LUND UNIVERSITY
DEPARTMENT OF LINGUISTICS
General Linguistics
Phonetics

WORKING PAPERS

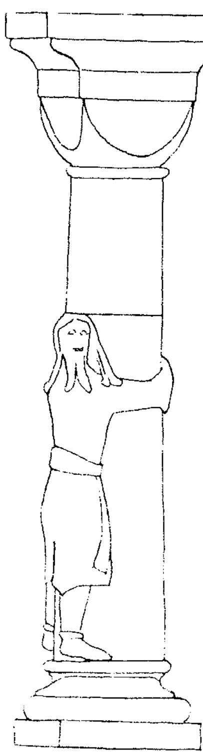
28.1985

LUND UNIVERSITY
DEPARTMENT OF LINGUISTICS
General Linguistics
Phonetics

WORKING PAPERS

28.1985

ISSN - 0280 - 526X



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

This issue has been edited by Eva Gårding and Eva Körner

ANNUAL REPORT

The past year has been marked by some memorable events. The most positive one was perhaps Humanistdagarna (the Days of Humanities) when our department contributed a day-long program (see below). This was an effort to show the public what we are doing and what we stand for in this time of economic squeeze and declining interest in higher education. The situation is further aggravated by the absence of a natural basis for recruitment. Only prospective logopeds have phonetics and general linguistics as a regular part of their training program. But we still have hope that our subjects will be recognized as a regular part of the training of language teachers.

The current interest in information technology, however, may give phonetics and linguistics better possibilities to realise their potential for applications. This has already happened in other European countries and courses entitled 'Informatik' and 'Informatique' are beginning to appear in the regular phonetics and linguistics curricula.

Despite the financial and administrative difficulties, our activities have continued as before which may be reflected by our seminar programs below, the contents of our Working Papers and the report on project work and dissertations listed at the end of the book.

Friday seminars

Sept 7	Reports from summer meetings	
Sept 21	Constancy and variation in Standard Chinese	Eva Gårding
Sept 28	Creak in Peking speech	Paul Kratochvil (Cambridge)
Oct 5	Perception of stress in Greek	Antonis Botinis
Oct 12	Tone 3 versus Tone 4. Design of perceptual experiments	Jan-Olof Svantesson
Oct 19	Preliminary results from the Sblos project (The reading and spelling of Children with delayed speech development	Kerstin Nauclér and Eva Magnusson
Oct 26	Focus in French and Swedish: A phonetic comparison	Paul Touati
Nov 9	The perception of word stress in prefocal position in Greek	Antonis Botinis
Nov 16	Perception of F_0-categories	David House

Nov 23	Rhythm	Gösta Bruce
Nov 30	Prosody in Cairo Arabic	Kjell Norlin
Dec 7	The relation between prosody, syntax and semantics in Greek	Antonis Botinis
Dec 14	Is the right hemisphere the location of emotions, attitudes and communicative functions?	Christina Dravins
Jan 17	Reports from a workshop: Methods of storing informatiom for documentation	Anne-Christine Bredvad-Jensen
Feb 15	Diphthongs in the languages of the world, and in Hausa in particular	Mona Lindau Webb
March 3	Intonation contours in ten Swedish sentences	Dieter Huber
March 8	Rehearsal for the Speech-Sound-Hearing conference in Göteborg	
March 29	Language and brain	Christina Dravins
April 12	Tone and stress in Peking Chinese	Paul Kratochvil (Cambridge)
April 26	F_0 and the perception of movement	David House
May 3	Markedness in linguistic knowledge	Manfred Bierwisch (Berlin)
May 10	Tour de France	Gösta Bruce
May 24	Some problems in the analysis of word stress in Hindi	Manju Ohala (San José)
June 17	The perception of speech in children and adults	Mary Smith (Dartmouth)

The general seminar, Language, Speech, Sound, Hearing, is now being held in cooperation with the Department of General Linguistics, The Centre of Voice and Speech Training, the Department of Logopedics and Phoniatrics and the Centre of Child Language Research.

Sept 10	Rhetoric in our time	Jørgen Fafner (Copenhagen)
Sept 17	How are words represented in memory	Morris Halle (MIT)
Oct 1	Modern Chinese and linguistic change	Paul Kratochvil (Cambridge)
Oct 15	The meaning of sentences	Staffan Hellberg (Göteborg)
Oct 22	From communication as language to language as communication	Ragnhild Söderbergh
Nov 12	Synthetic speech in communication and edcuation	Karoly Galyas (Stockholm) and Ewacarin Holmqvist
Nov 26	The project 'Segementation of text for speech'	Jan Svartvik
Feb 4	Smiles, sex and the symbolic use of speech sounds	John Ohala (Berkeley)
Feb 25	The languages of Africa	Mona Lindau Webb

March 6	Humaniora över gränserna	Open house - Program:
	Arabic phonology Greek prosody How Swedes speak French	Kjell Norlin Antonis Botinis Paul Touati
	Basque peculiarities The language of dogs and wolves Language and brain	Jana Geggus Gisela Håkansson & Mona Lindau Christina Dravins
	Reading and writing difficulties Analysis of foreign accents The deterioration of Scanian	Kerstin Nauclér & Eva Magnusson Robert Bannert Gösta Bruce
	Can a language be just anything?	Mona Lindau Webb & Eva Gårding Thore Pettersson, Elisabet Engdahl
March 18	The location of sentence stress	Merle Horne
March 25	Modularity in sentence Processing	Lyn Frazier (Amherst)
April 10	Subject or theme in Chinese	Paul Kratochvil (Cambridge)
April 22	An intonation model and its anchors	Eva Gårding
May 6	Structure and process in language production and the nature of speech errors	Manfred Bierwisch (Berlin)
May 20	Thoughts on speech training	Gabriella Stenberg-Koch
June 3	A model for analysis of dialogue between adults and children	Ragnhild Söderbergh and Anne-Christine Bredvad-Jensen

The seminars on research in language teaching, chaired by Robert Bannert, and on problems of reading and writing, chaired by Kerstin Nauclér and Eva Magnusson, are followed by an increasing number of participants.

There where the following courses open to advanced students:

How to use the ILS program	Sidney Wood
The production and perception of speech	Anders Löfqvist
Models and measurements	Bengt Sigurd, Department of General Linguistics
The phonetics and phonology of prosody	Gösta Bruce

Longer visits

Mona Lindau Webb stayed two months for work on the project 'Phonetic description of some non-European languages'. We also profited greatly from visits by Paul Kratochvil, Fac. of Oriental Studies, Cambridge. Janina Petecka, Swedish lecturer in Gdansk, who held a Swedish Institute scholarship, spent most of the spring term with us. The present issue of the Working Papers bear witness to their work while they were here.

Projects

The project 'Phonetic descriptions of some important non-European languages' (HSFR)[1] is now being concluded. The project Reading and spelling difficulties (HSFR) is entering its second year. A new project 'From text-to-prosody' (RJ)[2] started on February 1.

Lund in June 1985

Eva Gårding

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

2) RJ, The Bank of Sweden Tercentenary Foundation.

CONTENTS

Eva Gårding	In defence of a phrase-based model of intonation	1
Eva Gårding	Constancy and variation in Standard Chinese tonal patterns	19
Eva Gårding, Paul Kratochvil, Jan-Olof Svantesson and Jialu Zhang	Tone 4 and Tone 3 discrimination in Modern Standard Chinese	53
David House	Implications of rapid spectral changes on the categorization of tonal patterns in speech perception	69
David House	Sentence prosody and syntax in speech perception	91
Dieter Huber	Swedish intonation contours in text-to-speech synthesis	109
Kerstin Nauclér and Eva Magnusson	Language disordered children's reading and spelling: Preliminary results	127
Kjell Norlin	Pharyngalization in Cairo Arabic	139
Janina Petecka	A study of question intonation in Polish	151

In defence of a phrase-based model of intonation [1]

Eva Gårding

I will do two things:
1. Give a brief presentation and evaluation of the model I use for comparison of intonation across languages
2. Discuss objections to the basic tenets of the model

PRESENTATION

The model has grown out many years' work in intonation in various languages. My own starting point was a comparison between Swedish and English intonation that I did for strictly pedagogical purposes when I was still a school teacher (Gårding 1962). Many of my students, collaborators and critics have helped give the model its present shape.

One principle of the model is that global intonation stretching over a phrase or a sentence is separable from local intonation bearing on lexical accents and tone. Long ago, Ernst Meyer was very well aware that his kymoghrapic records of fundamental frequency patterns of accent manifestations from 100 Scandinavian dialects contained ingredients both from the word accents and the sentence intonation but he thought that it was next to impossible (kaum möglich) to separate the two (Meyer 1937 p. 41). However, this separation can be done by analysing sentences in which global and local properties have been varied systematically (Gårding and Lindblad 1973, Bruce 1977). It then turns out that the local features, the accent humps, seem to be added to or superimposed on a global phrase intonation component and that the timing of these humps is dialect-specific (Bruce & Gårding 1978). The superposition principle was used explicity and quantitatively for the first time in intonation analysis by Sven Öhman (1968) and later by Carlsson and Granström (1973), Robert Mc Allister (1971), and Nina Thorsen (1983). Superposition has been a guiding principle in my own work and it has been useful in all the languages that I have studied.

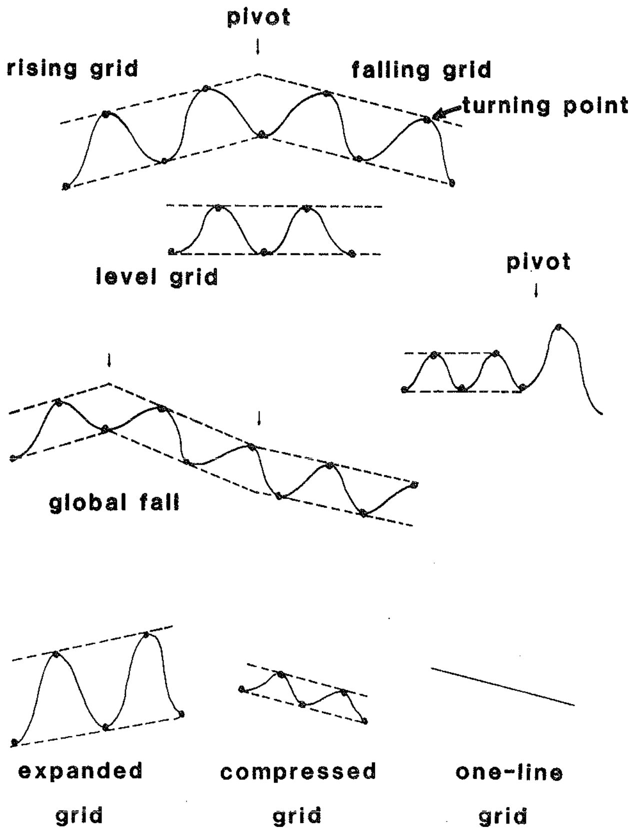
Another important principle of the model is to base the local analysis, i.e. accents and tones, on the notion of turning-point fixation. This principle can be explained in the following way. Characteristic of an intonation curve, not matter what language it comes from, is that is has an undulating shape and like all undulating curves it can be efficiently described by interpolation between local maxima and minima which I call turning points. A very important observation is that some of these turning points have a rather fixed position relative to specified acoustic segments. From this it follows that certain falls and rises also are kept in fixed positions relative to the segments. And it also follows that giving the positions of the turning points in time and frequency is an economic way of describing an intonation curve. This is precisely the principle we have used in the generative part of the model.

The third principle is to base the global analysis of phrase and sentence intonation on the large-scale pattern formed by the concatenated turning points. In long sentences, the local up and down structure of an intonation curve usually repeats itself within longer periods corresponding to phrase-like units. This larger pattern is expressed by the tonal grid. In the ideal case a grid is obtained by joining consecutive local maxima and minima separately. That part of the grid where the direction or width is changed or where the grid takes a jump is called a pivot. It marks the boundaries of what we might call prosodic phrases.

Figure 1 summarizes in schematic form the descriptive framework of the model. These concepts can be seen as part of a phonological inventory of prosodic features general across languages. Together with an account of their functions, which may be language specific particularly in the lexicon, they constitute a kind of prosodic grammar.

Examples

I shall now present some examples from real life and I will start with my favourite example from Standard Chinese, Sòng Yǎn mài niúròu 'Song Yan sells beef' (Fig 2). This sentence has tones which are in order from the start, falling, rising, falling, rising, falling and it has been uttered in different prosodic patterns. The broken lines are grids. The upper broken line of a grid connects the maxima and the lower broken line the minima of the fundamental frequency curve. Part of these turning points are fixed points in the sense that they are relatively fixed to the segments independently of the



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

Fig.1 Concepts of the model and their communicative functions.

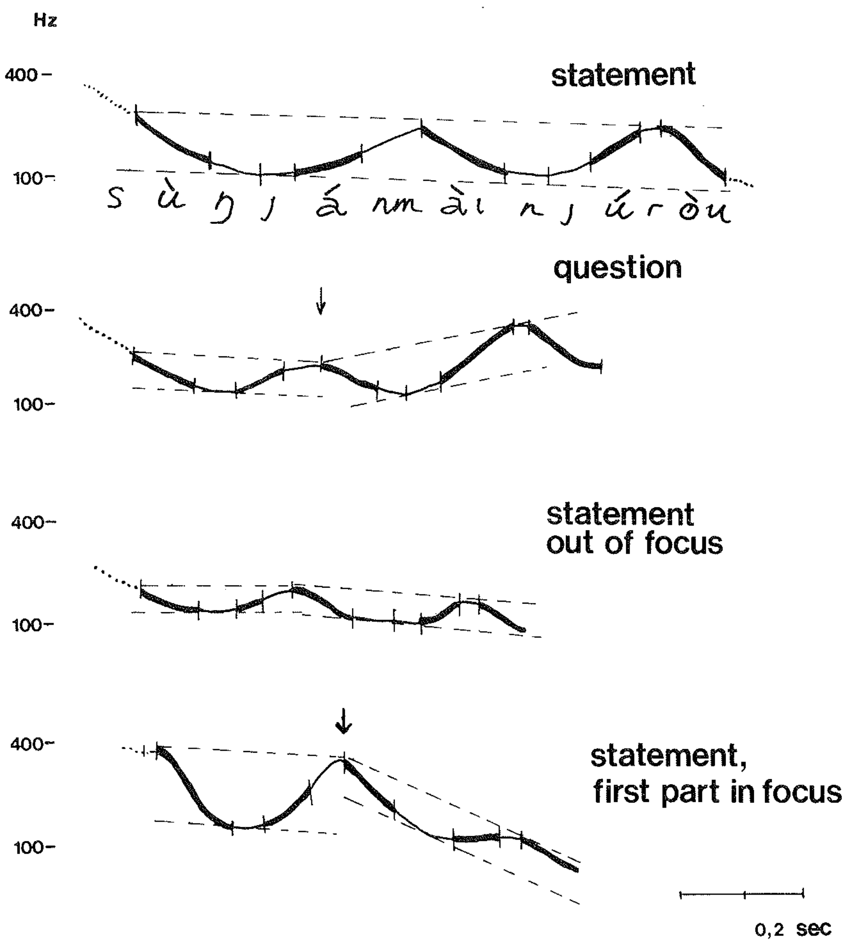


Fig.2 F0 curves from Standard Chinese <u>Sòng Yán mài niúròu</u> 'Song Yan säljer oxkött' in different prosodic patterns. A thick line marks the F0 curve over the vowels, a thin line marks the consonants. Broken lines are <u>grids</u>. An arrow marks <u>pivot</u>. From Gårding, Zhang & Svantesson, 1983.

sentence prosody. The direction of the grids shows the type of speech act and their width shows information weight. Note the expanded width in connection with focus and the compressed grid after focus. The pivot, marked by an arrow, is optional and dependent on style, e.g. factors like tempo and phrasing.

To sum up the most important aspect of the figure:
You see a basic lexical pattern which is deformed in various ways by sentence intonation. What is constant in these deformations? I claim that the relation of turning points to the segments and to the grid is an important invariant (Gårding 1984a). This rising tone over niũròu, for instance, is rising in the falling intonation of the last utterance of the figure if you see it from the lower grid line but constant in terms of absolute fundamental frequency.

I will now show you examples of my analysis of English intonation. Figure 3a shows yes-no questions of varying length, from the top to the bottom, a long line, a minimal line and a millionmile-long line. The voice is Mark Liberman's, the grids are mine.[3] The adjective is focussed in all three sentences. The grid helps us see that the turning points of the accented syllables in the rising part of the sentence are low. In the corresponding statements (Fig 3b) with falling intonation the grid shows that the turning points are high. This is in fact a general rule of English and German. For these languages only the accent location is distinctive, not the accent type. The turning points are fixed in location relative to the segment but they alternate between high and low in relation to the grid depending on the sentence intonation.

It is also evident that the slope of the rising grid is dependent on the length of the utterance. This suggests that it is the level of departure and the level of arrival which remain constant and carry the communicative function, not the steepness of the slope.

Another point worthy of observation is that the phrase-internal accented items have no autonomous tonal movements. This is in accordance with a very general rule that reduces the tonal component of phrase-internal accents. We shall see from my next examples that this rule operates also in Swedish. So let us now compare English and Swedish in a similar set of Swedish yes-no questions with focus in three different places (Fig 4a).

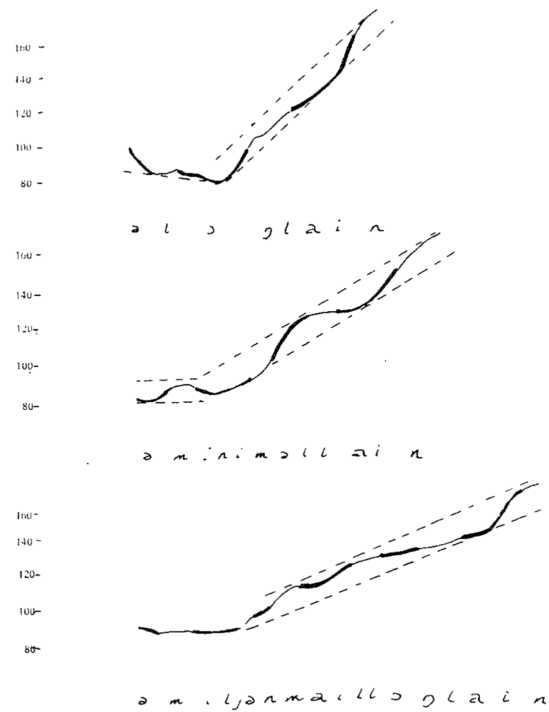


Fig. 3a F0 curves from American English yes-no questions of varying length with focus on the adjective.

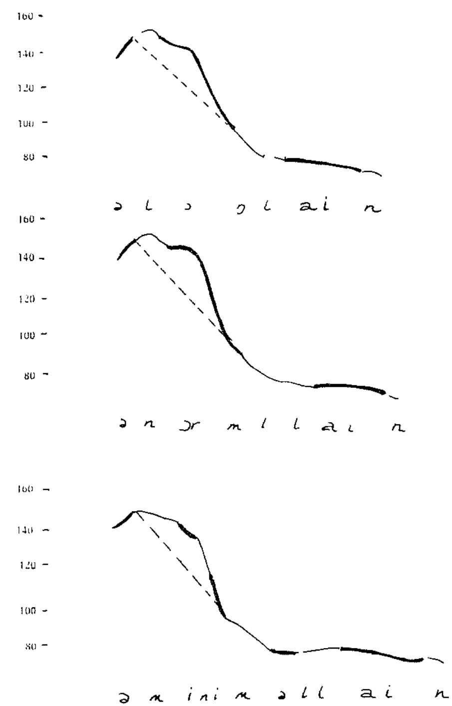


Fig. 3b Statements of varying length with focus on the adjective

Focus is expressed by a pivot and it is the direction of the grid after focus which signals the question. Again it is clear that the slope of the grid depends on the length of the part of the utterance following the focus. The accents are grave ones (Accent 2) and since the dialect is South Swedish, the turning points occur at the end of the accented syllable. (With acute ones the turning points would occur at the beginning of the vocalic segments.) Now, probably since Swedish has distinctive accent types, the turning points retain their locations not only in relation to the segment but also in relation to the grid regardless of the sentence intonation. A high remains a high, a low remains low as is shown by the comparable statements of Figure 4b.

My final examples are from Arabic (Fig 5). The statement has a falling grid over the predicate and a rising one over the subject with extra weight and an expanded grid for Munir. The question has a rising grid over the predicate and a falling one over the subject, still with extra weight for Munir. In the third sentence both the head and the modifier of the subject are focussed in an expanded grid.

It is interesting to note that also in this language which does not have distinctive accent types like Swedish, the turning points are fixed, even more so than in Swedish.

Using our descriptive framework (Fig 1) we can observe that similarities between languages are great for global features, thos expressing speech act, including focus. And we can assert that the fact that intonation sounds so different from language to language must depend on the fact that languages use intonation lexically in different ways and that speakers use their registers differently. Our Chinese speaker, for instance, used two octaves in a declamatory style, whereas our other informants only used one. Another important similarity of intonation in different languages is that speech acts can be expressed globally and/or locally. A question, for instance, may be given global expression by an increase of overall frequency and tempo, features which may be combined with a local rise at the end of the utterance. The speech act information is well protected by this redundancy.

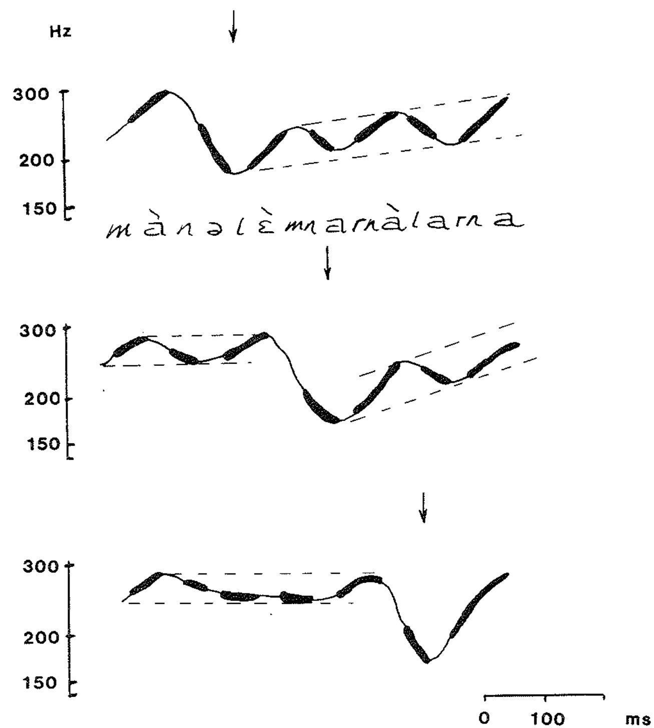


Fig. 4a F0 curves from South Swedish Manne lämnar nallarna 'Manne gives up the teddy bears'.
Yes-no questions with focus in different positions.

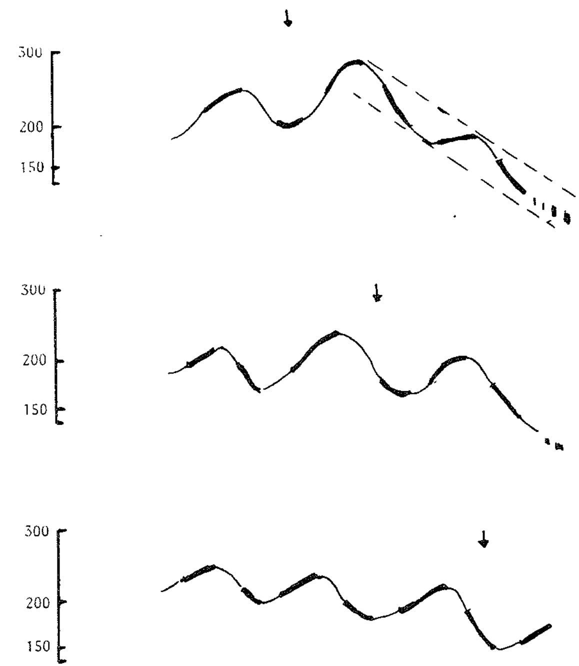


Fig. 4b Statements from South Swedish. From Anne-Christine Bredvad-Jensen (1984).

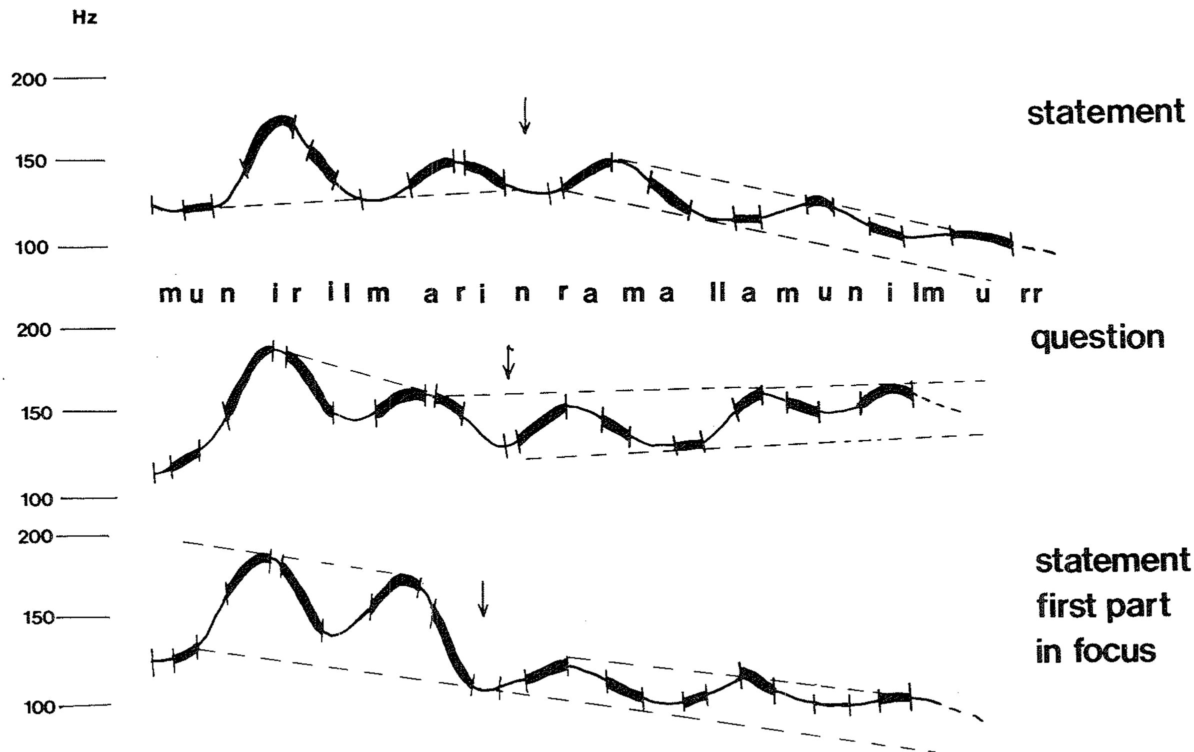


Fig. 5 F0 curves from Cairo Arabic <u>Munír ilmárin ráma llamún ilmúrr</u> 'The lithe Munir throw the bitter lemons away'. From Kjell Norlin

Generative aspects

So far pure description. I will now show you how this descriptive model can be used to generate intonation curves (Fig 6).
The input is a text with lexical phonological markings, i.e. markings for accents and tones, markings for speech act, dialect and syntactic boundaries. The main principle is that the global expression, the grid, is generated first. Then the turning points for accents and tones are inserted as specified by the dictionary in terms of highs and lows or combinations of highs and lows. Since these specifications are lexical, i.e. corresponding to citation forms, they are provided with statement intonation and have to be adjusted both for frequency and time depending on the position in the phrase and the syntactic and pragmatic context.
Finally the fundamental frequency curve is obtained by smooth interpolation over the voiced segments between the turning points.

Evaluation

It is said about models that they should be general and possess explanatory power (see e.g. in Lars Gårding, 1977). I have tried to show that this model is general by applying it to four languages with different prosodic systems. In addition it has been applied with some success to French, Greek (Gårding, Botinis, Touati, 1982) and Hausa (Lindau Webb 1982). There may be some reasons for this success. It is natural that the distinctive properties of words, i.e. accents and tones, should be retained in current speech when needed but subordinated to the global movement which is the basic speech act. In this process, intonation uses a common principle for combining movements from different sources, in particular small movements added to large movements, superposition. Superposition may also play a role in perception. When we interpret intonation we make use of our ability to refer a local movement to a global one.[4]

It is also natural that the turning points should be fixed to the words or morphemes and be part of their lexical representations in the brain. Further it is natural that we have learned to identify these patterns also when they are deformed by phrase or sentence prosody. With this view we get a credible and

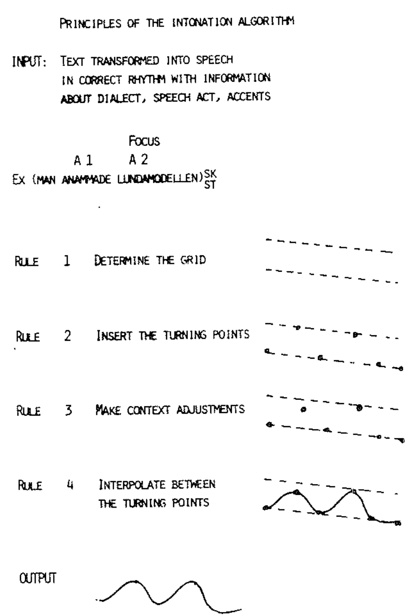


Fig. 6 The main principles of the generative scheme.

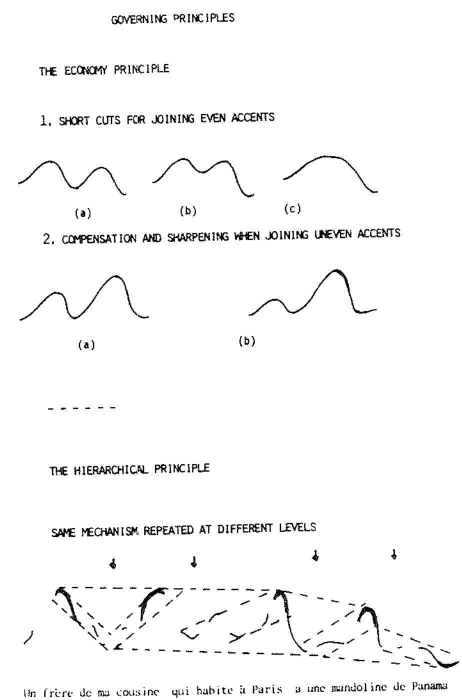


Fig.7 Some governing principles.

simple description of the existing variability of prosodic patterns. The number of basic elements in the model can be kept small.

As for the context rules, it is easy to see that the necessary adjustments express the tension between the speaker's wish to bring out the message and his urge to do it as economically as possible.

Figure 7 shows some examples of the economy principle. This principle explains why the joining of two accented words fem fem 'five five' can be as (a), (b) or (c) in the figure.

(a) has the meaning of fem, fem separated in a falling intonation

(b) corresponds to fem fem in a single phrase with the same intonation as before and

(c) is the intonation of a compound, other things being equal.

In other words, why go a long way when there is a shortcut? And why not use these three ways for communicative purposes so that (b) and (c) express an increasing degree of a coherence as compared to (a)? Another case governed by the economy principle is anticipation, exemplified under 2. A normal grid range is compressed before an expanded range. One can explain this in the following way. Let us save energy by compensating for a strong accent by making its neighbour weaker. At the same time we introduce a sharper contrast between the information weight of the words. These alternatives, 2a and 2b, can also be used to express different degrees of coherence, e.g.

(a) fråga Lund (b) frågalund 'ask Lund'

In 7.2.b the phrase (a) has been lexicalised as the title of a popular TV program.

An overall impression emerging in analysis of intonation by means of the model is that intonation is hierarchically organized in the following sense. A mechanism first observed on the lexical level appears also on the phrase and sentence level. I have an excellent example from French where turning points appear on different levels. In spite of the fact that accents are considered to be levelled out in a French phrase, there are traces of them in the acoustic records. Figure 7 is a case in point. It displays the fundamental frequency curve derived from the sentence Un frère de ma cousin qui habite à Paris a une mandoline de Panama.[5]

We see, according to my analysis, turning points in the narrowest parts of the grid lying above the smallest phrase constituents, un frère and ma cousine. These parts of the grid have equal ranges, indicating that equal weight has been given to the two constituents. The boundaries of the constituents are marked by a pivot, which turns the direction of the intonation. Then follows a jump-down pivot that introduces the relative clause and another jump-down pivot marks the end of it. The grid over the verb phrase has a larger range over mandoline than over Panama indicating that mandoline is the head of the phrase and that the following attribute is subordinated.

We may perhaps be able to relate the model to certain types of brain damage. It has been said that the right hemisphere is responsible for prosody and that this should explain why people with damage in this hemisphere have monotonous speech without emotional colouring. In a project supported by the Humanistic Research Council, Christina Dravins has worked out a comprehensive test with the aim of testing whether the ability to use intonation is tied to the right hemisphere. Her results indicate that persons with right side brain damage do have difficulties with the global aspect of intonation and that their accents have small ranges. It is not far fetched to imagine that such levelled intonation should have some connection with the diminished ability of spatial orientation which is characteristic of such patients. According to Christina Dravins the common factor is perhaps the inability to apprehend and control a sequence of events globally and locally at the same time.

OBJECTIONS

I should like to finish with some points for discussion in connection with the model that I have presented.

1. Superposition

The subordination of the lexical intonation movements into global movements is a principle reflected in the acoustic records. The principle can also be used to generate fundamental frequency curves and to explain intonation structure didactically. But it would be an unpermitted simplification to see superposition as a direct reflex of the laryngeal system being superimposed on the subglottal system. Here, as is well known, there is an intricate interaction between the two.

2. The grid

A question I am often asked is the following: What rules are there for drawing grids when one generates fundamental frequency curves? At this point I must say that so far we have only constructed grids for certain model sentences. We do not know enough about how grids vary for the same prosodic context for different speakers, for different emotions, and for connected speech.

Another question that has to do with the grid runs like this: In a given record, how can you find the grid with its varying width, slope and length and how do you find the pivots? To do this one must know a great deal about the intonation and the prosodic structure of the language one is analysing. The purpose of the grid is to reproduce in a simple way the prosodic phrase structure of an utterance. This ambition is not the same as having a simple recipe, much less a computer program. So far we can only give approximate rules.

