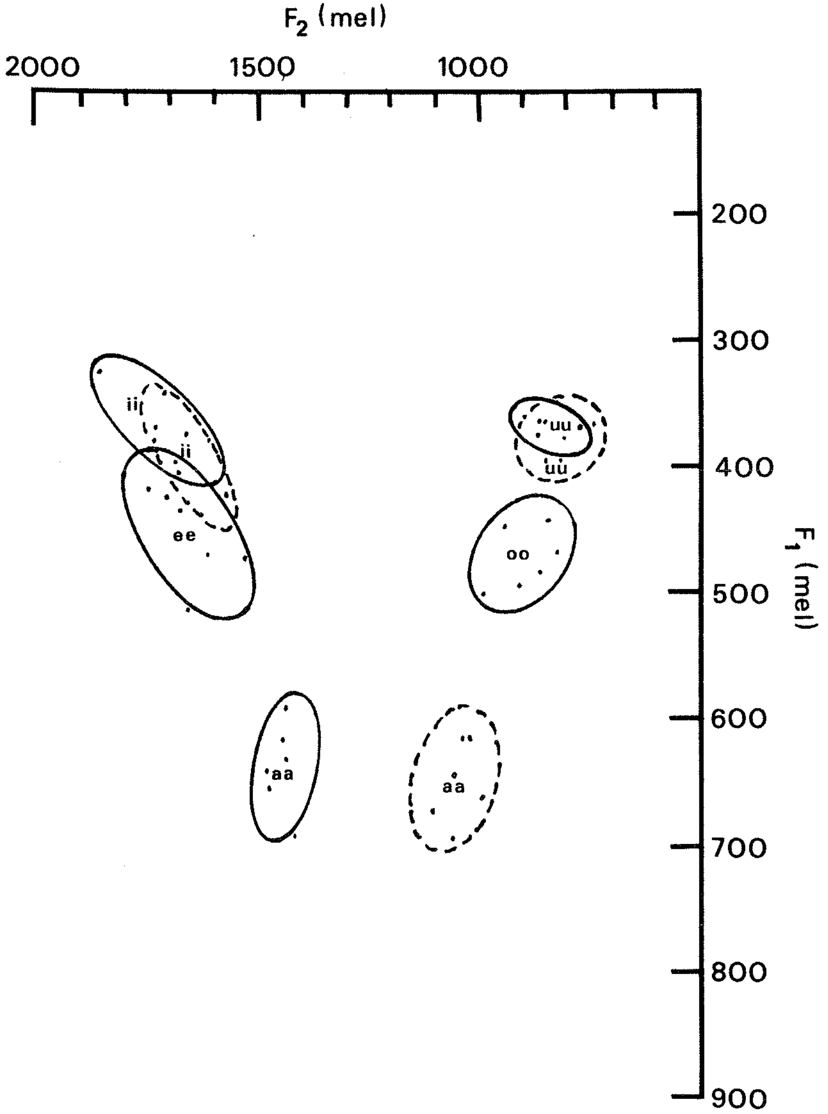


Figure 7. Long plain and pharyngalized vowels.

—— plain vowels

- - - pharyngalized vowels

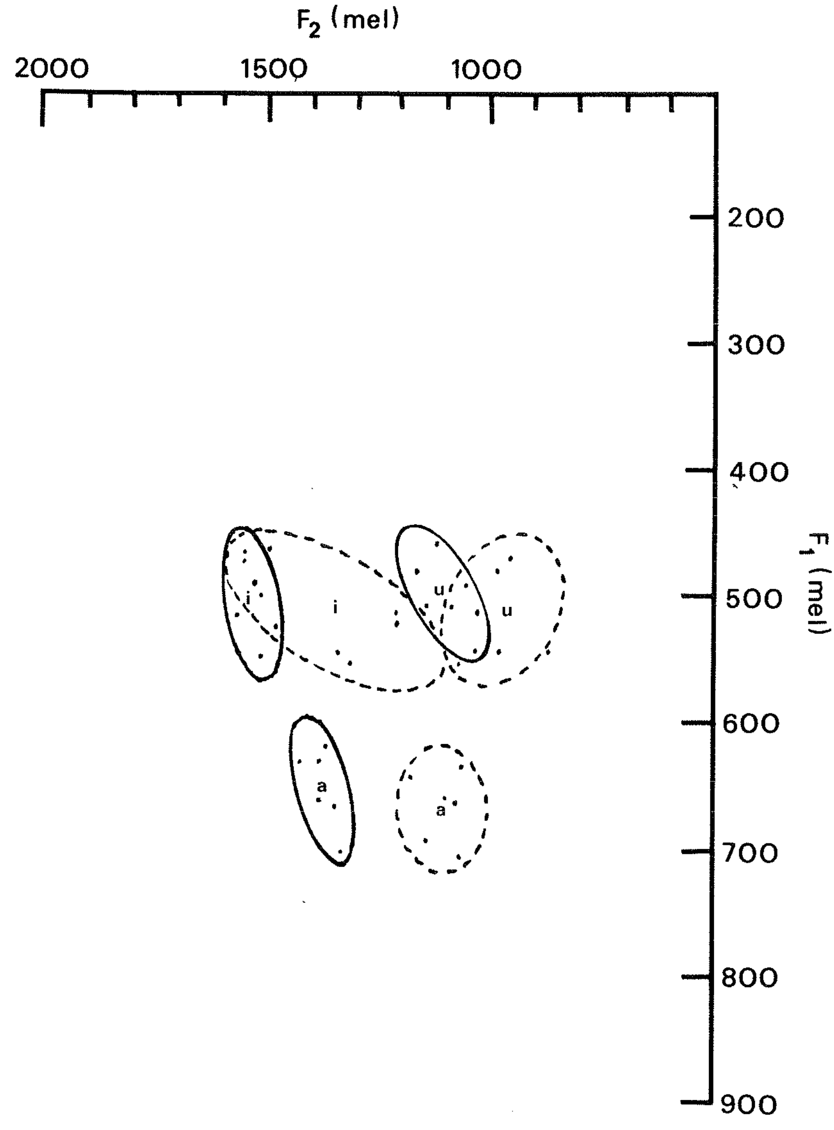


Figure 8. Short plain and pharyngalized vowels.

—— plain vowels

- - - pharyngalized vowels

Figure 9

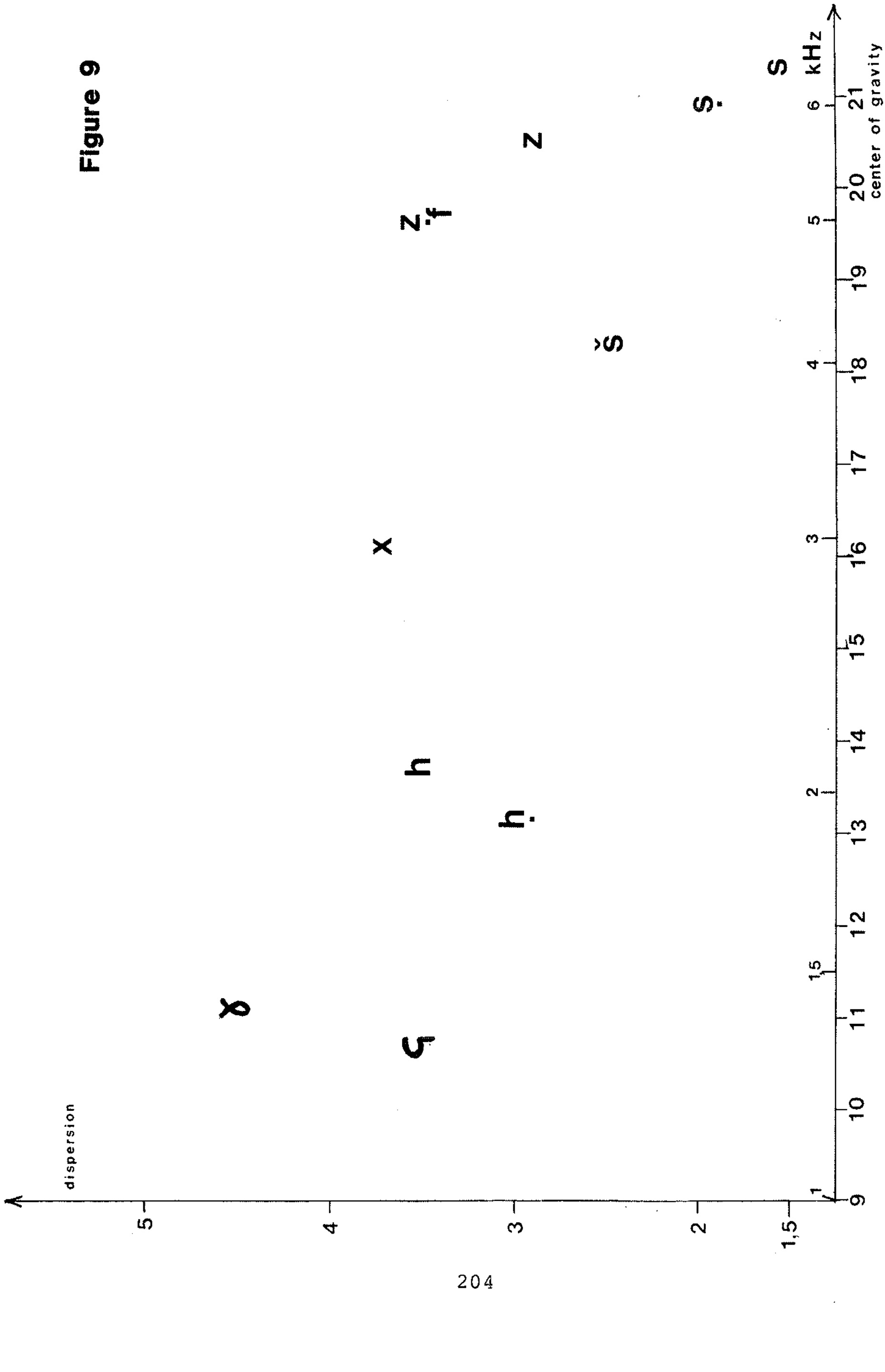
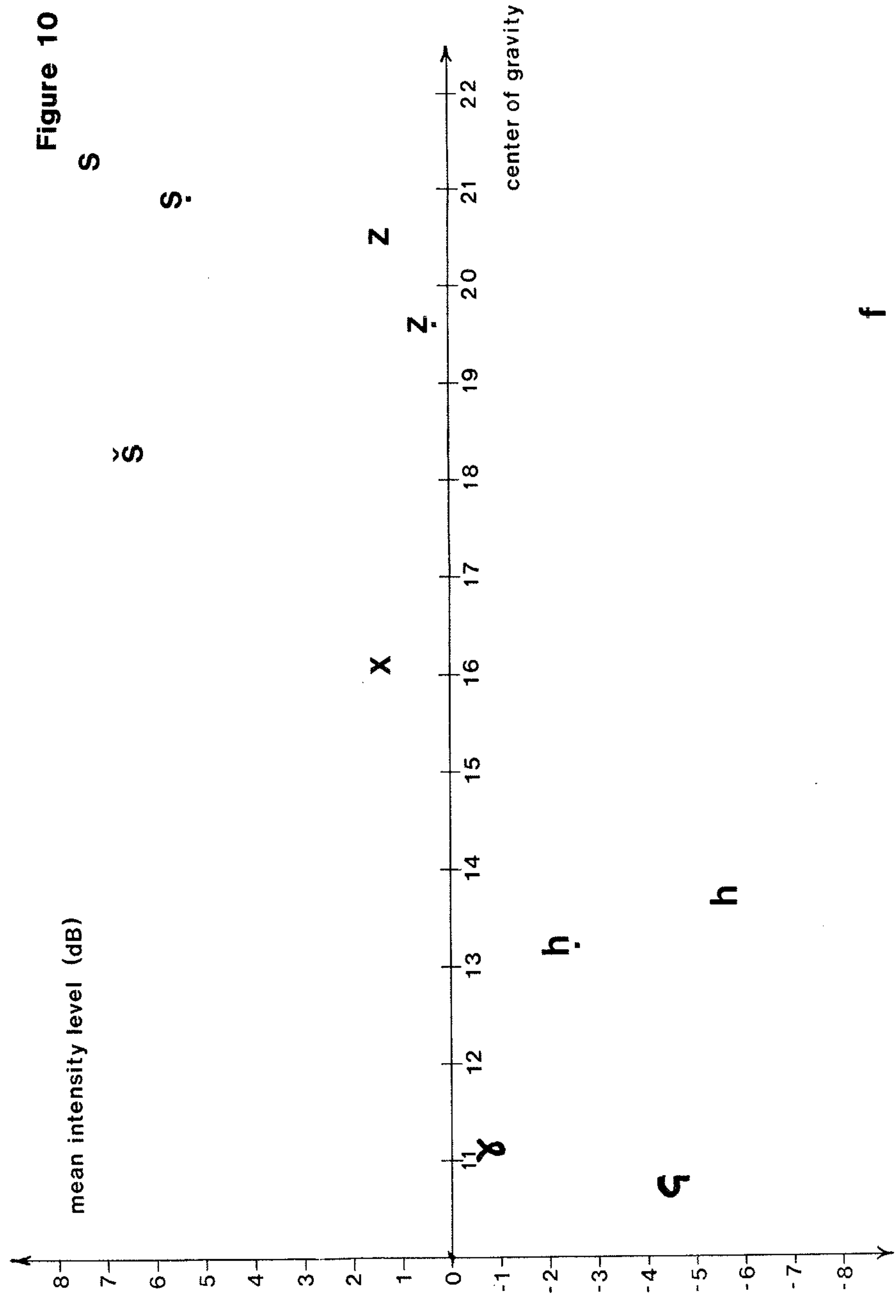

dispersion
center of gravity
kHz
5
4
3
2
1,5
1
9
10
11
12
13
14
15
16
17
18
19
20
21
1,5
2
3
4
5
6
x
h
z
s