3. Turning point fixation

Not even in a language like Swedish with distinctive accent types or a language like Chinese with distinctive tones are the turning points present in all prosodic situations. Some of them disappear in fast speech or are masked in creaky speech. How then are we to recognize the words? From the context, of course, the prosodic one (the grid) as well as the semantic one.

There are certain displacements of turning points with increased speech effort and increased intensity which seem to be general in the languages that we have studied. This calls for further research and is a reminder of our lack of knowledge about the interaction between what happens under, at, and over the vocal cords and also a lack of knowledge about the role of intensity in intonation. So far we have only a rough approximation of varying intensity, namely grid width. We have also avoided disturbances from articulation by constructing materials with sonorant segments.

4. Quantification

The model presented here is qualitative and covers many situations. It has been criticized for not being quantitative. A full quantization would be a computer program for a text-to-speech system which follows the generative scheme of the model and produces natural sounding intonation from a text with phonological markings. We have made a start in this direction (Bruce and Gårding, 1978, Huber, 1985). In this connection I would like to add that without the background of some general qualitative model, quantitative modelling may not be so terribly interesting.

Smooth interpolation, for instance, has been the object of great interest among those working with synthetic speech. One can ask why, when straight interpolation works as well. The more interesting question would be, why do straight interpolations work so well. It is possible that this has something to do with the width of the perceptual time window.

5. Orientation

The model is production oriented. An important complement would be to orient the model perceptually and give a description containing elements necessary for perception. These ideas have been inspired by Carlsson and Granström's work with auditory spectrograms (Carlsson and Granström, 1982).We have now started some experiments with Chinese to that end.

6. Phonology/Phonetics

It has been objected that the model does not distinguish clearly between phonology and phonetics. In particular Fig 1 provokes this question. My answer is here that inasmuch as phonology means abstractions based on communicative function, then the concepts of Fig 1 belong to phonology. They are the underlying patterns which appear in different shapes on the surface depending on various rules which are either phonological or phonetic in nature. They might be exchanged for more compact labels such as High Low and combinations of High Low, first at the lexical level, then at the phrase level and last at the sentence level. And the pivots might be exchanged for Arrows in different directions of the kind used by the structuralists. The reason why I prefer my system is that I want the analysis to be substance-based.

7. Generality

We are sometimes accused of having a model which is so general that it says nothing. Here I disagree categorically.

The main concepts of the model, grid, turning point fixation and pivot are not trivialities. They are the firm anchors of an intonation model which we have found useful for every language that we have analysed.

Notes

1 A revised version of talks given at symposia in Göteborg (March 1985) and in Stockholm (May). The model has been described earlier (Bruce & Gårding, 1978, Gårding 1978, 1981, 1983, 1984).

2 This informant is what John Ohala called a blessed speaker, by which is meant a speaker who conforms to the ideas of the experimenter. Materials from three more speakers are analysed in Gårding 1985. They use a one-octave voice register but have similar characteristics as the informant presented here.

3 I thank Mark Liberman for interesting material which he contributed when I was a consultant at Bell Laboratories and which was used in a different context (Gårding & Liberman 1977).

4 The same faculty characterizes visual perception. Watching a skier running down a slope we can sense both his general direction and his turns.

5 Un frère de ma cousine undoubtedly the same as mon cousin!

References

Bredvad-Jensen, A-C.1984. Tonal geography. Geographical variation in declarative and interrogative intonation along the west coast of Sweden. In C.-C. Elert, I. Johansson, & E. Strangert (eds.) Nordic Prosody III, Papers from a symposium. Kungälv Umeå Studies in the Humanities 59, Almqvist & Wiksell

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

Bruce, G.& Gårding, E. 1978. A prosodic typology for Swedish dialects. In Nordic Prosody, Travaux de l'institut de linguistique de Lund, XIII

Carlsson, R. & Granström,B. 1982. Towards an auditory spectrograph. In The representation of speech in the peripheral auditory system, Elsevier Biomedical. Amsterdam, New York, Oxford.

Carlsson, R. & Granström, B. 1973. Word accent, emphatic stress and syntax in a synthesis by rules scheme for Swedish. Speech Transmission Laboratory QPRS 2/3, p. 31-35

Dravins, C. Prosodi - språk, känslor och attityder - en stüdie i höger hemisfärs roll för kommunikativa funktioner (prel titel). Forthcoming.

Gårding, E. 1962. Olikheter mellan svensk och engelsk intonation i påståendesatser. Moderna språk 56, p. 269-280. English teaching abstracts. British Council 2:3.

Gårding, E. 1978. Satsintonation i svenska. In Svenskans beskrivning 11. Josephson, Strand & Estman (eds.), Stockholm University, Dept of Nord. Lang. p. 81-93.
Sentence intonation in Swedish, Phonetica 36: 207-215

Gårding, E. 1981. Contrastive Prosody: A model and its application. Studia Linguistica Vol 35, no 1-2: 146-166

Gårding, E. 1983a. A generative model of tone and intonation, In Prosody: Models and Measurements, Berlin. Cutler A & Ladd D R (eds.), Springer Verlag

Gårding, E. 1983b. Intonation units and pivots, syntax and semantics, Abstracts. Tenth Int. Congress of Phonetic Science, Foris Publ. P. 99-105

Gårding, E. 1984a. Superposition as an invariant feature of intonation. MIT conference on Variability of speech signals. In press.

Gårding, E. 1984b. Comparing intonation, WP 27, Dept of Linguistics, Lund University

Gårding, E. 1985. Constancy and variation in Standard Chinese tonal patterns, WP 28, Dept of Linguistics, Lund University

Gårding, E. Botinis, A. & Touati, P. 1982. A Comparative Study of Swedish, Greek and French Intonation, WP 22. Dept of Linguistics, Lund University

Gårding, E. & Liberman. M. 1977. English intonation by rules. Unpublished manuscript. Bell Telephone Laboratories.

Gårding, E. & Lindblad, P. 1973. Constancy and Variation in Swedish Word Accent Patterns, WP 7, Dept of Linguistics, Lund University

Gårding, E., Zhang, J., & Svantesson, J-O. 1983. Tone and intonation in Standard Chinese, WP 25, Dept of Linguistics, Lund University

Gårding, L. 1977. Encounter with mathematics. Springer Verlag

Huber, D. 1985. Swedish Intonation Contours in Text-to-Speech Synthesis, WP 28, Dept of Linguistics, Lund University

Mc Allister, R. 1971. Predicting physical aspects of English stress. Speech Transmission Laboratory, QPSR, p. 20-29

Meyer, E.A. 1937. Die Intonation im Schwedischen I: Die Sveamundarten. Studies Scand. Philol.10. University of Stockholm

Norlin, K. Fonologisk och fonetisk beskrivning av kairoarabiska. Forthcoming

Thorsen, N. 1983. Two issues in the prosody of standard Danish. In Prosody: Models and Measurements, Berlin. Cutler A & Ladd D R (eds.), Springer Verlag

Constancy and variation in Standard Chinese tonal patterns

Eva Gårding

INTRODUCTION

For some time now we have been studying Chinese intonation in Lund. This work is part of a project sponsored by the Swedish Research Council which aims at giving phonetic descriptions of some non-European languages, Chinese, Arabic, and Hausa.[2]

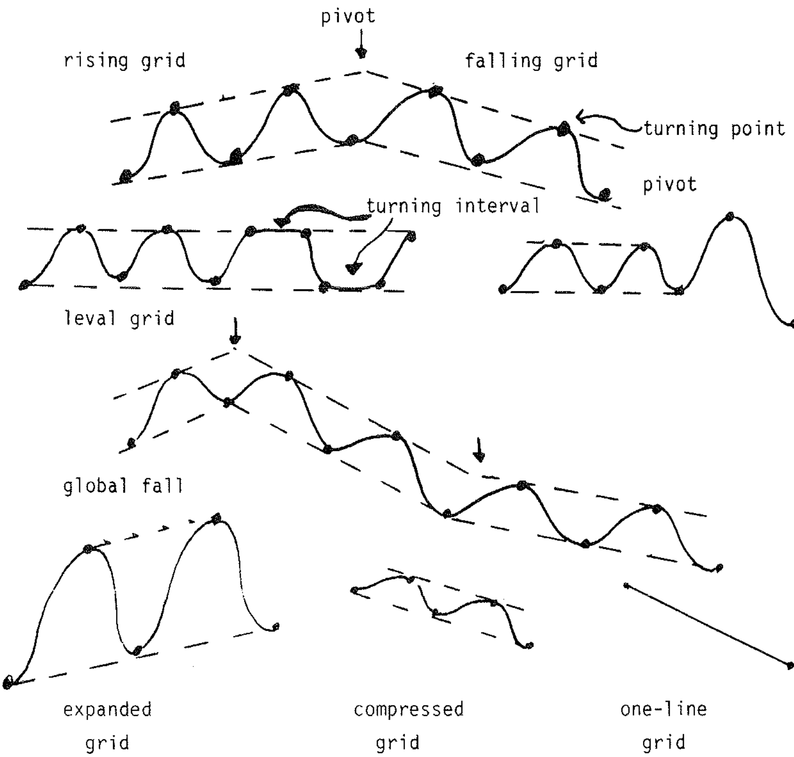
My personal interest in the project is primarily to test if the intonation model developed for Swedish dialects could be used for diverse prosodic systems (Bruce & Gårding, 1978, Gårding, 1981).

Our first study of Chinese tone and intonation was based on data from one Standard Chinese speaker (Gårding et al., 1983, Gårding, 1983). In this paper I shall present data from three more speakers of the same dialect.

Descriptive frame

Figure 1 illustrates the main features of the descriptive frame we have used, in particular the turning points, the grid and the pivot.

With good accuracy, an intonation curve can be reconstructed by smooth interpolation over the voiced segments between its turning points, i.e. the local maxima and minima and the beginning and end points. Turning points may also be end points of flat intervals called turning intervals. All these points are part of a global pattern, the grid, which is most easily seen if the main maxima are connected by a topline and the main minima by a baseline. In ideal cases, as in the figure, the grid appears



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

Fig. 1 Concepts of the model and their communicative functions, illustrated by schematic F0 curves connecting turning points over sonorant segments. Grids are marked by broken lines, pivots by vertical arrows.

as a sequence of units which are clearly rising, falling or level. A grid can be normal, expanded or compressed, even to the extent of being best represented by one line only. Grids of careful speech may have four lines, an upper and a lower topline and an upper and a lower baseline.

The part of the grid where the direction or width is changed or where the grid takes a jump is called a pivot. The lower part of the figure shows the communicative function of these features. The local turning points signal words or morphemes. The pivots serve as semantic and syntactic boundaries. The general direction of the grid of a sentence, often in combination with the direction of its final part, is associated with speech act type and sentence type. The levels (lines) of the grid vary according to style of speech, and the use of the levels differs across idiolects, dialects and languages. The width and position of the grid signal its information weight relative to other intonation units (Gårding, 1983c).

When the model is used for generative purposes, the global features, the grid, are generated first. Then the local maxima and minima pertaining to lexical accents or tones are inserted as points into the grid according to specific rules which state how the points are aligned relative to the segments. The fact that this alignment is relatively constant makes this arrangement of the rules convenient. In the last step the fundamental frequency curve is obtained by smooth interpolation. The generative part of the model is not the topic of this paper, but there will be some comments on its feasability in the light of the new data.

Material

Our material, presented in Figure 2, has been selected to investigate the interaction of tone and intonation.

A Chinese morpheme of the standard dialect can have one of four tones: Tone 1 called 'high' or 'level', Tone 2 'rising', Tone 3 'low' or 'dipping' and Tone 4 'falling'. There is, in addition, phonemic stress and the nuclear stress falls on the last stressable item if nothing else is indicated (Kratochvil, 1968).

MATERIAL

1. Wāng Yī chōu xiāngyān
   Wang Yi smokes cigarettes
2. Sòng Yán mài niúròu
   Song Yan sells beef
3. Chěn Lǐ mǎi yǔsǎn
   Chen Li buys an umbrella
4. Wāng Lǐ chuān yǔyī
   Wang Li wears a raincoat

PROSODIC CONTOURS

1. Focus free statement
2. Focus free yes-no question

   Statements:

3. after focus
4. focus left
5. focus right

There are 3 utterances of each sentence called a, b, c.

Code:

Li: 1.2.a means Li's production of tonal pattern 1 in prosodic contour 2, utterance a.

Fig. 2

Four tonal patterns have been chosen, three consisting of so-called contour tones, i.e. rises and falls, and one composed of level tones only.

The tonal patterns occur in a syntactic subject/predicate frame and have been pronounced as statements and questions. The statements have four different focus arrangements. Test sentence 1 represents a statement without any focussed part, an intonation used for example when the sentence introduces a short story. Test sentence 3 is a statement intonation occurring after a strongly focussed shì 'yes'. It expresses confirmation of known circumstances. Numbers 4 and 5 have been elicited by questions calling for focus to the left and right respectively. Number 2 is a yes-no question. The questions have not been asked in different focus arrangements.

Four speakers have been analysed. Each sentence has been pronounced three times as an answer to a question or elicited by a well defined and well described situation. The order of the test items has been randomized.

Informants

All of our informants represent Standard Chinese (Putonghua). Chen was born in Suzhou, Southern China, in 1948, and moved to Beijing at the age of six. He was an elementary school teacher until 1980, when he came to Sweden.

Li was born in Beijing in 1947 and studied languages and literature at a university there. He works as a translator.

Shi was born in Beijing in 1965. He grew up in Beijing where he completed his high school education. At the time of the recording he was a student at the Royal Institute of Technology in Stockholm.

Zhang was born in 1931 in the Liaoning province north-east of Beijing. At the age of 15 he came to Beijing where he pursued his university studies. He is professor of speech acoustics at the Academia Sinica, Beijing.

An impressionistic ranking of the speech style of the speakers from formal to informal would be: Zhang, Chen, Li, Shi.

Outline

The first part of the paper describes the material produced by Chen, who has been chosen as the prototype.[3] In the second part of the paper I shall summarize the most important similarities shared by all of our four speakers and also comment on some differences between them.

CONSTANCY AND VARIATION: ONE SPEAKER

Figure 3 shows one of the tonal patterns, Sòng Yán mài niúròu, 'S.Y. sells beef', produced by Chen, our prototype speaker. The tonal pattern is falling, rising, falling, rising, falling, in five different prosodic contours uttered three times in each. The three utterances have been lined up with the end of each utterance as a common time reference.

The uppermost statement shows a slightly falling general trend all through the sentence. The interrogative contour, that of a

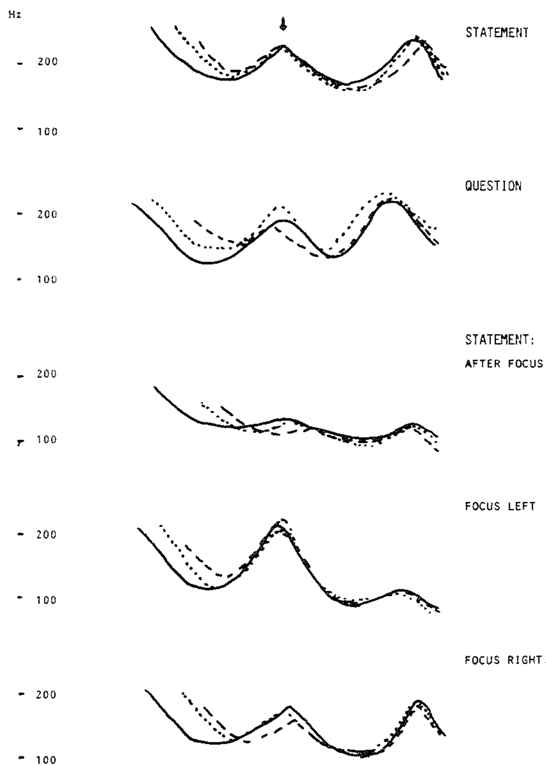


Fig.3 Chen: Sòng Yấn mài niúròu. 3 utterances in 5 prosodic patterns. Arrow marks C/V boundary of mài.

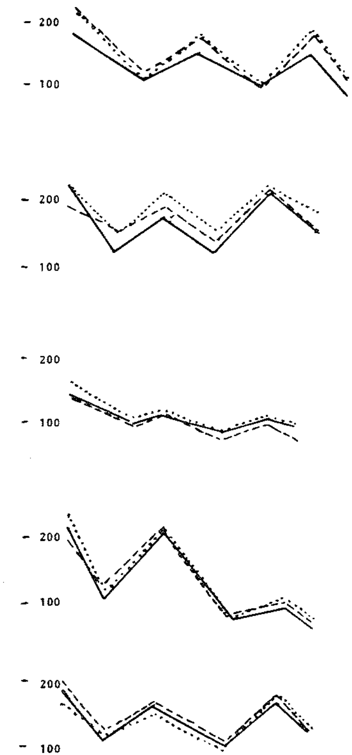


Fig.4 Chen: Sòng Yấn mài niúròu, normalized for duration (see text).

yes-no-question, is level over the subject and rising over the predicate.

The third contour, a statement referring to previously known circumstances, occurred after a strongly focussed shi 'yes' meaning 'yes of course'. The contour is falling as in the first statement but the level of the upper turning points has dropped half an octave and the range of the rising and falling intervals is compressed by half. The fourth contour, also a statement, is level or rising over the focussed subject and strongly falling over the predicate. The last one is falling over the subject, and the focussed predicate reaches a slightly higher level. This contour is similar to the first one, which reminds us of the rule giving nuclear stress to the last stressable item (see Introduction). We notice in passing that there is the same asymmetry in the pitch movements before and after focus as in other languages in that before focus there is much more fundamental frequency movement than after focus.

We can picture the prosodic contours of Figure 3 as global deformations of a basic tonal pattern. This illustrates in a neat way the aim of this part of the analysis: to describe the basic patterns and the deformations imposed by sentence intonation.

Temporal features

Let us now look at the durations. With the ending as a common time reference we notice that the temporal variation is largest at the beginning of an utterance. Some of the temporal variation is contour bound. The first statement contour, the one expressing new information, is the longest one. The question is somewhat shorter and the patterns with the whole sentence or part of the sentence unfocussed are the shortest ones.

Even within each group the durations vary quite a bit. In general the first utterance of the speaker is longer than the other ones. The temporal variation is somewhat irregular but it is clear, even if it is less conspicuous in this speaker than in the others, that the variation is not due to a uniform stretching and shrinking. What may happen when a speaker goes from a faster to a slower tempo is that the rate of fall and rise is kept rather constant and a turning point is exchanged for a turning interval. The presence and width of such turning intervals seems to be connected with consonants and syntactic boundaries. (Similar observations were made for Swedish by Gårding, 1975.) Variation of fundamental frequency patterns due to tempo changes are being studied separately.

Fundamental frequency features

To describe the variation of intonation contours we need to measure similarities and differences in some way. Among all possible alternatives, I have chosen to study the variation of fundamental frequency values of turning points (max and min), including beginning points and end points. This is particularly appropriate for sequences consisting of contour tones. In Figure 4, I have normalized the patterns for overall duration and durations between consecutive turning points. From what we can see here the curves are roughly parallel. This means that there is covariation of the FO values of these points and hence a certain stability in the configuration of the contour. A natural interpretation of this covariation is that it is the direction (or directions) of the configuration that is important for the speech act type and that it is possible to move it up and down on the frequency scale without changing its speech-act information.

Turning points related to segments

In Figure 5, I have marked the position of the turning points relative to the segments which have been normalized to a certain duration for the vowels and another for the consonants and consonant clusters.

It is quite obvious that the position of the turning points is remarkably constant all through the productions of all the prosodic contours. The falling tone starts at the beginning of the vocalic segment. There is one utterance in which the turning point has crept forward a bit into the vocalic segment. We may be able to correlate this displacement to increased speech effort.[4] The fall of the falling tone is completed before the beginning of the following vowel which carries a rising tone. Also the rising tone is connected with the beginning of the vowel. We draw the natural conclusion that it is the vocalic segment that carries the pertinent tone movement whether it is falling or rising. This is in agreement with certain statements

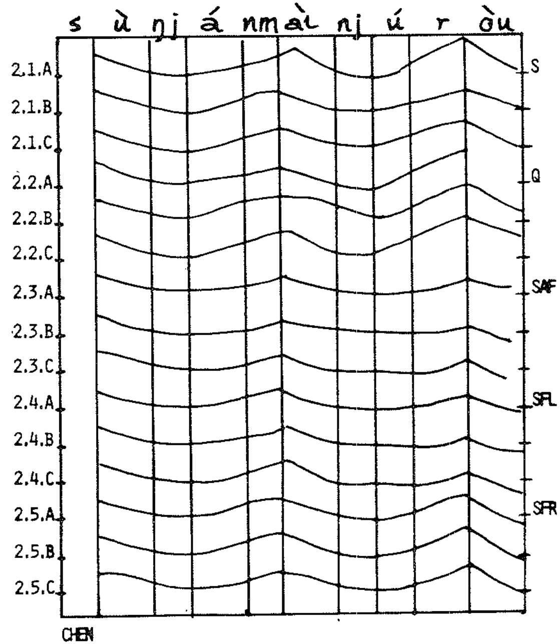


Fig.5 Location of turning points related to segments
Duration of segments normalized (see text).

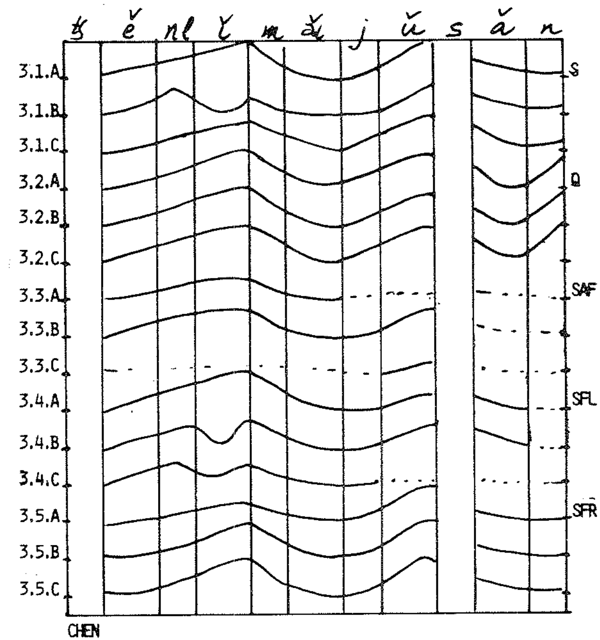


Fig.6 Location of turning points related to segments
Durations of segments normalized (see text).

in the literature (e.g. Howie, 1974, and Kratochvil, 1968) and in disagreement with others (Chao, 1968).

Figure 6 shows the location of turning points in another tonal sequence, uttered by the same speaker. In this sentence, mǎi meaning buy with a low (dipping) tone, appears in a similar position and context as mài with a falling tone meaning sell, which occurred in Figure 5.

We notice that the speaker has fixed the turning point all through the contours and utterances to the beginning of the consonant m. The relevant tone pattern, which we have labelled low, is reached in the middle or latter half of the vocalic segment.

There is also a variation of tone movements over the subject phrase. These are varying results of a sandhi rule, which will be commented on presently.

Considering the alleged importance of turning points, it is interesting to see what a tonal pattern consisting of only level tones looks like and how it behaves in the different contours. Figure 7 shows an all high (level) tonal pattern. Wāng Yī chōu xiāngyān, 'Wang Yi smokes cigarettes'. The curves of the utterances of the same prosodic pattern are rather parallel. As before, the tonal pattern has a constant shape and this shape has a certain leeway on the time scale as well as the frequency scale.

One can ask about the size of this leeway. How far down can a speaker place this all-level pattern without running the risk of being misunderstood?

It turns out that the problem of keeping an all-level pattern distinct from e.g. an all-low one never becomes critical. The reason is that the all-low one has a sandhi rule attached to it which changes any low-low sequence into a rising-falling one. (In this way the low tone really appears 'dipping' as one of its names indicates.) This is the reason why the all-low pattern rises and falls the way it does in Figure 6. The sandhi rule says T3 → T2/ - T3, but this rule does not apply over deep syntactic boundaries, at least not in the style of speech used here (cf. Cheng, 1973). We notice that speaker Chen adheres to this rule (as do the other speakers). After the boundary between noun-phrase and verb-phrase, măi never becomes rising. The noun-phrase, on the other hand, has a variety of manifestations, rising or level all through or rising-falling-level (Fig. 6).

Pivots

Figure 7 also shows that there are bends in the all-high pattern. Some of them are caused by disturbances of the laryngeal tone in connection with the consonants, but others are pivots, associated with the syntactic/semantic structure of the sentence (see Introduction). In the figure pivots occur in connection with the boundary between the noun-phrase and the verb-phrase.

Interaction of tonal patterns and intonation contours

Figure 8 completes the story of Chen. It shows four tonal patterns on top of each other for the five different prosodic contours. Each curve has been chosen as the most representative one out of the speaker's three productions of each sentence (see Appendix).

The tonal patterns are represented as follows.

Unbroken line ——— level tones
Dotted line ..... falling-rising tones
Dashed line --- low tones, by sandhi rule converted to rises and falls
Dashed and dotted line -·-·-·-· level and low tones, phonetically rising and falling with a lingering on high frequencies and the low frequencies often accompanied by creak, which is marked by vertical pieces of line.

From this figure it is possible to grasp the different impositions of the speech acts on the tonal patterns.

In most contours the level tone pattern seems to represent an intonational mean around which the other curves move. The overall tendency for statement intonation is falling. It seems that a corresponding grid should have four levels for the contour tones but the material does not permit precision at this point.

The interrogative intonation is rising except for the low-tone sequence, in which the question is signalled by a local terminal rise, lengthening the pattern of a contour which otherwise

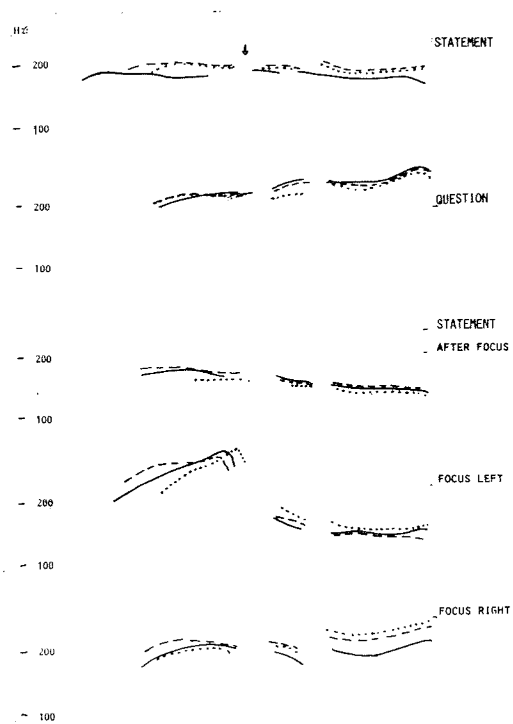


Fig.7 Chen: Wāng Yī chōu xiāngyān. 3 utterances in 5 prosodic patterns. Arrow marks boundary between subject and predicate.

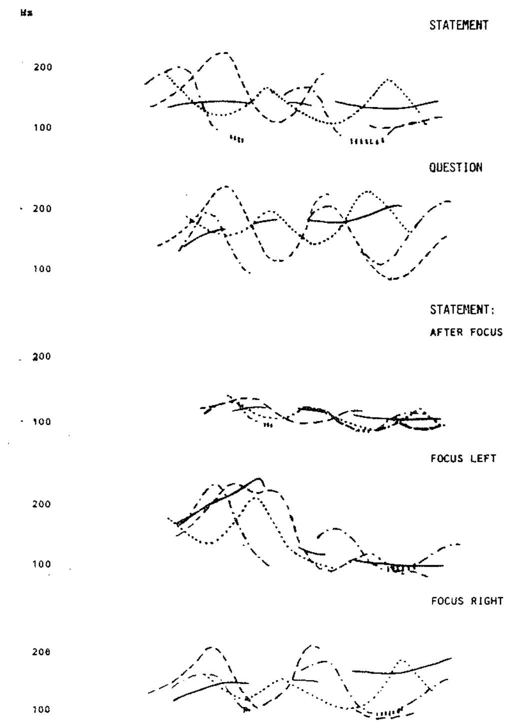


Fig.8 Chen: 4 tonal patterns in 5 prosodic contours. For code, see text.

looks like a statement. In our material, this is the only exception to the general rule that tonal patterns adjust to an intonation contour.

The statement after focus has characteristics similar to the first statement but here all the tonal events occur in a compressed grid. Even in this compressed state there are traces of the four levels, however.

The last two patterns will be treated together. One striking feature is the asymmetry between the effect of focus on the first phrase as compared to the focus effect on the last one. In both cases focus is manifested by an expanded grid, but the compression is considerably greater after focus than before. This feature seems to be general across languages (Gårding, 1983. For Swedish, see Bruce, 1977 and 1978). With focus to the left, the level tone pattern leaves its mean position and reaches the same high level as the other tonal patterns. This may be the optimal way of bringing out focus in a sequence of level tones where the speaker cannot expand a local up and down tonal movement.

CONSTANCY AND VARIATION: SEVERAL SPEAKERS

Tonal patterns and intonation contours

The figures in the appendix (A1-A12) show the test sentences produced by Li, Shi and Zhang and the remaining ones by Chen. The voices of the speakers have different ranges from Zhang's close to two octaves to Shi's less than one.[5] The tonal patterns are the same for all speakers in all situations except for Shi's after-focus statement. Here the last part of the utterance is creaky.

The intonation contours of the speakers have the same overall trends as before with falling statements and rising questions. The focus patterns are also similar with different degrees of compression depending on the position of focus.

As earlier the end of the utterances has been used as a common time reference, because, as was remarked already, the variability is much larger in the beginning of an utterance than in the end. There is for these speakers the same covariation of FO values of consecutive turning points as found for Chen. This gives stability to the configuration of the contour.

Interaction

Figures 9-11 show four tonal patterns on top of each other for the five different prosodic contours.

The following comments made in connection with Chen's intonation patterns remain valid:

The level-tone sequences represent some mean around which the contour-tone patterns move up and down.

The statements are in general falling and the questions rising.

The manifestation of focus is a combination of expansion and compression.

Li's high-level tone in the first part of the statement seems to be an isolated occurrence. Shi's statements are falling according to the general rule but his questions are level rather than rising. On the other hand they are much shorter than his statements. (According to Kratochvil it should be possible to produce a question just by quickening the tempo.)

The interaction between tone and intonation is clarified by the introduction of the descriptive terms of our model, especially grids and pivots.

Li and Zhang follow the focus rule for grids in all cases. Shi replaces a compressed grid in the latter part of the sentence by creaky voice just as he did in the after-focus statement.

The normal variability of repeated renderings of the same prosodic pattern makes it necessary to replace the idealized lines of a theoretical grid by preferably non-overlapping zones, which are nevertheless approximately parallel. The greater variability in fundamental frequency values of the high-frequency turning points of our material compared to the low ones (cf. figures in the Appendix) makes it natural to let the high zones be broader than the low ones.

Figure 12 shows grids for the questions of all our speakers. The scale is logarithmic for speaker Zhang to facilitate comparison between his two-octave voice and the one-octave voices of the other speakers. When looking at the lines of the grid, one has to bear in mind that they are supported by the most representative utterance and that in real life they correspond to

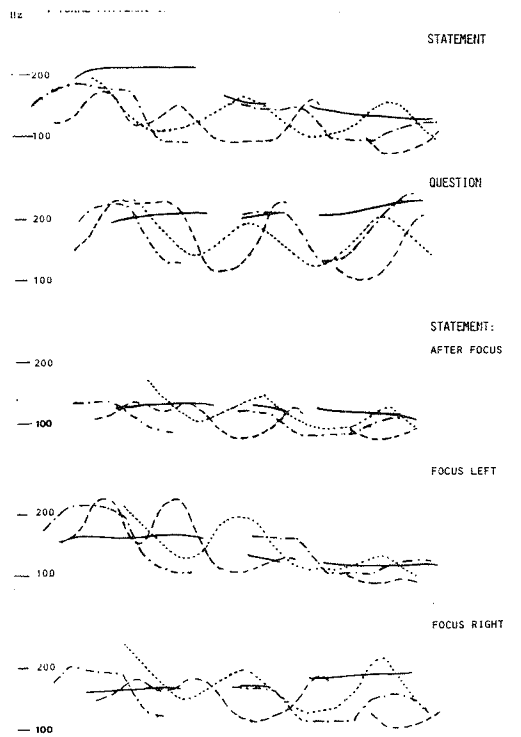


Fig.9 Li: 4 tonal patterns in 5 prosodic contours.
For code, see text.

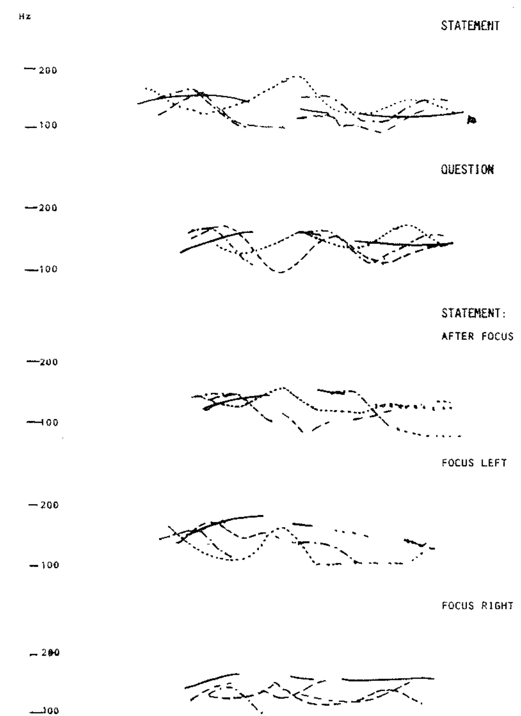


Fig.10 Shi: 4 tonal patterns in 5 prosodic patterns
For code, see text.