Figure 10


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

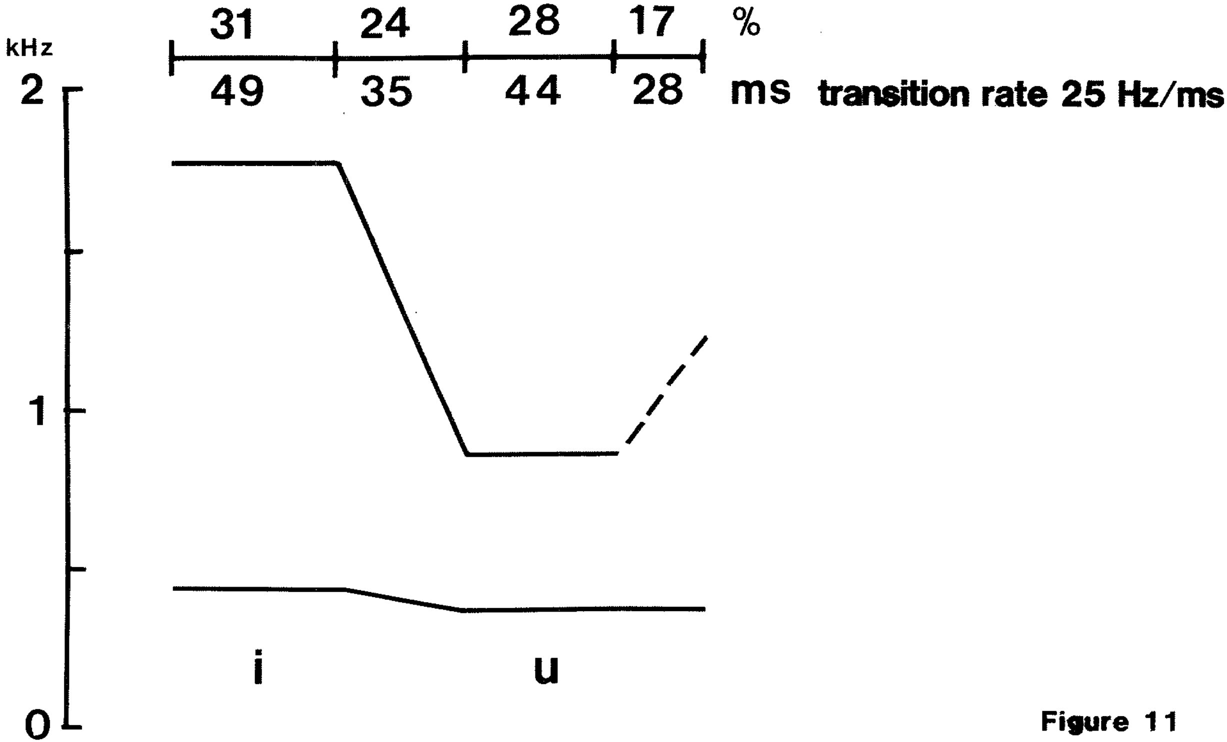


Figure 11

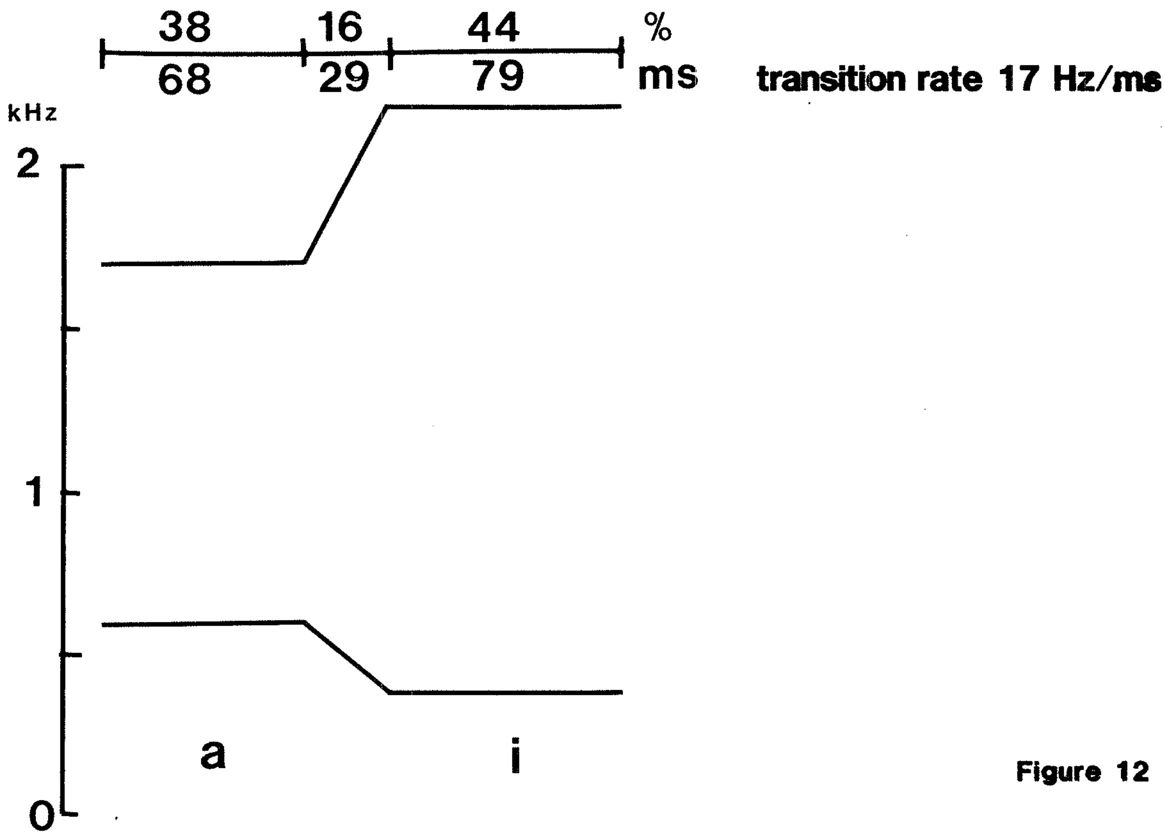


Figure 12

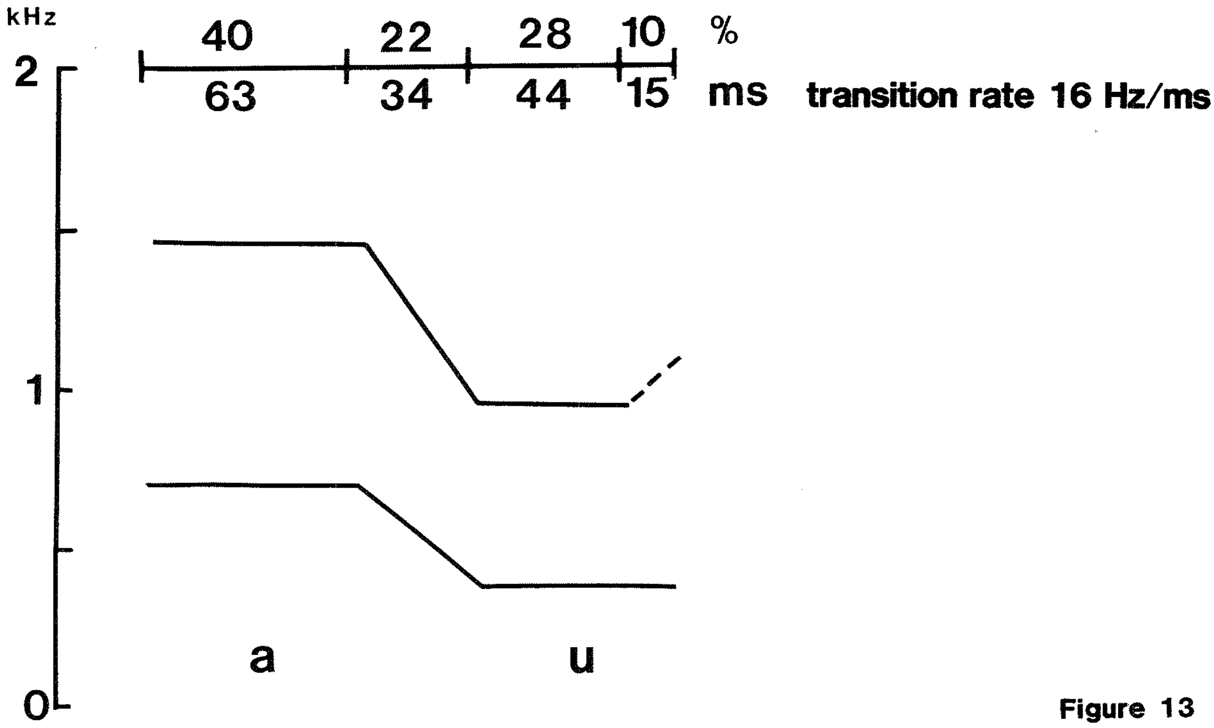


Figure 13

Vowels and Diphtongs in Standard Chinese

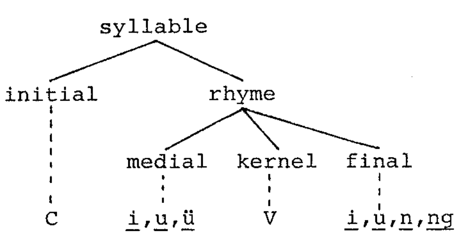
Jan-Olof Svantesson

In this article, the acoustic properties of Standard Chinese (pǔtōnghuà) vowels and diphthongs are described. This is one of the most interesting areas of Chinese phonetics, since there are only five monophthongic vowel phonemes, which form an unusual system, but as many as eleven diphthongs, and also two triphthongs. The diphthongs exemplify different types of timing of steady states and transitions between them, and it will be seen that not only the formant frequencies of the steady states and their relation to the vowel goals, but also the timing of the transitions betweeen the steady states is important, and differs between different Chinese diphthongs and also differs from the "same" diphthong in other languages.

Pīnyīn spelling (underlined) is used throughout, except in the section on phonology, where a more phonemic transcription is sometimes used (within /.../).

1. PHONOLOGY

A Standard Chinese syllable can be analyzed into an initial consonant and a rhyme. The rhyme has a kernel vowel which can be preceded by one of the medials i, u or ü, and followed by a final, which is either one of the vowels i or u, or one of the consonants n or ng:



(In traditional Chinese phonology, the medial is not considered a part of the rhyme.)

Because of the large amount of interaction between the vowels and both the preceding and the following consonant (if any), it is possible to analyze the phoneme system in several different ways, and this has also been done, see e.g. Chao 1934, Hartman 1944, Hockett 1947, Cheng 1973. In particular, the phonemic status of [ɿ] and [ʅ], i.e. if they are the allophones of a separate phoneme, or are allophones of /i/ (as assumed here) has been analyzed differently by different authors.

Here the following vowel phoneme system will be assumed:

i	ü			u	
			e		
			a		

The vowel /i/ has the allophone [ɿ] after dental sibilants (s, z [ts] and c [tsh]), the allophone [ʅ] after postdental sibilants (sh [ʂ], zh [tʂ], ch [tʂh] and r [ʐ]), and is otherwise [i].

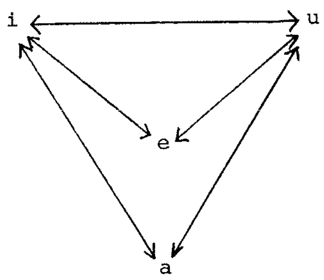
There is no contrast between (phonetically) different mid vowels, so they will be regarded as allophones of a phoneme written /e/. It has the allophone [ɤ] as a single-vowel rhyme (written e in the pīnyīn spelling),but in other rhymes it has allophones ranging from [o] to [e].

In the pīnyīn spelling there is a vowel o, which occurs only after labial consonants (b, p, f and m). Acoustically, o is very similar to the diphthong uo, which is in complementary distribution with o, so o will be regarded as a notational variant of uo. o is also written in the diphthongs ou, uo and ao, which are phonemicized as /eu/, /ue/ and /au/ (see below).

The following diphthongs and triphthongs occur:

iu	/iu/	ui	/ui/
ia	/ia/	ai	/ai/
ua	/ua/	ao	/au/
ie	/ie/	ei	/ei/
uo	/ue/	ou	/eu/
üe	/üe/		
iao	/iau/	uai	/uai/

The system of diphthongs is rather symmetrical, and with the exception of üe all the diphthongs can be obtained by going from one of the four vowels /i/, /u/, /e/ or /a/ to any other (except that */ea/ and */ae/ are not found). Also the triphthongs /iau/ and /uai/ are symmetric to each other:



There is also a syllable consisting of the vowel er [ɚ], which is usually analyzed as /er/. In the regular syllable inventory (as written by Chinese characters), there is only this single syllable (in three different tones) with the final r, and this rhyme cannot be preceded by an initial consonant. It can be added, however, to other syllables as a suffix, with the phonetic result of an r-colouring of the syllable, with somewhat different effect on different rhymes. It is not entirely clear if this "erization" (érhuà) is a feature of Standard Chinese, even though it is a common feature of Běijīng pronunciation, since there is a tendency to regard erization as a vulgarism and to avoid it in Standard Chinese. Erization will not be treated in this article.

The following rhymes occur:

i			in	ing
ü			ün	
u	ui		un	ong /uŋ/
iu				iong /iuŋ/
e	ei	ou /eu/	en	eng
ie				
üe				
uo /ue/				
a	ai	ao /au/	an	ang
ia		iao /iau/	ian	iang
			üan	
ua	uai		uan	uang
er				

2. PROCEDURE

Four speakers of Standard Chinese were recorded. Two of the speakers (B and C) were born and raised in Běijīng, one (A) was born in Sūzhōu and moved to Běijing when he was six years old, and one (D) is from Liáoníng and has lived in Běijīng since he was 12.

For each speaker, syllables containing each rhyme were recorded in a sentence frame (wǒ bǎ ___ zì xiě hǎo), and each sentence was read twice. The syllable initial was chosen as a dental (d when possible), and the syllables were in the high (first) tone whenever possible, and otherwise in the rising (second) tone.

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

For each syllable, wide-band spectrograms were made on a Kay Digital Sona-Graph 7800. The frequencies of the first three formants, and also the durations of the vowels were measured on the spectrograms.

3. FORMANT FREQUENCIES

Formant frequencies of Standard Chinese vowels have also been published by Howie 1976 (for one speaker), Brotzman 1963 (reported by Howie), and Wú and Cáo 1979 (showing only charts of average F1 and F2 values).

The formant frequencies as measured in the middle of monophthongic vowels in the context C___# are given in Table 1.

The formants of the five main allophones of the vowel phonemes (i [i], ü [y], u [u], a [a] and e [ɤ]) are plotted on Figure 1, and the formants of er [ɚ] and the /i/ allophones [ɿ] and [ʅ] are plotted on Figure 2.

The vowels [ɿ] and [ʅ] are usually described as vocalic [z̩] and [ʐ̩]. According to Cheng 1973:13, X-ray studies by Zhōu and Wú 1963 (not available to me) show that compared to [i], the highest point of the tongue is slightly more front and the back of the tongue is slightly higher for these vowels. (The non-IPA symbols [ɿ] and [ʅ], which are generally used in Chinese linguistics were introduced by Bernhard Karlgren, who took [ʅ] from the Swedish dialect alphabet, where it denotes the "Viby i" occurring in Swedish dialects. This alphabet was widely used in Swedish dialectology, and its main inventor, J.A. Lundell, was Karlgren's teacher.)

The vowel pairs i and ü and e and er do not differ much in F1 or F2, but are clearly separated by F3, the second member of each pair having much lower F3 than the first.

The first two formants of vowels before nasals (i.e. in the contexts C___n and C___ng) are given in Table 2 and on Figure 3. The main differences as compared to open-syllable vowels are: i is lowered in nasal contexts, e and a are fronted before n, and u is considerably lowered before ng (where it is written o in the pīnyīn spelling) and fronted-lowered before n.

For the diphthongs and triphthongs, the first two formants for each steady state in the spectrograms were measured, as well as the duration of each steady state and the duration of the transition between the steady states. The formant frequencies were measured in the middle of each steady state.

These results are shown in Table 3 (diphthongs) and Table 4 (triphthongs). Steady state formant frequencies and duration data for diphthongs before a nasal (n or ng) are given in Table 5.

In Figure 4, schematic drawings of average diphthong and triphthong formant frequency movements are shown on a F1-F2-diagram.

The endpoints of diphthongs which do not involve the phoneme /e/ are rather close to the respective vowel phoneme average (represented by a star on Figure 4), while the startpoints differ more, so that for instance ao /au/ and ai start from positions higher than a, and ui starts from a (acoustically) much more central position than u. Diphthongs which contain the phoneme /e/ (realized monophthongically as [ɤ]), i.e. ie, üe, ei, ou /eu/ and uo /ue/ contain [e] ~ [ɛ] or [o]-like allophones of /e/.

As Figure 4:2 shows, the final a component of the diphthongs ua and ia is much fronter before the nasal n than before ng [ŋ].

4. DURATIONS

In Standard Chinese, there is no phonemic length distinction for vowels, but there has been some discussion in the literature about vowel quantity, in the context of tonal phonology. Woo 1969 represents contour tones (e.g. three out of four Standard Chinese tones) as sequences of level tones, and this presupposes that contour tones are assigned to sequences of more than one voiced segment. This causes no problem for rhymes which consist of diphthongs, triphthongs or a vowel followed by a nasal, but for monophthongic vowels in open syllables it means that they must be represented as a cluster of two identical vowels. To justify this, Woo presents acoustical data which shows that vowels are longer in the context C___# than when followed by a nasal or when included in a diphthong, and says that "It is generally assumed that all pure vowels are normally long, and that vocalic clusters, which are diphthongs, consist of two "short" members" (Woo 1969:25). Walton 1983:174 doubts that there is such a general agreement, but their discussion concerns phonological interpretation rather than the physical properties of the sounds.

Also this investigation shows that the different components of a diphthong, and also vowels before nasals, are shorter than single vowels in open syllables (see Figure 5).

On the other hand, diphthongs and vowel-nasal rhymes are in most cases longer than single-vowel rhymes. Thus Woo's statement (1969:27) that "the duration of the syllabic nucleus appears to be a constant also, irrespective of whether it is a long vowel [i.e. a single vowel in an open syllable], a diphthong, or a vowel + nasal cluster" is not confirmed by this study (Woo's data came from syllables said in isolation, however).

It is well-known that the duration of Standard Chinese rhymes is dependent on the tone of the syllable (see e.g. Kratochvîl 1968, Woo 1969:24-30), and thus both the tone and the segmental composition affect the duration of a rhyme. A preliminary investigation (Nordenhake and Svantesson 1983) shows that the effects of the different tones on the duration vary with the position of a syllable within a sentence, so that for instance the falling (fourth) tone has the shortest duration of all tones in sentence final position, while it is the longest tone in sentence medially.

In this investigation, high-tone syllables have been used whenever possible. (In a few cases, syllables with rising (second) tone were used; duration data from such syllables are marked with a star in the tables and figures, since they are not comparable with the other (high tone) data.) The question how the tones affect the duration and the vowel quality - especially the quality of the diphthongs seem to be somewhat dependent on the tones - will thus not be taken up here, but will be made the subject of a special study.

Figure 5 shows average duration values for all speakers. The durations of open syllable vowels are given in Table 1, and in Table 2, durations of vowels in rhymes with final nasal are given, together with the duration of the nasal. For monophthongs followed by a nasal, the vowel is generally shorter than the nasal, and also shorter than the same vowel in an open syllable, but also here the duration of the entire rhyme is longer than for an open syllable.

For the diphthongs of a language, not only the goal values and the way the start and end values of the diphthong relates to these goals (which are here assumed to be vowel phonemes of the language) are important, but also the dynamics of the diphthong, i.e. the way the formant frequencies change with time. This can be quantified in different ways; the way chosen here is to measure the formant frequencies of the steady states and the durations of the steady states and of the transition between them, and to calculate the ratio between the transition duration and the total duration. (It would also be possible to calculate the velocity with which the formant frequencies (especially F2) change during the transition.) These data are given in Table 3, and are plotted on Figure 6.

This kind of analysis reveals differences between the "same" diphthong in different languages, e.g. [ai] in Standard Chinese, Hausa and Arabic (these two languages have been analyzed with the same methods as used for Chinese). In Hausa (data from Mona Lindau) and Arabic (Norlin 1984), these diphthongs can be regarded as a succession of two vowels [a] and [i], which are nearly identical to the short [a] and the [i] of the respective language, both as regards quality (formant structure) and quantity (duration). Thus, a speaker of Hausa or Arabic first makes an [a], then goes quickly to [i] and produces that vowel. So there are two steady states, each with about the same length as a short vowel, and a short transition in between.

In Chinese, this diphthong is more gliding, with relatively short steady states, and a long transition (average ratio of transition to total duration is 50.5% for this diphthong). Furthermore, the total duration of a diphthong is usually longer than that of a monophthongic vowel (see Figure 6), but not about twice as long (as is the case in Arabic and Hausa).

ACKNOWLEDGEMENT

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

REFERENCES

Brotzman, Robert. 1963. "Research on Mandarin phonology: Vowel formant values". Project on Linguistic Analysis, Report No. 6, 7-18. Columbus: Ohio State University Research Foundation.

Chao Yuen-Ren. 1934. "The non-uniqueness of phonemic solutions of phonetic systems". Bulletin of the Institute of History and Philology (Academia Sinica) 4, Part 4, 363-97. Also in Joos, ed. 1957, 38-54.

Cheng Chin-chuan. 1973. A synchronic phonology of Mandarin Chinese. The Hague: Mouton.

Hartman, Lawton. 1944. "The segmental phonemes of the Peiping dialect". Language 20, 28-42. Also in Joos, ed. 1957, 116-23.

Hockett, Charles 1947. "Peiping phonology" . Journal of the American Oriental Society 67, 253-67. Also in Joos, ed. 1957, 217-228.

Howie, John. 1976. Acoustical studies of Mandarin vowels and tones. Cambridge: University Press

Joos, Martin, ed. 1957. Readings in linguistics. Washington: American Council of Learned Societies.

Kratochvíl, Paul. 1968. The Chinese language today. London: Hutchinson.

Nordenhake, Magnus and Jan-Olof Svantesson. 1983. "Duration of Standard Chinese word tones in different sentence environments". Working Papers (Department of Linguistics, Lund), 25, 105-11.

Norlin, Kjell. 1984. "Acoustic analysis of vowels and diphthongs in Cairo Arabic". In this volume.

Walton, Ronald. 1983. Tone, segment and syllable in Chinese: A polydimensional approach to surface phonetic structure. Ithaca: Cornell University China-Japan Program.

Woo, Nancy. 1969. Prosody and phonology. Ph.D. diss., MIT. Reproduced by Indiana University Linguistics Club 1972.

Wú Zōngjì and Cáo Jiànfēn. 1979. "Shíyàn yǔyīnxué zhīshí jiǎnghuà [Lectures on experimental phonetics]". Zhōngguó yǔwén 1979:4, 314-20.

Zhōu Diànfú and Wú Zōngjì. 1963. Pǔtōnghuà fāyīn túpǔ.[Spectrograms of Standard Chinese pronunciation]. Běijīng.

Table 1. Formant frequencies and duration of Standard Chinese monophthongic vowels in the context C ___ #.

Vowel		Speaker	F_1	F_2	F_3	duration (ms)
i	[i]	A	200	2370	3430	175
			240	2400	3400	165
		B	220	2040	2960	180
			340	2320	3270	170
		C	240	1800	3360	150
			400	1830	3390	195
		D	200	2420	3600	185
			230	2360	3510	200
		mean	259	2192	3365	177
ü	[y]	A	210	2140	2580	150
			220	2220	2490	160
		B	270	2150	2340	140
			460	2070	2630	170
		C	360	1820	2450	180
			380	1900	2670	155
		D	220	2200	2510	150
			220	1890	2340	175
		mean	292	2040	2501	160
u	[u]	A	360	810	2460	150
			240	760	2730	150
		B	430	640	2430	165
			330	720	2610	150
		C	240	940	2280	140
			310	700	2620	145
		D	450	760	2750	150
			280	760	2720	150
		mean	330	761	2575	150
a	[a]	A	770	1200	2530	145
			930	1290	2600	185
		B	770	1180	2360	190
			930	1340	2620	190
		C	650	1340	2640	155
			960	1500	2530	120
		D	860	1370	2800	195
			920	1450	2810	185
		mean	849	1334	2611	171

Table 1 (cont.)

Vowel	Speaker	F_1	F_2	F_3	duration
e [ɤ]	A	340	1170	2550	220*
		330	1130	2600	235*
	B	510	1080	2500	205*
		500	1120	2570	225*
	C	380	1360	2310	160*
		380	1430	2200	190*
	D	500	1260	2580	240*
		480	1400	2560	215*
	mean	428	1244	2484	211*
er [ɚ]	A	400	1480	1890	225*
		500	1480	1820	335*
	B	490	1420	1750	250*
		600	1380	1760	280*
	C	430	1430	1710	195*
		440	1370	1760	240*
	D	440	1320	1630	435*
		450	1340	1720	330*
	mean	469	1402	1755	261*
i [ɿ]	A	240	1160	2700	85
		270	1170	2800	120
	B	370	1200	2710	155
		420	1210	2790	190
	C	400	1240	2620	290
		440	1380	2700	145
	D	490	1220	2600	135
		480	1280	2620	140
	mean	389	1232	2692	140
i [ʅ]	A	430	1750	2300	115
		280	1970	2510	130
	B	480	1690	2510	155
		450	1710	2580	170
	C	470	1600	2620	135
		440	1590	2760	130
	D	510	1710	2220	140
		510	1700	2470	125
	mean	446	1715	2496	137

Table 2. Formant frequencies and durations of monophthongic vowels before nasals.

Rhyme	Speaker	F_1	F_2	Duration (ms) vowel	nasal	total
<u>in</u>	A	260	2200	110	150	260*
	B	470	2230	100	155	255*
	C	380	1900	90	140	230*
	D	240	2400	130	115	245*
	mean	337	2182	107	140	247*
<u>ün</u>	A	260	2110	95	175	270
	B	450	1880	100	150	250
	C	380	1800	80	145	225
	D	240	2050	95	125	220
	mean	332	1960	92	149	241
<u>un</u>	A	240	1080	75	190	265
	B	440	1130	80	165	245
	C	350	1100	85	155	240
	D	500	1150	95	130	225
	mean	382	1115	84	160	244
<u>en</u>	A	490	1500	80	155	235
	B	570	1520	75	160	235
	C	490	1440	65	135	200
	D	680	1720	75	135	210
	mean	557	1545	74	146	220
<u>an</u>	A	820	1610	135	145	280
	B	840	1420	110	130	240
	C	750	1550	100	100	200
	D	870	1590	145	95	240
	mean	820	1542	122	117	240
<u>ing</u>	A	450	2230	110	145	255
	B	410	2310	70	140	210
	C	330	2140	95	135	230
	D	460	2320	105	120	225
	mean	412	2250	95	135	230
<u>ong</u>	A	480	890	55	155	210
	B	430	760	45	160	205
	C	520	830	100	140	240
	D	490	780	85	125	210
	mean	480	815	71	145	216
<u>eng</u>	A	500	1410	80	165	245
	B	430	1200	85	170	255
	C	520	1470	90	150	240
	D	470	920	70	165	235
	mean	480	1250	81	162	244

Table 2 (cont.)