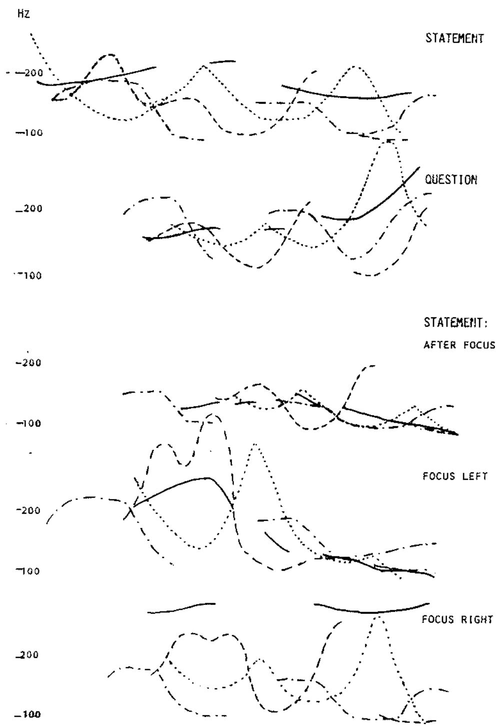


Fig.11 Zhang: 4 tonal patterns in 5 prosodic contours.
For code, see text.

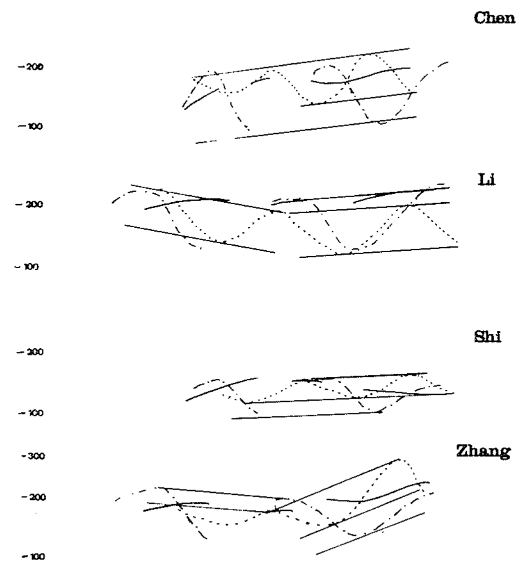


Fig.12. 4 tonal patterns in the question of 4 speakers.

zones due to the variability of the position on the frequency scale of the contours.

When constructing these grids I have not included the all-high (level) patterns, which, as I have mentioned earlier, seem to represent an average of a speaker's voice register. I have also left the all-low pattern aside because for this tonal pattern, all speakers, except Zhang, have used a statement contour combined with a local terminal rise to signal question. In our material this is the only case where a tonal pattern seems to influence the choice of an intonation contour. As a result the grids drawn in Figure 12 are only derived from the falling-rising and high-low tonal patterns. The grids have been obtained in the following way: I have connected the maxima and minima of the rising-falling pattern (dotted line) with the purpose of obtaining two parallel lines. This produces an 'interior grid' which seems to give a good indication of the general direction of the contour. We can see how Chen and Shi use rising grids, whereas Li and Zhang have falling-rising ones with a change of direction, i.e. a pivot, in connection with the subject/predicate boundary. It is obvious that the bidirectional grid is connected with a slower rate of speech. Hence the pivot is not obligatory.

With the parallel grid lines we also notice that the falling-rising subject phrase of Chen and Zhang has a compressed grid as compared to a more expanded grid enclosing the predicate phrase. This indicates that these speakers have given more weight to the predicate phrase in accordance with the nuclear stress rule of their dialect.

The credibility of these grids is perhaps so far a matter of judgement. In any case there are two tonal patters in this contour with three clear turning-point zones, as demonstrated in Figure 12.

Turning points and turning intervals

Figures AT2, AT3 and AT4 at the end of the appendix present the positions of turning points in relation to segments across speakers for the three sequences in which this feature is relevant. (The level tonal pattern has been left out for natural reasons.)

In the sequence Sòng Yǎn mài niúròu (Fig. AT2) the turning points for the rising and falling tones occur near the C/V boundary. The relevant tone movement, then, seems to start with the vocalic segment. A junction between a falling and a rising tone tends to produce turning intervals whereas a junction between a rising and a falling one seems to produce pointed peaks.

In the sequence Shěn Lǐ mǎi yǔsǎn, the sandhi rule invariably produces a rise for the first syllable of the compound yǔsǎn, starting from a low turning point at the beginning of the vocalic segment (Fig. AT3). The morpheme mǎi introducing the predicate phrase always has a fall from a high turning point in the beginning of the consonant [m] which continues to a low turning point in the first half of the following vocalic segment. In many cases this low point is followed by a turning interval. In the subject phrase, on the other hand, the sandhi rule has different results depending on both speaker and intonation. One result is a rising intonation over the whole phrase regardless of the individual word tones.

In the sequence Wāng Lǐ chūan yǔyī (Fig. AT4) the turning points have clear locations in most cases. The high and low tones seem to be well represented by high and low frequency levels in the vocalic segments. The exceptions, namely the slight rise over the first vowel, the incomplete tonal segments before the voiceless affricate and the end of the utterance are clear coarticulation effects. In many cases a low is accompanied by a creaky segment. The characteristic feature of the low tone, then, is an interval of low frequencies in the vowel after a sharp fall.

The observations above agree well with those made earlier for the prototype speaker. The constancy of the turning-point feature permits an easy specification of the alignment rules of accents and tones onto the grids and a convenient arrangement of the rules of the pitch algorithm with the global features first and the local ones later (see Introduction).

Perceptual testing, work in progress

The two sentences containing the sequence mai have been the object of a special study.

Figure 13 illustrates fundamental frequency curves from our speakers' productions of the sequence mai with a low or a

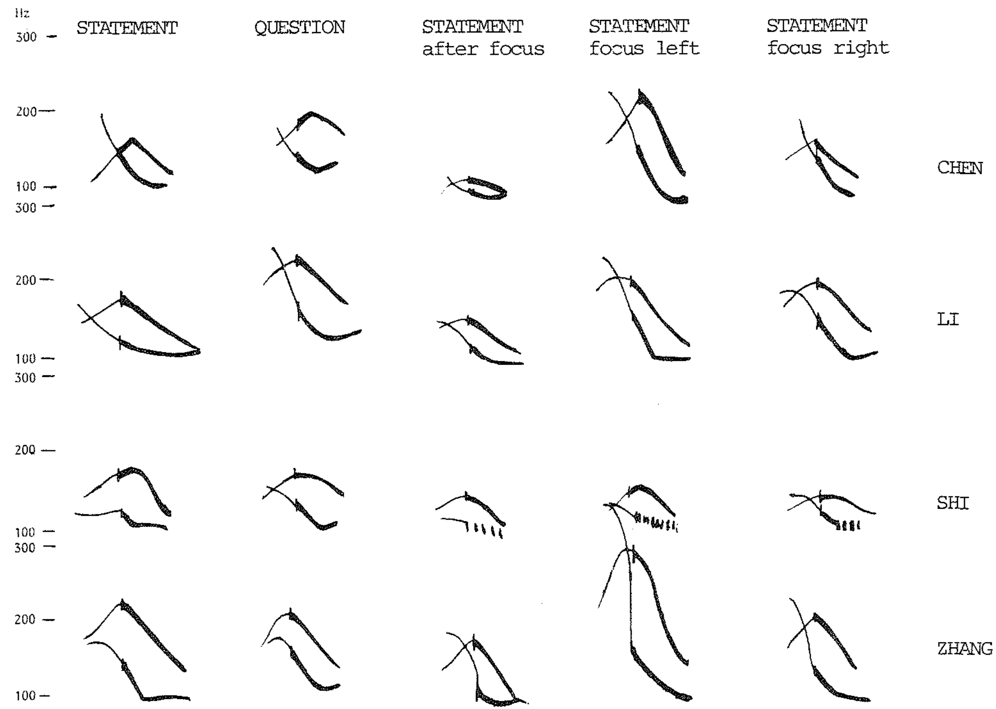


Fig. 13 Superimposed sequences of <u>mai</u> derived from 4 speakers' productions of five prosodic patterns. The falling tone 4 is above and the low tone 3 below. The C/V boundary is the time reference for each pair. The thick line marks the vocalic segment.

falling tone in all five prosodic patterns. The productions have been taken from a comparable context and are superimposed with the beginning of the vocalic segment as a common time reference.

The two tones have distinct manifestations regardless of speaker and prosodic pattern. The common feature of Tone 4 can be described as a fall from a rather high level at the beginning or near the beginning of the vowel. This fall is preceded by a rise including the syllable-initial consonant.

The common characteristic of Tone 3 is a fall reaching low level and sometimes accompanied by creak in the latter half of the vocalic segment. This fall is in most cases a continuation of a fall which started in the preceding consonant. In the last three prosodic patterns, Zhang's low-tone fall shows a discontinuity at the beginning of the vocalic segment as if the vocal cords went into another mode of vibration.

At this point we might perhaps venture the hypothesis that the perceptual characteristic of Tone 4 is a fall prepared by high frequencies in the beginning of the syllable (mainly in the consonant) and that of Tone 3 is low frequencies in the latter half of the syllable prepared by a fall.

Using the ILS program we have tried to change a falling tone into a low one by various maneuvers, e.g. by translating the falling movement of Tone 4 backwards in time or by introducing a simulated creak into the falling tone. It has been claimed that the creak or glottal stop is THE distinctive feature of the low tone. The reactions of my two Chinese-speaking collaborators do not support this claim (Kratochvil, Svantesson). A test tape has been evaluated by a group of listeners in Beijing who respond in a similar way. The results are presented in a separate paper (1985, Gårding, Kratochvil, Svantesson, Zhang).

This experiment marks a new direction in our research. We would like to give our intonation model, which is production oriented, a perceptual complement. And with all the experience we will gain from our analyses of Chinese, Arabic and Hausa we will return to Swedish for reconsidered analysis.

SUMMARY OF RESULTS AND CONCLUSIONS

Finally I would like to sum up our main findings.

1. The four speakers use the same tonal patterns with turning points very much fixed relative to the segments. As a consequence the main falls and rises are also fixed.

2. The tonal patterns have overall shapes characteristic of the five situations which have been used to elicit the utterances. In terms of the model we can say that the grids are similar except that the use of pivots may differ from speaker to speaker, obviously depending on tempo and style. However, when pivots do occur,they are in the same location.

3. In contrast to this constancy there is individual variation in voice range. One speaker has a two-octave voice in a declamatory speech style and the others use about one octave. There is also some individual variation in stress patterns. When focus has not been explicitly provoked to fall on a particular part of the utterance as in the first statement and the question, stresses are either equally distributed over the sentence or the last phrase has been given stronger stress in accordance with the nuclear stress rule.

4. There is also variability in the manifestation of the sandhi rule for Tone 3 but all four speakers agree in letting the rule be blocked by a deep constituent boundary.

5. The model we proposed for Chinese tone and intonation based on one speaker has been validated in at least the following sense: It serves as a convenient frame for further explorations of the interaction of tone and intonation in Chinese. At the same time we have strengthened our claim that the model gives a general frame for any prosodic system and that it sheds light on the structure of intonation in general.

NOTES

1) Modified version of talk given to the Prosody Club of the Department of Linguistics, Stockholm University, 15.11.1984.

2) My collaborators are for the Chinese part, Jan-Olof Svantesson and for Arabic and Hausa, Mona Lindau Webb and Kjell Norlin. I should also like to thank Jiālù Zhāng, Acoustics Institute, Academia Sinica, Beijing, and Paul Kratochvil, Faculty of Oriental Studies, Cambridge University, for help and advice. Jan-Olof Svantesson and Jíālu Zhāng composed the material and made the recordings.

3) The material was recorded in speech laboratories in Lund, Stockholm and Beijing and analysed by Spectrograph Kay Digital Sona-Graph 7800 and by the ILS-program implemented on VAX 730.

4) The effect of speech effort is being studied in a special material.

5) Zhāng has used a declamatory speech style which is common in recitals.

REFERENCES

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

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

Chao, Y.R. 1968. A grammar of spoken Chinese

Cheng, Chin-Chuan. 1973. A synchronic phonology of mandarin Chinese

Gårding, E. 1975. The influence of tempo on rhythmic and tonal patterns in three Swedish dialects. Working Papers 12, Department of Linguistics, Lund University, 71-84.

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

Gårding, E. 1983. 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. 1984. Chinese and Swedish in a generative model of intonation. In Nordic Prosody III, ed. Elert, University of Umeå

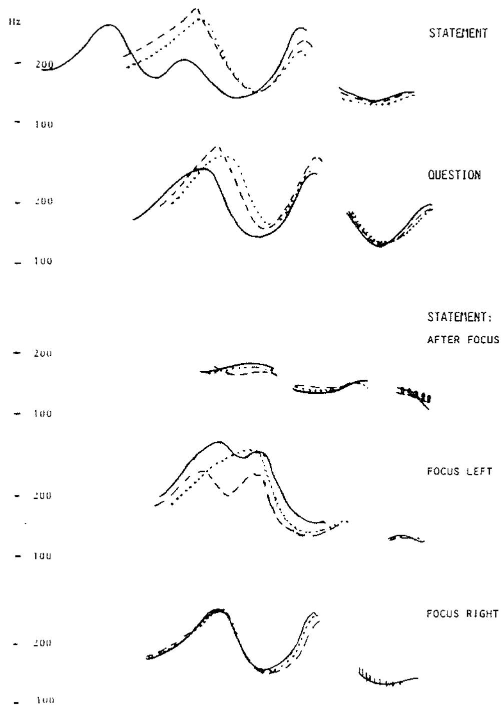
Gårding, E. 1984. Comparing intonation. Working Papers 27, Department of Linguistics, Lund University: 75-96

APPENDIX

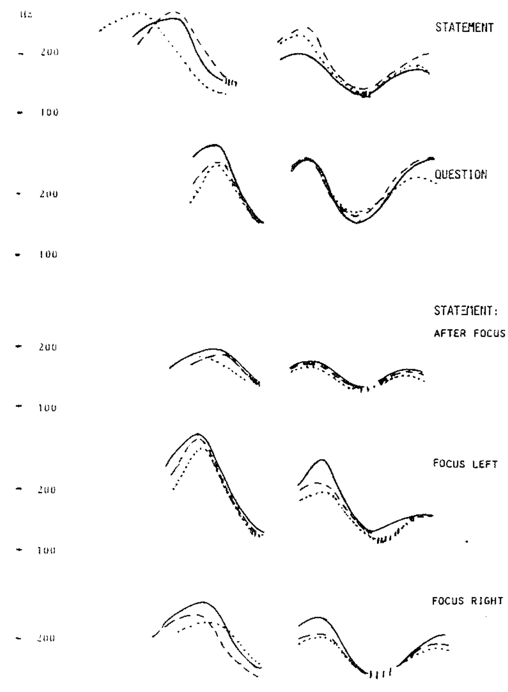
Here the complete material is collected as drawings of FO and diagrams of turning-point locations from spectrograms except parts given in the text (Chen 1 and Chen 2).

CODE

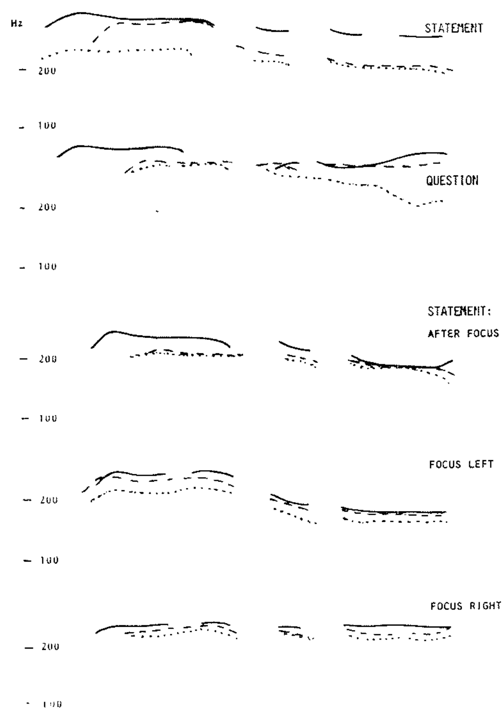
A = Appendix
1 = Wāng Yī chōu xiāngyān
2 = Sòng Yǎn mài niúròu
3 = Chěn Lǐ mǎi yǔsǎn
4 = Wāng Lǐ chuān yǔyī
T = Turning-point diagram
S = Statement
Q = Question
SAF = Statement after focus
SFL = Statement focus left
SFR = Statement focus right



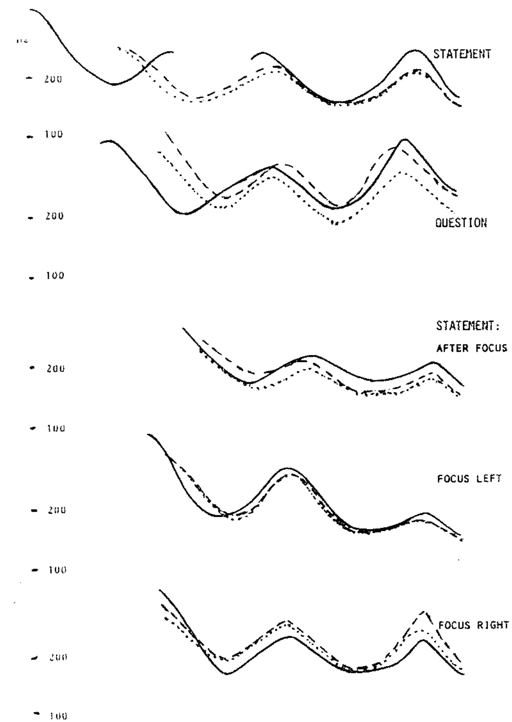
Chen A3: Chěn Lǐ mǎi yǔsǎn. 3 utterances in 5 prosodic patterns



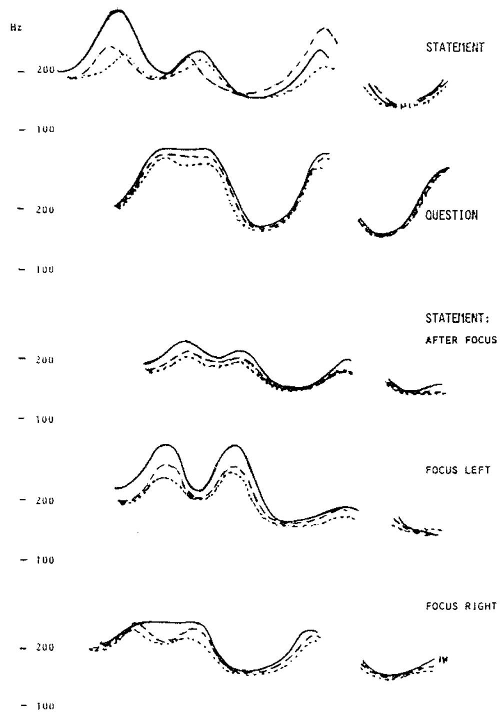
Chen A4: Wāng Lǐ chuān yǔyī. 3 utterances in 5 prosodic patterns



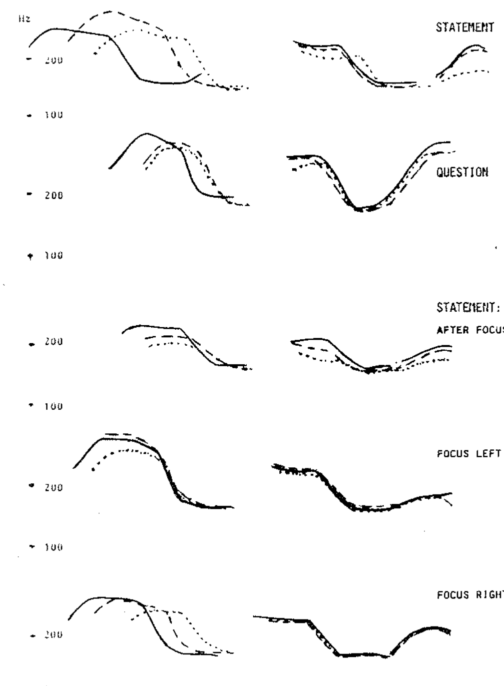
Li A1: Wāng Yī chōu xiāngyān. 3 utterances in 5 prosodic patterns



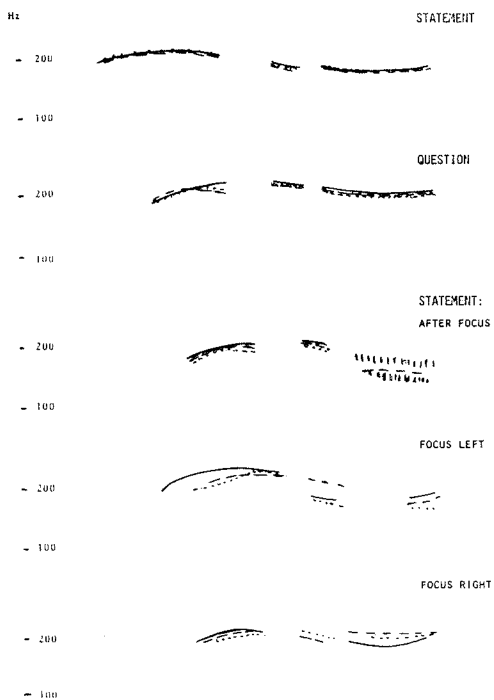
Li A2: Sòng Yǎn mài niúròu. 3 utterances in 5 prosodic patterns



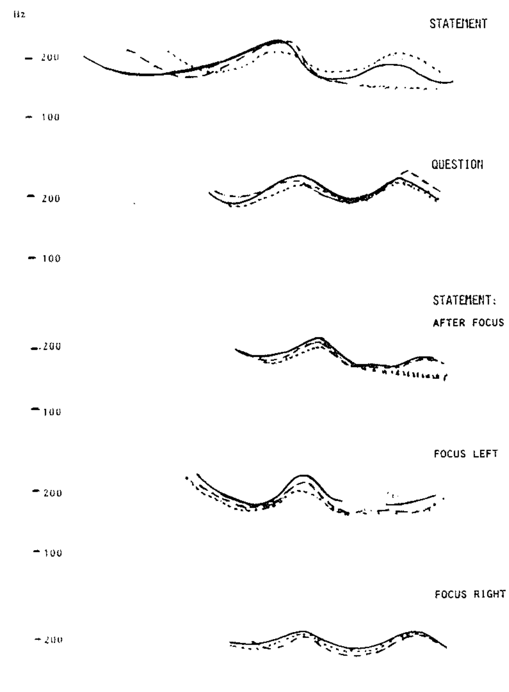
Li A3: Shěn Lǐ mǎi yǔsǎn. 3 utterances in 5 prosodic patterns



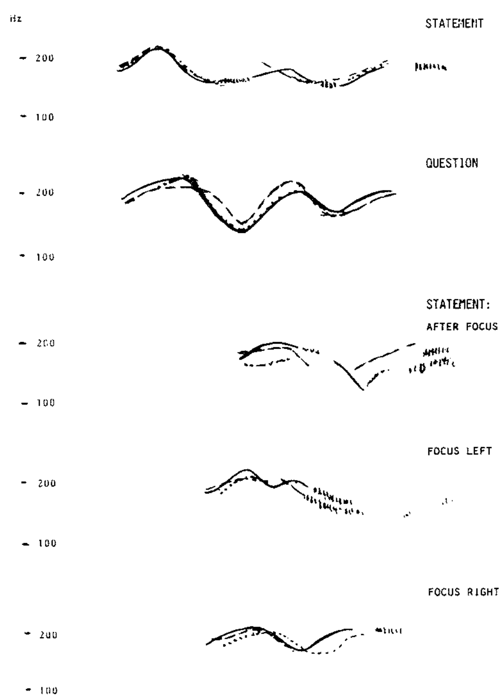
Li A4: Wāng Lǐ chuān yǔyī. 3 utterances in 5 prosodic patterns



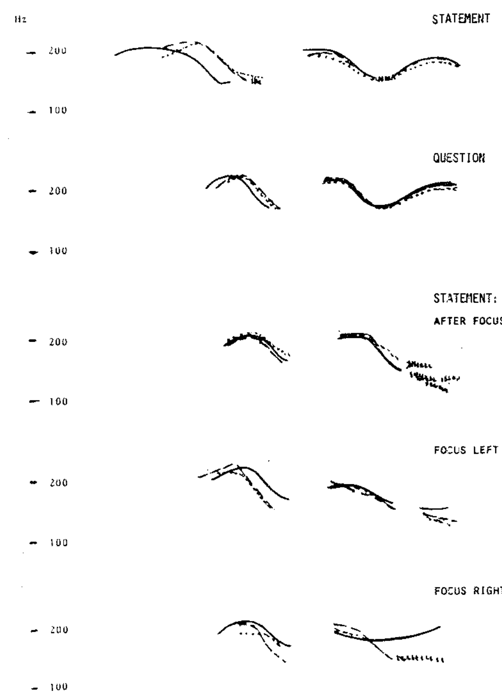
Shi A1: Wāng Yī chōu xiāngyān. 3 utterances in 5 prosodic patterns



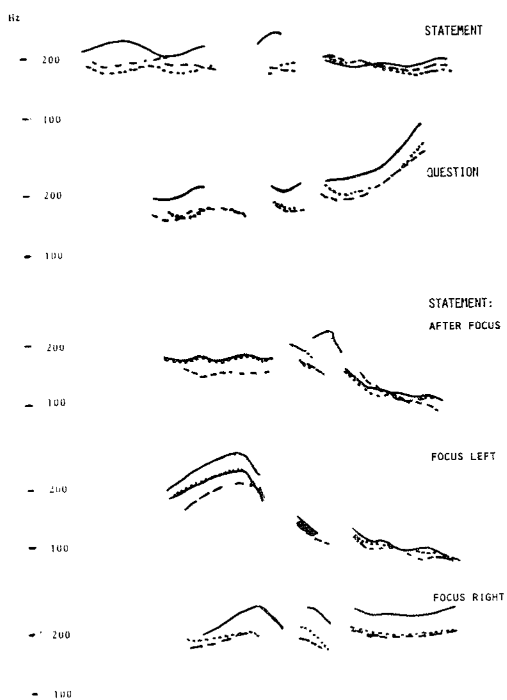
Shi A2: Sòng Yǎn mǎi niúròu. 3 utterances in 5 prosodic patterns



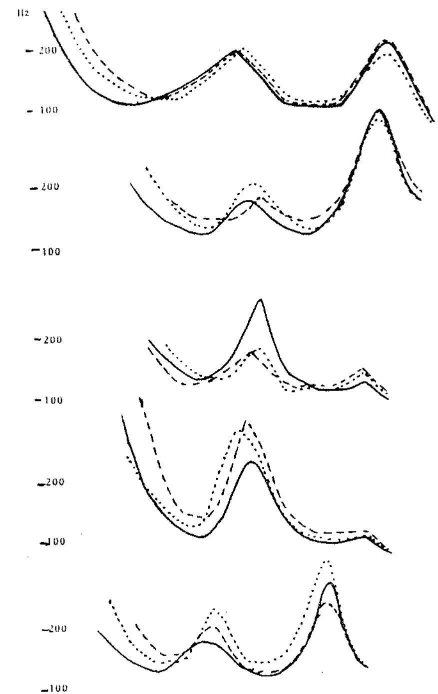
Shi A3: Chěn Lǐ mǎi yǔsǎn. 3 utterances in 5 prosodic patterns



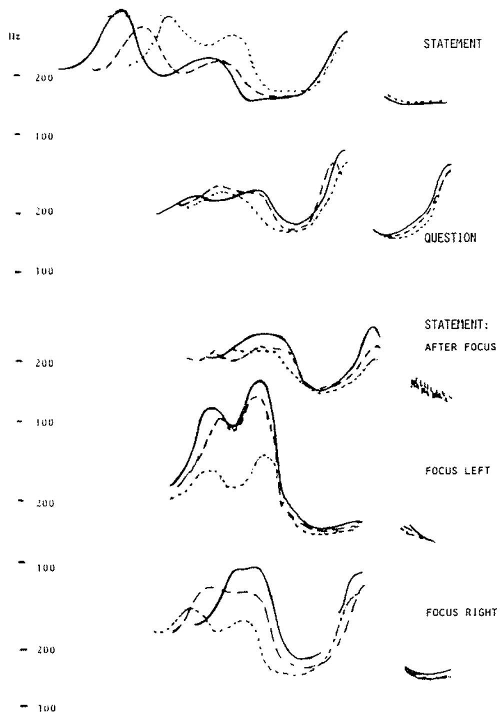
Shi A4: Wāng Lǐ chuān yǔyī. 3 utterances in 5 prosodic patterns



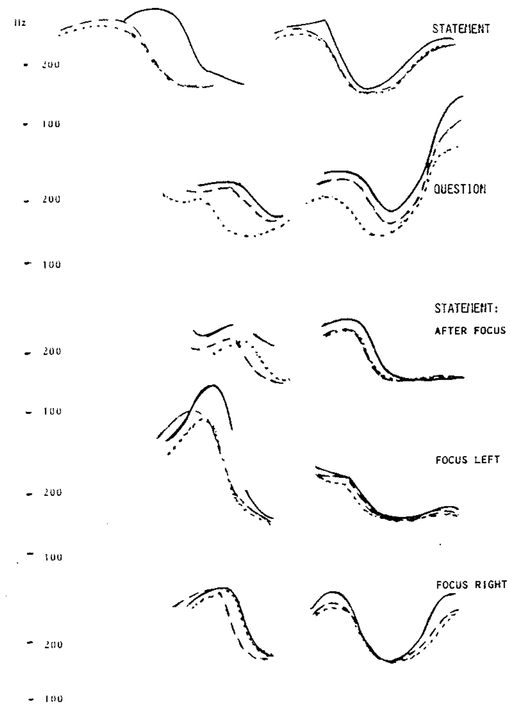
Zhang A1: Wāng Yī chōu xiāngyān. 3 utterances in 5 prosodic patterns



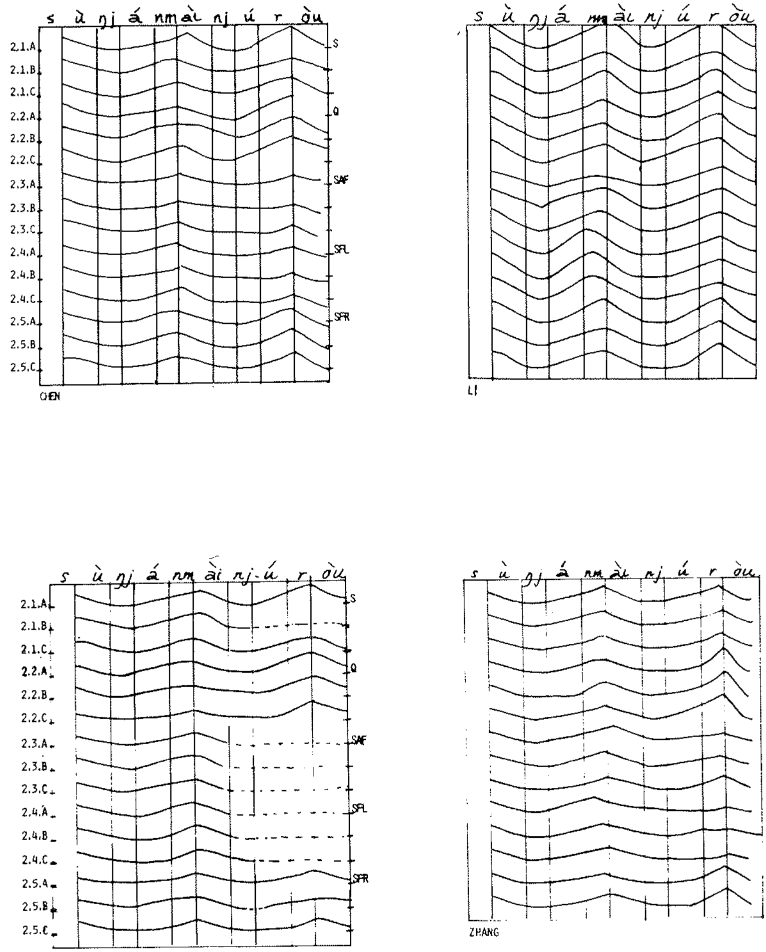
Zhang A2: Sǒng Yán mǎi niúrǒn. 3 utterances in 5 prosodic patterns



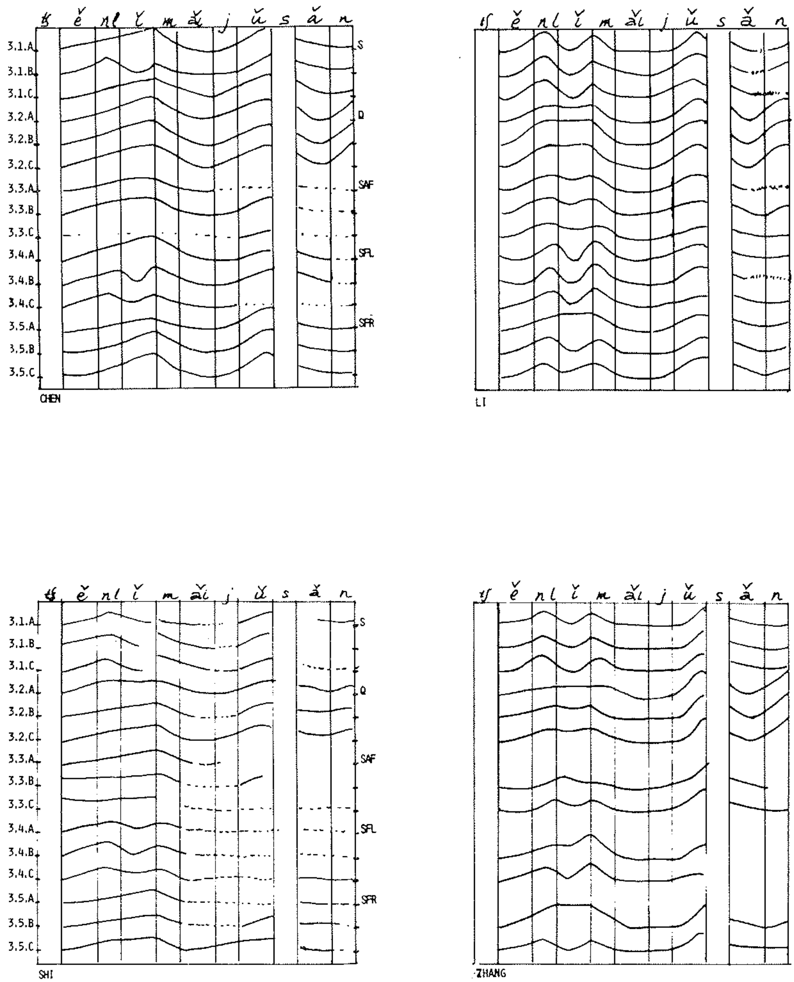
Zhang A3: Chěn Lǐ mǎi yǔsǎn. 3 utterances in 5 prosodic patterns



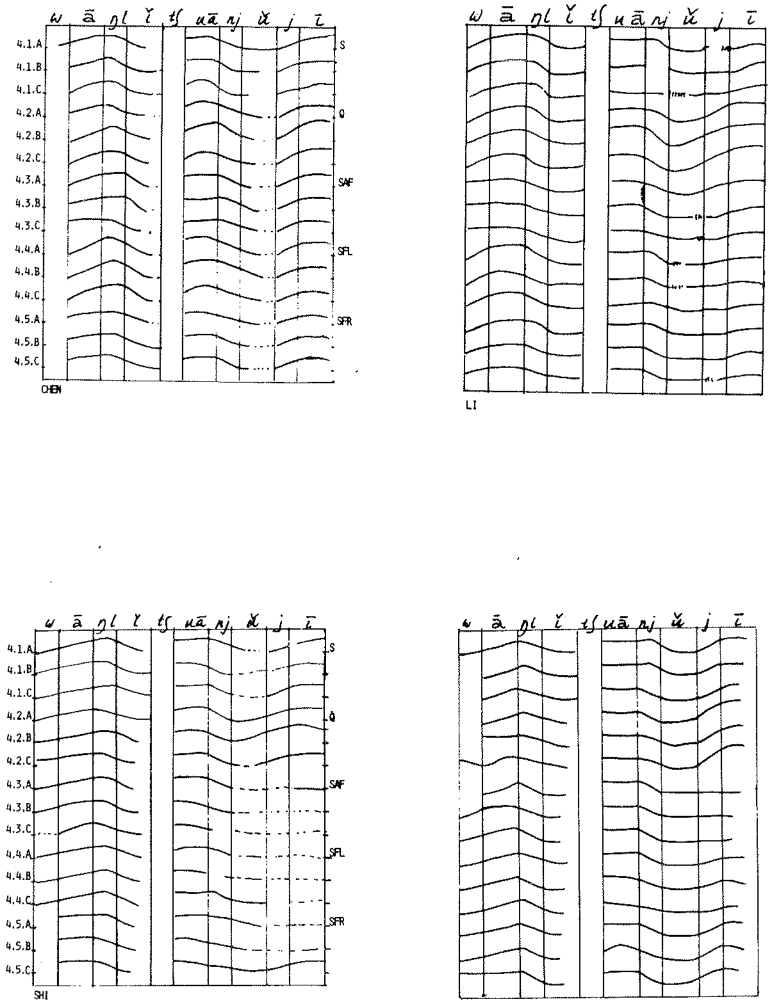
Zhang A4: Wāng Lǐ chuān yǔyī. 3 utterances in 5 prosodic patterns



AT2: Sòng Yán mài niúròu. Turning points related to segments. 4 speakers



AT3: Chěn Lǐ mǎi yǔsǎn. Turning points related to segments. 4 speakers



AT4: Wāng Lǐ chuān yǔyī. Turning points related to segments. 4 speakers

Tone 4 and Tone 3 discrimination in Modern Standard Chinese

Eva Gårding, Paul Kratochvil, Jan-Olof Svantesson and Jialu Zhang

INTRODUCTION

Unless affected by tone sandhi (see Cheng 1973:46-53 and Kratochvil 1985), the pitch contour of Tone 3 in the mid position in Modern Standard Chinese declarative sentences is generally falling, the same as that of Tone 4, but the pitch of the two tones differs in such respects as the relative level and the shape of the contour, as seen in the existing descriptions of Standard Chinese tones, such as Howie 1976, and in our own material (see Gårding, Zhāng and Svantesson 1983 and Gårding 1985). These two tones may also differ as regards the presence or absence of creak; see Hockett 1947:256, where glottalization (i.e. creak) is referred to as one of the properties of Tone 3.