Rhyme	Speaker	F_1	F_2	Duration vowel	nasal	total
ang	A	830	1310	125	145	270
	B	830	1270	120	130	250
	C	670	1200	115	140	255
	D	900	1340	100	150	250
	mean	807	1280	115	141	256

Table 3. Formant frequencies and durations of steady states in diphthongs in the context C__#

	Speaker	F_1	F_2	F_1	F_2	Duration[1] t_1	t_2	t_3	tot.
iu	A	220	2330	250	740	55	60	75	240
	B	410	2080	410	820	30	60	80	210
	C	370	1930	410	720	45	55	75	175
	D	490	2360	500	830	45	60	85	245
	mean	372	2175	392	778	44	59	79	216
ui	A	250	1480	270	2340	70	25	100	210
	B	350	1530	330	1900	40	30	70	165
	C	390	1260	480	2050	40	50	50	190
	D	460	1400	480	2160	50	35	90	205
	mean	362	1418	390	2112	50	35	78	192
ia	A	410	2050	800	1350	40	90	100	235
	B	350	1900	860	1280	20	65	100	210
	C	410	1910	700	1260	25	65	100	190
	D	400	2270	900	1890	20	80	130	230
	mean	392	2032	815	1445	26	75	107	216
ai	A	830	1710	240	2330	45	110	20	185
	B	710	1700	420	1840	50	100	50	205
	C	660	1720	400	1960	30	100	50	180
	D	810	1760	470	2300	70	65	30	170
	mean	752	1722	382	2108	49	94	38	185
ua	A	480	1040	720	1320	90	20	125	245
	B	460	910	710	1180	40	50	110	200
	C	480	900	830	1310	45	35	110	210
	D	450	850	980	1320	50	20	170	245
	mean	468	925	810	1282	56	31	129	225
ao	A	640	1160	400	770	55	55	80	235
	B	690	1120	520	820	120	30	50	235
	C	520	890	350	800	70	50	60	200
	D	880	1200	480	850	120	25	60	235
	mean	682	1092	438	810	91	40	62	226
ie	A	220	2300	550	2050	95	15	30	160
	B	350	2200	620	1880	95	20	40	160
	C	290	1920	540	1640	55	30	45	150
	D	210	2470	500	2170	80	20	50	180
	mean	268	2222	552	1935	81	21	41	162
ei	A	420	1650	320	2270	25	115	20	180*
	B	500	1760	340	2050	20	60	90	170*
	C	- - - -							
	D	- - - -							
	mean	460	1705	330	2160	22	87	55	175*
uo	A	480	770	550	1110	120	30	45	205
	B	420	810	620	1220	110	25	30	215
	C	260	720	520	1080	100	20	50	190
	D	420	720	470	920	110	20	50	230
	mean	395	755	540	1082	110	24	44	210

1. t_1 = duration of first steady state; t_2 = duration of transition between the steady states; t_3 = duration of second steady state.

Table 3 (cont.)

	Speaker	F_1	F_2	F_1	F_2	Duration t_1	t_2	t_3	tot.
<u>ou</u>	A	480	1000	280	800	35	20	115	200
	B	360	950	350	780	25	20	120	200
	C	500	1250	320	1090	50	40	80	175
	D	540	1090	420	780	45	25	100	205
	mean	470	1072	342	862	39	26	104	195
<u>üe</u>	A	260	1940	510	1850	60	15	60	140
	B	450	1920	580	1700	50	20	80	160
	C	270	1840	520	1610	55	25	65	145
	D	300	2250	500	1740	60	25	85	170
	mean	320	1987	527	1725	56	21	73	154

Third formants of the diphthongs <u>ie</u> and <u>üe</u>:

		<u>ie</u>		<u>üe</u>	
Speaker:	A	3070	2670	2240	2600
	B	2790	2680	2280	2520
	C	2710	2460	2210	2420
	D	3210	2730	2500	2640
	mean	2945	2635	2307	2545

Table 4. Formant frequencies of triphthong steady states.

	Speaker	F_1	F_2	F_1	F_2	F_1	F_2	duration
<u>iao</u>	A	420	2270	440	1250	440	900	250
	B	350	1940	590	1080	510	940	250
	C	390	1840	520	1240	400	980	225
	D	430	2460	640	1400	520	980	260
	mean	398	2128	548	1242	468	950	246
<u>uai</u>	A	360	1270	620	1720	370	2180	205
	B	390	1230	620	1700	400	1930	195
	C	480	1350	590	1680	500	1850	180
	D	470	1180	740	1510	480	1970	220
	mean	425	1257	642	1652	437	1982	200

Table 5. Formant frequencies and durations for diphthongs before nasals.

Rhyme	Speaker	F_1	F_2	F_1	F_2	duration vowel	duration nasal	duration total
<u>ian</u>	A	290	2320	900	1690	185	115	300
	B	310	2200	630	1750	120	130	250
	C	370	1900	670	1650	130	105	235
	D	460	2390	500	1920	185	90	275
	mean	357	2202	675	1752	155	110	265
<u>üan</u>	A	310	2080	500	1720	165	110	275
	B	300	1820	550	1330	145	110	255
	C	450	1800	730	1620	130	115	245
	D	280	2100	630	1630	170	90	260
	mean	335	1950	602	1575	152	106	259
<u>uan</u>	A	450	910	690	1410	155	110	265
	B	380	1000	580	1390	135	140	275
	C	500	1130	630	1510	130	130	260
	D	430	930	700	1320	185	85	270
	mean	440	992	650	1407	151	116	267
<u>iong</u>	A	250	2110	260	1500	100	135	235
	B	420	1710	400	940	70	150	220
	C	330	2100	330	820	65	140	205
	D	250	2160	270	980	90	120	210
	mean	312	2020	315	1060	81	136	217
<u>iang</u>	A	430	2280	720	1180	140	110	250*
	B	600	2170	860	1240	120	125	245*
	C	540	2100	700	1110	150	110	260*
	D	290	2360	850	1370	140	100	240*
	mean	465	2227	782	1225	137	111	249*
<u>uang</u>	A	500	960	730	1090	115	135	250
	B	370	930	560	1180	120	140	260
	C	510	840	620	1110	110	135	245
	D	490	910	700	1280	125	115	240
	mean	467	910	652	1165	117	131	249

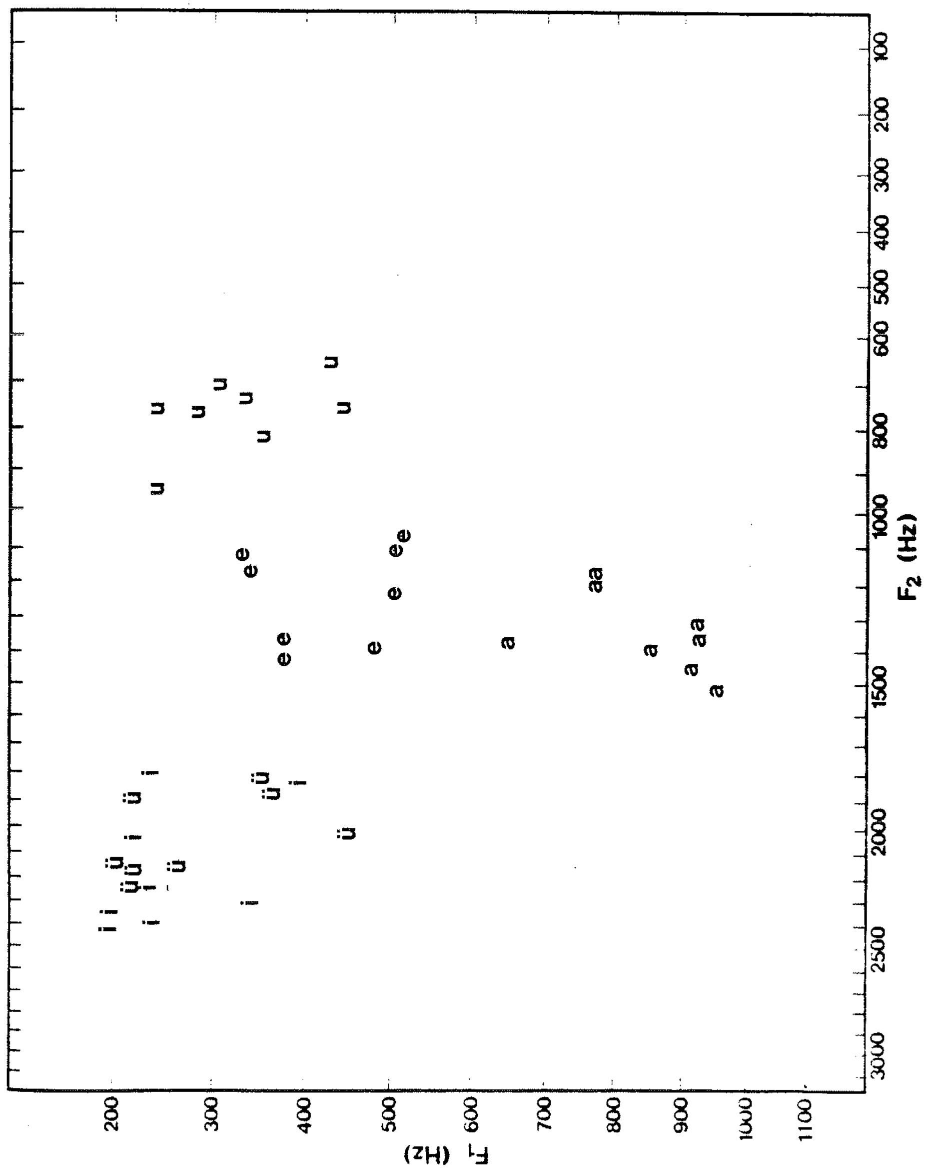


Figure 1:1. F_1-F_2-diagram for the five vowel phonemes.

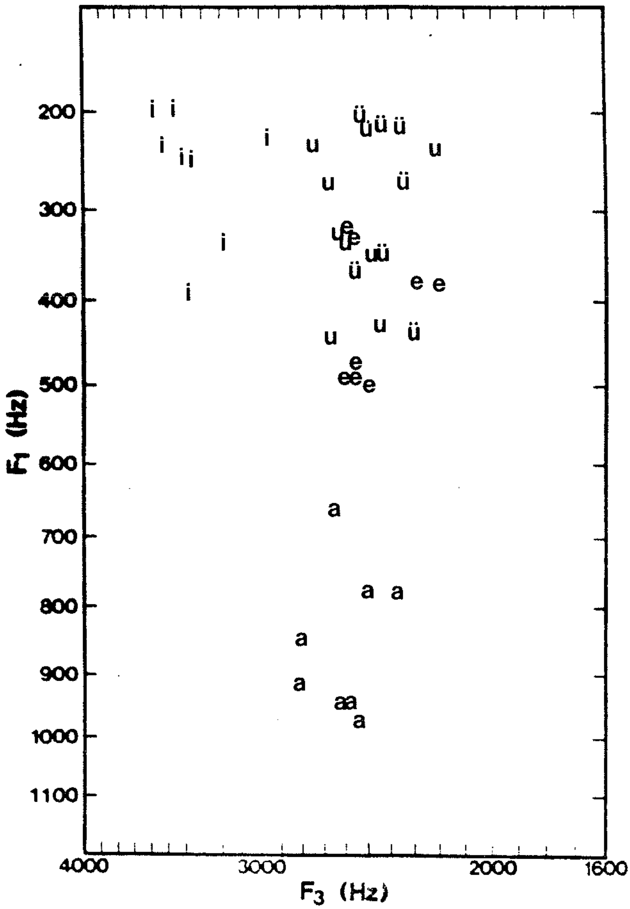


Figure 1:2. F_1-F_3-diagram for the five vowel phonemes.

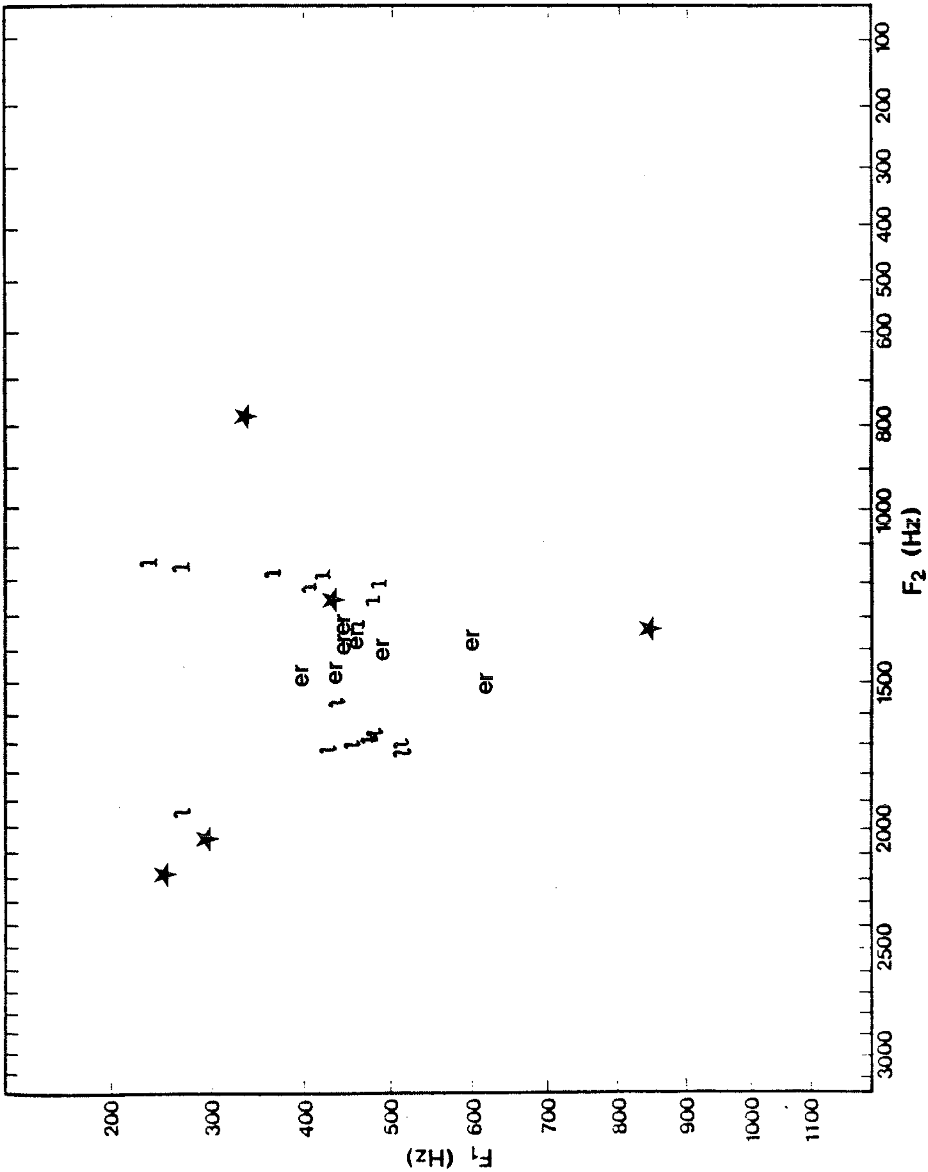


Figure 2:1. F_1-F_2-diagram for [ɚ], [ɭ] and [ʅ]. The stars represent the averages for the five main allophones of the vowels.

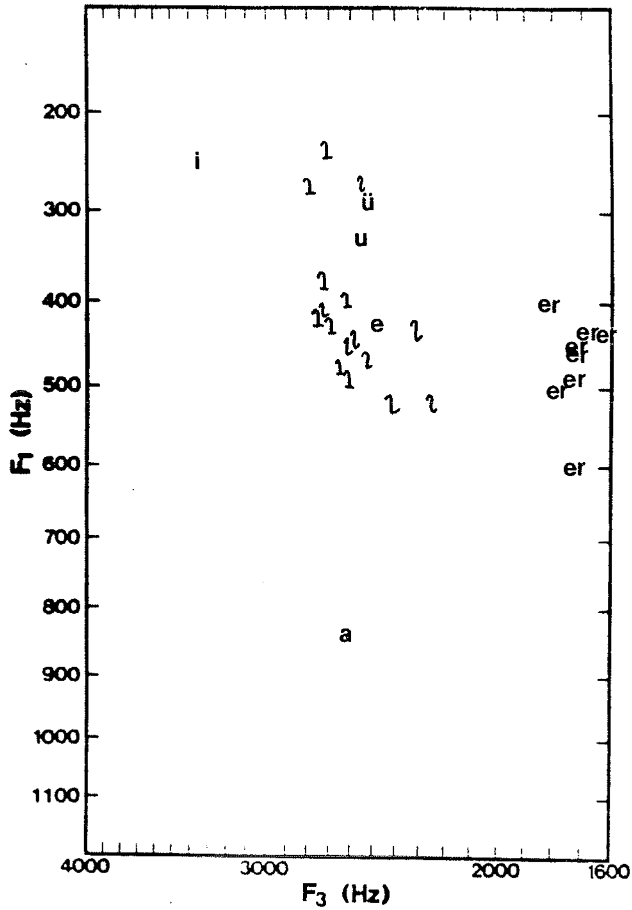


Figure 2:2. F_1-F_3-diagram for [ɚ], [ɿ] and [ʅ]. The letters i, ü, u, e and a represent average formant values for these vowels in open syllables.

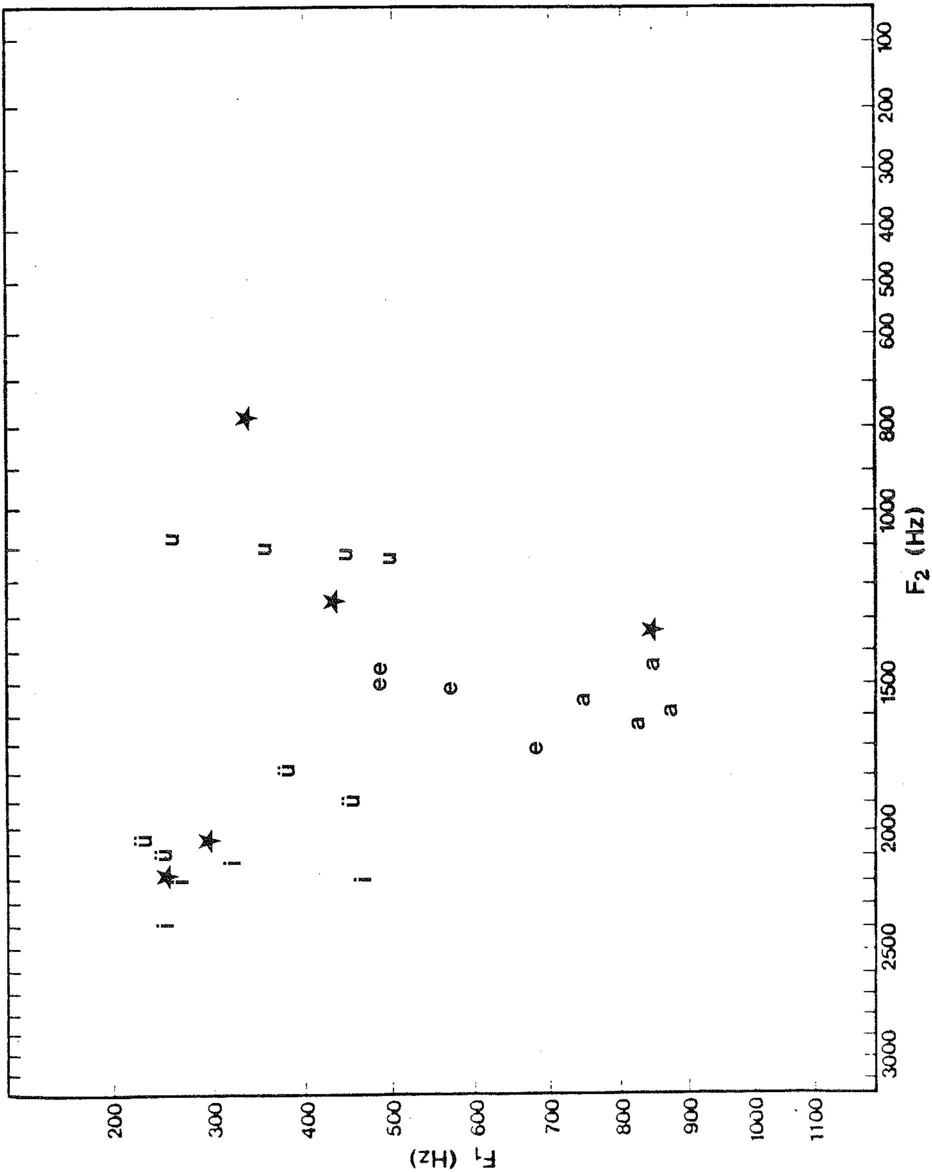


Figure 3:1. F_1-F_2-diagram for vowels before n. The stars represent average formant values for vowels in open syllables.

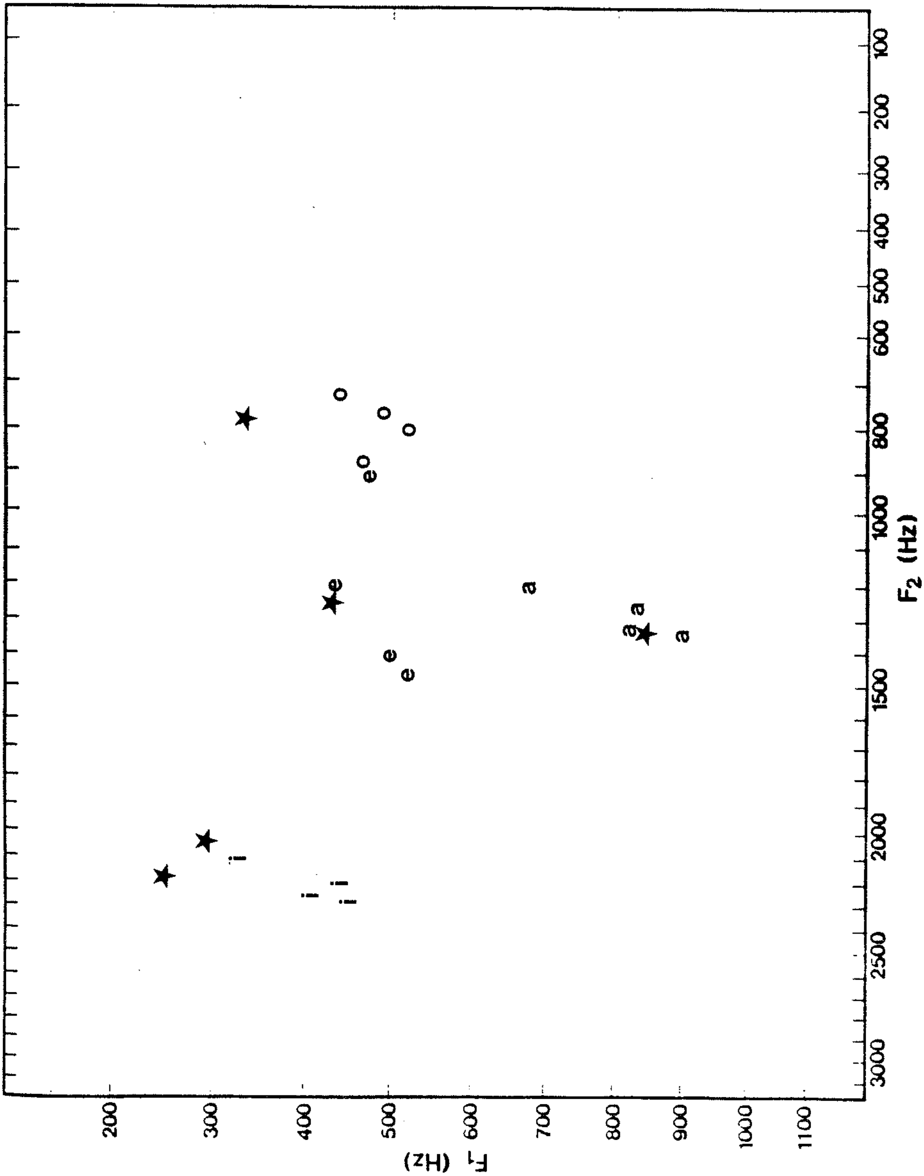


Figure 3:2. F_1-F_2-diagram for vowels before ng.

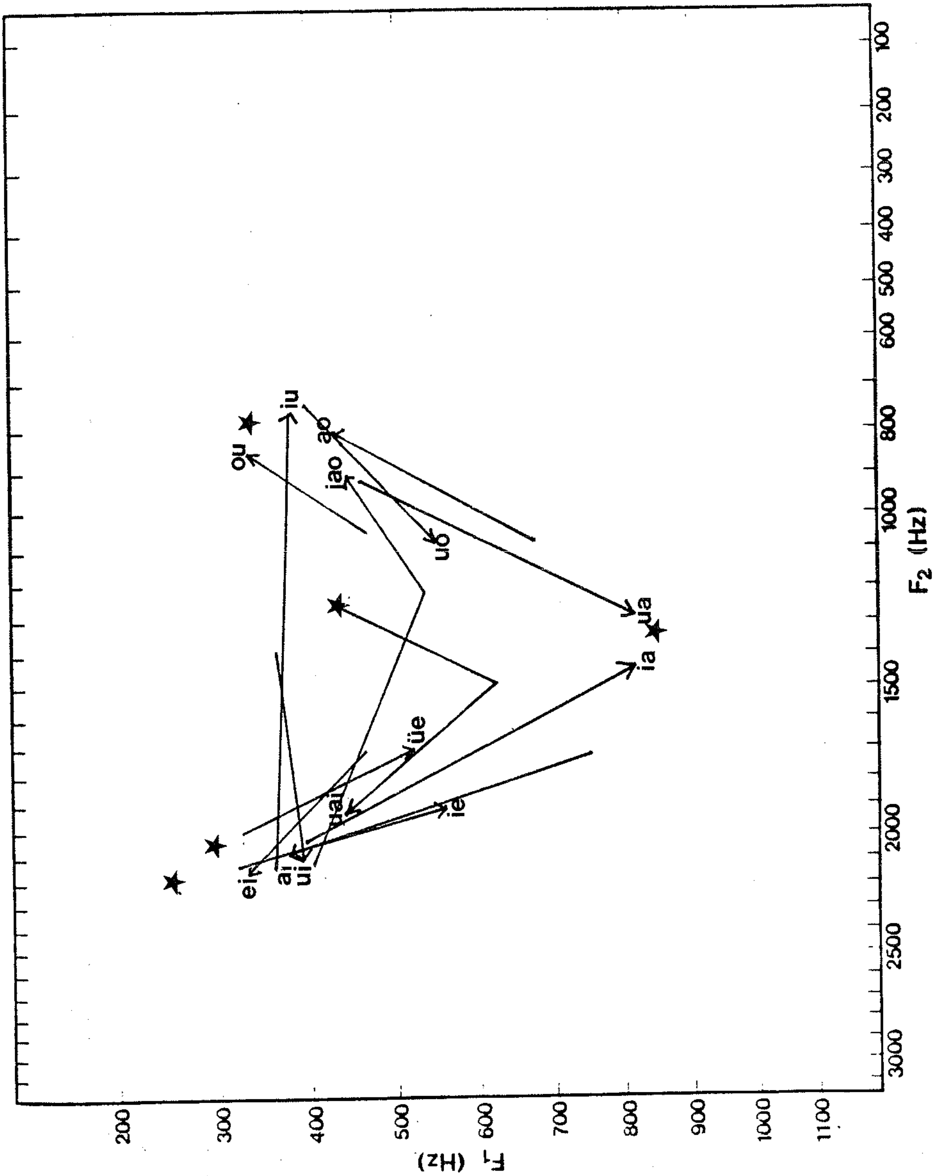


Figure 4:1. Diphthongs before Ø.