It has, however, neither been known which, or the combination of which of these differences were distinctive, nor have the differences been understood clearly in quantitative terms. The aim of the experiment described in this paper was to examine the pitch characteristics distinguishing from each other Tone 4 and Tone 3 in such circumstances. It was hoped that the experiment would also contribute to the understanding of tone discrimination in continuous Standard Chinese speech in general.

Figure 1 shows the F0 contours of these two tones on

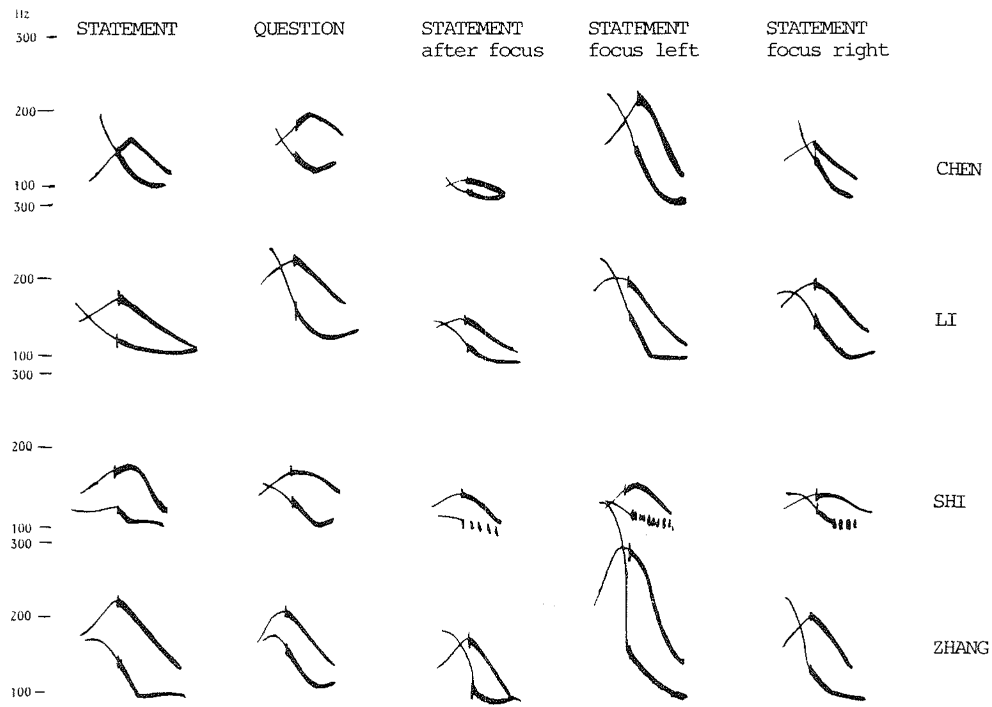


Fig. 1 Superimposed sequences of mai derived from 4 speakers' productions of five prosodic patterns. The falling tone 4 is above and the low tone 3 below. The C/V boundary is the time reference for each pair. The thick line marks the vocalic segment.

segmentally identical syllables (mài "sell" and mǎi "buy"), as said by four male speakers of Standard Chinese in five prosodic patterns. These syllables were said in comparable statements with stress equally distributed over the phrase.

Comparison of Tone 3 and Tone 4 in our material and in other studies shows that although the two tones are realized in different ways by different speakers, certain features seem to be invariant. Tone 4 generally starts with or quickly reaches a peak, from which it falls gradually towards a minimum which is often not reached until in the beginning of the next syllable. The third tone generally has a relatively low pitch level throughout the second half of the vowel. This can be achieved either by starting lower than for Tone 4, or by making the pitch fall quicker. These observations suggested various manipulations for changing Tone 4 into Tone 3 in our experiment.

BASIC DATA

The basis for the experiment was one of the isolated sentences Sòng Yán mài niúròu "Sòng Yán sells beef", mentioned above. Since the sentence is voiced throughout after the initial voiceless fricative [s] in sòng, its alternating Tones 4 and 2 gave its pitch contour the shape of a regular sequence of peaks and troughs. The key syllable mài "sells" coincided with the top of the middle peak (see Figure 2).

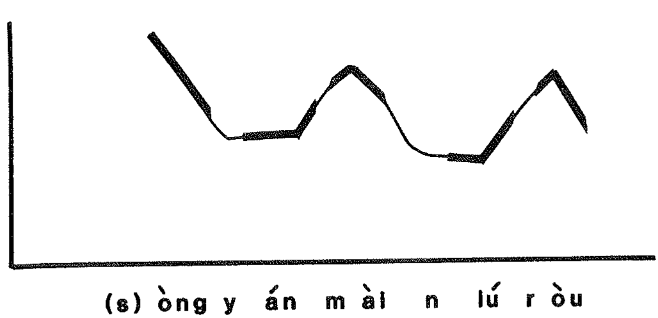


Figure 2. Stimulus 1, segmented.

This and the immediately adjoining parts of the sentence pitch curve were modified in various controlled ways, in order to find out what modification or modifications made the syllable become perceived as carrying Tone 3, i.e. as mǎi "buys".

The processing of the original sentence incorporating the modifications was carried out by using the programs of the ILS package. Although only the F0 in the relevant part of the sentence was manipulated, this resulted in some changes in the corresponding amplitude values due to the automatic adjustments carried out by the given program.

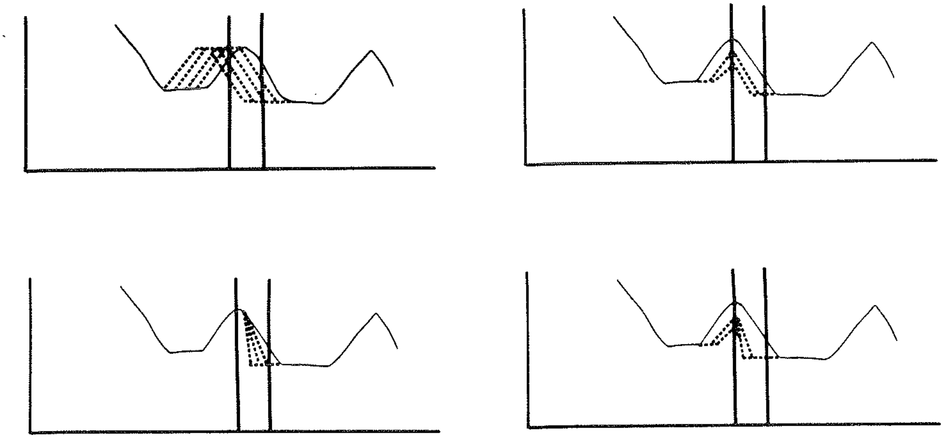
The F0 curve of the whole voiced part of the sentence was first extracted and segmented by referring to the waveform. In particular, the borderlines of the vocalic and nasal components in the area of the transition from the syllable yán to the syllable mài were established as precisely as possible. F0 has a maximum reached about 30 ms after the nasal-vowel borderline in mài, and then it falls to a minimum reached in the vocalic part iu of the following syllable niú.

The periodicity of the waveform was undisturbed throughout the whole syllable, and there was no sign of the kind of irregularities in it which signal creak.

MODIFICATION OF THE DATA

The F0 curve of the original sentence was manipulated in various ways, and each time the respective variant of the sentence was recorded on a magnetic tape, so that it could serve as a stimulus in the subsequent perceptual test. In all, 18 different stimuli were prepared.

The manipulation followed three main strategies as suggested from inspection of the realizations of the tones in our material. Figure 3 summarizes these modifications, which are labelled as follows:



Stimulus	Peak shifted	Peak lowered	Steepness increased	Creak added
1				
2	32 ms			
3	64 ms			
4	96 ms			
5	128 ms			
6				+
7	32 ms			+
8	64 ms			+
9	96 ms			+
10	128 ms			+
11		20 Hz		
12		40 Hz		
13			32 ms	
14			64 ms	
15			96 ms	
16		20 Hz	32 ms	
17		20 Hz	64 ms	
18		40 Hz	32 ms	

Figure 3. Design of the stimuli.

1. Peak shifted.

The middle peak was moved to the left (that is, in the direction of the beginning of the sentence) in four steps of 32 ms each (each step corresponding to 5 frames in the ILS program terms), without changing its general shape and the shape of the terminal peaks. The movement was thus achieved only at the cost of shortening the trough preceding the middle peak and lengthening the trough following it. See Figures 3 and 4 for the four steps of the movement to the left (stimuli 1-5).

2. Peak lowered.

The peak was lowered by 20 Hz (stimulus 11) and by 40 Hz (stimulus 12) without changing its shape and steepness, and without proceeding below the level of the surrounding troughs. The movement thus prolonged the troughs. Stimulus 2, i.e. step 1 of the leftward movement by which the middle peak was made to coincide with the nasal-vowel borderline in mài, was used as the starting point. See Figures 3 and 5 for the first step of the leftward movement and the two steps of lowering the peak (stimuli 2, 11, 12).

3. Steepness increased.

The original drop of the curve to the level of the following trough was made steeper in three steps (stimuli 13-15), by aiming the curve to points which were 32 ms, 64 ms and 96 ms closer to the middle peak (again made to coincide with the nasal-vowel borderline of mài first). The movement prolonged the trough following the middle peak. See Figures 3 and 6 for the first step of the leftward movement and the three steps of making the fall of the F0 curve steeper (stimuli 2, 13-15).

2+3. Combination of peak lowered and steepness increased.

In addition, three stimuli were prepared which combined the last two strategies. The first and the second of these stimuli (16-17) combined lowering of the peak by 20 Hz with making the following curve drop to points 32 ms and 64 ms closer to the peak respectively, and the third (stimulus 18) combined

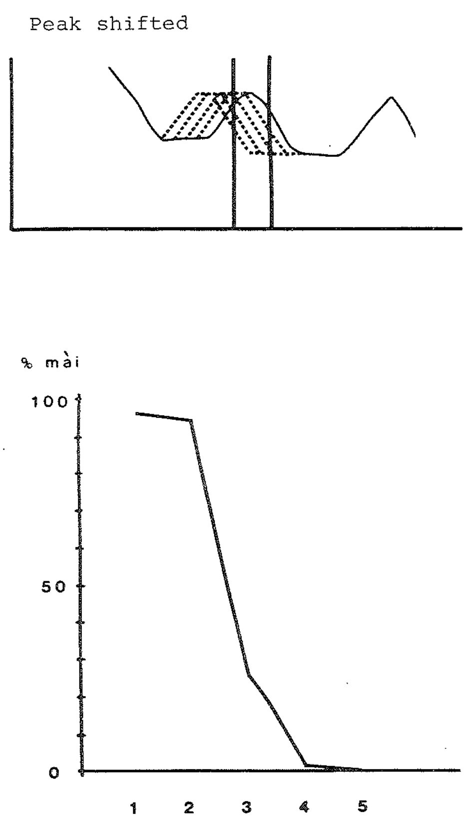


Figure 4. F0 curves and results for stimuli 1-5.

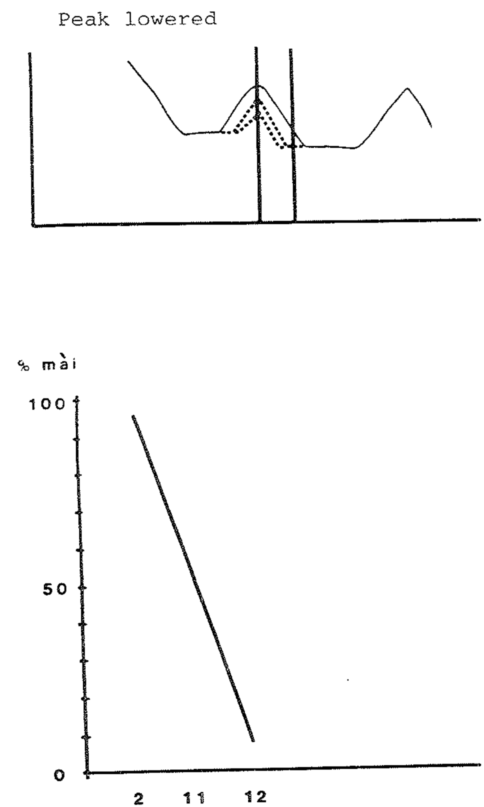


Figure 5. F0 curves and results for stimuli 2, 11 and 12.

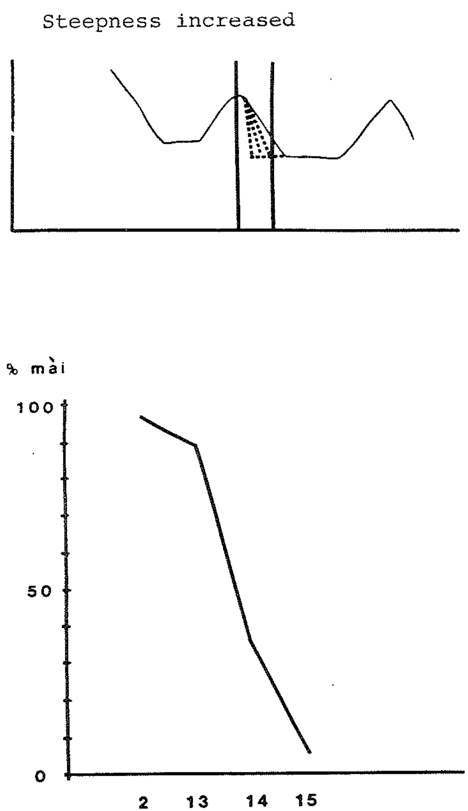


Figure 6. F0 curves and results for stimuli 2 and 13-15.

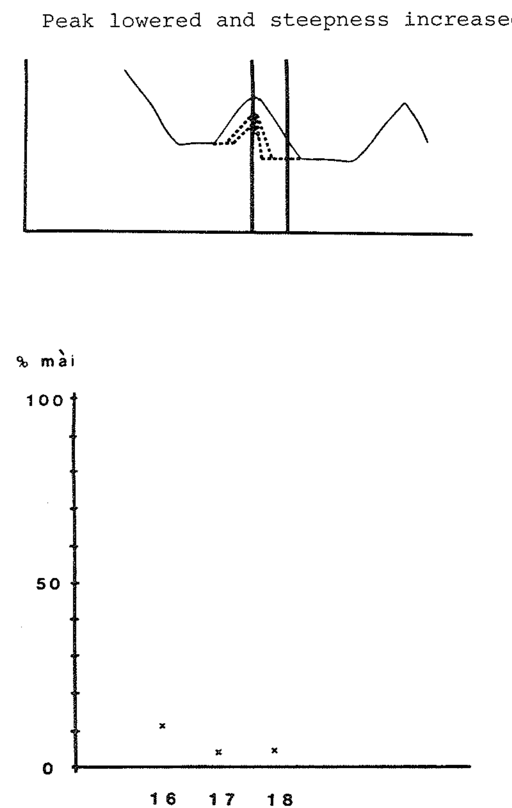


Figure 7. F0 curves and results for stimuli 16-18.

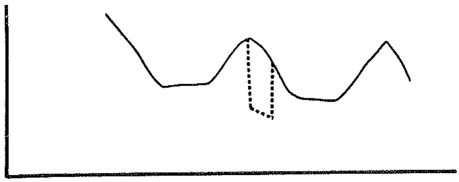


Figure 8. Stimulus 6. Creak added.

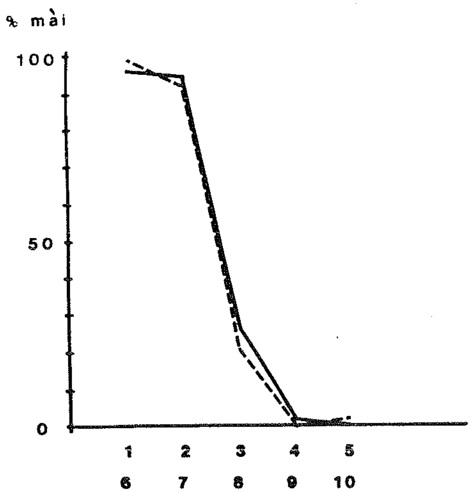


Figure 9. Result of added creak (stimuli 6-10, dashed line) as compared to the same stimuli without creak (stimuli 1-5, solid line).

lowering of the peak by 40 Hz with dropping the curve to the point 32 ms closer to the peak. (See Figures 3 and 7; stimuli 2, 16-18).

4. Creak added.

Finally, five stimuli (6-10) were prepared, which were identical to stimuli 1-5, except that simulated creak was added. Creak was simulated by abrupt halving of F0 over the middle part of the vowel in mài. (Halving of F0 brought about by biphasic phonation is one of the most common features of creak in Standard Chinese; for biphasic phonation see Lehiste 1970:59.) See Figure 8 for the F0 curve of stimulus 6.

PERCEPTUAL TEST

The perceptual test in which the subjects were exposed to the individual stimuli and simply asked whether they heard Sòng Yán mài niúròu "Sòng Yán sells beef" or Sòng Yán mǎi niúròu "Sòng Yán buys beef" in each case, was carried out in the Institute of Acoustics, Academia Sinica, in Běijīng. There were 18 subjects, all native speakers of Běijīng dialect. Altogether 60 stimuli recorded on magnetic tape were presented to the subjects. Of these stimuli, the first six were used with the intention of getting the subjects accustomed to the data, and responses to them were subsequently discarded. The remaining 54 stimuli were the 18 stimuli described above, each repeated three times in the corpus, and randomly ordered.

RESULTS AND DISCUSSION

The results of the perceptual test are found in Table 1 and shown in the diagrams in Figures 4-7 and 9.

The results indicate that all the three strategies according to which the F0 curve was manipulated (leftward shifting of the turning-point in the syllable mài, decreasing of the F0 value at the turning-point, and increasing of the steepness of the fall after the turning-point) proved successful in changing Tone 4 to Tone 3.

A common denominator of these changes is a decrease of the F0 average over the segments [ai] of the syllable mài. Figure 10 shows the mài percentage plotted against the F0 average for the stimuli which did not have simulated creak. It shows that listeners separated the stimuli into three distinct groups:

1. Stimuli which were perceived as carrying Tone 3 (stimuli 4, 5, 12, 15-18).

2. Stimuli which were perceived as carrying Tone 4 (stimuli 1, 2, 13).

Stimulus	Responses			F0 average
	mài	mǎi	% mài	(Hz)
1	52	2	96	184
2	51	3	94	170
3	14	40	26	148
4	1	53	2	126
5	0	54	0	112
6	53	1	98	-
7	49	5	91	-
8	11	43	20	-
9	0	54	0	-
10	1	53	2	-
11	28	26	52	146
12	4	50	7	129
13	47	7	87	161
14*	18	35	34	147
15	2	52	4	133
16	6	47	11	133
17	2	52	4	122
18	2	52	4	120

*One response is missing for this stimulus

Table 1. Responses to the different stimuli and F0 average over [ai].

3. Stimuli for which the results did not indicate any clear preference for one of the tones (stimuli 3, 11, 14).

The stimuli of group 1 have a low, rather flat interval at the end of the vocalic segments.

Group 2 have a high, rather flat or rising interval at the beginning of the corresponding segments.

With this description in mind, it is interesting to inspect the dubious cases of group 3, which have percentages between 20 and 50. Number 3 (26%) is falling throughout the pertinent segments, number 11 (52%) is also falling with a very short

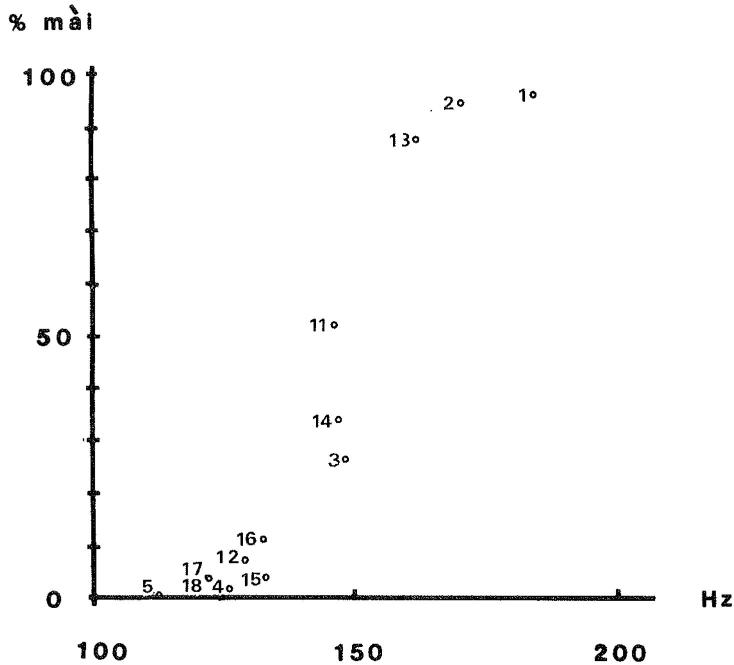


Figure 10. Tone 4 percentage plotted against F0 average.

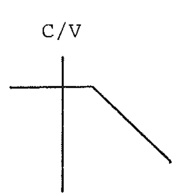
	First half	Second half
Tone 4	m=188 s=1.5	m=155 s=21.6
Tone 3	m=143 s=14.1	m=108 s=2.6

Figure 11. Mean values and standard deviations (in Hz) for F0 over the first and second half of the segments ai for stimuli judged as carrying Tone 4 or Tone 3.

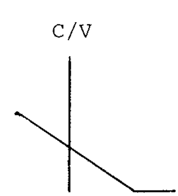
flat part at the end, and number 14 (34%) is falling with flat parts at both ends.

If we assume the hypothesis that a flat or rising part at the beginning of a fall is characteristic of Tone 4 and a flat or rising part at the end of a fall is characteristic of Tone 3, one case above, number 14, has conflicting cues, and the others (3 and 11) have no distinctive tonal cues at all. Hence our hypothesis fits the three dubious cases. Also our production data give support to this hypothesis. Figure 2, for instance, shows how the falling tones are preceded by rising intervals (except the first, beginning with voiceless [s]) and the rising tones are preceded by flat ones.

Another feature which strengthens the hypothesis is indicated by Figure 11, which shows the mean fundamental frequency values and Standard deviations for the first and the second halves of the vocalic segments of the synthetic stimuli. For those judged as Tone 4, the values of the first half have very little variability compared to those of the second half. For Tone 3 the situation is reversed. Here the second half has little variability compared to the first half. One may perhaps conclude that the change to a fall from a certain level is important for the identification of Tone 4, and a change from a fall to a certain low level is important for Tone 3. If we use the notion of turning points[1] in our interpretation of the results, we might say that Tone 4 has an early high turning point introducing a fall in the vocalic segments [ai], whereas Tone 3 has a late low one ending a fall in the same segments[2]. The following schematic figure illustrates the situation:



Tone 4



Tone 3

Our hypothesis may not be without general perceptual significance. A turning point marks the change from one clearly perceived mode (falling, rising, level) into another clearly perceived mode and should therefore be a salient perceptual event.

The fact that manipulation of F0 was sufficient to change Tone 4 into Tone 3 shows that creak is not a necessary perceptual cue for Tone 3. It also turned out that the addition of simulated creak had very little or no effect (see Figure 9). This indicates that the presence of creak is not an important cue for the identification of Tone 3. However, it is also possible that creak simulated by greater disturbance of the periodicity of the waveform would have a greater effect.

NOTES

1. Turning points are basic concepts in the generative model of intonation developed at Lund.

2. As the F0 properties of the prevocalic part in the syllable mài had no effect on the interpretation of the tone the syllable carried, the results indicated support to the hypothesis about the domain of tone in Standard Chinese formulated by Howie 1974:129-48. According to this hypothesis, the tone-carrying part of the syllable in Standard Chinese is its voiced part with the exclusion of the voiced non-vocalic component which may occur at the beginning of the syllable.

REFERENCES

Cheng, C.C. 1973. A synchronic phonology of Mandarin Chinese. The Hague.

Gårding, E., Zhāng Jiālù and J-O. Svantesson 1983. A generative model for tone and intonation in Standard Chinese. Working Papers 25, 53-65. Department of Linguistics, Lund University.

Gårding, E. 1985. Constancy and variation in Standard Chinese tonal patterns. Working Papers 28, Department of Linguistics, Lund University.

Hockett, C.F. 1947. Peiping phonology. Journal of the American Oriental Society 67.

Howie, J.M. 1974. On the domain of tone in Mandarin. Phonetica 30.

Howie, J.M. 1976. Acoustical studies of Mandarin vowels and tones. Cambridge.

Kratochvil, P. 1985. The case of the third tone. In Papers presented to Wang Li on his eightieth birthday. Hong Kong.

Lehiste, I. 1970. Suprasegmentals. Cambridge, Mass.

Implications of rapid spectral changes on the categorization of tonal patterns in speech perception

David House

ABSTRACT

In order to study what kinds of categorization features can be used in the perception of tonal contours in speech, and to determine how these features interact with spectral changes, four perception experiments were carried out. The stimuli were comprised of synthetic /a/ vowels with different tonal patterns, and listeners were presented with an ABX-test configuration. In the first experiment, steady-state vowels were used. In the three other experiments, a gap in the tonal pattern consisting of an intensity drop preceded and followed by formant transitions was introduced in different places in the vowel. The results of the tests suggest that tonal movement in vowels can be categorized in terms of pitch movement such as rise-fall and fall-rise or in terms of pitch levels. Listeners who categorized in terms of pitch levels demonstrated tendencies which can be interpreted as heightened attention to pitch frequency immediately following spectral changes. These findings can have implications for production and perception models of tone and intonation.

1. INTRODUCTION

In studying the functions and tasks of the auditory system in speech perception, at least two levels of frequency dependent analysis can be distinguished. A first order frequency analysis is carried out based on the mechanical properties of the basalar membrane and characteristic frequencies and temporal responses of auditory-nerve fibers. This first-order analysis can be seen as providing the raw materials for a second-order analysis of pitch and timbre (Plomp, 1976; Green, 1976; Young and Sachs, 1979; Gelfand, 1981).

Current psychoacoustic research and physiological modelling experiments are generally in agreement that this second-level analysis, which resolves spectral information and pitch, involves some degree of central processing (Plomp, 1976; Delgutte, 1982). Although many models of pitch perception presuppose a spectral resolution of the lower harmonics from which pitch can then be derived (Wightman, 1973; Goldstein, 1973; Terhardt, 1974) it is still unclear as to what degree of spectral resolution is necessary for pitch perception and exactly what type of information (spatial or temporal) is used (Sachs, Young, and Miller, 1982; Delgutte, 1982). A further question involves the interaction between spectral cues and pitch cues in speech perception.

The present study is concerned with two issues related to fundamental frequency perception and spectral analysis in speech. The first issue is an attempt to define short-term memory categorization features in speech-like pitch movements. The second issue is an attempt to study how such categorization features of pitch might interact with spectral changes during pitch movement.

Two candidates for pitch movement categorization features can be proposed. One would be a continuous pattern storage involving categories such as rise-fall and fall-rise. The other possibility would be the storage of discrete pitch frequencies at given time intervals with movement then being interpolated after the pitch analysis.

Where interaction with spectral changes is concerned, categorization features based on continuous pattern storage might be sensitive to disturbances by such rapid spectral changes as formant transitions. On the other hand, spectral changes associated with consonants could provide perceptual boundaries which, when related to linguistic structure, might facilitate discrete pitch frequency resolution.

2. METHOD

A. Stimuli design and synthesis

Fundamental frequency cues in speech such as syllable stress, word accents, word tones, etc. are tightly related to lexical items and are therefore, as categories, acquired and processed centrally as are other features having linguistic functions. To attempt to separate lexical function from intonation contours, question-statement categories have often been used (Hadding-Koch and Studdert-Kennedy, 1964; Fourcin, 1980). In the present study, however, an ABX test design was used to force listeners to create categories instead of using presupposed linguistic categories. The basic categories expected to be created by the listeners were rise-fall and fall-rise.

A Klatt software synthesizer and a VAX digital computer were used to synthesize a Swedish /a/ vowel with formant frequencies of 600, 925, 2540, and 3320 Hz. (Klatt, 1980; Fant, 1973). Vowel duration was 300 msec. including 30 msec. intensity onset and offset. Fundamental frequency was systematically varied to create 18 different stimuli (Fig. 1). The Fo contour for stimulus A (also stimulus 1), designed to elicit rise-fall categories, rose from 120 Hz to a turning point of 180 Hz and then fell to an end point of 100 Hz. The Fo contour for stimulus B (stimulus 12), designed to elicit fall-rise categories, began at 120 Hz falling to 80 Hz and then rising to 160 Hz. The difference in end-point frequency was designed to test the effect of end-point variation on the rise-fall, fall-rise categories, i.e. movement pattern recognition versus discrete frequency analysis. The 18 stimuli were constructed by

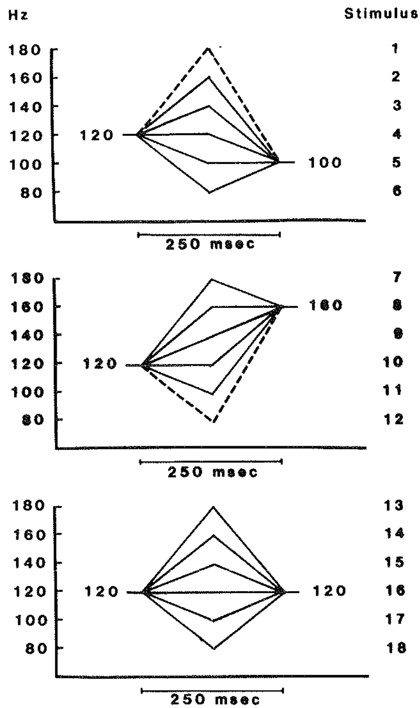


Figure 1. Stylized Fo contours for the first set of stimuli, steady-state vowel. The dashed lines (stimuli 1 and 12) represent stimuli A and B of the ABX configuration.

systematically varying the turning point in steps of 20 Hz from 80 Hz to 180 Hz with three different end-point configurations: 100 Hz, 160 Hz and 120 Hz.

Varying the Fo parameter in the Klatt synthesis program, as is often the case in natural speech, causes concomitant intensity variations. An informal listening test using LPC versions of selected stimuli with corrected intensity parameters demonstrated that the difference between corrected stimuli and uncorrected stimuli was marginal and in most cases could not be perceived at all. In view of the marginal perceptual nature of the intensity variation when compared to Fo variation, the intensity variations in the stimuli used in the experiments were not corrected. This is not to say, however, that intensity, especially in conjunction with Fo variation, is not perceptually relevant, but simply that intensity variation due to Fo changes in the context of these experiments is viewed as a concomitant feature of Fo.