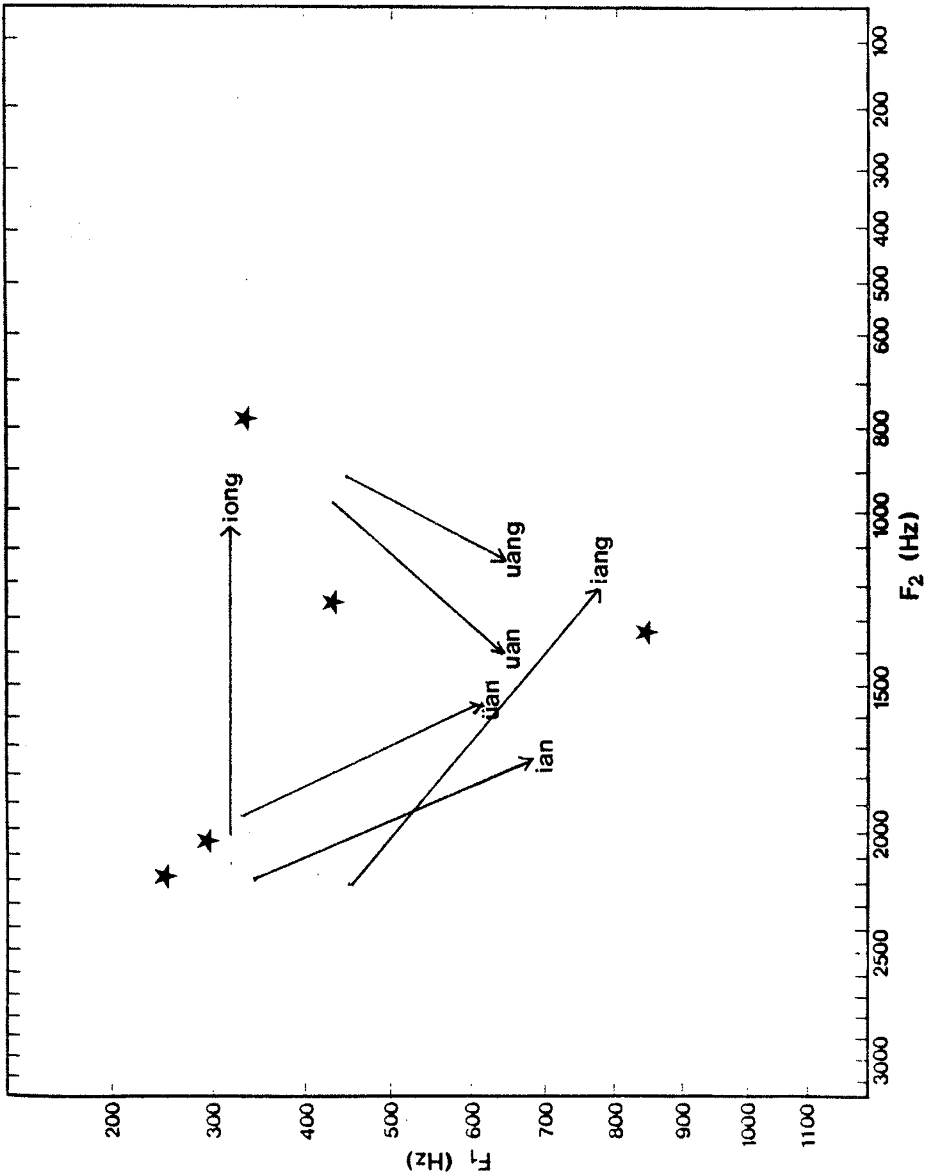


Figure 4:2. Diphthongs before nasals.

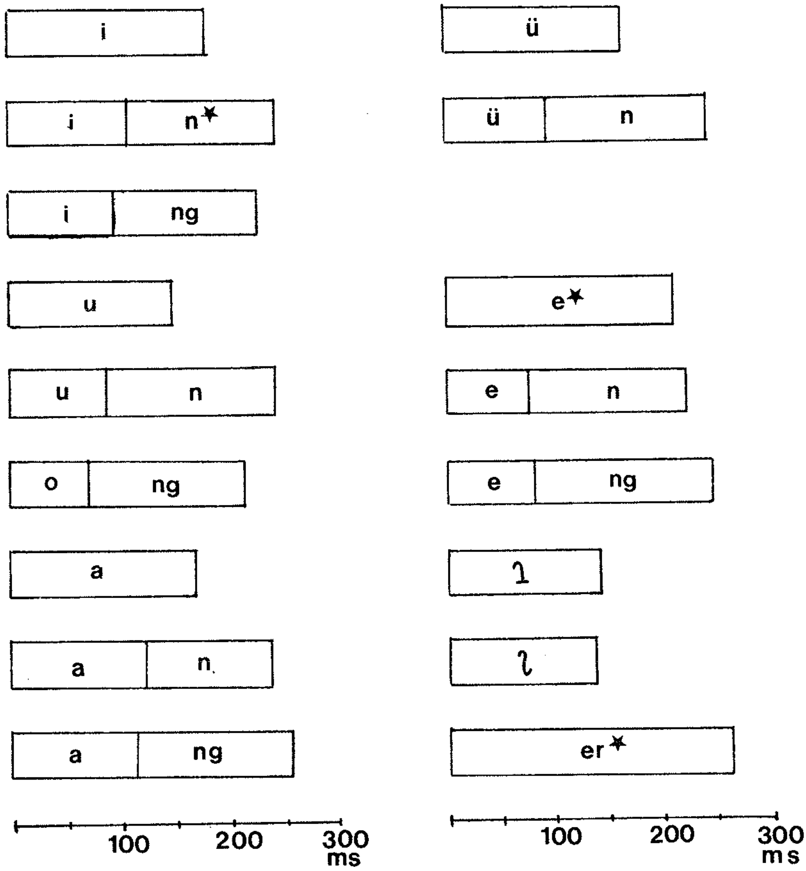


Figure 5:1. Duration of monophthongs.

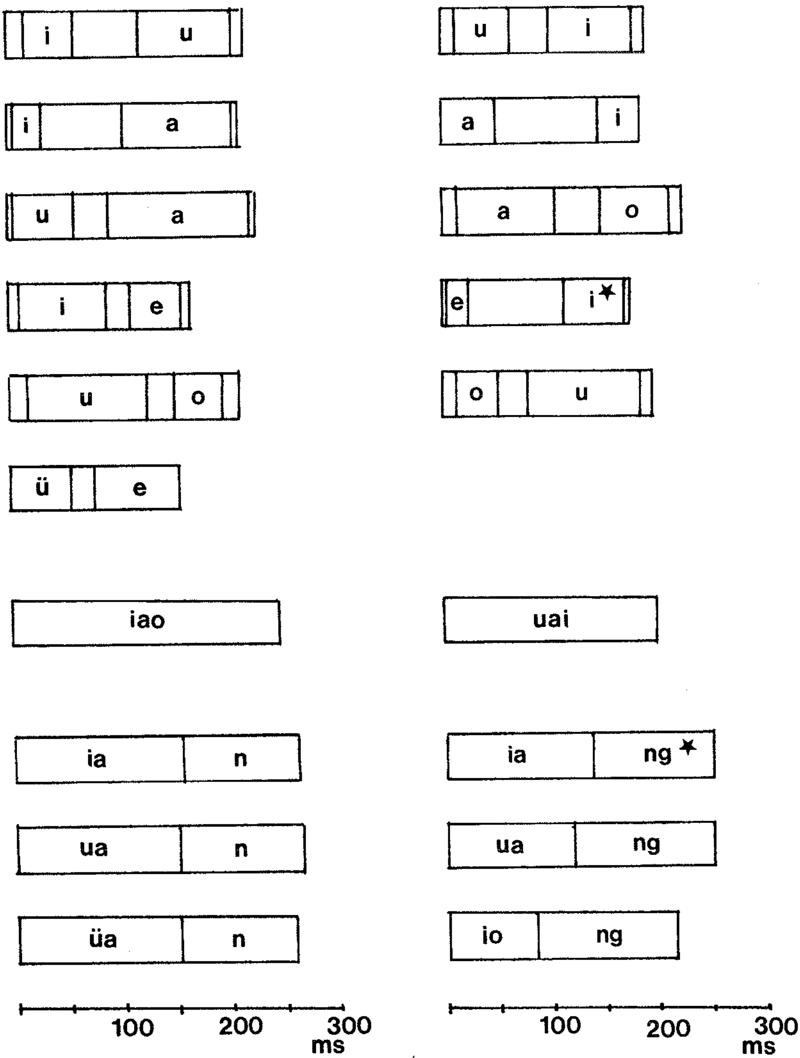


Figure 5:2. Duration of diphthongs and triphthongs.

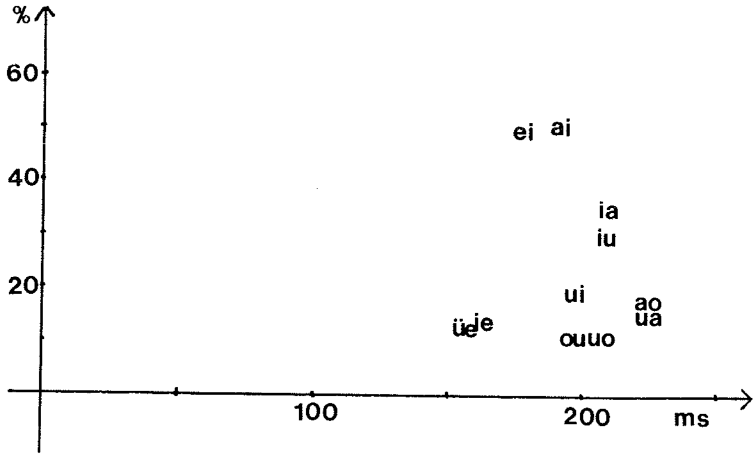


Figure 6. Transition percentage plotted against total duration for the diphthongs.

Perceptual Studies in Chinese Intonation Using Foreign Listeners

Jiālü Zhāng

At the Department of Linguistics and Phonetics, Lund University, studies of tone and intonation in Standard Chinese have been carried out by using both natural and synthetic speech. The declination as an intrinsic feature of speech pitch has been investigated by Cohen et al. 1982. In this paper, synthetic speech is used to investigate how the baseline and the topline of the pitch contour influence the perception of sentence intonation and how focus is expressed in a sentence.

Using the LPC speech synthesis program, we can change any part of the pitch contour but preserve the spectrum parameters. One tone contour could also be changed into another one by changing only the fundamental frequency. However, we discovered that intensity is another important factor in intonation, not only for whispered, but also for normal speech, so care must also be taken to study the interaction between fundamental frequency and intensity in intonation.

Synthesized Chinese sentences in which the pitch contour was changed were presented to a group of Swedish listeners and ABX tests were carried out.

TESTING MATERIAL

The following three test sentences designed for the study of tone and intonation in Standard Chinese were used for the perceptual experiment (the pīnyīn transcription is used):

1. Wāng Yī chōu xiāngyān. ("Wāng Yī smokes cigarettes.")

2. Sòng Yàn qùguò Dàqìng. ("Sòng Yàn went to Dàqìng.")

3. Lǐ Xiǎobǎo xiě jiǎngyǎn gǎo. ("Lǐ Xiǎobǎo writes lecture notes.")

(Tone marks: ¯ stands for high level tone, ` for falling tone, and ˇ for low dipping tone.)

All these sentences were processed on a microcomputer with the LPC program, and the pitch contours were abstracted. The pitch contour was changed on the screen of a graphic terminal, and using the new contour the test sentences were synthesized. Sentences 1 and 2 were used for ABX tests and sentence 3 was evaluated only by the author.

All pitch contours of sentences 1 and 2, which were presented to listeners in synthesized sentences, are shown in Figures 1 and 2. Every pitch contour of a test sentence is accompanied by a test number followed by the highest frequency of the ordinate and a target mark (S, Q or F). Test sentences marked S have a target sentence which is the answer to a yes-no question with focus on the word <u>shì</u> "yes" (such as <u>Shì, Wāng Yī chōu xiāngyān</u> "Yes, Wāng Yī smokes cigarettes"). The focussed word <u>shì</u> was omitted. Q stands for a yes-no question (segmentally the same as a statement), and F for a sentence with focus at some place. In some cases, where the change in pitch was difficult to see on the contours, frequency values are indicated above the curves.

TESTING METHODS

1. <u>ABX Tests</u>

In order to investigate how changing only the pitch contour influences intonation, a special testing crew which consisted of 16 persons (10 males and 6 females, aged from 20 to 61 years, most of them native speakers of Swedish) was organized. The listeners were teachers or students of phonetics, who knew no Chinese. The instruction words: "You should make a decision based only on the intonation for every test item" were given to all listeners before the tests. Thus no syntactic or semantic information besides the pitch contour was utilized by these foreign listeners. In this experiment, A is the reference sentence - a neutral (focus free statement) sentence -, B is the target sentence, and X is the tested sentence. Stimuli A and B were presented in random sequence and X followed them. A pause of about 10 seconds was made between the tests to allow the listeners to make their responses. The listeners were seated in an acoustically treated classroom and listened via a loudspeaker.

2. Subjective evaluation

The third test sentence (Lǐ Xiǎobǎo xiě jiǎngyǎn gǎo) was evaluated only by the author (who is a native speaker of Standard Chinese).

RESULTS AND DISCUSSION

The listeners' responses for test sentences 1 and 2 under different testing conditions (see Figures 1 and 2) are listed in Tables 1 and 2. In these tables, A or B stands for the listener's response to the test sentence X, and the percentage shows the relative number of responses made for the target sentence, i.e. B. If the target response percentage is greater than 50%, the effect of the pitch change is considered as positive and denoted "yes", and if the percentage is less than or equal to 50%, the effect is considered negative and denoted "no". These responses were based only on the pitch contour, because none of the listeners had any knowledge of Chinese. In my opinion, using foreign listeners is a good way to exclude the influence of syntactic and semantic factors in speech, while preserving the intonation.

The results of this experiment can be summarized as follows:

1. From tests No. 1, 7, 8 and 9 in Table 2, it can be seen that the final part of a pitch contour (a word or a syllable), and especially its baseline, is very important for a question sentence. Test No. 1 in Table 2, which shows that a perfect question sentence can be formed by shifting only the baseline of the last syllable qìng upwards, while

preserving the topline, is convincing evidence. This has a similar effect as test No. 14 in Table 2, where the whole pitch contour was changed into the shape of a question sentence.

2. However, if the baseline of each syllable in a sentence is shifted upwards while preserving the topline (test No. 2 in Table 2), then there is no effect. The same phenomenon can be observed in test No. 1 in Table 1, in which the topline of xiāngyān was shifted upwards. Comparing test No. 2 to test No. 9 in Table 2, we find that the response percentage increases rapidly when both the topline and the baseline are shifted upwards. So it can be concluded that high pitch level at the end of the sentence and both topline and baseline with positive slope (maybe with a steeper slope in the baseline) are distinctive characteristics of question sentences.