An ABX-type test was constructed with a 1.3-second pause between stimulus A and B and a 2.2-second pause between stimulus B and X. Between stimulus X and the following A stimulus there was a 4.6-second pause. The X stimuli were randomized by a computer program and the test was divided up into blocks of 10 stimuli each. The test consisted of 100 ABX stimuli, the first block of 10 being a buffer block to acquaint the listeners with the test. Each of the 18 different stimuli occurred five times in the remaining 90 stimuli. All 18 stimuli were presented once before any stimulus was repeated. The same test configuration was also used for the second, third and fourth versions of the test. The stimuli were recorded on tape using a Revox PR99 tape recorder. A pause of about 10 seconds was included between each block. The total duration of the test was about 15 minutes.

To test the effects of rapid spectral changes on the categorization of the tonal patterns, four different versions of the test were produced creating four experiments. The first version consisted of the steady-state /a/ vowel as described above. In the second version, a gap, consisting of an intensity drop preceded and followed by formant transitions, was introduced in the tonal pattern (Fig. 2). The first set of formant transitions extended from 60 msec. into the vowel to 75 msec. The second set extended from 135 msec. to 150 msec. into the vowel. The intensity drop began 80 msec. into the vowel and ended at 130 msec. This created /abaa/-like stimuli. Otherwise the tonal patterns were the same as in the first version (Fig. 3). In the third version, the gap was placed in the last half of the vowel; the transitions beginning and ending at 150 - 165 msec. and 225 - 240 msec. respectively, and the intensity drop extending from 170 to 225 msec. (Fig. 4). This created /aaba/-like stimuli. In both second and third versions the actual turning point frequency was present. In the fourth version, however, the gap was placed in the middle of the vowel; transitions beginning and ending at 105 - 120 Hz and 180 - 195 Hz respectively, with the intensity drop extending from 120 to 180 Hz. (Fig. 5). These stimuli were /aba/-like.

These manipulations were done to test which parts of the tonal pattern are most susceptible to possible disturbances caused by rapid spectral changes. All other elements of the test configuration were held constant throughout the four different versions.

B. Subjects

Five members of the secretarial staff at the Department of Linguistics and Phonetics, Lund University and 31 first-term students in phonetics participated in the experiments. All of the subjects had normal hearing and were native speakers of Swedish. To avoid fatigue effects, each of the four tests was administered on a separate occasion.

The five staff members participated in all four tests while each individual student participated in only one of the four tests. The total number of listeners for each test was 11 for test 1, 11 for test 2, 10 for test 3, and 14 for test 4.

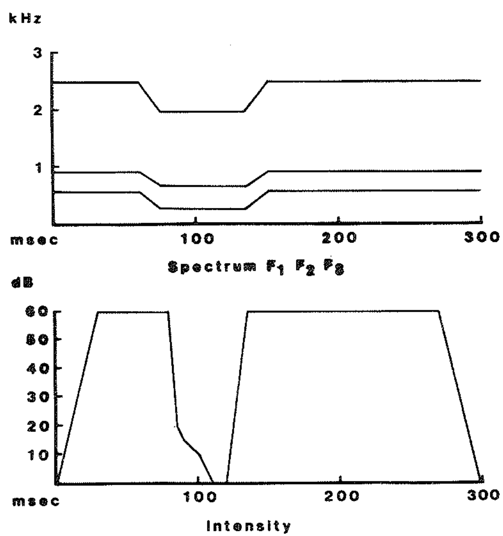


Figure 2. Spectral and intensity manipulations in the second version of the stimuli (abaa-like stimuli).

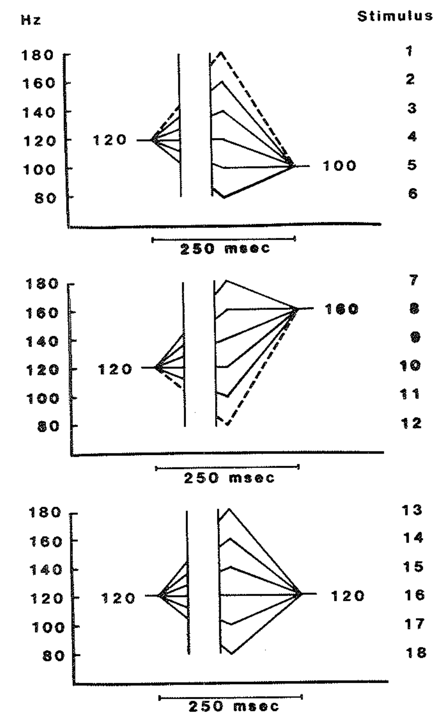


Figure 3. Stylized Fo contours of the second version of the test (abaa-like stimuli). The dashed lines (stimuli 1 and 12) represent stimuli A and B of the ABX configuration. The gaps represent spectral and intensity manipulations.

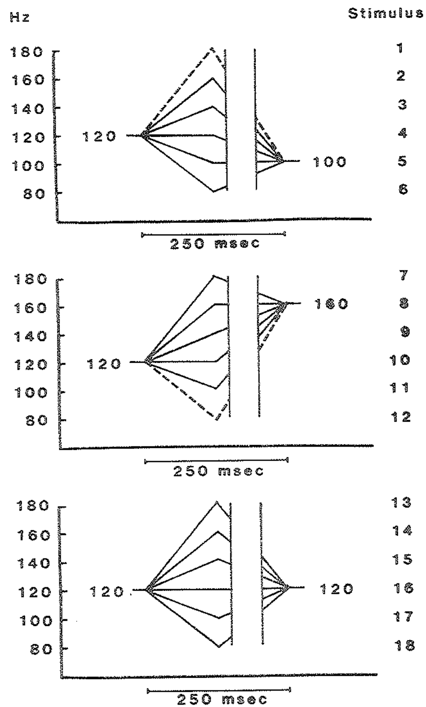


Figure 4. Stylized Fo contours of the third version of the test (aaba-like stimuli). The dashed lines (stimuli 1 and 12) represent stimuli A and B of the ABX configuration. The gaps represent spectral and intensity manipulations.

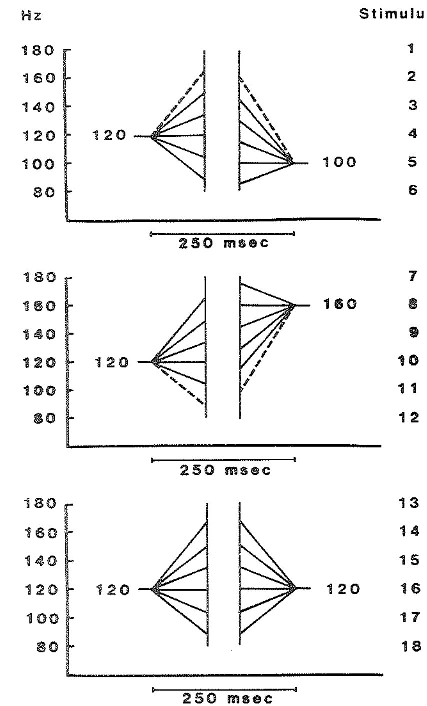


Figure 5. Stylized Fo contours of the fourth version of the test (aba-like stimuli). The dashed lines (stimuli 1 and 12) represent stimuli A and B of the ABX configuration. The gaps represent spectral and intensity manipulations.

C. Procedure

The test tapes were presented in a sound-treated perception lab via a Revox A77 tape recorder and Burwen PMB6 orthodynamic headphones. Each listener was given a printed instruction sheet and an answer sheet. The instructions were also read aloud, the listeners being instructed to listen to the first two sounds of each test item and, upon hearing the third sound, to decide if it was most like the first or the second sound and to circle the corresponding number (1 or 2) on the answer sheet. The tape was stopped after the practice block and the listeners were allowed to ask questions. The listeners were also given a rest pause after half the test.

3. RESULTS

A. Steady-state vowel

The results of the test version consisting of the steady-state vowel were uniform for all listeners. The test was easy to perform, and all listeners categorized the stimuli on the basis of movement pattern recognition, i.e. rise-fall or fall-rise disregarding the differences in end-point frequency.

Figure 6 shows that stimuli 1-4 were categorized as rise-fall (most like stimulus 1) while stimuli 5 and 6 were heard as fall-rise (most like stimulus 12) despite the fact that the end-point frequency was the same in all 6 stimuli. In the group of stimuli where the end-point frequency was 160 Hz, stimuli 7-9 were heard as rise-fall and stimuli 10-12 as fall-rise. The same pattern of results was obtained in stimuli 13-18 with the constant stimulus 16 chosen as most like stimulus 1. No listener chose categories on the basis of end-point frequency, although the difference in responses for stimuli 4 and 10 could be interpreted as being influenced by end-point frequencies.

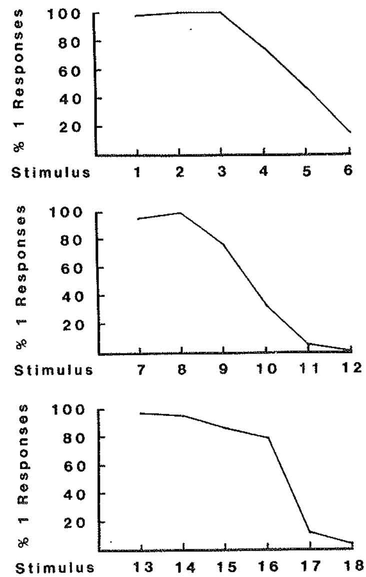


Figure 6. Averaged responses for test 1, steady-state vowel, showing rise-fall and fall-rise categorization.

B. Spectral change early in the vowel.

The averaged results for the test version where spectral changes were placed early in the first part of the vowel show a considerable ambiguity concerning stimulus categories (Fig. 7). In contrast to the results obtained in the steady-state vowel version, stimulus 4 was perceived as most like stimulus 12, stimulus 10 as ambiguous and stimulus 16 as most like stimulus 12. These differences can not be explained in terms of end-point frequencies as the categories seem to be formed contrary to the frequencies of the end-points, i.e. there were

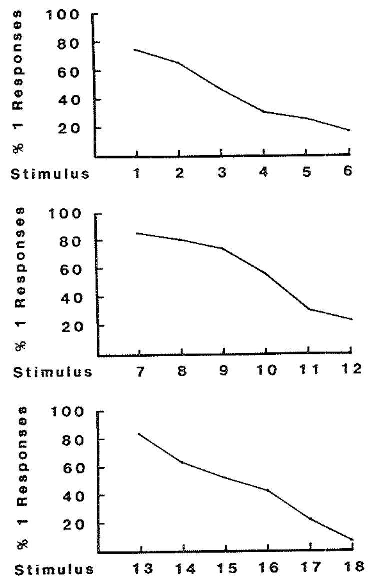


Figure 7. Averaged responses for test 2, /abaa/ stimuli.

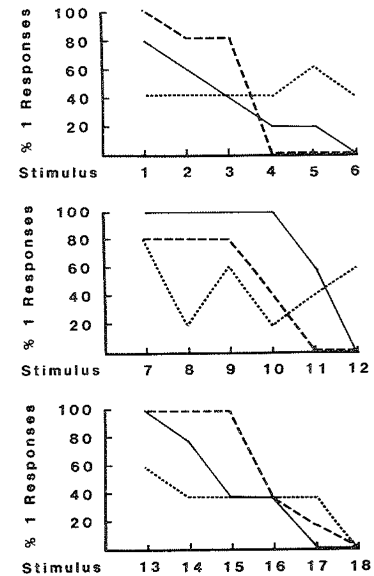


Figure 8. Individual responses (3 listeners) for test 2, /abaa/ stimuli. The solid line represents reverse end-point frequency categorization, the dashed line represents rise-fall, fall-rise categorization, and the dotted line represents no perceived categories.

more stimulus 12 responses to stimuli 1-6 and more stimulus 1 responses to stimuli 7-12. When the end-point frequency was 120 Hz, stimuli 13-18, the averaged responses showed rise-fall, fall-rise categorization, but with much more ambiguity than in the first test.

There was also considerable individual variation among the listeners. Basically, responses fell into three classes of roughly equal size: 1.) those who seemed to perceive categories contrary to end-point frequencies, 2.) those who perceived the same tonal categories as in the first version, and 3.) those who could no longer perceive the categories (Fig. 8).

C. Spectral changes late in the vowel.

In this version of the test, results were similar to those of the second version with the exception that the listeners who perceived categories which were not based on movement pattern recognition tended to choose categories based on the end-point frequency. For example, stimulus 3 which was ambiguous in test 2, was perceived as most like stimulus 1. Stimulus 9, clearly most like stimulus 1 in test 2, was perceived as most like stimulus 12 in this version. Although individual variation was again considerable, the averaged responses show a tendency towards end-point categories (Fig. 9) as there were more stimulus 1 responses to stimuli 1-6 and more stimulus 12 responses to stimuli 7-12. This tendency is more clearly seen in individual responses (Fig. 10).

D. Spectral changes in the middle of the vowel.

In the fourth version of the test, the results were almost identical to the results obtained from the third test where the spectral changes occurred late in the vowel (Fig. 11). There was, however, more of a tendency for listeners to perceive categories based on end-point frequencies in this version than in any other version of the test as exemplified by Fig. 12.

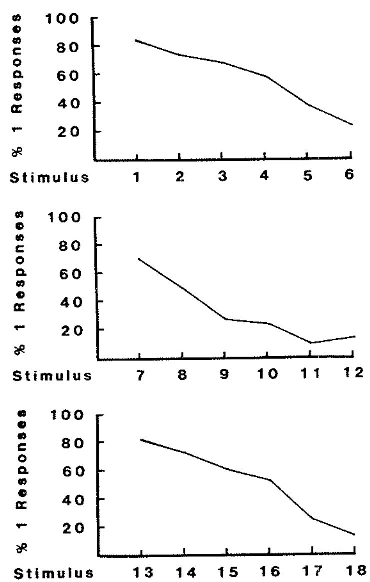


Figure 9. Averaged responses for test 3, /aaba/ stimuli.

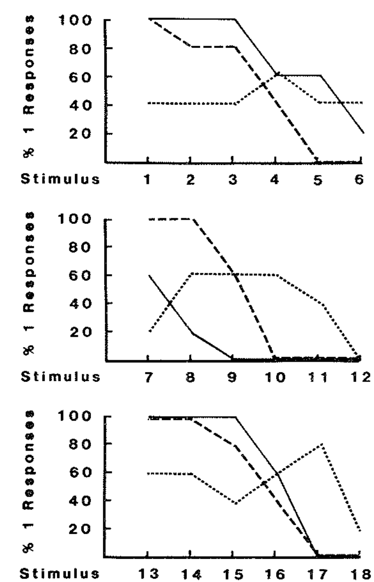


Figure 10. Individual responses (3 listeners) for test 3, /aaba/ stimuli. The solid line represents end-point frequency categorization, the dashed line represents rise-fall, fall-rise categorization, and the dotted line represents no perceived categories.

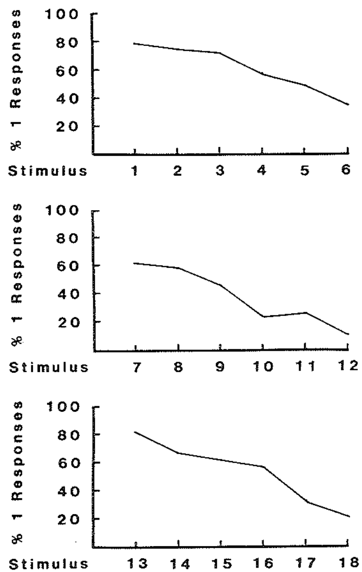


Figure 11. Averaged responses for test 4, /aba/ stimuli.

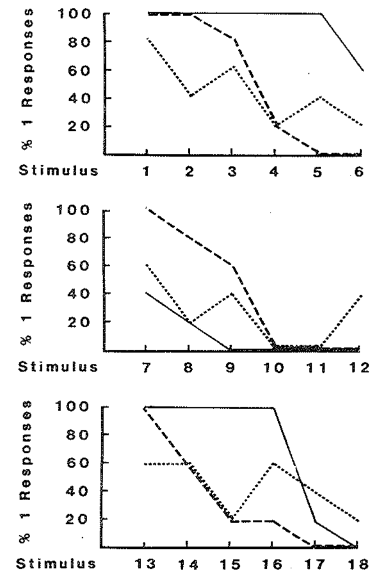


Figure 12. Individual responses (3 listeners) for test 4, /aba/ stimuli. The solid line represents end-point frequency categorization, the dashed line represents rise-fall, fall-rise categorization, and the dotted line represents no perceived categories.

E. Summary of the results

In the first version of the test, comprising steady-state vowels, listeners performed uniformly, categorizing the stimuli in terms of movement direction; rise-fall or fall-rise. In the three other versions, where spectral changes were introduced in the vowel, responses fell into three classes: ambiguous responses, categorization based on movement direction, and categorization based on final pitch frequency. However, in the stimuli where spectral changes occurred early in the vowel, the group of listeners who did not use movement direction but still performed categorizations, seemed to do so based on the reverse of the end-point frequencies.

DISCUSSION

A. Movement pattern as categorical feature.

The first experiment was designed to determine if pitch movement can be used as a categorization feature for tonal patterns in speech perception. The results for stimuli 3,4,9 and 15, all categorized as stimulus 1, and for stimuli 5, 11, and 17, which were categorized as stimulus 12 (Fig. 6), seem to indicate that listeners can use the movement features rise-fall and fall-rise to categorize tonal patterns. The results of the first experiment could also be explained in terms of turning-point configuration where a change in the direction of the pitch movement results in a convex or a concave configuration. It could be that the perceptual mechanism is particularly sensitive to changes in pitch movement direction whereby an increasing number of periods per time unit followed by a decreasing number or vice-versa would be registered by the perceptual mechanism as an event which could then form the basis for categorization. Whatever the mechanism, the results indicate that in the absence of spectral changes, movement pattern rather than end-point frequency level is selected by listeners to form the basis for categorization in this type of forced choice test.

Although it is a large step from synthetic vowel stimuli to actual speech stimuli, these results suggest the possibility that languages could make use of movement direction in the vowel as a distinctive perceptual feature. Furthermore, the ability of about one-third of the listeners (12 listeners) to perform the same movement categories in the stimuli containing rapid spectral changes indicates that some listeners are able to interpolate the movement and arrive at the same kinds of categories as in the continuous pattern categorization. However, the diversity of the responses to the stimuli containing rapid spectral changes indicates that spectral changes do in fact influence pattern categories.

B. Pitch level as categorical feature.

In tests 3 and 4, those listeners who perceived categories which were not based on movement patterns, seemed to perform the categories based on end-point identification. Had it not been for the conflicting results of test 2 where listeners seemed to perform categories in direct opposition to end-point frequencies, the results would seem to point toward some kind of movement masking effected by the rapid spectral changes. This masking would lead listeners to use end-points for forming categories. The results of test 2, however, complicate matters.

In their classic study of pitch discrimination for synthetic vowels, Flanagan and Saslow (1958) found slightly more acute discrimination of changes in fundamental frequency in vowels than in a pure tone. This could mean that listeners use the relatively larger changes in the harmonics present in the vowels as an aid in discriminating Fo changes. Klatt (1973) substantiated these results but found that discrimination performance deteriorated when a linear ramp fundamental frequency contour was used in the place of a monotone. t'Hart, in his study of just noticable differences in pitch movement (1981), found that falls were more difficult to judge than rises and that many subjects compared final pitches instead of sizes of movements. He also reported from Nabelek (reference omitted) a separate concentration on the initial or the final pitch if there was a pause between the low and high frequency parts of a stimulus rather than a continuous glide.

These findings could be used to help explain the present results. If rapid spectral changes are introduced in the vowel, the load on both the processor and memory is drastically increased. In some cases, an economy measure may be necessary whereby the pitch movement is recoded into pitch levels. In tests 3 and 4, the final pitch levels correspond to the end-point frequencies. If the pitch levels are defined as the average frequency some 30 to 50 msec. after the vowel onset following the stop occlusion, stimulus 1 in test 2 (see Figure 3) would be stored as low-high and stimulus 12 as high-low. Stimuli 3-6 might have therefore been classified as high-low (i.e. most like stimulus 12) while stimuli 7-10 might have been classified as low-high (i.e. most like stimulus 1). This would account for the difference in results between test 2 and tests 3 and 4. Categorization in these tests would therefore not be based on end-point frequency but on averaged frequency level following the vowel onset after the stop occlusion.

C. Interaction with spectral changes.

The results of these experiments indicate that spectral changes occurring during a tonal movement can affect the perception of the tonal contour. During the time period in which rapid spectral changes occur and immediately following this period, the perceptual mechanism may be somewhat less sensitive to pitch movement than during a longer steady-state portion of the vowel. Whether this insensitivity is a result of an increased load on the peripheral mechanism or a result of language acquisition or both remains an open question. However, this insensitivity to movement could require the pitch perception mechanism to increase attention to averaged pitch frequency during this time period. The averaged frequency would then be stored in memory as a pitch level, e.g. High, Low or Mid. These levels could then be used linguistically as categorization features.

It is possible, then, that rapid spectral changes such as a stop release or any consonant release followed by a vowel can serve as a perceptual boundary marker. This boundary marker would inhibit perception of continuous pitch movement and enhance perception of discrete pitch frequency. The perceptually important feature of the tonal contour immediately following this boundary would be, therefore, a tonal level. Continuous movement as an important feature of the tonal contour would then need to be placed in a longer, steady-state portion of a presumably stressed vowel.

D. Possible linguistic implications

The possible perceptual division of the tonal contour into sections of differential sensitivity for levels and movement might have implications for both production and perception models of tone and intonation. Knowledge of such sections and how they relate to the segmental structure of an utterance could, for example, facilitate the exact placement of Lows and Highs in production models. An aid for perception models would be in determining what part of the tonal contour is relevant for memory storage.

The results of these experiments and an explanation involving differential sensitivity seem to agree with the results obtained in perception tests aimed at studying the perceptual cues necessary to distinguish tone 3 from tone 4 in Modern Standard Chinese, (Gårding, et al. 1985, this issue). It could be that tone languages are prone to use tonal patterns which fit pitch sensitivity differences in the perceptual periphery. These patterns would then be linguistically relevant reinforcing the selective pitch sensitivity.

Finally, the idea of selective pitch sensitivity could be applicable to speech synthesis and speech recognition. In synthesis, the placement of the tonal contour in relation to the segmental structure could be facilitated, and in recognition, segmental boundaries could be used as markers for extracting critical parts of the tonal contour.

E. Further issues

Some further issues concerning the perception of pitch and spectral change lying beyond the scope of the present study involve perception of pitch immediately prior to spectral change. Another issue could be the use of non-speech noise to see if the perceptual boundary is constrained to spectral changes produced by the articulators. Finally, more work needs to be done using real speech material to more fully understand the processes involved in the perception of the tonal contour in speech.

ACKNOWLEDGEMENTS

I would like to thank Gösta Bruce and Eva Gårding for valuable discussion and comments concerning this study.

REFERENCES

Delgutte, B. 1982. Some Correlates of Phonetic Distinctions at the Level of the Auditory Nerve. In Carlson and Granström (Eds.) The Representation of Speech in the Peripheral Auditory System. Elsevier Biomedical Press, Amsterdam, New York, Oxford, pp. 131-149.

Fant, G. 1973. Speech sounds and features. The MIT Press. Cambridge, Mass.

Flanagan, J.J., and Saslow, M.G. 1958. Pitch discrimination for synthetic vowels, J. Acoust. Soc. Am. 30, 435-442.

Fourcin, A.J. 1980. Speech pattern audiometry. In Beagley, H.A. (Ed.) Auditory investigation: the scientific and technological basis.

Gårding, E., Kratochvil, P., Svantesson, J.O., & Zhang, J. 1985. Tone 4 and Tone 3. Discrimination in Modern Standard Chinese. Working Papers 28 (This issue), Department of Linguistics and Phonetics, Lund University.

Gelfand, S.A. 1981. Hearing, an introduction to psychological and physiological acoustics. Marcel Dekker, Inc. New York and Basel, Butterworths, London.

Goldstein, J.L. 1973. An optimum processor theory for the central formation of the pitch of complex tones. J. Acoust. Soc. Am. 54, 1496-1516.

Green, D.M. 1976. An introduction to hearing. Lawrence Erlbaum Associates, Hillsdale, New Jersey.

Hadding-Koch, K., and Studdert-Kennedy, M. 1964. An Experimental Study of Some Intonation Contours. Phonetica 11, 175-184.

t'Hart, J. 1981. Differential sensitivity to pitch distance, particularly in speech. J. Acoust. Soc. Am. 69, 811-821.

Klatt, D.H. 1973. Discrimination of fundamental frequency contours in synthetic speech: Implications for models of pitch perception. J. Acoust. Soc. Am. 53,8-16.

Klatt, D.H. 1980. Software for a formant synthesizer. J. Acoust. Soc. Am. 67, 971-995.

Plomp, R. 1976. Aspects of Tone Sensation, A Psychophysical Study. Academic Press, London.

Sachs, M.B., Young, E.D., & Miller, M.I. 1982. Encoding of Speech Features in the Auditory Nerve. In Carlson and Granström (Eds.) The Representation of Speech in the Peripheral Auditory System. Elsevier Biomedical Press, Amsterdam, New York, Oxford, pp. 115-130.

Terhardt, E. 1974. Pitch, consonance, and harmony. J. Acoust. Soc. Am. 55, 1061-1069.

Wightman, F.L. 1973. The pattern-transformation model of pitch. J. Acoust. Soc. Am. 54, 407-416.

Young, E.D. & Sachs, M.B. 1979. Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers. J. Acoust. Soc. Am. 66, 1381.

Sentence prosody and syntax in speech perception

David House

ABSTRACT

In order to test the use of sentence prosody related to syntax in sentence processing, a listening test was constructed consisting of marked and unmarked prosodic versions of sentences with various syntactic structures. Results of a pilot test showed considerable listener differences in sentence interpretations. While listeners were able to use prosody as an aid in sentence processing, they also relied heavily upon other cues such as simplicity strategies and morphological restrictions supplied by the lexicon. When the cues provided conflicting information, listeners, on an individual basis, tended to select the most salient cue to determine syntax.

1. INTRODUCTION

One of the functions of sentence prosody is to provide the listener with acoustic cues which represent aspects of syntactic structure. This is accomplished primarily by means of fundamental frequency excursions (intonation) and manipulation of the time dimension such as final lengthening and pauses. These aspects of the speech signal can be used by the speaker to emphasize certain structural relationships within a sentence and by the listener to confirm underlying and surface structures or to resolve structural ambiguities.

In a study of the production and perception of ambiguous sentences, Lehiste (1973) found a correlation between surface constituent structure and success in disambiguation. Speakers were able to convey structural differences and listeners were able to perceive them. Ambiguous sentences in which a difference in meaning was not correlated with a difference in surface constituent structure were generally not disambiguated. There were, however, considerable differences between speakers and between listeners concerning success in disambiguation. Nonetheless, it seems clear that when available, intonation, final lengthening and pauses in the acoustic signal can impinge on the syntactic structure of the sentence.

One question in this respect is to what extent listeners can make use of the acoustic signal to aid in the processing of syntax. The goal of this paper is to investigate relationships between the use of prosody in sentence perception and the nature of the sentence structure in terms of complexity and frequency of occurrence.

A reasonable hypothesis would be that generally the use of prosodic cues in perception increases as a function of syntactic complexity with certain reservations for frequency differences between structures of varying complexity. A correlate to this hypothesis would then be that the use of such cues in perception decreases as a function of syntactic cues available through lexical subcategorization frames (Ford, Bresnan, & Kaplan, 1982).

To test the hypothesis, listeners could be presented with sentences having a reduced relative structure where the first verb is locally ambiguous between the simple sentence structure and the more complex reduced relative structure. An example of this type of sentence in English is the classic "garden path" sentence:

1.) The horse raced past the barn fell.

where the listener is led down "the garden path" until the final verb is understood at which point the sentence must be reparsed.

Clark & Clark (1977) propose a complexity strategy by which a listener will assume the first clause to be a main clause provided it is not marked at or prior to the main verb as something other than a main clause. In accordance with this strategy, if sentence prosody is also ambiguous, listeners would be expected to begin processing by assuming the first clause to be a main clause thus incorrectly resolving the ambiguity of the verb by interpreting it as active. The listener would therefore end up with a structure which does not match the input upon input of the final verb. The sentence would then have to be reparsed lengthening processing time. If, however, markedly appropriate prosodic cues are supplied in the acoustic input, e.g. a final fall-rise in intonation at the end of the subject combined with a pause following the subject marking the second clause as a relative clause prior to the verb, the correct structure should be projected earlier. This would in turn lead to a shorter processing time. Reaction times to the point at which the sentence is understood could be used as a rough gauge of processing time.

To test the correlate hypothesis, i.e. that the use of prosodic information decreases as more syntactic information is available from lexical redundancy rules, "subcategorization features" or morphological restrictions, a second sentence type such as:

2.) The horse driven past the barn fell.

could be used. Here, the syntactic structure can be determined by lexical information and the listener would be able to tell upon processing the verb that it belongs to a reduced relative clause. Thus the listener should be able to parse the sentence correctly considerably earlier solely on the basis of segmental cues and linguistic competence. In this case, marked prosodic information should not contribute to the speeding-up of processing time as much as when the verb is ambiguous if at all.

The differences in processing time between the two sentence types could be manifested experimentally by differences in reaction times. Reaction times to sentence type 1. with marked prosodic cues should be considerably shorter than without marked prosody. Reaction times to sentence type 2. with marked prosodic cues should only be slighty shorter than without marked prosody.

There is then a proposed interplay between sentence complexity, prosodic features, and lexical-morphological syntactic information. This interplay should be manifested by processing time for sentence parsing during speech perception.

2. SPEECH MATERIAL AND METHOD

Swedish sentences were chosen for the test such that plural subject-verb agreement corresponded to sentence type 1. (ambiguous verb), and singular subject-verb agreement corresponded to sentence type 2. (unambiguous verb) by virtue of the fact that the plural morpheme for passive in weak verbs is identical to the imperfect, while for singular they are differentiated. Two different sentence structures were chosen. The first structure contained a single passive verb, the preposition and the object followed by the active verb:

3a) Männen, ritade av flickan, rökte.
(The men, drawn by the girl, smoked.)

3b) Mannen, ritad av flickan, rökte.
(The man, drawn by the girl, smoked.)

In this structure, the preposition "av" can be misparsed as a particle which follows an active verb. This would alter the meaning of sentence 3a to "The men drew the girl, smoked". Sentence 3b should not be misparsed because "ritad" can only be passive. When "av" is used as a particle, however, it takes primary stress manifested prosodically by rising intonation, lengthening and increased intensity. It was feared that without these prosodic cues, "av" would be immediately parsed as a preposition leading to a correct reparsing earlier than the final verb.

As a control of the possible effects of this stressed-unstressed distinction in the first structure, a second structure was also included in the test. This second structure contained a compound passive verb with the subject implied followed by the active verb:

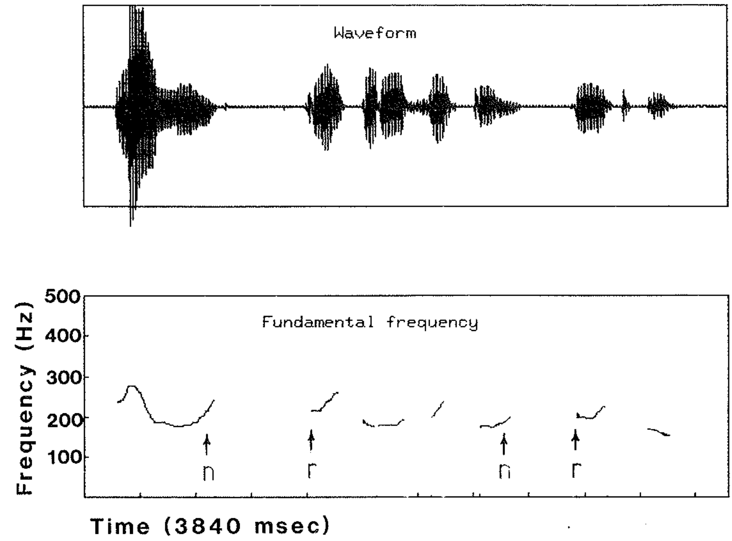
4a) Killarna, rakade och tvättade, sjöng.
(The boys, shaved and washed, sang.)
4b) Killen, rakad och tvättad, sjöng.
(The boy, shaved and washed, sang.)

In this structure, correct reparsing would not be expected until the final verb is perceived.

A listening test was constructed comprising three pairs of each sentence type plus four pairs of unambiguous filler sentences of the following type:

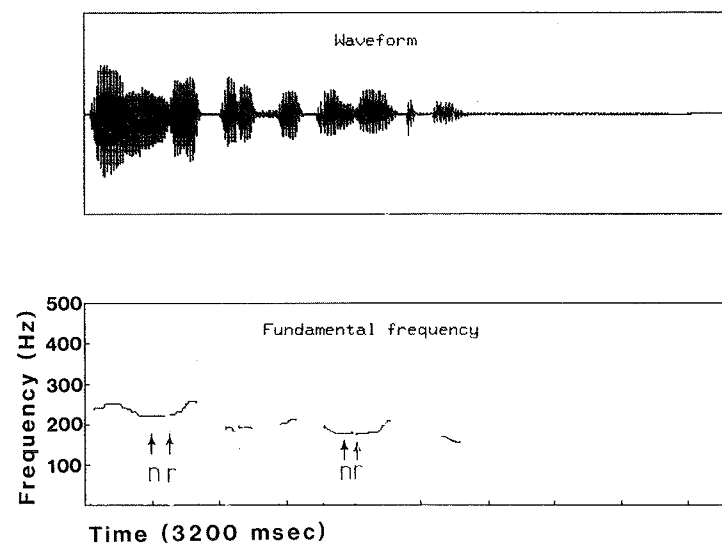
5a) Killarna sa, att bilen skulle tvättas.
(The boys said, that the car should be washed.)
5b) Killen sa, att bilen skulle tvättas.
(The boy said, that the car should be washed.)

Each sentence was recorded by a native female speaker of Southern Swedish in two versions, one version with marked intonation and pauses and one version without pauses and with unmarked intonation (Fig. 1 and 2). The forty test sentences were then ordered systematically so that different versions of different sentences appeared before either version or sentence type was repeated. (See appendix for test sentences and sentence order.) This type of test construction would enable different test versions to be presented to different listeners with the results being compiled across listeners to eliminate learning effects. Four filler sentences were presented as a buffer before the test began.



Männen, ritade av flickan, rökte.

Figure 1. Waveform and fundamental frequency contour of test sentence 1A (marked intonation and pauses).



Männen ritade av flickan rökte.

Figure 2. Waveform and fundamental frequency contour of test sentence 1C (unmarked intonation, no pauses).

Two listeners trained in linguistics and two naive listeners participated in a pilot version of the test. The sentences were presented over a loudspeaker in a sound-treated room where the listeners were tested individually. They were instructed to rest their index finger on a button and to press it as soon as they had understood the sentence. They were requested to paraphrase each sentence after pressing the button.

Reaction times were measured by a micro-computer triggered at the beginning of each sentence by an envelope detector. The paraphrased versions of each sentence were recorded by the experimenter.

3. PILOT TEST RESULTS

A. Reaction time measurements

The expected reaction time results, i.e. that sentence versions A and B would elicit faster reaction times than sentence versions C and D and that the difference between reaction times for A and C would be greater than for B and D were not obtained. Instead, reaction times varied from around 500 msec. preceding the end of the sentence to around 500 msec. following the end of the sentence. The variation for all four listeners appeared to be completely random. This is most probably due to the rather coarse-grained nature of the task. To press a button when "you have understood the sentence" does not appear to be a very precise measure of reaction time, and probably does not correlate well at all with actual processing time.

B. Sentence interpretations

Results obtained from the listeners' paraphrases were much more illuminating. The two listeners with training in linguistics (listeners 1 and 2) were immediately aware of the structures involved and paraphrased each sentence correctly. The naive listeners (listeners 3 and 4), on the other hand, differed in their interpretations of the sentences. Both listeners, however, correctly parsed the sentences in more instances in the marked prosody versions (42%) than in the unmarked versions (25%) (Fig. 3).

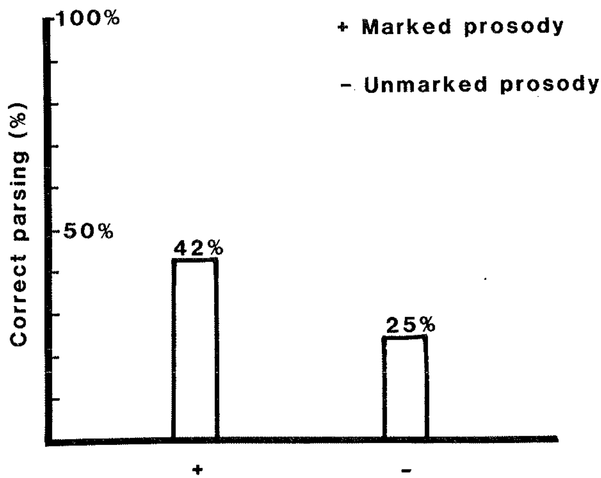


Figure 3. Correct parsing in percent for listeners 3 and 4 (combined results).

Misparsings of the sentences by both naive listeners conformed to the complexity strategies outlined above. Sentences having the structure as in 3a and 3b (sentences 1 to 3 in the test, see appendix) were most often parsed as two NP-VP's joined by "och" (and). Thus test sentence 1A "Männen, ritade av flickan, rökte," (The men, drawn by the girl, smoked) was interpreted as "Männen ritade och flickan rökte," (The men drew and the girl smoked). Listener 4, however, parsed several of these sentences as "Männen ritade av flickan och hon rökte," (The men drew the girl and she smoked). In these instances, listener 4 disregarded the unstressed prosodic features of the preposition "av". Even sentences where the syntax was specified by lexical subcategorization frames were similarly misparsed.

Sentences having the structure of 4a and 4b (test sentences 4 to 6) were misparsed by both naive listeners as NP-compound VP by adding an additional "and" between the final two verbs. It is interesting to note, however, that both sentences containing the phonetically more salient allomorph /-at/ (test sentences 5B and 5D) were correctly parsed by both listeners, and that these two sentences were the only two with the exception of 6B that were correctly interpreted by listener 4 (Fig. 4).

Correct parsings, listener 3

A	B	C	D
1A			
	2B		
3A	3B		
4A	4B	4C	4D
	5B	5C	5D
6A		6C	

Correct parsings, listener 4

A	B	C	D
	5B		5D
	6B		

Figure 4. Sentences in which correct parsing was performed (naive listeners, 3 and 4)

4. DISCUSSION

A. Reaction times

The difficulties involved in obtaining reaction times which could be correlated to actual sentence processing times rendered impossible the testing of those aspects of the hypothesis concerning the relationships between speed-of-processing, preliminary misparsing and reparsing. The fact that many of the sentences were not reparsed upon comprehension of the final verb, but rather the misparsings were maintained to conform to a simple structure, seems to indicate that in those sentences processing time might have been faster than in the sentences which received correct parsing.

Factors other than syntax and prosody might also have influenced reaction times. As previously mentioned, measuring reaction times where comprehension of an entire sentence is concerned, at least judging from this pilot test, is a very rough tool for investigating the complex process of sentence comprehension. The results indicate that perhaps more success can be attained by constructing material which can be misinterpreted and then using listener interpretations to investigate the processing of syntactic structures.

B. Syntactic complexity strategies and listener differences

The awareness of structure on the part of listeners 1 and 2 (linguistically trained) reflect the necessity of using naive listeners if misinterpretation results are to be used to investigate syntax and prosody. The difference in interpretations between listeners 3 and 4 (naive) reflect the individual nature of syntactic competence when processing complex structures.

Basically, however, these results support the complexity strategy described above. The naive listeners tended to interpret the first verb as the logical verb of the initial NP even in cases where this interpretation conflicted with phonetic form. In other words, the processing mechanism can override phonetic form to make it fit a simple sentence syntax. There seem to be, however, certain limitations depending on the phonetic saliency of the morphology (see D. below). There was not a strict priority order for choosing simple syntax over phonetic form or linguistic competence, but rather if the phonetic form of the sentence and linguistic competence would allow the misinterpretation, the simple syntactic structure was adhered to.

C. Effects of prosody on syntax

The preliminary results of the pilot test indicate that marked prosody can be used in speech perception to parse a sentence in accordance with a complex syntactic structure. Prosody then fits into the complexity strategy as it supplies information concerning sentence structure (a fall-rise in Fo and a pause) prior to the main verb.

The results also indicate, however, that such prosodic cues are not always needed for correct parsing depending on the linguistic competence of the listener and the saliency of the segmental phonetic aspects of the sentence. In a like manner, prosody is not always used leading to misparsings of sentences.

D. Morphological saliency and phonetic form

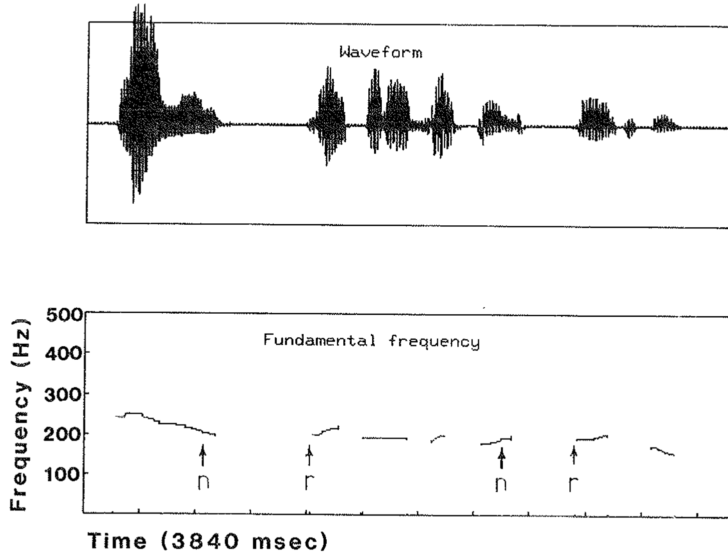
There seems to be a rather complex interplay between phonetic form (both segmental and prosodic), morphology as manifested by mapping between lexical items, and syntax. As evidenced by the domination of the allomorph /-at/ over complexity strategies and unmarked prosody, the use of information specified by the lexicon as discussed by Bresnan (1978) and Wasow (1977) can depend greatly on the phonetic saliency of the surface form. The suffix /-at/ with its clearly perceived high-frequency burst is immediately distinguished from the suffix /-ade/ whereas /-ad/+V is much more easily interpreted as /-ade/+V.

The results of this experiment can be seen as preliminary evidence for a model of syntactic processing which operates by simultaneously integrating aspects of the acoustic signal, syntactic complexity strategies involving linguistic competence, and morphological restrictions in the lexicon. This model of syntax perception would be characterized by a mechanism which latches on to the most prominent aspect or cue in the sentence which can determine syntax. The important point here is that cue dominance is not determined by the linear order in which the cues appear in the sentence, but rather dominance is affected by a number of different factors such as phonetic saliency, linguistic competence, and relative frequency of structure. It could also be conjectured that cue dominance leads to a certain kind of perceptual masking of other cues in an intricate interplay of processing.

E. Further experiments with synthetic stimuli

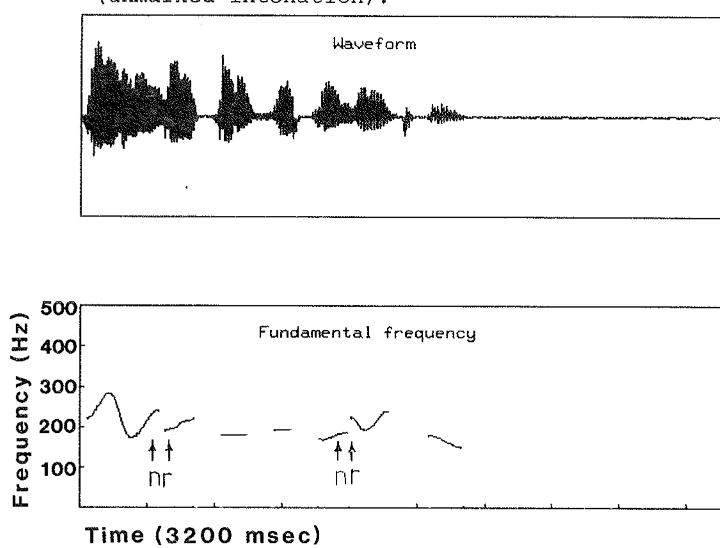
A more satisfactory test of the hypothesis and model proposed in this paper would entail the synthetic manipulation of the prosodic parameters of intonation and pauses. A much larger number of listeners would of course also be needed to support the trends indicated by the pilot study.

An informal experiment was carried out in which the intonation contour of test sentence 1A "Männen, ritade av flickan, rökte" was replaced by the intonation contour of 1C "Männen ritade av flickan rökte", i.e. unmarked intonation replaced marked intonation and vice versa (Fig. 5 and 6). Simply switching intonation, however, without changing the timing aspects produced extremely artificial sounding sentences. A systematic alteration of both intonation and pauses would need to be undertaken to more successfully test the degree of prosodic representation necessary to serve as the dominating cue in the perception of sentence syntax.



Männen, ritade av flickan, rökte.

Figure 5. Waveform and fundamental frequency contour of test sentence 1A with pitch coefficients edited to conform to the contour of test sentence 1C (unmarked intonation).



Männen ritade av flickan rökte.

Figure 6. Waveform and fundamental frequency contour of test sentence 1C with pitch coefficients edited to conform to the contour of test sentence 1A (marked intonation).

ACKNOWLEDGEMENTS

I would like to thank Helen Goodluck for much help, discussion and comments. Thanks also to Gösta Bruce and Eva Gårding for discussion and comments and to Anders Löfqvist for help with the computer program.

REFERENCES

Bresnan,J. 1978. A Realistic Transformational Grammar, in Halle, Bresnan, Miller (Eds.) Linguistic Theory and Psychological Reality. The MIT Press. Cambridge, MA.

Clark H. & Clark E. 1977. Psychology and Language, Harcourt Brace Jovanovich, New York.

Ford, M., Bresnan, J., & Kaplan, R. 1982. A Competence- Based Theory of Syntactic Closure. In Bresnan, J. (Ed.) The Mental Representation of Grammatical Relations. The MIT Press, Cambridge, MA.

Lehiste, I. 1973. Phonetic disambiguation of syntactic ambiguity. Glossa, 7, 107-122.

Wasow, T. 1977. Transformations and the Lexicon. In Culicover, Wasow, Akmajian (Eds.) Formal Syntax. Academic Press, New York.

APPENDIX

Test sentences. Commas indicate marked prosody.

1A. Männen, ritade av flickan, rökte.
1B. Mannen, ritad av flickan, rökte.
1C. Männen ritade av flickan rökte.
1D. Mannen ritad av flickan rökte.

2A. Kvinnorna, målade av mannen, log.
2B. Kvinnan, målad av mannen, log.
2C. Kvinnorna målade av mannen log.
2D. Kvinnan målad av mannen log.

3A. Flickorna, tecknade av pojken, skrattade.
3B. Flickan, tecknad av pojken, skrattade.
3C. Flickorna tecknade av pojken skrattade.
3D. Flickan tecknad av pojken skrattade.

4A. Killarna, rakade och tvättade, sjöng.
4B. Killen, rakad och tvättad, sjöng.
4C. Killarna rakade och tvättade sjöng.
4D. Killen rakad och tvättad sjöng.

5A. Barnen, tvättade och torkade, lekte.
5B. Barnet, tvättat och torkat, lekte.
5C. Barnen tvättade och torkade lekte.
5D. Barnet tvättat och torkat lekte.

6A. Kvinnorna, tecknade och målade, log.
6B. Kvinnan, tecknad och målad, log.
6C. Kvinnorna tecknade och målade log.
6D. Kvinnan tecknad och målad log.

7A. Killarna sa, att bilen skulle tvättas.
7B. Killen sa, att bilen skulle tvättas.
7C. Killarna sa att bilen skulle tvättas.
7D. Killen sa att bilen skulle tvättas.

8A. Pojkarna sa, att flickorna skulle spela med.
8B. Pojken sa, att flickorna skulle spela med.
8C. Pojkarna sa att flickorna skulle spela med.
8D. Pojken sa att flickorna skulle spela med.

9A. Flickorna sa, att pojkarna skulle måla huset.
9B. Flickan sa, att pojkarna skulle måla huset.
9C. Flickorna sa att pojkarna skulle måla huset.
9D. Flickan sa att pojkarna skulle måla huset.

10A. Killarna sa, att huset skulle målas.
10B. Killen sa, att huset skulle målas.
10C. Killarna sa att huset skulle målas.
10D. Killen sa att huset skulle målas.

Test Sentences (Translation). Commas indicate marked prosody.

1A. The men, drawn by the girl, smoked.
1B. The man, drawn by the girl, smoked.
1C. The men drawn by the girl smoked.
1D. The man drawn by the girl smoked.

2A. The women, painted by the man, smiled.
2B. The woman, painted by the man, smiled.
2C. The women painted by the man smiled.
2D. The woman painted by the man smiled.

3A. The girls, sketched by the boy, laughed.
3B. The girl, sketched by the boy, laughed.
3C. The girls sketched by the boy laughed.
3D. The girl sketched by the boy laughed.

4A. The boys, shaved and washed, laughed.
4B. The boy, shaved and washed, laughed.
4C. The boys shaved and washed laughed.
4D. The boy shaved and washed laughed.

5A. The children, washed and dried, played.
5B. The child, washed and dried, played.
5C. The children washed and dried played.
5D. The child washed and dried played.

6A. The women, sketched and painted, smiled.
6B. The woman, sketched and painted, smiled.
6C. The women sketched and painted smiled.
6D. The woman sketched and painted smiled.

7A. The boys said, that the car should be washed.
7B. The boy said, that the car should be washed.
7C. The boys said that the car should be washed.
7D. The boy said that the car should be washed.

8A. The boys said, that the girls should play along.
8B. The boy said, that the girls should play along.
8C. The boys said that the girls should play along.
8D. The boy said that the girls should play along.

9A. The girls said, that the boys should paint the house.
9B. The girl said, that the boys should paint the house.
9C. The girls said that the boys should paint the house.
9D. The girl said that the boys should paint the house.

10A. The boys said, that the house should be painted.
10B. The boy said, that the house should be painted.
10C. The boys said that the house should be painted.
10D. The boy said that the house should be painted.

Sentence order in the listening test

1. (10A)	11. (3C)	21. (6A)	31. (9C)
2. (10B)	12. (6D)	22. (9B)	32. (1D)
3. (10C)	13. (9A)	23. (1C)	33. (4A)
4. (10D)	14. (1B)	24. (4D)	34. (7B)
5. (1A)	15. (4C)	25. (7A)	35. (2C)
6. (4B)	16. (7D)	26. (2B)	36. (5D)
7. (7C)	17. (2A)	27. (5C)	37. (8A)
8. (2D)	18. (5B)	28. (8D)	38. (3B)
9. (5A)	19. (8C)	29. (3A)	39. (6C)
10.(8B)	20. (3D)	30. (6B)	40. (9D)

Swedish intonation contours in text-to-speech synthesis

Dieter Huber

ABSTRACT

The purpose of this study is to analyse the intonation contours of ten Swedish sentences pronounced by three different native speakers, in order to define a set of generative rules that can be useful in text-to-speech synthesis. Intonation in this context is understood in terms of interactions between word accents, sentence accents, initial juncture and terminal juncture, but without any reference to intrinsic qualities, segmental conditioning factors or various emotional and idiolectal features.

Out of a number of already existing intonation models for Swedish, the procedure developed by Eva Gårding and her collaborators at Lund University is adopted in a simplified version to synthesize 'idealized' F_0-contours for each of the test sentences in the pronunciation characteristics of each speaker (Bannert, 1984; Bruce, 1977; Gårding, 1977, 1979, 1981, 1983, 1984). The obtained results are compared with the original pitch sequences, mapping matches and mismatches. Some minor adjustments and modifications are proposed to improve the model. Finally, validity testing demonstrates the efficiency of the applied procedure and its underlying concepts with regard to the analysed text material, at the same time indicating areas for further research.

1. PRESENTATION

1.1 Text Material

The following ten Swedish sentences are analysed:

- Bussens förare fick körkortet indraget.
- Isen kan omöjligt bära en vuxen.
- Torpet hade blommor och gräs på taket.
- Många trivs med att vandra i fjällen.

- Sikten är ganska skymd i kurvan.
- Skorna var nya och alldeles för trånga.
- Lingonen brukar mogna i september.
- Pumpen på gården hade rostat på vintern.
- Vågorna slog högt över bryggan i stormen.
- Dagen firades med klang och jubel.

These sentences were chosen from lists compiled by Margareta Korsan-Bengtsen (Distorted Speech Audiometry, Acta Ota-Laryngologica, Suppl. 310, Göteborg 1973) and constitute statements without any prominent contextual features. They were recorded in an echoless sound studio at the Acoustics Laboratory of the Swedish Telephone Company in Stockholm under equally controlled conditions for each speaker.

1.2 Test Speakers

The three speakers comprise a 37-year old man, a 32-year old woman and a 11-year old boy, all of them living in the Stockholm area and speaking Swedish without noticeable particularities. They were instructed to read the ten statements one after another, with short intervals, in a normal and colloquial fashion as if talking to another person. As a matter of convenience I have chosen to apply the labels DIAMAL, DIAFEM and DIACHI throughout the course of this study whenever I refer to the male, the female or the child's voice respectively.

1.3 Vocoded Pitch Contours

A copy of the original tape recording was analysed at Chalmers University of Technology in Göteborg with respect to voicing determination, pitch value extraction and allophonic segmentation, using the vocoder system developed by and available at the Department of Information Theory (Hedelin, 1981). Pitch extraction was performed by the autocorrelation method (Hess, 1983; Rabiner & Schafer, 1978). Allophone labelling was carried out manually with judgement based on both visual and auditory evidence. Word accents are marked only for syllables receiving primary stress. Word boundaries and lower levels of stress are disregarded, as well as phrase accents and phrase boundaries. This analysis resulted in a set of thirty vocoded intonation contours, the first three of which are demonstrated in figure 1.

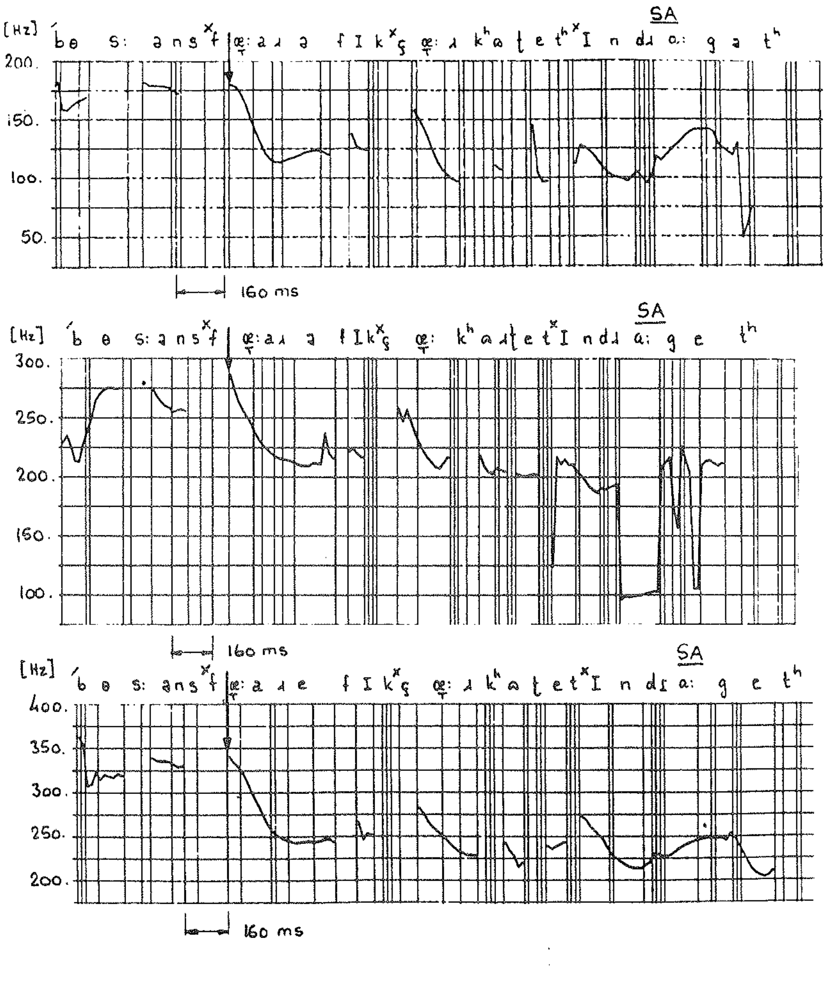


Figure 1. 'BUSSENS FÖRARE FICK KÖRKORTET INDRAGET'
Vocoded F_0-contours for DIAMAL, DIAFEM and DIACHI.
The line-up point is at the CV-boundary of the stressed syllable in 'FÖRARE'.

1.4 Generated Pitch Contours

A first superficial glance at the lined-up text material reveals both similarities and dissimilarities in the pitch fluctuations of the various F_o-contours. Some of the irregularities can be easily accounted for in terms of individual variation. Others might prove more difficult to explain. One first point of interest is that we obviously have to consider large-scale pitch movements over relatively wide ranges of frequency and duration as well as comparatively minute vacillations that are added to or superimposed upon the larger strctures without, however, changing their overall course of direction.

One possible approach to the investigation of the problems involved would be by way of a detailed description of the pitch sequences in the presented text material. Thus correlating the data obtained for DIAMAL, DIAFEM and DIACHI and viewing them in the broader light of established phonetic theory, we might hope to unearth recurrent regularities that can be formulated into a set of generative rules and corroborated by further systematic research. Such an analysis method requires, however, large and variable text material from many different sources if it is to yield reliable results, and does therefore not seem feasible within the limited scope of this study. Instead I have chosen an analysis-by-synthesis procedure, applying a simplified version of the Lund model of sentence intonation to synthesize 'idealized' pitch contours of all ten sentences in the pronunciation characteristics of the three test speakers, which I will then compare with the 'real' (vocoded) F_o-contours.

The Lund model of sentence intonation comprises in its most comprehensive application nine consecutive steps which gradually transform a given string of phonemic symbols representing the utterance in common alphabetic writing (INPUT) into its concomitant F_o-contour (OUTPUT). For a more detailed description of the entire procedure see Gårding, 1984.

Applying this model to the text material and using a tonal grid defined by four parallel lines at a distance of 20/40/20 Hz from each other, with a overall fall of 50 Hz, I obtained thirty generated intonation contours, one of which is shown in figure 2.

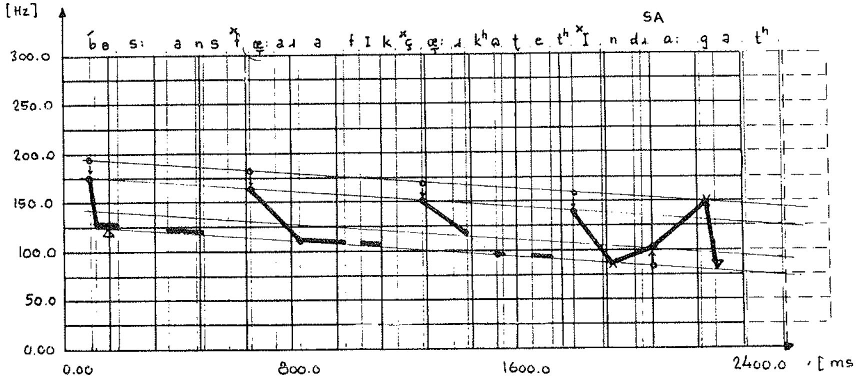
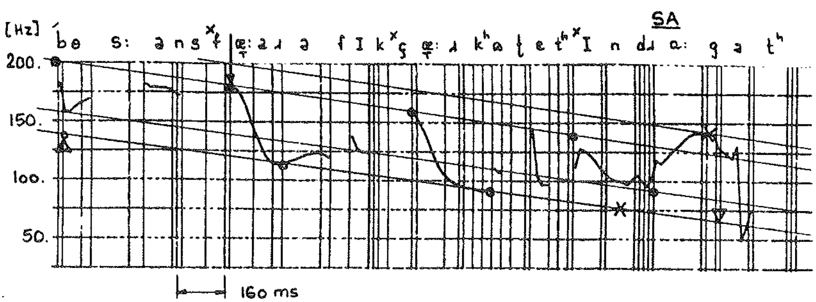


Figure 2. 'BUSSENS FÖRARE FICK KÖRKORTET INDRAGET'
Generated F_0-contour for DIAMAL.

2. COMPARISON

In order to facilitate comparison of the 'idealized' with the 'real' sequences I have transferred the generated HIGHs and LOWs directly into the vocoded contours, using prominent turning points in the latter to establish the grid. This was not always possible under the conventions suggested in paragraph 1.4. Compromise was sometimes necessary with respect to both parallelism and internal stratification. It should be noted, however, that construction of a tonal grid obeying some kind of regularity was possible without major difficulties in all the analysed sentences.

Figure 3. 'BUSSENS FÖRARE FICK KÖRKORTET INDRAGET'
Vocoded F_0-contours for DIAMAL, DIAFEM and DIACHI with superimposed grid and generated HIGHs and LOWs.



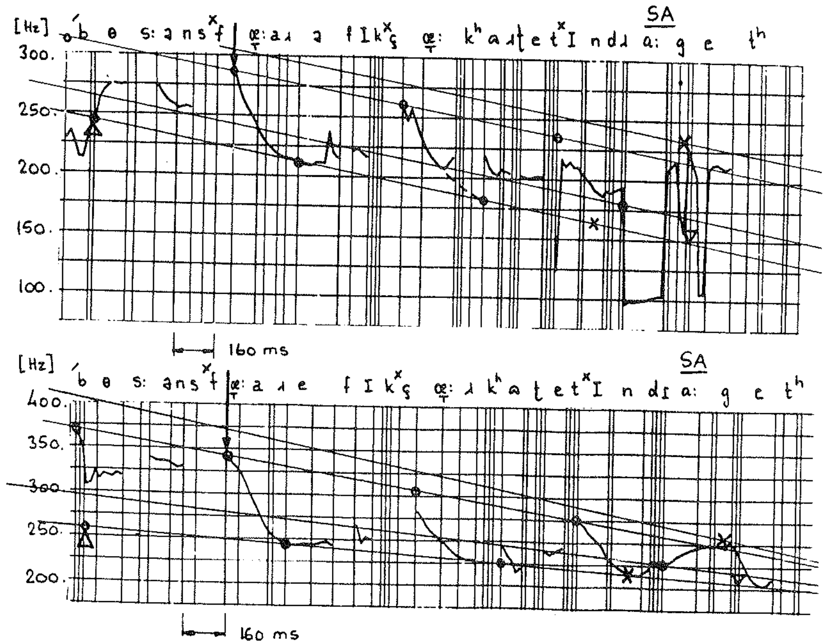


Figure 3 cont'd.

2.1 Matches and Mismatches

Comparing a total of thirty intonation contours representing ten Swedish sentences in the pronunciations of one male, one female and one child speaker, we are confronted with a bewildering number of both striking matches and blatant mismatches. Before entering a detailed discussion of the accumulated material I have tried to condense the results into a systematic arrangement which I hope will contribute to reveal some of the underlying regularities.

Table I. Matches and mismatches between vocoded and generated contours (Explanatory notes on page 8).

		GRID					INITIAL JUNCT.								WORD ACCENT							UNSTRESSED WORDS AND SYLLABLES							SENTENCE ACCENT						TERMINAL JUNCT.							
		o	▲	▼	↕	m	o	↑	↓	←	→	↗	↘	m	o	↑	↓	←	→	m	t	o	↑	↓	↗	↘	~	m	o	↑	↓	←	→	m	o	↑	↓	←	→	↗	↘	m
BUSSENS...	M	+	-	-	-	-	-	-	+	-	-	+	-	-	7	1	-	-	-	-	8	-	-	+	-	+	+	-	1	-	1	-	-	-	-	+	-	+	-	+	-	-
	F	+	-	-	-	+	+	-	-	-	-	+	-	-	6	2	-	-	-	-	8	-	-	+	-	+	+	-	1	-	1	-	-	-	-	+	-	+	-	+	-	-
	C	-	-	+	-	-	-	-	+	-	-	+	-	-	6	1	1	-	-	-	8	-	-	+	-	+	+	-	2	-	-	-	-	-	-	-	-	+	-	+	-	-
IGEN...	M	+	-	-	-	-	+	-	-	-	-	+	-	-	6	1	1	-	-	-	8	-	-	+	+	+	+	-	-	-	-	1	-	-	-	-	-	+	-	+	-	-
	F	-	-	+	-	-	-	-	+	-	-	-	+	-	1	4	3	-	-	-	8	+	-	+	+	+	+	-	2	-	-	-	-	-	-	-	+	-	-	+	-	-
	C	+	-	-	-	-	-	-	+	-	-	+	-	-	4	2	2	-	-	-	8	-	-	+	+	+	+	-	-	-	-	1	-	-	-	-	+	+	-	+	-	-
TORPET...	M	+	-	-	-	-	-	-	+	-	-	+	-	-	4	3	1	-	-	-	8	-	-	+	-	-	+	-	1	1	-	-	-	-	+	-	-	-	-	+	-	-
	F	+	-	-	-	-	-	-	+	-	-	-	+	-	3	2	3	-	-	-	8	-	-	+	+	+	-	-	2	-	-	-	-	-	-	-	-	-	-	-	+	-
	C	-	+	-	-	-	-	-	-	+	-	-	-	-	1	1	4	2	-	-	8	-	-	+	+	+	+	-	1	-	-	-	-	-	-	-	+	-	-	-	+	-
MÅNGA...	M	+	-	-	-	-	+	-	-	-	-	+	-	-	4	-	-	-	2	-	6	-	-	+	+	+	+	-	1	-	-	-	-	-	-	-	-	+	-	+	-	-
	F	-	-	-	+	+	-	-	+	-	-	+	-	-	1	2	3	2	-	-	8	-	-	+	-	+	+	-	-	1	-	-	-	-	-	+	-	-	-	+	+	-
	C	+	-	-	-	-	-	-	+	-	-	+	-	-	5	1	2	-	-	-	8	+	-	+	+	-	+	-	1	-	-	-	-	-	+	-	-	-	-	+	-	-
SIKTEN...	M	-	-	+	-	-	-	+	-	-	-	+	-	-	2	3	3	-	-	-	8	-	-	+	-	-	+	-	1	1	-	-	-	-	+	-	-	-	-	-	-	-
	F	+	-	-	-	+	-	-	+	-	-	-	-	-	3	3	2	-	-	-	8	+	-	+	+	+	+	-	1	1	-	-	-	-	-	-	+	+	-	+	-	-
	C	-	+	-	-	-	-	-	+	-	-	-	-	-	2	1	5	-	-	-	8	-	-	+	+	+	+	-	-	1	-	-	-	-	-	-	+	-	-	-	-	-
SKORNA...	M	+	-	-	-	+	+	-	-	-	-	+	-	-	6	1	1	-	-	-	8	-	-	+	+	-	+	-	1	-	-	-	-	-	-	-	+	+	-	+	-	-
	F	+	-	-	-	+	-	-	+	-	-	-	-	-	5	-	3	-	-	-	8	-	-	+	+	+	-	-	0	-	-	-	-	-	-	+	-	-	-	+	-	-
	C	-	-	+	-	-	-	-	+	-	-	-	-	-	5	-	3	-	-	-	8	-	-	+	-	-	+	-	0	-	-	-	-	-	-	-	+	-	-	-	-	-
LINGONEN...	M	-	+	-	-	+	+	-	-	-	-	-	-	-	6	1	1	-	-	-	8	+	-	-	-	-	-	-	-	1	-	-	-	-	+	-	-	-	-	+	-	-
	F	+	-	-	-	+	+	-	-	-	-	-	-	-	4	1	2	1	-	-	8	+	-	-	-	-	-	-	0	-	-	-	-	-	-	+	-	-	-	+	-	-
	C	-	+	-	-	-	+	-	-	-	-	+	-	-	4	-	4	-	-	-	8	+	-	-	-	-	-	-	0	-	-	-	-	-	+	-	-	-	-	+	-	-
PUMPEN...	M	+	-	-	-	-	+	-	-	-	-	+	-	-	5	1	2	-	-	-	8	-	-	+	-	-	+	-	1	-	-	-	-	-	-	-	-	+	-	+	+	-
	F	+	-	-	-	+	-	-	+	-	-	-	-	-	4	3	1	-	-	-	8	+	-	-	-	-	-	-	0	-	-	-	-	-	-	+	-	-	-	+	+	-
	C	+	-	-	-	-	+	-	-	-	-	-	-	-	2	2	4	-	-	-	8	+	-	-	-	+	-	-	1	-	-	-	-	-	-	-	+	-	-	-	-	-
VÅGORNA...	M	-	-	+	-	-	+	-	-	-	-	+	-	-	2	2	3	1	-	-	8	-	-	+	-	-	+	-	1	-	-	-	-	-	-	-	+	-	-	+	-	-
	F	+	-	-	-	+	+	-	-	-	-	+	-	-	5	1	2	-	-	-	8	+	-	-	-	-	-	-	0	-	-	-	-	-	+	-	-	-	-	+	-	-
	C	+	-	-	-	-	+	-	-	-	-	+	-	-	3	2	3	-	-	-	8	-	-	+	+	+	+	-	1	-	-	-	-	-	-	-	+	-	-	+	-	-
DAGEN...	M	+	-	-	-	-	+	-	-	-	-	+	-	-	3	3	2	-	-	-	8	-	-	+	-	-	+	-	1	-	-	-	-	-	-	-	-	+	-	+	-	-
	F	+	-	-	-	+	-	-	-	-	+	+	-	-	3	3	2	-	-	-	8	-	-	+	-	-	+	-	-	1	-	-	-	-	+	-	-	-	-	+	+	-
	C	-	-	+	-	-	-	-	+	-	-	-	+	-	4	1	3	-	-	-	8	-	-	+	+	+	+	-	-	-	-	1	-	-	-	-	+	+	-	+	+	-
Total		19	4	6	1	10	13	1	14	1	1	18	3	0	116	48	66	6	2	0	238	9	0	24	13	15	22	0	20(6)	7	2	3	0	0	7	6	11	11	0	24	7	0

For code, see next page.