3. By shifting down the baseline only (test No. 4 in Table 2) or the topline only (test No. 15 in Table 2), we cannot make a neutral sentence into a focus left statement sentence. For forming a focus left statement sentence it is necessary to shift both the topline and the baseline downwards, giving them a negative slope, much steeper for the topline than for the baseline (see test No.6 in Table 1).

4. Focus is formed by expanding the pitch range of the focussed lexical unit and compressing the pitch in other parts of the sentence, especially in the part of the sentence which follows the focus. For all tones with the exception of the dipping (third) tone, shifting the topline upwards is sufficient for forming a focus. For the dipping tone, however, the baseline has to be shifted downwards and stay low for a long time (sometimes with a rising end). A perfect focus was formed on the subject of sentence 1 (Wāng Yī) by shifting the topline of this word upwards and shifting it down for other words (see test No. 8 in Table 1). If only the topline of the subject in sentence 2 (Sòng Yàn) is shifted upwards and the rest is preserved, this does not form a focus (test No. 11 in Table 2).

5. Besides the fundamental frequency, intensity may be another important factor in the perception of intonation. All

efforts to change a neutral sentence in which all syllables have the dipping tone (such as sentence 3) into a question sentence by shifting the baseline and the topline upwards, as was done for sentences No. 1 and 2, were defeated. Segmenting the speech sounds in a synthesized sentence (test sentence 3), I found that even the baseline of the final syllable <u>gǎo</u> could be shifted greatly upwards without it being heard, because of the masking effect of the preceding part with higher intensity and lower frequency.

6. By changing the tone contour only, we can change any tone into any other tone, with the exception of the dipping tone at the end of a sentence.

ACKNOWLEDGEMENTS

I wish to express my gratitude to Professor Eva Gårding, who heads the project, and to Engineers Bengt Mandersson and Karl-Gustav Jönsson for their help in finishing this work.

REFERENCES

Cohen, A., R. Collier and J. ʼt Hart. 1982. "Declination: Construct or intrinsic feature of speech pitch?". <u>Phonetica</u> 39, 254-73.

Table 1. Listeners' responses for ABX tests.
Test sentence: Wāng Yī chōu xiāngyān.
Date: 25 March 1983.

Test						L i s t e n e r										Correct	
No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	%	
1	A	B	A	A	A	A	A	A	A	B	A	A	A	A	A	B	19	No
2	A	B	B	A	A	B	B	A	A	A	A	A	A	A	A	B	32	No
3	A	B	A	B	A	A	A	A	A	A	B	A	B	A	A	B	32	No
4	B	B	B	B	B	B	A	A	A	B	A	A	B	A	B	B	62	Yes
5	B	A	B	B	B	B	B	B	B	B	A	B	B	B	A	B	81	Yes
6	B	A	B	B	B	B	B	B	B	B	B	B	B	B	B	B	94	Yes
7	B	A	B	A	B	A	B	B	B	B	B	B	B	A	B	B	75	Yes
8	B	B	B	B	B	B	B	B	B	B	B	B	A	B	B	B	94	Yes
9	B	A	B	B	B	A	B	B	B	B	B	B	B	B	A	B	81	Yes

Table 2. Listeners' responses for ABX tests.
Test sentence: Sòng Yàn qùguò Dàqìng.
Date: 25 March 1983.

Test						L i s t e n e r										Correct	
No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	%	
1	B	B	B	A	B	B	B	B	B	B	B	B	B	B	B	A	87	Yes
2	B	A	A	A	B	B	A	A	A	B	B	B	B	B	B	B	62	Yes
3	B	B	B	B	B	A	B	A	B	A	B	B	B	A	B	A	62	Yes
4	A	A	B	A	B	A	A	A	A	A	A	A	A	A	A	A	13	No
5	B	A	A	B	B	B	A	B	A	B	A	B	B	B	B	A	62	Yes
6	A	B	B	B	B	B	B	B	A	B	B	B	B	A	B	B	81	Yes
7	A	B	B	B	B	A	A	B	B	B	B	B	B	A	B	B	75	Yes
8	A	B	B	B	B	B	B	B	A	A	B	B	B	B	A	A	68	Yes
9	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	100	Yes
10	B	B	B	A	B	B	B	B	B	B	B	A	A	A	B	A	68	Yes
11	A	B	A	A	B	B	B	A	A	B	A	B	A	B	A	A	38	No
12	B	B	B	B	B	A	A	A	A	B	A	B	A	B	A	A	50	No
13	B	B	B	B	B	B	B	B	B	A	B	A	B	A	B	B	81	Yes
14	B	A	B	B	B	B	A	B	B	B	B	B	B	B	B	B	87	Yes
15	A	B	A	B	B	B	A	A	A	B	A	B	B	A	A	B	50	No
16	B	B	B	B	B	B	B	B	B	B	B	B	B	A	B	A	87	Yes

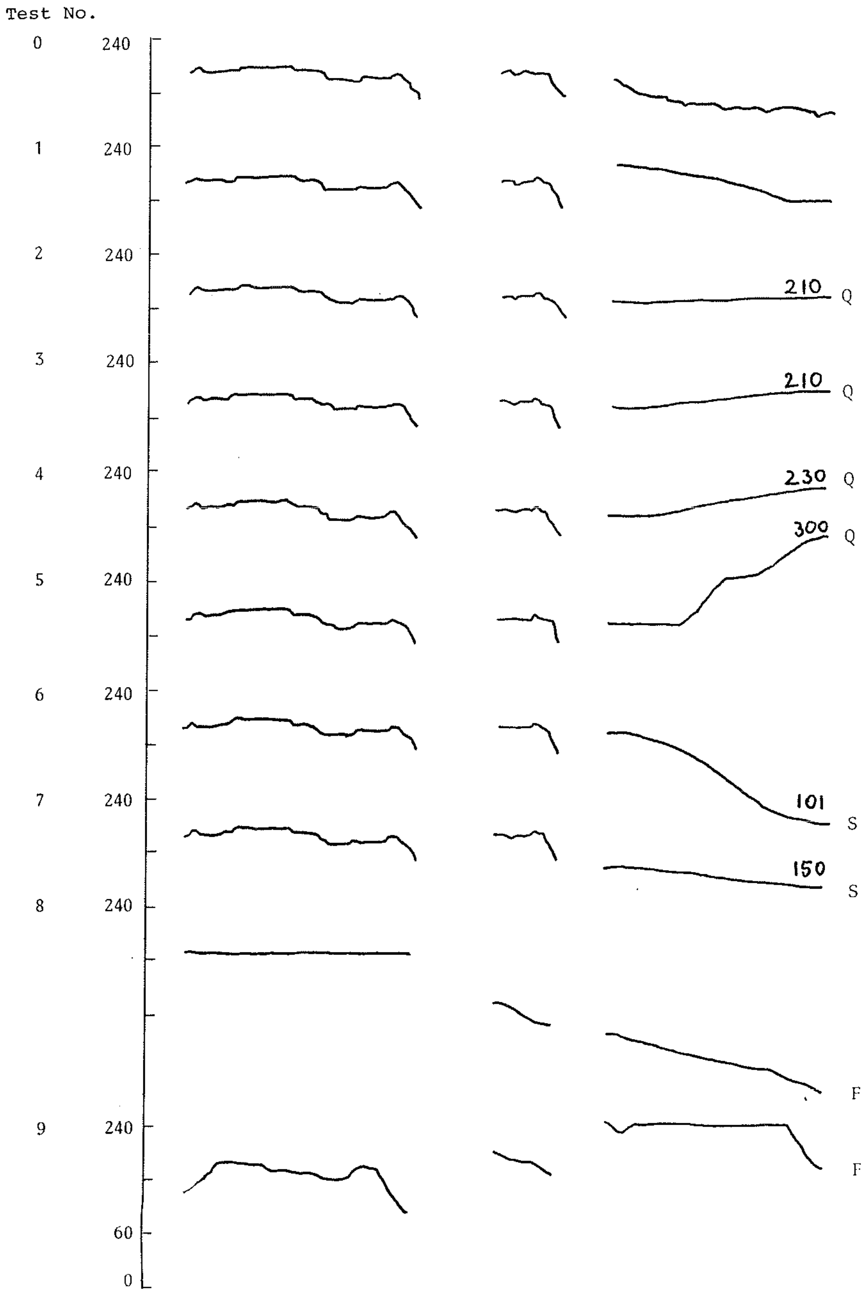


Figure 1. Test sentence: Wāng Yī chōu xiāng yān.

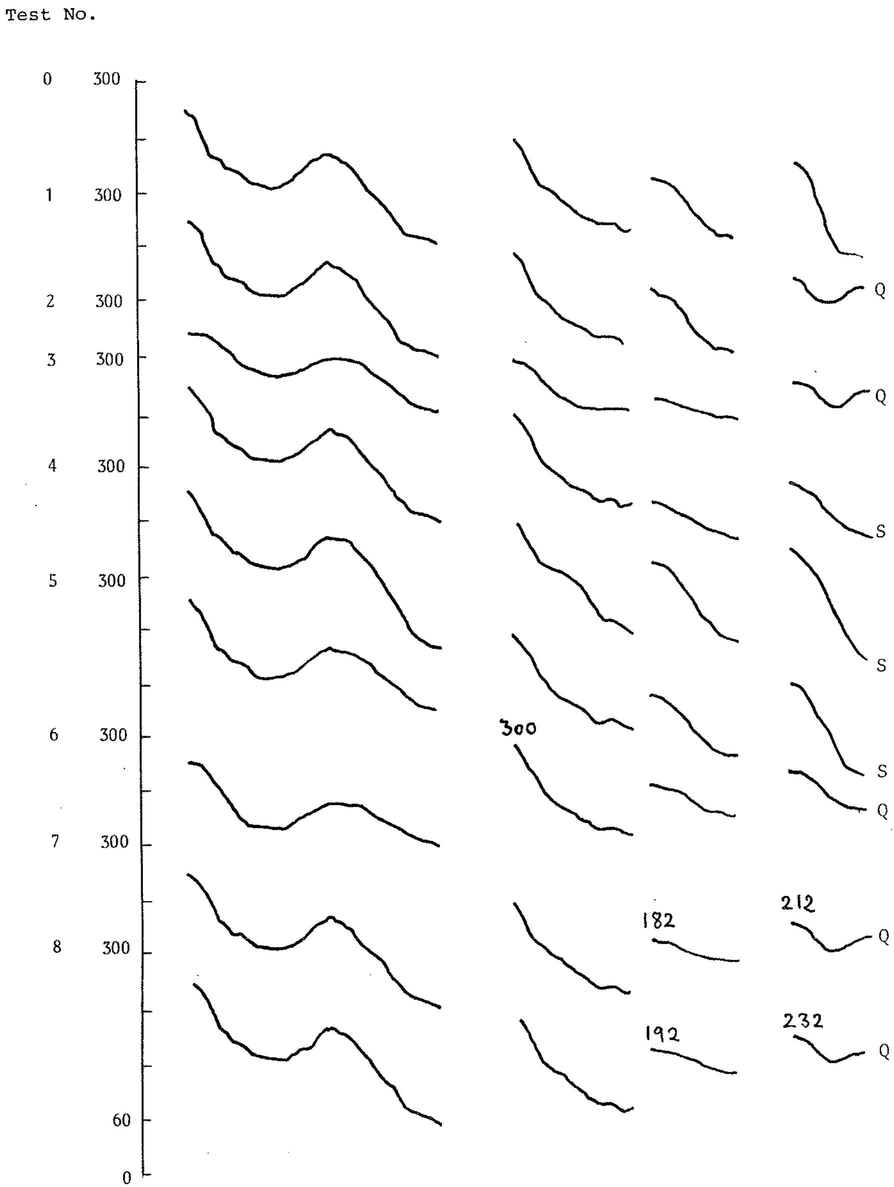


Figure 2. Test sentence: Sǒng Yǎn qùguò Dàqǐng.