M	=	DIAMAL
F	=	DIAFEM
C	=	DIACHI
o	=	complete correspondence between vocoded and generated aspect
▲	=	diverging grid
▼	=	converging grid
↕	=	misfit with respect to internal stratification
↑	=	generated accent/juncture mark to high
↓	=	generated accent/juncture mark to low
←	=	generated accent/juncture mark to early
→	=	generated accent/juncture mark to late
↗	=	F_o-rise missing in generated version
↘	=	F_o-fall missing in generated version
t	=	total number of accent HIGHs and LOWs
m	=	other misfits
~	=	F_o-level in vocoded version

Code to preceding table

2.2 Discussion

2.2.1 The Tonal Grid

Permitting individual adjustments with respect to width, range and rate of fall, the model predicts correct grids in nineteen out of thirty occurrences. This figure may need some modification, taking into account those incidences (marked in the m-column) where considerable parts of the intonation contour fall one octave below the otherwise regular tonal surroundings. This happens eight times for DIAFEM and twice for DIAMAL. A closer review reveals 'higher-harmonic detection' or 'higher-harmonic tracking' (if continued for a longer time) of the

subharmonic at $F_0/2$ as the probable cause. Errors of this kind are also referred to as octave errors (Hess 1983) and can be treated in terms of individual aberrations on the phonation level with negligible impact on the intonation sequence as a whole. Therefore they do not need to be taken into consideration when drawing the grid.
Of the remaining 'non-correct' patterns in the presented text material, four can be described as diverging and six as converging tonal grids, all ten of them displaying accurate internal stratification. I am hesitant to dismiss those cases simply as not conformable, or rather, I expect that further advances in the research of the general concept of the grid and its applications to different languages and sentence types, including emotional as well as attitude features, will sooner or later provide us with the tools to ascribe correctness even to these kinds of deviations. One possible lead in this direction might be that non-parallelism in our material occurs nine out of ten times for male speakers (three times for DIAMAL, six times for DIACHI) but only once in the case of DIAFEM. Much larger text material has to be analysed, however, before these observations may be established as reliable facts.

One example, test sentence number four in the DIAFEM-version, displays a grid which clearly deviates from the internal 1-2-1 stratification principle, at the same time differing markedly even with respect to the overall intonation slope. Far from proffering any explanation I wish to stress that the vocoded intonation contour even in this case, however, provided enough clues to establish some kind of grid pattern without any greater difficulties.

2.2.2 Initial Juncture

The placement of the initial juncture LOWs has been predicted correctly in thirteen out of thirty cases in the presented text material. Faulty judgements are confined almost exclusively to the frequency scale with an overwhelming tendency towards underrating (14:1 ratio too low versus too high). Temporal defects are comparatively rare (once too early and once too late).

If the model does very nicely with respect to at least the temporal prediction of the initial juncture LOWs, it does not succeed very well in generating the F_0-rises that typically accompany those LOWs. Our text material displays eighteen such F_0-rises (versus three F_0-falls and nine F_0-levels), none of which is predicted by the generated sequences.

It will be discussed later, if theses F_0-rises are to be treated as intrinsic juncture features or rather as belonging to concatenation.

2.2.3 Word Accents

The thirty Swedish sentences constituting the text material for this study contain 238 word accent markers (HIGHs and LOWs), which are distributed in pairs over 119 words. 61 of these words receive acute accent (ACCENT I). The remaining 58 words have grave accent (ACCENT II).

116 of the 238 word accent markers, which means roughly half of them (48,7 %), are synthesized at their proper locations in the vocoded contours. 48 (20,2 %) are placed too high and 66 (27,7 %) too low, which adds up to a total of 114 word accent markers (47,9 %) which are positioned correctly on the temporal scale with deviation in pitch determination only. The remaining 8 word accent markers (3,4 %) are located inaccurately on the temporal scale. Half of them (1,7 %) are likewise faulty as to their frequency level and can thus be regarded as totally misplaced.

To sum up the results so far it can be stated that with regard to the presented text material, the model proves

- highly successful in the temporal prediction of word accent markers (96,6 %)
- considerably less reliable in frequency ranging (50,4 %)

with a total failure rate as low as 1,7 %.

Defects in correct F_0-labelling are distributed rather evenly between overshooting and undershooting values, which excludes inaccurate grid placement as possible cause. Improved results might be achieved by introducing phrase structure, by adjusting the algorithm for pitch generation to variable sentence types or by altering the concatenation rule.

2.2.4 Unstressed Words and Syllables

The Lund model of sentence intonation defines the F_o-sequences of unstressed words and syllables by way of concatenation, using copy and join rules to establish both levels and directions. Applying these conventions to our text material we usually obtain F_o-levels which are about half an octave below the vocoded ones. The table in paragraph 2.1 reveals nine correct occurrences versus 24 where the contour is placed too low. There is no incidence where the generated conour is located too high.

One further discrepancy between the synthesized and the vocoded contours is found in the realization of small scale F_o-rises (in 13 cases) and F_o-falls (in 15 cases) in unstressed words and syllables, which the model does not distinguish at all. Even though most of them may simply be 'ripples on waves on swells on tides' in Dwight Bolingers analogy (1964) and thus irrelevant to our perception of sentence intonation, I am still hesitant if it is wise to neglect them as is generally done in speech synthesis.

2.2.5 Sentence Accent

One first observation has to deal with the absence of sentence accent in six out of thirty analysed sentences, four of them being pronounced by DIAFEM, two by DIACHI. The appropriate contours were thus generated totally disregarding any kind of sentence accent commands, producing correct results in all six cases.

Including these six occurrences in the o-column, there is an allout number of 26 accurate sentence accent markers (68,4 %) versus 7 too high (18,4 %), 2 too low (5,3 %) and 3 too early (7,9 %).

Summing up these results under the same prerogative as in paragraph 2.2.3, it can again be noted that the conventions -

here for sentence accent marking - prove

- highly successful in the temporal prediction (92,1 %)
- less reliable in frequency ranging (76,3 %)

with a total failure rate all the way down at 0 %.

Two of the three markings which are placed to early, are specially interesting, as they seem to reveal a plausible explanation for the delays. The sentence accent HIGHs in both the DIAMAL and DIACHI version of 'ISEN KAN OMÖJLIGT BÄRA EN VUXEN' together with the adjoining terminal juncture LOWs determine the ordinates of two large-scale intonation contour falls, which according to the rules would have to be implemented entirely within the limits of voiceless speech segments. What the vocoded text material displays, however, is that the F_0-falls are placed not where the model predicts but in the voiced speech areas following immediately after. In other words, the whole F_0-contour confined by the sentence accent plus terminal juncture markers is in both cases transferred into the adjoining voiced section without changing neither the pitch levels nor the falling rate.

Similar occurrences of pitch contour reorganisation in connection with voiceless segments have been described by Rapp (1971), Eriksson (1973), Bannert & Bredvad-Jensen (1975) and others. The procedure seems both reasonable and logical. It would be difficult to imagine how large-scale pitch movements within the limits of unvoiced speech could be performed otherwise by the human voicing mechanism without simply truncating them. Two incidences out of thirty are, however, not enough to establish reliable evidence and have to serve in the context of this limited study as mere observations.

2.2.6 Terminal Juncture

The placement of the terminal juncture LOWs has been predicted correctly in seven out of thirty cases in the presented text material. Contrary to the results obtained for the initial juncture, inaccurate predictions occur both with respect to frequency and time.

F_0-rises after terminal juncture LOWs are even more common here (24 occurrences) than they were for initial junctures. The model does not predict any one of them. Again the question arises if these features are to be treated as an outcome of concatenation or rather as intrinsic juncture quality.

2.3 Modifications

It can reasonably be assumed that many of the shortcomings of the synthesized contours will be avoided once there are more adjustable rules and conventions on how to establish the grid to accommodate different contextual surroundings. This kind of improvement is specially to be expected with regard to frequency level determination in both stressed and unstressed parts of the speech utterance, but also when it comes to include phrase structure into the general model. I will therefore not deal with the problems of frequency ranging in this paragraph, leaving possible solutions to future research.

Some minor adjustments, however, I would like to suggest already here, in order to ameliorate the results obtained for the text material analysed in this study. One of these supplementary rules deals with the construction of the grid, two with initial juncture and terminal juncture respectively, and the two remaining ones are amendments to the concatenation process.

1. Permit diverging and converging patterns as well as parallelism when drawing the grid.
2. Connect initial juncture LOW with the next following word accent HIGH by direct interpolation.
3. Add F_0-rise to terminal juncture LOW.
4. Move large-scale F_0-movements into adjacent voiced areas without changing their properties, if otherwise they would be located entirely within voiceless sections.
5. Connect F_0-turning-points representing

different pitch levels by cosine interpolation.

Application of these amendments to our text material of thirty sentences would score the following improvements:

1. Ten more 'correct' grids, rendering a total of 29 out of 30 accurate predictions.
2. Considerable improvements both with respect to initial juncture rises (18 incidences) and prediction of unstressed words and syllables (27 cases).
3. 24 correct predictions out of 30.
4. Refers to sentence 'ISEN KAN OMÖJLIGT BÄRA EN VUXEN' (see paragraph 2.2.5).
5. Reflects the general impression of the F_o-contours in all thirty text sentences.

3. VALIDITY TESTING

One question left unanswered in this study so far concerns our perception of intonation contours. Do the simplified contours generated by the Lund model of sentence intonation actually produce acceptable human intonation when used in speech synthesis?

Using again the digital equipment available at the Department of Information Theory at Chalmers University of Technology in Göteborg, I replaced the vocoded contours of the first three sentences gradually by their respective generated ones, first without the modifications suggested in paragraph 2.3, later with them included. The final result for sentence 1 spoken by DIAMAL is demonstrated in figure 4.

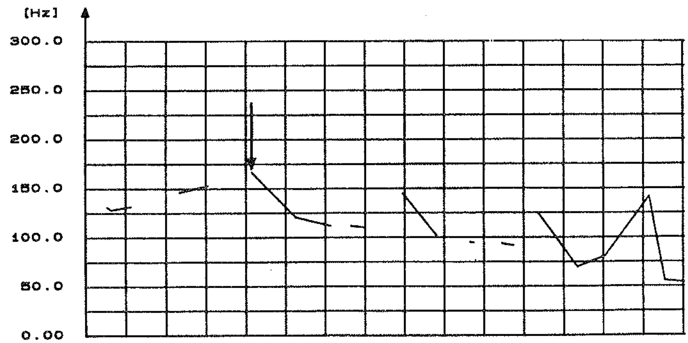


Figure 4. 'BUSSENS FÖRARE FICK KÖRKORTET INDRAGET'
Synthesized F_0-contour for DIAMAL in the modified final version. The arrow marks the CV-boundary of the stressed syllable in 'FÖRARE'.

The concomitant sound tracks were played to a group of listeners after every introduced change and were continually critisized by them as to the naturalness of the perceived intonation.

3.1 Results

Speech synthesis based on the unmodified contours only produced curiously base pitch level impressions in the first part of all sentences, whereas the later sections sounded quite acceptable. Sentence number two (ISEN KAN OMÖJLIGT BÄRA EN VUXEN) was beyond that characterized by the total absence of emphasis, which did not coincide with the impression from the original recording.

The first defect could be completely remedied in all incidences by applying modification number two, which means by linear interpolation between the initial juncture LOWs and the next following word accent HIGHs without clinging to the baseline.

Application of modification number four reestablished sentence accent in the voiced part of the last word in sentence two, which thus replicated the tonal contour of the original vocoded version.

Both rules number two and number four proved thus highly productive in the limited context of the synthesized material.

The equally proposed modification number three (adding F_0-rises after terminal juncture LOWs) did on the other hand not produce any easily discernable improvements and was judged negligible for the purpose of this study.

The remaining two modification rules number one and number five were not included in the test procedure at all, number one out of lack of appropriate features in the synthesized material (all three sentences show parallel tonal grids in the vocoded versions), number five because linear interpolation already produced highly satisfactory results.

BIBLIOGRAPHY

Bannert R. 1984. Towards a model for German prosody. Working Papers 27, 1-36. Department of Linguistics, Lund University.

Bannert R. and A.-C. Bredvad-Jensen. 1975. Temporal organization of Swedish tonal accents: The effects of vowel duration. Working Papers 10, 1-36 and 15 (1977), 133-138. Phonetics Laboratory, Lund University.

Bolinger D.L. 1964. Around the edge of language: Intonation. Harvard Educational Review 34, 282-293.

Bruce G. 1977. Swedish word accents in sentence perspective. Dissertation Lund University.

Erikson Y. 1973. Preliminary evidence of syllable locked temporal control of F_o. STL-QPSR 2-3, 23-30.

Gårding E. 1977. The Scandinavian word accents. CWK Gleerup, Lund.

Gårding E. 1979. Sentence intonation in Swedish. Phonetica 39, 288-301.

Gårding E. 1981. Contrastive prosody: A model and its application. Studia Linguistica vol 35, no 1-2, 146-166.

Gårding E. 1983. A generative model of intonation. In Prosody: Models and Measurements, Cutler&Ladd (eds.), 11-25, Springer Verlag, Berlin.

Gårding E. 1984. Comparing intonation. Working Papers 27, 75-99. Phonetics Laboratory, Lund University.

Hedelin P. 1981. A tone-oriented voice excited vocoder. Proceedings of the 1981 IEEE International Conference on Acoustics, Speech and Signal Processing, Atlanta GA, 205-208.

Hess W. 1983. Pitch determination of speech signals. Springer Verlag, Berlin.

Korsan-Bengtsen M. 1973. Distorted speech audiometry. Acta Ota-Laryngologica, Suppl. 310, Göteborg.

Rabiner L.R. & R.W. Schafer. 1978. Digital processing of speech signals. Prentice Hall, Englewood Cliffs NJ.

Rapp K. 1971. A study of syllable timing. STL-QPSR 1, 14-20.

Language disordered children's reading and spelling: Preliminary results [1]

Kerstin Nauclér and Eva Magnusson

A great number of children who have been diagnosed as language disordered during their pre-school years will have reading and spelling difficulties at school (Rabinovitch, 1968, Bergendal, 1969, Bruce et al, 1978). This is so for a majority of the language disordered children, even if some of them seem to acquire written language in much the same way as normal children do. For a few exceptional ones it has been noticed that learning to read and write is followed by a sudden increase in their speech production ability.

All language disordered children do not show the same kind of language disturbances. For some children speech perception problems dominate, while for others different types of speech production problems are the most prevalent. It should thus be possible to identify subgroups within the larger group of language disordered children and to relate these subgroups to children's later reading and spelling success. By doing so, it would be possible to find out if children with one type of language disorders are more likely to become poor readers and/ or spellers than children with another type of disorders.

It has been reported in several studies (e.g. Liberman et al, 1977, Lundberg et al, 1980) that the best predictor of pre-school children's future reading and spelling success is their linguistic awareness. In studies of normally developing children it has been shown that linguistic awareness increases with age

1) Paper presented at the 7th World Congress of Applied Linguistics, Brussels, Belgium, August 5 - 10, 1984

(Bruce, 1964, Calfee et al, 1973, Liberman, 1973).

Results from studies with language disordered children (e.g. Curtiss, 1977, Magnusson, 1983) indicate that linguistic awareness is not necessarily directly related to the linguistic level as manifested in children's speech production. In this perspective, it is of interest to study the relationship between linguistic awareness and language developmental level in preschool children, (especially in language disordered pre-school children) and their later reading and spelling success.

We hypothesize that different linguistic and metalinguistic factors are differentially relevant for learning to read and spell. We are now going to report on a study where we have tried to identify some such linguistic and metalinguistic factors and to evaluate their importance as predictors for language disordered children's later reading and spelling success.

The subjects in this investigation are taken from an earlier study by Magnusson (1983) where the phonology of 32 language disordered children was described. The children had been diagnosed by speech pathologists but had not yet been enroled in any language programs. The age of the subjects ranged from 3:9 to 6:6. Twentyfive of these children agreed to take part in this follow-up study. The recordings from the first study were made six years ago. The children's speech production was (at that time) registered by means of a naming test and in spontaneous speech. A rhyming test was used to test their linguistic awareness. From these data subclassifications were made, based on

- phonological patterns
- phonological deviance score based on values for developmental status, range and frequency values for each phonological process
- rhyming scores
- syntactic production evaluated in spontaneous speech
- lexicon calculated from the naming task

From the production data, the subjects were divided into four subgroups with different phonological patterns consisting of

1 - children whose speech was nearly normal

2 - children in whose speech implicational patterns could be found. The degree to which the subjects deviated from the norm differed, but they were classed together since the same implicational ordering could be observed in the patterns they exhibited, i.e. the subjects had reached different developmental levels though following the same developmental order.

3 - children who had one dominating segmental problem. The problematic segment was not among the types that are the latest to be acquired by normally developing children.

4 - children whose speech was characterized by word patterns. Restriction of word structure seemed to be a more important determinant of their speech than substitutions or cluster reduction patterns.

The deviance scores ranged from 3 to 158 and were divided into three degrees where

1	covers the range	0 - 49
2	" " "	50 - 99
3	" " "	100 -

The rhyming test consisted of nine tasks. Based on statistical criteria the subjects were regarded as

good rhymers (5) (six correct choices or more)
poor rhymers (5) (five correct choices) and
non-rhymers (14) (four correct choices or less)

(One subject was not tested.)

The syntactic production was assigned values from 1 to 4, where

1 indicated normal performance for the age,
2 nearly normal,
3 below the norm and
4 far below the norm for the age.

Lexical scores were calculated from the naming task. The results were arranged into three groups where

1	corresponds to	99% - 90%	correct naming
2	"	89% - 81%	" "
3	"	80% and less	

When first tested, the age of the subjects ranged from 3:9 to 6:6 as mentioned before. At the time of the present investigation, the children have attended school for 3 to 5 years. The subjects' parents and teachers were interviewed about the children's reading and spelling, their general academic success and their adjustment at school, as well as about the children's enrolment in special teaching programs and/or language programs.

According to the parents, 46% of the children had had or still had reading problems and 75% had spelling difficulties. According to the teachers, 55% of the children were poor readers and only a few of them performed above the average. 40% were poor spellers and about the same number were judged as average spellers. Thus, in the parents opinion, spelling was more problematic than reading, while the teachers considered reading a bigger problem than spelling. Although the judgements of the parents and the teachers are not in complete agreement, the number of poor readers and spellers is much higher than in a group of students of the same age with no history of language disorder.

Two third of the children had been or still were enroled in special teaching programs or language programs. Earlier reports on the high frequency of reading and spelling problems among language disordered children are thus supported.

Our aim was to evaluated not only the reading and spelling performance but the children's present linguistic and metalinguistic ability as well. Both spoken and written performance in perception as well as production was to be considered.

Apart from the conventional <u>reading</u> and <u>spelling tests</u>, <u>verbal comprehension</u> was evaluated by means of the <u>Token test</u> (de Renzi et al, 1962). The test consists of tokens of various geometrical forms, sizes and colours to be manipulated according to verbal instructions.
The <u>comprehension of logic-grammatical relations</u> was measured by means of a test based on Luria's theories and developed for linguistically normal children by Askman et al (1982). It contains among other things inverted constructions, double negations and double comparisons.

Comprehension and production of syntactic structures were investigated by encouraging the children to model syntactic structures of various kinds.

Oral production was investigated by asking the children to tell a story about a picture and written production by asking them to write a composition about the same picture.
Finally, the children's phonological awareness was evaluated by means of a rhyming test, the same one as was used when the subjects were tested as pre-school children. The children were asked to choose rhyming word pairs out of sets of rhyming and non-rhyming words.

Here are the preliminary results from the tests:

Reading and spelling performance

Nine of the 25 subjects made a lot of reading errors and two of them did not manage to read more than half of the text. This equals roughly the number of children judged by parents and teachers as having reading problems. Only four of the subjects read the text with less than five errors. Spelling performance was somewhat better as six subjects made none or only one error but ten made a high amount of misspellings.

As can be seen in table 1, all subjects with few reading errors made few spelling errors, but not all subjects with few spelling errors made few reading errors. Seven of the 11 subjects who made lots of reading errors also made lots of spelling errors, while the other four made less spelling errors. Three subjects were better readers than spellers but six subjects were better spellers than readers.

Language comprehension

The Token test was managed fairly well by the majority of the subjects. Six of them scored 100% correct. A certain relation to reading performance was found, as the two subjects who did not manage to read the whole text both scored far below the others (13/22) and five of the six subjects making more than 4 errors were among the worst readers.

Tabel 1. Number of good (I), average (II), and poor (III) readers and spellers.

	READING I	READING II	READING III
SPELLING I	4	2	
SPELLING II		5	4
SPELLING III		3	7

Grammatical comprehension

The logic-grammatical test turned out to be more difficult than the Token test, and only one subject scored 100% correct. This is what could be expected from the testing of linguistically normal children (cf Askman et al, 1982) who also found some of the tasks difficult to master. However, there seemed to exist a certain connection between poor performance on the test and poor reading, since all nine subjects with low scores (< 18/22) also scored low on the reading test.

Comprehension and production of syntactic structures

This was also a difficult task for most subjects. None of them scored higher than 9 out of 11. No obvious relation to reading performance could be found by using only the raw data. The qualitative analysis still remains to be done.

Spoken and written production

The linguistic analysis of the children's spoken and written production when they were telling a story about a picture has not yet been completed. It was obvious, however, that there was no phonological deviance in the children's production anymore.

Metalinguistic ability

The metalinguistic ability or more precisely the phonological awareness of the subjects as measured by the rhyming test was compared with the children's spelling scores since it has been argued that phonological awareness predicts spelling performance better than reading performance. We found that all good spellers obtained top scores on the rhyming test. So did some of the avarage and poor spellers as well, but the poor rhymers were all poor spellers (and readers). This indicates that phonological awareness (rhyming ability) is a necessary but not sufficient prerequisite for spelling.

Reading strategies

By means of a detailed linguistic analysis of the reading errors, taking into account not only error types such as deletion or addition but also what kind of linguistic unit and what linguistic level that was affected by the error, it could be shown that all readers did not use the same reading strategy.
The errors made by the best readers were almost exclusively made on meaningful units, i.e. morphemes or words, the errors made by the worst readers on meaningless units as phonemes and syllables. Four poor readers, however, showed the same reading strategy as the good readers in that the majority of their errors included and affected meaningful units. Those four were all better spellers than readers. This strongly suggests that when investigating reading and spelling proficiency it is insufficient to account for the quantitative aspects only.

We started this investigation with the hypothesis that various linguistic and metalinguistic factors are differentially relevant for learning to read and spell. The aim was to identify such factors and evaluate their importance as predictors for language disordered children's reading and spelling success. The subjects' reading and spelling was therefore examined in relation to pre-school data of the following types:

1 - phonological patterns
2 - phonological deviance scores
3 - rhyming scores
4 - syntactic production
5 - lexicon

This was done in order to evaluate to what extent the following predictions based on the pre-school data are valid: Value 1 on a certain factor was taken to predict good performance, value 2 average performance and value 3 poor performance. Predictions based on phonological patterns imply good performance for subgroups 1 and 3, and poor performance for subgroup 4. No prediction is made for subgroup 2 besides what can be predicted from the deviance scores.

All factors are regarded as equally potent predictors for reading and spelling at this point. Each value is considered in relation to reading and spelling performance with regard to whether the prediction is met or not. By this procedure the following ranking lists for factors predicting reading and spelling correctly were set up:

READING		SPELLING	
syntax	.68	syntax	.60
group	.50	rhyming	.44
deviance score	.48	group	.42
rhyming	.44	deviance score	.36
lexicon	.40	lexicon	.36

The factors predicted reading correctly to a somewhat greater extent than spelling. The best predictor for both reading and spelling was syntax. (As mentioned before, this is the same correlation as we found when comparing reading and spelling with the results from the syntactic testing in the follow-up study.) The least predictive factor was lexicon.

Rhyming is the second best predictor for spelling, while for reading rhyming is placed lower in the ranking list although the prediction value is the same. This is as could be expected from many observations made by both educators and researchers that phonological awareness (here measured by a rhyming task) predicts spelling performance better than reading performance.

The next step was to check how well reading and spelling respectively were predicted by the five factors for each subject. When all factors indicated the same outcome the prediction was considered reliable. When one factor at the most contradicted the others the prediction was considered less reliable, but still a prediction. Using these principles, the reading performance for nine subjects is reliably predicted, and for another seven subjects less reliable predictions are made. Thus, reading performance of 16 of the 25 subjects can be predicted with fairly good reliability. Spelling performance is also reliably predicted for nine subjects and less reliably for another six, thus accounting for 15 of the 25 subjects.

When both reading and spelling are considered in the same individual, reliable predictions are made for six subjects and less reliable ones for seven subjects. Thus, it is possible to make correct predictions for 13 subjects. When both reading and spelling are taken into account, seven subjects showed reading and spelling scores that were contradictory to what was predicted from the factors. Four subjects had developed better reading and spelling than predicted, and three less good than predicted. If syntax were considered a more powerful predictor than the other factors, the three subjects that developed less well than predicted and one of the subjects that developed better than predicted could be accounted for.

When reading and spelling proficiency both have the same value (good, poor or average), the predictions are more reliable than when reading and spelling performance have different values, e.g. poor spelling and avarage reading, or good reading and poor spelling. This is so because the same factors have been used for predicting both reading and spelling and, furthermore, have been assigned equal power for predicting both reading and spelling. Some of the results suggest that some factors have differentially predictive values for reading and spelling, as e.g. rhyming, which is a better predictor for spelling than for reading. A further analysis of differentially predictive values seems to be promising in the quest for more reliable predictions when reading and spelling are performed at different levels.

On the whole, poor reading and spelling is more often reliably predicted than good reading and spelling. Five out of eight subjects with poor reading and spelling are correctly predicted while only one subject out of five with good reading and spelling is correctly predicted. For clinical purposes the more frequent predictions of poor readers and spellers, i.e. the identification of the at-risk children, is more important than the correct identification of the other group.

TO SUM UP

Pre-school children with language disorders will have more problems with reading and spelling in school than linguistically normal children. This investigation has shown that

1. Reading and spelling performance in language disordered children can be predicted from pre-school data.
2. The most important factor for predicting both reading and spelling is syntactic ability, the least important for both is lexicon.
3. One factor seems to predict spelling better than reading, i.e. phonological awareness as measured by rhyming ability.
4. Poor reading and spelling is more correctly predicted than good reading and spelling.

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

REFERENCES

Askman, C., Hansson, K., Johansson, N.& Åström, B. 1982. Är en myra större än en elefant? Testning av normala barns förståelse av logisk-grammatiska konstruktioner. Examensarbete i logopedi. Institutionen för logopedi och foniatri, Lunds universitet

Bergendal, B-I. 1969. Språklig oförmåga manifesterad i tal och skrift. Lic.avhandling i fonetik, Lunds universitet

Bruce, D.J. 1964. Analysis of word sounds by young children. British Journal of Educational Psychology 34, 158 - 169

Bruce, B., Persson, I. & Santesson, R. 1978. Läs- och skrivförmågan hos språkförsenade barn. 60-poängsarbete i fonetik, Lunds universitet

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

Curtiss, S. 1977. Genie. A psycholinguistic study of a modern-day "wild child". Academic Press, New York

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

Liberman, I.Y., Shankweiler, D., 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 the first school years predicted from phonemic awareness in the kindergarten. Scandinavian Journal of Psychology 21, 159 - 173

Magnusson, E. 1983. The phonology of language disordered children. Production, perception, and awareness. Travaux de l'institut de linguistique de Lund XVII, Gleerups, Lund

Rabinovitch, R.D. 1968. Reading problems in children: definitions and classifications. In Keeny & Keeny (eds.), Dyslexia. Diagnosis and treatment of reading disorders. The C.V.Mosby Company, Saint Louis

deRenzi, E. & Vignolo, L.A. 1962. The Token test: a sensitive test to detect receptive disturbances in aphasics. Brain 85, 665 - 678

Pharyngalization in Cairo Arabic

Kjell Norlin

1. INTRODUCTION

This paper reports an investigation of pharyngalization in Cairo Arabic, one of the most important Arabic dialects.

A characteristic feature of Arabic is that it has few contrasting vowel phonemes, but a large number of consonants, Fig.1. Among the consonants there is another striking feature, which is typical of most Semitic languages, namely the forming of extra series of consonants by using different secondary articulations. In Arabic pharyngalization is used to form extra series of stops and fricatives, a very rare way to produce phonemic contrast in the languages of the world.

The Arabs themselves consider pharyngalization to be a very characteristic feature of the language and generally ascribe it to the consonant. There has been much discussion, however, about what segment is primarily pharyngalized, the consonant or the surrounding vowels. Surrounding vowels are namely strongly affected.

2. PROCEDURE

The investigation is based on six speakers who have recorded real words in a sentence frame with consonants in word initial position, surrounded by the vowel [a]. The pharyngalized and non-pharyngalized stops are disregarded in this context. The voiceless ones do not differ from each other in length of aspiration or duration. The voiced ones do not differ from each other in duration or wave-form.

3. ANALYSIS

1. For the investigation of the consonant segments FFT-spectra of each fricative to the limit of 10 kHz were made. These FFT-spectra in logarithmic scale were transformed to critical band spectra according to the method which has been worked out by

Manfred Schroeder el al. (1979). According to this formula, the spectrum is described as 24 bands. It is drawn as a histogram with each critical band as a bar with constant breadth where the heigth of the bar indicates the average intensity level in dB within each critical band. Each band corresponds to the same distance on the basilar membrane of the ear and gives a more correct auditory representation than the original FFT-spectra. For practical reasons only bands 2-24 were used. To differentiate between the fricatives three measures were used, namely the center of gravity in the critical band spectra, the dispersion in the same spectra and the average level of intensity. The center of gravity is a measure of the overall pitch level of the spectrum and the dispersion is a measure of its flatness.

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

4. RESULTS FROM CRITICAL BAND SPECTRA

A comparison between critical band spectra of /s/ and /ṣ/ shows that both are characterized by a peak in the high frequency range with a sharp fall towards lower frequencies, with /ṣ/ having a more flattened peak than /s/, Figure 2.
/z/ and /ẓ/ both have a peak in the lowest bands in addition to the peaks in the high frequency range which characterize /s/ and /ṣ/ where the voiced pair has lower intensity as compared with the voiceless counterparts, Figure3 . Plotting the center of gravity in critical band spectra against dispersion shows that pharyngalized fricatives have a lower center of gravity and greater dispersion than their non-pharyngalized counterparts, Figure 4. Plotting center of gravity against level of intensity shows the pharyngalized fricatives to have lower intensity than the non-pharyngalized counterparts, Figure 5.

5. RESULTS FROM SPECTROGRAMS

The mentioned differences are based on average values of six speakers and in the individual case they do not need to be particularly great. In one case the relation between /s/ and /ṣ/ was even the opposite as compared with Figure 2. Therefore the difference does not always seem to be sufficient or reliable to allow a differentiation between pharyngalized and non-pharyngalized pairs of sibilants. As a consequence it is also necessary to look at the effects of pharyngalization on the following vowel. Measurements on spectrograms show strong influence on formant transitions of the following vowel, but there are also differences between pharyngalized and non-pharyngalized vowels in the steady state of the vowel. Regarding formant transitions, F2 in particular is affected and the beginning of F2 is rather drastically lowered for all vowels except /uu/, Figure 6. /uu/ is raised by 100-150 Hz instead, Figure 7. This is a consequence of the pharyngal tounge constriction. Uvular-pharyngal tounge constrictions have these effects on F2 as shown by Gunnar Fant (1968). The formant frequencies were also measured in their steady-states, however, to clarify the differences in vowel quality. A comparison between long pharyngalized and non-pharyngalized vowel pairs shows that /ạạ/ differs strongly in F2 and is further back than /aa/. /ii/ and /ịị/ overlap to a great extent, but t-test shows that the difference is significant. /uu/ and /ụụ/ overlap altogether and t-test shows that the difference is not significant, Figure 8. The short vowel pairs on the other hand are always different, where all short vowels are further back in pharyngalized surroundings, Figure 9.

6. DISCUSSION

In the end the picture of pharyngalization turns out to be rather complicated. On the one hand pharyngalized and non-pharyngalized sibilants differ, but not in an altogether clear way. On the other hand vowels also differ, but in a more complex way. Pharyngalized and non-pharyngalized low vowels show great difference in the F2 dimension with no difference between long and short vowels, with pharyngalized vowels further back. Long, high vowels show no or small difference, but short high vowels are always further back. Since pharyngalization is phonemic in Arabic it is only natural to ask what factor is the most important one in this connection. There are very few investigations done in this field, but one made with synthetic speech by Dean Obrecht (1968) shows that the formant transitions are the most important cue to perceive pharyngalization.