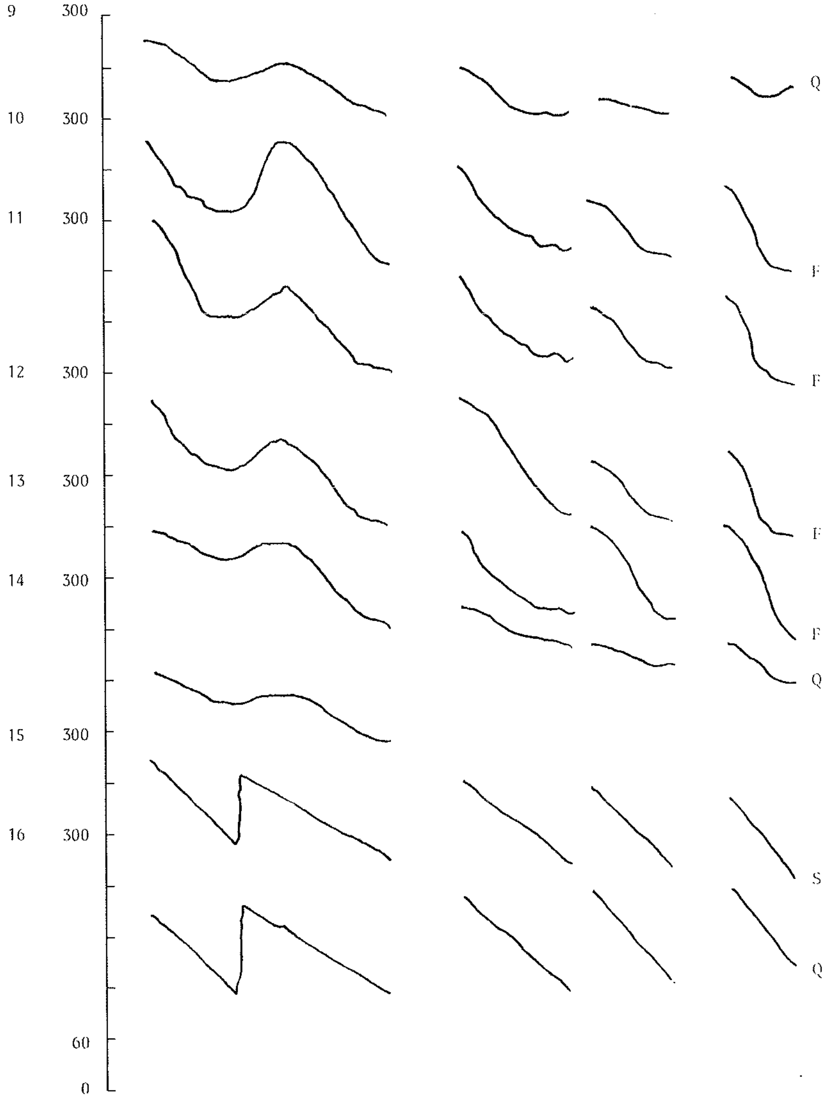


Figure 2 continued

RECENT NUMBERS AND CONTENTS

WP 22 1982	ROBERT BANNERT	Temporal and tonal control in German
	ANTONIS BOTINIS	Stress in modern Greek: An acoustic study
		Stress in modern Greek: A physiological study
	GÖSTA BRUCE	Developing the Swedish intonation model
	EVA GÅRDING	Prosodic expressions and pragmatic categories
	EVA GÅRDING, ANTONIS BOTINIS & PAUL TOUATI	A comparative study of Swedish, Greek and French intonation
	DAVID HOUSE	Identification of Nasals: An acoustical and perceptual analysis of nasal consonants
	ANDERS LÖFQVIST, BJÖRN CARLBORG & PETER KITZING	Initial validation of an indirect measure of subglottal pressure during vowels
	BENGT MANDERSSON	Resolution of convolved signals in biomedical applications
	KERSTIN TEVAJÄRVI	Intonation in Finland-Swedish: Word and sentence stress in the Helsinki dialect
WP 23 1982	SIDNEY WOOD	X-ray and model studies of vowel articulation
WP 24 1982	GUSTAV JÖNSSON, BENGT MANDERSSON & BENGT SIGURD	A microcomputer pausemeter for linguists
	ELISABET ENGDAHL	Interpreting sentences with multiple filler-gap dependencies
	KERSTIN NAUCLÉR	Spoken and written language - relationships between two linguistic means of expression
	BENGT SIGURD	Changes in Swedish address terms described by power-solidarity diagrams

LECTURES AND PUBLICATIONS ELSEWHERE

1982

Håkansson, G. Quantitative studies of teacher talk. Scandinavian Working Papers on Bilingualism 1:1982, 52-71, Stockholm University, Institute of Linguistics, Dept of Research on Bilingualism

Magnusson, E & Ek, I. Ge orden liv! Språk och samtal hos vuxna utvecklingsstörda. Socialstyrelsen redovisar 1982:14

Nauclér, K. Spoken and written language - relationships between two linguistic means of expression. Ninth World Congress on Reading, Dublin, August

1983

Bannert, R. Prosody and intelligibility of Swedish spoken with a foreign accent. Nordic Prosody III, 7-18. Acta Universitatis Umensis 59, Elert C. - C, Johansson J, and Strangert E. (eds), Umeå

Bannert, R. Intelligibility of foreign accent. Poster-presentation at the 10th International Congress of Phonetics Sciences, Utrecht, August

Bannert, R. Förståelighet i brytning. AScLA-symposium "Nordens språk som målspråk", Stockholm, December

Bruce, G. Accentuation and Timing in Swedish. Folia Linguistica. Tomus XVII / 1-2, pp. 221-238

Bruce, G. Textual aspects of prosody in Swedish. Abstracts of the 10th International Congress of Phonetic Sciences, Utrecht, Aug. 1-6, Cohen & v.d. Broecke (eds.), p. 579

Dahlbäck, H. Vowel variation in a Swedish dialect as a problem of intelligibility for L2 learners. Papers from the 7th Scandinavian Conference of Linguistics 2, 506-520. Karlsson, F. (ed).

Gårding, E. A generative model for tone and intonation. Institut für Phonetik und sprachliche Kommunikation, Universität München. June

Gårding, E. Intonation units and pivots, syntax and semantics. Contribution to Symposium 1. 10th International Congress in Phonetic Sciences. Utrecht. August

Gårding, E. Superposition as an invariant feature in intonation. Contribution to Symposium Invariance and Variability of Speech Processes. MIT. October

Gårding, E. Superposition som invariant fenomen i intonation. Institutionen för Lingvistik, Stockholm

Gårding, E. En intonationsmodell för svenska och dess användning. Kråków. Polen, November

Gårding, E. , J. Zhāng & J-O. Svantesson. A generative model for tone and intonation based in one speaker. Section paper. 10th International Congress of Phonetic Sciences. Utrecht. August

Horne, M. Review of Nigel Love, Generative Phonology: A case study from French. Studia Linguistica 37:2, 204-9.

House, D. Transitions in Speech Perception. 10th International Congress of Phonetic Sciences, Utrecht, August 1-6

House, D. Cue Interaction and Hearing Loss. FRN:s Sommarskola "Tal, Ljud, Hörsel," Tällberg, August 20-27

Håkansson, G. Makt och maktlöshet i klassrummet - en analys av topikval. Studies in Communication 6, 19-33, University of Linköping

Håkansson, G. En kvantitativ analys av språklärarspråket. LFI-nytt 3, 20-25

Hyltenstam, K. & Magnusson, E. Tondistinktionen hos klusiler: språktypologiska markeringsförhållandens prediktionsvärde för första- och andraspråksinlärning av svenska. THL-symposium, Stockholm, March

Magnusson, E., Nauclér, K., & Söderpalm, E. Språklig medvetenhet hos förskole- och lågstadiebarn. Utvecklingsarbete och fältförsök, rapport 8, Lärarhögskolan i Malmö

Magnusson, E., Nauclér, K., & Söderpalm, E. Kartläggning av orsaker till läs- och skrivsvårigheter. Språklig medvetenhet hos barn med och utan försenad språkutveckling. THL-symposium, Stockholm, March

Magnusson, E., Nauclér, K., & Söderpalm, E. Form or substance? The linguistic awareness of pre-school children and school children investigated by means of a rhyming test. Xth International Congress of Phonetic Sciences, Utrecht, August

Magnusson, E., Nauclér, K., & Söderpalm, E. Rhyming as a mean of investigating linguistic awareness in young children. XIX IALP-Congress, Edingburgh, August

Nauclér, K. Connections between spoken and written language. Evidence from three investigations on normal and pathological written performance. Journal of Pragmatics 7, 595-602

Norlin, K. Akustisk analys av frikativor i kairoarabiska. Finska Orientsällskapet, Helsingfors, March 30,

Touati, P. Profils temporels en suédois et en français (with an abstract in English) Nordic Prosody III. Umeå Studies in the Humanities 59: 231-243. Umeå

Touati, P Métrique pour l'évaluation des erreurs prosodiques. Abstracts of the Tenth International Congress of Phonetic Sciences. Utrecht, p.745.

Weimer, K. Finskspråkiga invandrarelevers språkfel vid skriftlig och muntlig produktion av svenska. AScLA-symposium " Nordens språk som målspråk", Stockholm, December 3-4

1984

Bruce, G. Rhythmic alternation in Swedish. Nordic Prosody III. Elert, Johansson & Strangert (eds), pp. 31-41. Umeå

Bruce, G. Timing and Accent in Swedish. Workshop on Accentology. Center for Study in Language and Information, Stanford University: March 13-15

Bruce, G. & P-K. Halvorsen. Overview of Scandinavian accent.

Bruce, G. On the phonetics of rhythm: evidence from Swedish. Acoustical Society of America, 107th Meeting, Norfolk, Virginia, May 6-10, 1984. JASA, Suppl.2 vol.75, Spring 1984, p. S40

Bruce, G. On the phonology and phonetics of rhythm: evidence from Swedish. Abstracts Phonologietagung 1984. Institut für Sprachwissenschaft. Fifth International Phonology Meeting. June 25-28, Eisenstadt, Austria, pp. 8-9

Håkansson, G. Lärarens lexikonanvändning - hur lärare talar till elever på olika nivåer. To appear in the Proceedings of the AScLA-symposium, Stockholm, December 1983

Norlin, K. Acoustic analysis of vowels and diphtongs in Cairo Arabic. 15th Conference on African Linguiatics, UCLA, Los Angeles, March 28-30

Weimer, K. Invandrarbarns väg till tvåspråkighet - parallellitet vid inlärning av uttal, stavning grammatik och semantik. 4th Nordic Symposium on Bilingualism. Uppsala, June 5-7

CURRENT RESEARCH PROJECTS

The immigrant and communication technology - understanding and being understood (The Bank of Sweden Tercentenary Foundation)

Robert Bannert

Verbal reforms (The Bank of Sweden Tercentenary Foundation)

Bengt Sigurd, Christer Platzack, Anna Orrling

Commentator, a computor system simulating verbal behavior (Swedish Council for Research in the Humanities and Social Sciences)

Bengt Sigurd, Jan Fornell

Language and brain (Swedish Council for Research in the Humanities and Social Sciences)

Eva Gårding, Christina Dravins

Central Features in Scandinavian Syntax (Nordic Committee for Coordinating Research in the Humanities)

Elisabet Engdahl

Phonological and Phonetic studies of vowel articulation in Bulgarian (Swedish Council for Research in the Humanities and Social Sciences)

Thore Pettersson, Sidney Wood

Phonetic description of some important languages in the third world (Swedish Council for Research in the Humanities and Social Sciences)

Eva Gårding, Mona Lindau, Kjell Norlin, Jan-Olof Svantesson

Language disordered children's reading and spelling (Swedish Council for Research in the Humanities and Social Sciences)

Kerstin Nauclér, Eva Magnusson

Reading and spelling difficulties. A search for causes (Swedish Council for Research in the Humanities and Social Sciences)

Kerstin Nauclér, Eva Magnusson

Diphthongization versus vowel duration in Southern Swedish

Kurt Johansson

Interactive text generation (STU)

Jan Fornell, Bengt Sigurd

PRAKTISK LINGVISTIK

Nr 1 - 1979	Eva Gårding & Robert Bannert Optimering av svenskt uttal (incl. cassette recording) out of print
Nr 2 - 1979	Kenneth Hyltenstam (Ed) Invandrarundervisning - aspekter på alfabetisering och den första nivån out of print
Nr 3 - 1979	Robert Bannert Ordprosodi i invandrarundervisningen 40:- (incl. cassette recording)
Nr 4 - 1979	Eva Wigforss Svensk pluralbildning hos finska invandrar-barn - en psykolingvistisk utvecklingsstudie 30:-
Nr 5 - 1981	Robert Bannert Svårigheter med svenskt uttal: inventering och prioritering 40:- (incl. cassette recording)
Nr 6 - 1981	Gösta Bruce (Ed) Tema prosodi - seminarieuppsatser i fonetik 30:-
Nr 7 - 1982	Don Kulick, Karina Vamling, Lars-Åke Henningsson Om tolkning 30:-
Nr 8 - 1983	Jan Fornell, Bengt Sigurd Commentator - ett mikrodatorbaserat forsknings-program för lingvister 30:-
Nr 9 - 1984	Thore Pettersson (Ed) Färgterminologi - seminarieuppsatser i allmän språkvetenskap 40:-
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.

XIX *Ulrika Nettelbladt.* Developmental Studies of Dysphonology in children. 1983.

Numbers V, VII-XII, XIV-XIX are available from the publisher, Liber International, Room 405, S-205 10 MALMÖ, Sweden.

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, Lotta Wogensen
Office	Eva Körner, Britt Nordbeck, Anneli Byskog, Gunilla Edsheim, Johan Rosengren

General Linguistics

Berthout, Jean-Jacques	BA; Assistant
Dahlbäck, Hans	Fil Kand; Research Assistant, extra teacher
Engdahl, Elisabet	PhD; Research Fellow
Fornell, Jan	Research Assistant
Geggus, Jana	Fil Kand; Assistant, extra teacher
Håkansson, Gisela	Fil Mag; Research Assistant, extra teacher
Malmberg, Bertil	Fil Dr; Professor Emeritus
Nordbeck, Britt	Secretary
Pettersson, Thore	Fil Dr; Docent; Lecturer
Sigurd, Bengt	Fil Dr; Professor; Head of Division
Viberg, Åke	Fil Dr; Research Fellow
Research Students:	Jean-Jacques Berthout, Hans Dahlbäck, Jan Fornell, Jana Geggus, Cecilia Hedlund, Lars-Åke Henningsson, Merle Horne, Gisela Håkansson, Christian Matthiesen, Ingegärd Mills, Barbara Prohovnik, Leila Ranta, Emilio Rivano-Fischer, Christopher Stroud, Jan-Olof Svantesson, Karina Vamling

Phonetics

Bannert, Robert	Fil Dr; Docent; Research Assistant
Bruce, Gösta	Fil Dr; Docent
Dravins, 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 Assistant; Phonetics Dept, UCLA
Löfqvist, Anders	Fil Dr; Docent; Research Fellow
Mandersson, Bengt	Tech Dr; Civ Ing; Computer Engineer
Nauclér, Kerstin	Fil Dr; Logoped; extra teacher
Nettelbladt, Ulrika	Fil Dr; Logoped; extra teacher
Norlin, Kjell	Fil Kand; Research Assistant
Norman, Lennart	Fil Kand; Voice and speech teacher
Svantesson, Jan-Olof	Fil Kand; Research Assistant
Wood, Sidney	Fil Dr; Research Fellow
Research Students	Catharina Andersson, Antonis Botinis, Anne-Christine Bredvad-Jensen, Christina Dravins, David House, Boris Larnert, Ann-Christine Ohlsson, Lili-Ann Rudberg, Paul Touati, Kjell Weimer, Ursula Willstedt

<u>Child Language Research</u>

Bredvad-Jensen, Anne-Christine	Fil Kand; Research Assistant
Mårtensson, Bodil	Secretary
Söderberg, Ragnhild	Fil Dr; Professor; Head of Division

Adress: Department of Linguistics and Phonetics
Lund University
Helgonabacken 12
S-223 62 LUND

Tryckericentralen
AVC Lund 1984