The conclusion is that the acoustic correlates of pharyngalization cannot be ascribed to one single segment. Its minimal domain is the syllable where formant transitions are the most important perceptual feature.

ACKNOWLEDGMENTS

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

REFERENCES

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

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

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

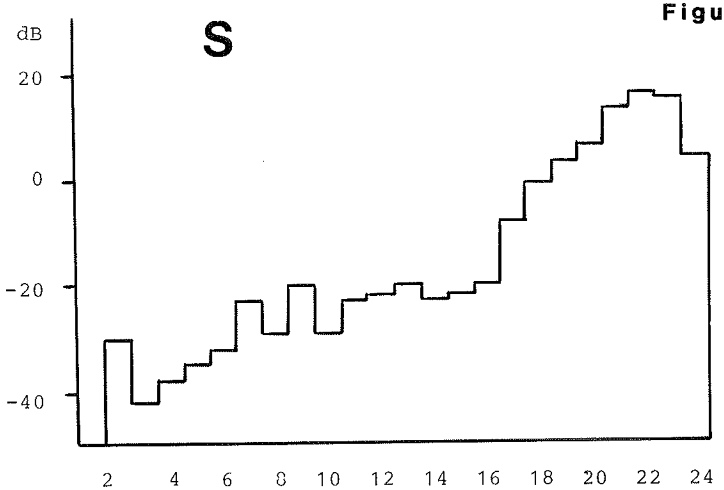
Figure 1

b	t d		k g	q		ʔ
	ṭ ḍ					
f	s z	š	x ɣ	ḥ	ʕ	h
	ṣ ẓ					
m	n					
	l					
	r					
w		y				

---

ii	uu	i	u
ee	oo		
aa		a	

Figure 2


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

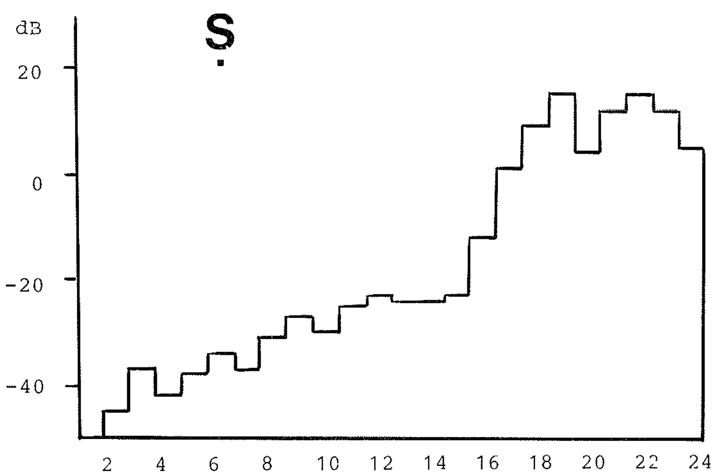
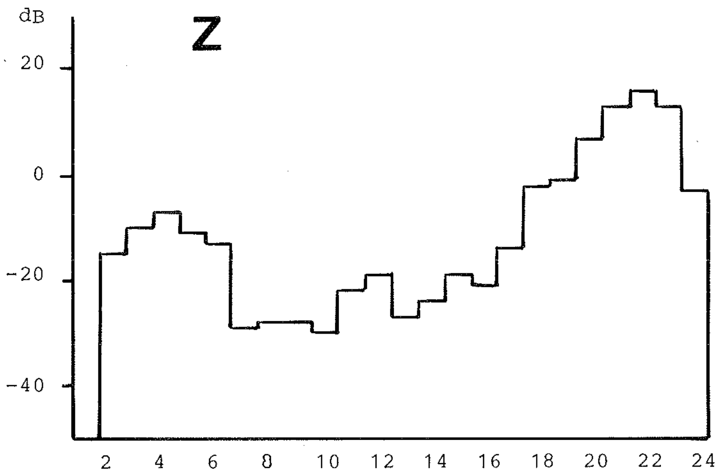
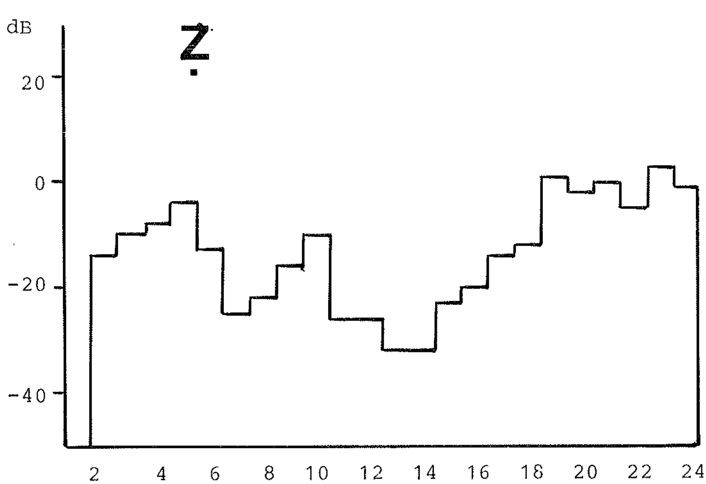

dB
Ṣ
20
0
-20
-40
2
4
6
8
10
12
14
16
18
20
22
24

Figure 3


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


dB
Ẓ
20
0
-20
-40
2
4
6
8
10
12
14
16
18
20
22
24

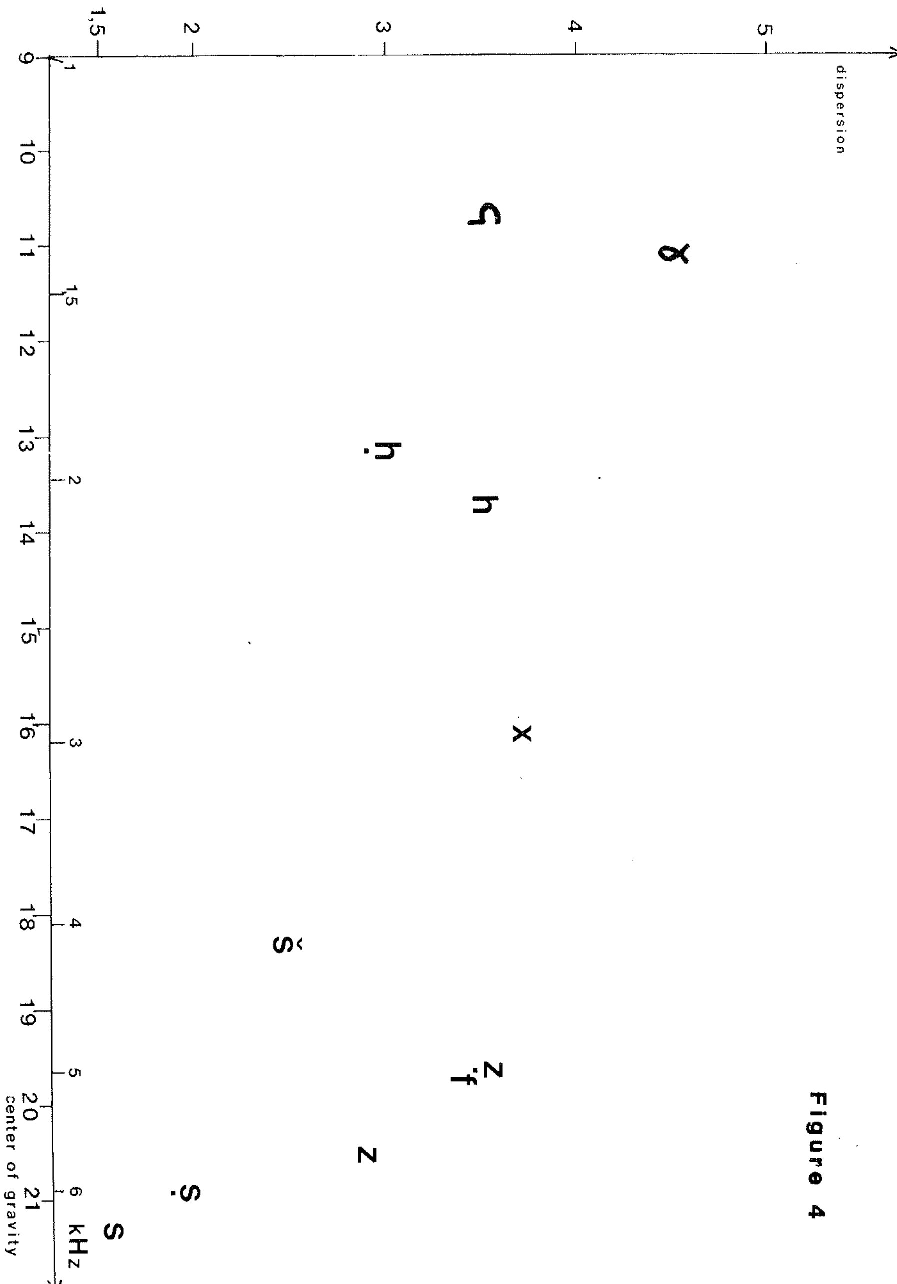

dispersion
1,5
2
3
4
5
center of gravity
kHz
9
10
11
12
13
14
15
16
17
18
19
20
21
1
1,5
2
3
4
5
6

Figure 4

Figure 5

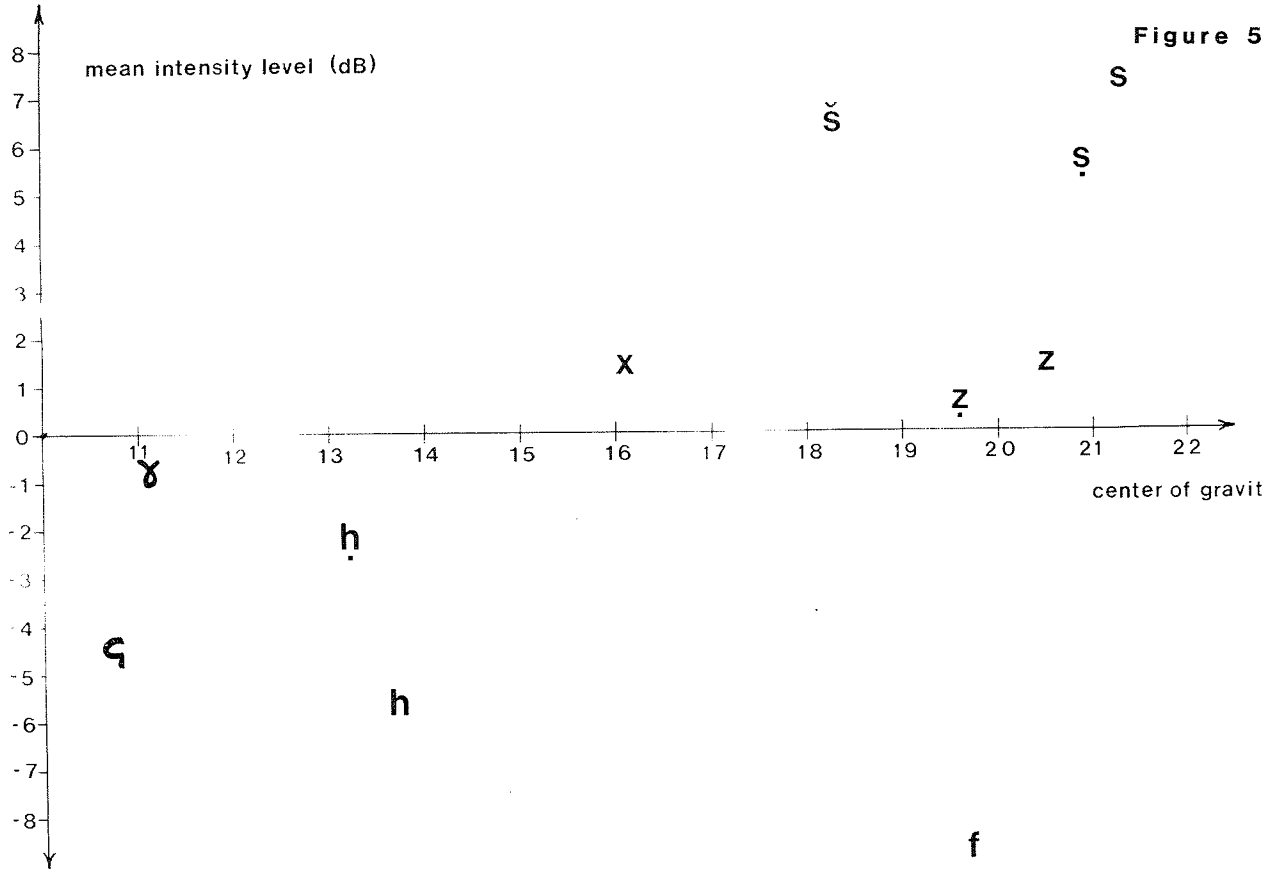
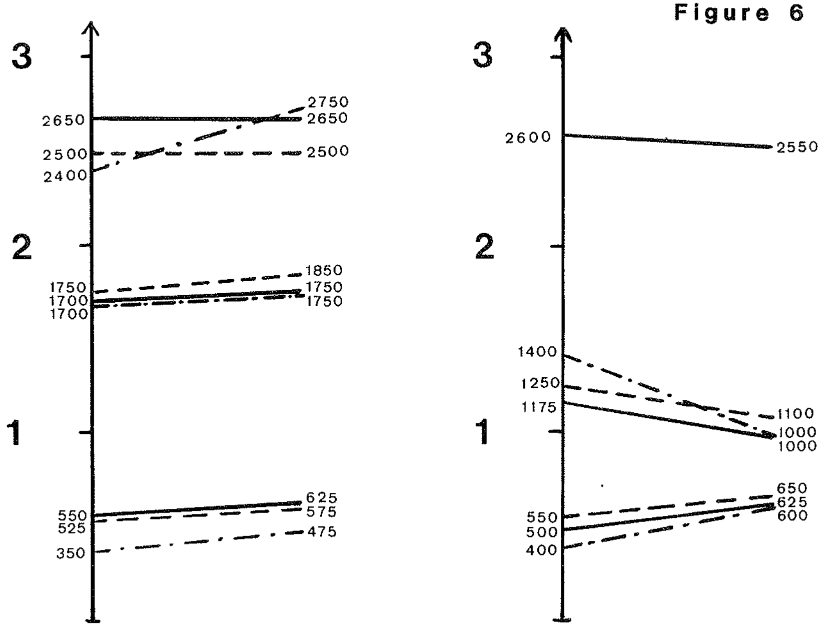

mean intensity level (dB)
8
7
6
5
4
3
2
1
0
-1
-2
-3
-4
-5
-6
-7
-8
11
12
13
14
15
16
17
18
19
20
21
22
center of gravity
s
š
ṣ
x
z
ẓ
γ
ḥ
h
f

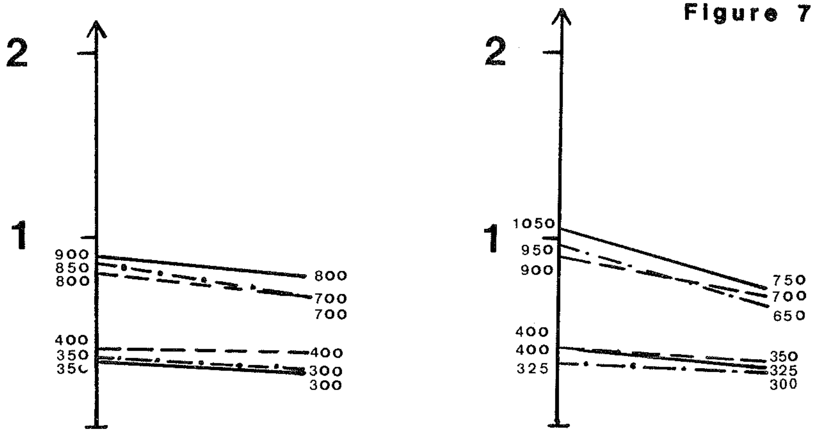
Figure 6



(s)ā(d i s)

(ṣ)ā(d i ʔ)

Figure 7



(s)ū(r)

(ṣ)ū(f)

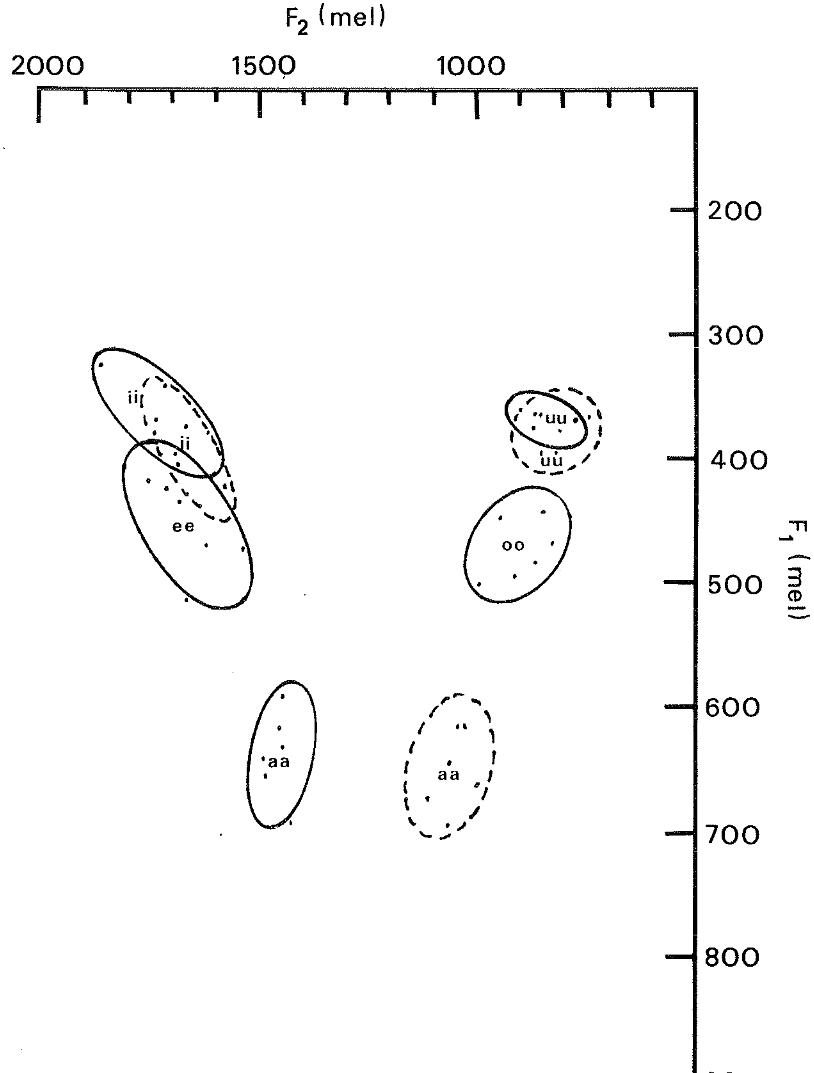


Figure 8. Long plain and pharyngalized vowels.

——— plain vowels

- - - pharyngalized vowels

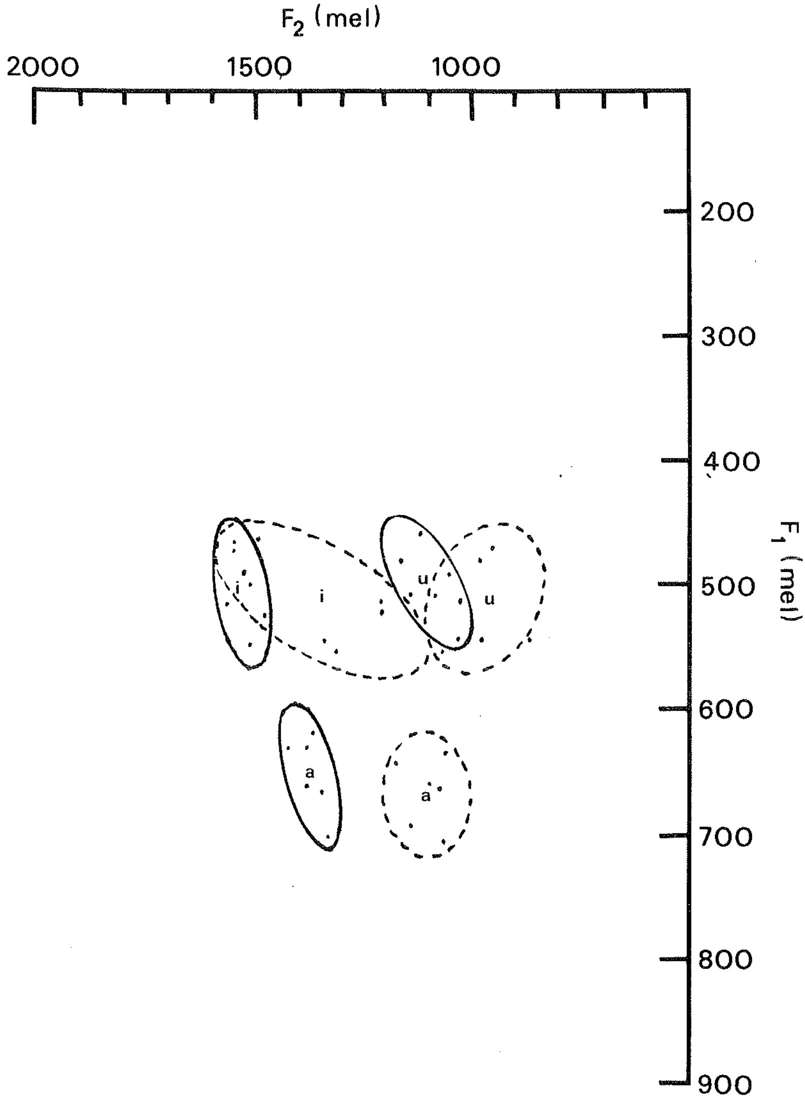


Figure 9. Short plain and pharyngalized vowels.

—— plain vowels

- - - pharyngalized vowels

A study of question intonation in Polish

Janina Petecka

THE PROBLEM

This investigation has been inspired by similar work carried out for Swedish and other languages by Eva Gårding and Gösta Bruce at the Department of Linguistics and Phonetics at Lund University. Audio tape recordings, oscillograms and spectrograms have been made with the help of the equipment of this department /Sona-Graph 7800, Spectrograph Kay Digital/.

The purpose of this paper is to try to add detail to the picture of question intonation in Polish. Intonation is used here as the tonal feature signalling sentence type,i.e. the fundamental frequency contour /F_o contour/.

In this paper yes/no questions and question-word-questions /wh-questions/ have been investigated. There are two kinds of yes/no questions in Polish: lexically marked and lexically unmarked questions.

A. <u>Lexically unmarked yes/no questions</u>.
A statement can be changed into a question only by altering the falling intonation contour into a rising one in the final tonal segment of an utterance.

B. <u>Lexically marked yes/no questions</u> are introduced by an interrogative particle.

C. <u>Wh-questions</u> investigated in this paper are introduced by an interrogative pronoun.

In earlier research it has been stated that yes/no questions are characterized by a rising F_o contour, wh-questions,on the other hand,have more or less the same falling F_o contour as statements (c.f. Dłuska 1947, II 1976, Wodarz 1962).

Gårding /1974/ points out that it is quite a common phenomenon with regard to question utterances that a lexically or grammatically marked question does not need to be further marked with the help of a tonal feature and can therefore have the same intonation as a statement.

These two assertions have been the point of departure for my investigation since both marked and unmarked question utterances exist in Polish.

It has been my intention to elicit the prosodic features that characterize intonation in natural, neutral questions.

My hypothesis is that in a given situation it is possible to have a neutral question intonation without any paralinguistic features.

The intonation can be seen as neutral even outside the VP where it is otherwise normally positioned (Gårding, 1974, Dukiewicz, 1977).

Sentence stress has been marked in the test material for two reasons:

- in order to position it as far as possible from the VP in some utterances,
- in order to investigate syllable length for some syllables which occur in the material several times in both focused and unfocused positions.

EARLIER RESEARCH

The correlation between stress and intonation in Polish was observed at the beginning of this century.

Benni /1916/ compares data evaluated by auditive and instrumental methods. His hypothesis is that a stressed syllable has a higher pitch in most utterances except questions, where the final, in Polish always unstressed, syllable receives a higher pitch.

Dłuska /1947, II 1976/ distinguishes two basic intonation contours: the falling (cadence) and the rising contour (anticadence). The falling contour characterizes statements and wh-questions, the rising contour characterizes other questions.

Jassem /1959 and 1962/ points out some regularities in pitch movements in Polish, postulating 6 nuclear tunes. Those are expanded by Staffen-Batogowa, 1966, to 26 intonation patterns. Jassem claims that Polish stress is tonal. In his both auditive and instrumentally based investigation Jassem examines four features: intensity, pitch, duration and quality. Jassem states that "relations in pitch are relevant for stress in Polish, relations in duration and intensity being incidental" (Jassem, 1959, 269).

Wodarz /1962/ discusses three types of intonation contours:

1) Terminal, as in statements and wh-questions
2) Continuative
3) Interrogative, as in yes/no questions and, when including speakers' attitudes, also in wh-questions.

In general the intonation curve falls on the last stressed syllable and lies below the level of the preceding syllables. In question utterances the final syllable, which is unstressed, is elevated and the intonation contour rises to a higher level than the other syllables within the question utterances.

Dukiewicz presents the results of a number of investigations on intonation in e.g. 1977, 1978, 1979, 1982. Concerning question intonation Dukiewicz examines turning points and asserts for yes/no questions a rising (or slightly falling-rising) F_0 contour and for wh-questions a strong falling-rising contour.

The correlation between stress and intonation was investigated furthermore by Dobrogowska /1978/ and Pluciński /1978/, the correlation between duration and intonation by Richter /1978 and 1980/.

Earlier research results show some differences in the description of question intonation especially concerning wh-questions.

The purpose of the present investigation is to try to give a more detailed picture of question intonation in Polish with regard to the above mentioned question type.

1. 4 yes/no questions and 4 wh-questions were examined regarding fundamental frequency.

2. By marking focus in the test material attempts were made to examine the effect of focus on the intonation pattern of yes/no questions.

3. The question of whether pitch and intensity peaks occurred in the same position as focus was also studied.

4. The syllable length was examined in three syllables which occurred both stressed and unstressed in the material.

Besides the basic question material, 6 statements were investigated /cf. below/.

MATERIAL

The test material consists of 6 statements and 8 questions. The phonematic structure is similar to the material used by Gårding and Bruce at the Department of Linguistics and Phonetics, Lund University (cf. Bruce and Gårding, 1978, Gårding, 1979, 1981, 1984).

The statements are composed of:

a) six syllabic vowels: i,ɨ,e,a,o,u

b) two nonsyllabic phonemes: j and w

c) four consonants: m,n,l,b

Furthermore, in the questions the following consonants and consonant clusters are present: k,t,tʃ and ts. All voiceless consonants are located initially in the interrogative particle or in interrogative pronouns. For a description of Polish phonemes see Wierzchowska /1980/.

All words occurring in the test material have the lexical stress on the penultimate syllable.

The six statements are meant to give a situational background to the questions. It has been stated that a little girl named Alina and her mother like raspberries, my small raspberries. The statements and questions in the Polish original version are presented in Appendix 1.

The basic structure (S V O) has been extended right and left from the verb. The number of syllables in each utterance is presented below (Table 1 a).

Statements ST-2 and ST-3 have the same number of syllables but the focus varies as shown in Table 2 a.

Tab. 1 a.

	number of syll.	S	V	O
ST-1	8	3	2	3
ST-2	10	2+3	2	3
ST-3	10	2+3	2	3
ST-4	12	2+3	2	2+3
ST-5	14	2+2+3	2	2+3
ST-6	16	2+2+3	2	2+2+3

Tab. 1 b.

Yes/no questions		number of syll.	S	V	O
A. unmarked	Q-1	8	3	2	3
	Q-2	10	2+3	2	3
B. marked	Q-3	9	1+3	2	3
	Q-4	13	1+2+3	2	2+3
C. WH-questions	Q-5	6	1	2	3
	Q-6	8	1	2	2+3
			(Q)	V	S
	Q-7	6	1	2	3
	Q-8	8	1	2	2+3

1=question particle or interrogative pronoun

All the fourteen utterances were written on cards. Focus was marked in order to achieve a maximal variation of sentence stress. The stressed word was underlined in red. Focus marking is presented here below.

Tab. 2 a.

	NUMBER of syll.	V
ST-1	8	-x- x- -x-
ST-2	10	x- --- -- ---
ST-3	10	.x., -- --- -- ---
ST-4	12	-- --- -- x- ---
ST-5	14	-- -- --- -- -- -x-
ST-5	16	-- -- --- -- -- -- ---

ST-3 : the stressed syllable preceeds ST-3

ST-6 : subjects marked focus by themselves

Tab. 2 b.

	NUMBER of syll.	V
Q-1	8	-x- -- ---
Q-2	10	x- --- -- ---
Q-3	9	- --- x- ---
Q-4	13	- -- --- -- x- ---
Q-5	6	x -- ---
Q-6	8	- -- x- ---
Q-7	6	x -- ---
Q-8	8	- -- x- ---

/cf. original versions, Appendix 1/

SUBJECTS

Two Polish speaking subjects, a male and a female, read the statements and questions at their normal rate of speech. Both were native Poles visiting Sweden for only a short time. Three other subjects, also native speakers of Polish, took part in the auditory test.

EXPERIMENTAL PROCEDURES

Recording I

First the two speakers were informed that there would be information about raspberries to be passed on. This information consisted of six sentences, each of them with additional information. To facilitate the reading the new information on some cards was marked in red. Next the subjects pretended that they were calling somebody in order to get the information about the raspberries themselves. They had to ask some questions, which, owing to e.g. noise interference, had to come in a given order.

Recording II

The speakers were asked to repeat the information to another person, but they were told that it did not matter in which order the statements were repeated. The order of the statement cards had been randomized. Then the speakers had to repeat the questions to check the information once again (the order of the question cards had also been randomized).

Recording III

Both the six ST-cards and the eight Q-cards were mixed at random. The speakers mixed all the cards by themselves and read them in order to examine if they had all the fourteen utterances. They were asked before reading to note whether there was a question mark or full stop. The test material was recorded as follows:

1. Statements /ST-1 to ST-6/
2. Questions /Q-1 to Q-8/
3. Random statements
4. Random questions
5. Random statements and questions

The auditory test

1. Three other subjects listened to the recorded utterances. The listeners had to indicate simple pitch patterns for each utterance showing the auditory impression. Only three marks were given: rising /↗/, falling /↘/ and level /→/.

2. Each utterance was numbered.

3. The syllable the listeners perceived as focused had to be marked with x. Each recorded utterance was repeated as many times as the listeners wished.

4. The listeners were asked to examine utterances from No 29-42, i.e. recording III, of both the male and the female speaker, and to note down a full stop or a question mark.

5. Finally the listeners were asked to define their impressions about the paralinguistic features regarding question utterances.

Instrumental analysis

For each recorded sample oscillograms were made and the 84 fundamental frequency contours were examined with regard to

1a. F_0 minimal and maximal values

b. F_0 values at the beginning and end of each utterance

2. The auditory pitch patterns were then compared with the pitch curves derived from the oscillograms, and the pitch contours for statements and questions were compared.

3. The difference between F_0 curves for yes/no questions and wh-questions was examined in more detail. The male voice in recording II was chosen (MR II) for its complete F_0 contour. The choice of recording II can be motivated as follows: It was expected that the most natural intonation would occur precisely in this part. Firstly: the questions were situationally based. Secondly: the questions came in random order so as to avoid the possibility of perhaps slightly unnatural intonation in recording I. Recording III, where statements and questions were mixed at random could not be accepted as natural speech acts. This recording, however, could be of interest for the evaluation of other aspects. It could be more important for comparing a) statement and question intonation and b) yes/no- and wh-question intonation.

4. The listeners' question patterns were transformed to oscillograms from which pitch curves and intensity curves were extracted.

5. Spectrograms were made for recording II MR.

6. Intensity peaks on oscillograms were compared to focus marks in the auditory patterns.

7. Duration of statements and questions was measured. Syllable length for three syllables was studied in both focused and unfocused positions. For the sake of comparison, additional spectrograms for four statements and three questions were made from recording III MR. The duration measurements with an accuracy of 0,01 sec. were made from oscillograms and compared to spectrograms.

RESULTS

The auditory test results

Full stops and question marks have been correctly placed. All the listening subjects have

- stated that the unmarked yes/no questions Q-1 and Q-2 express surprise, the other questions are described as neutral

- noted a rising pattern for both types of questions
- placed the focus mark x at the same syllable as positioned in the test material concerning questions.

Pitch curves and intensity curves have been extracted from the recordings judged auditively. The results are discussed below.

1. F_o contours and F_o intervals

1.1. The pitch curves show that all statements have a local falling, all questions a local rising fundamental frequency contour in the final part of the utterance /cf. Appendix 2A, B, C, D and 3A, B/.

1.2. In all the pitch curves, the F_o rises on the vowel in the final unstressed syllable. The spectrograms show the same. The F_o values in the beginning of the question utterances, the highest and lowest values of the final syllable compared to the whole question contour /recording II MR/ are presented below.

		a)	b)	c)
A.	Q-1	100-110	90-190	90-190
	Q-2	110-125	95-180	95-190
B.	Q-3	130-150	100-190	95-190
	Q-4	140-150	100-220	100-220
C.	Q-5	160-200	95-160	95-200
	Q-6	160-140-190	90-130	90-190
	Q-7	150-200	95-140	95-200
	Q-8	150-130-180	100-150	100-180

Tab. 3. a) F_o movement at the beginning of the question
b) F_o movement on the final syllable
c) $F_{o\ (min)}$ and $F_{o\ (max)}$ for the whole question utterance.

The local fall for the plosive b and the intervocalic glottal stop have not been included.

It has been found that the two types of questions are differentiated by interval values in the rising contour of the fundamental frequency in the final syllable.

For yes/no questions the values are between 85Hz and 120Hz thereby agreeing with the range for the whole question contour.

Wh-question curves show lower values than do yes/no questions: between 40Hz and 65Hz, and lower than the values for the whole question.

In wh-questions the highest values are in most utterances identical to the values for the vowel in the focused syllable. This, however, cannot provide sufficient support for the hypothesis in previous investigations that intonation contours in wh-questions and in statements are the same.

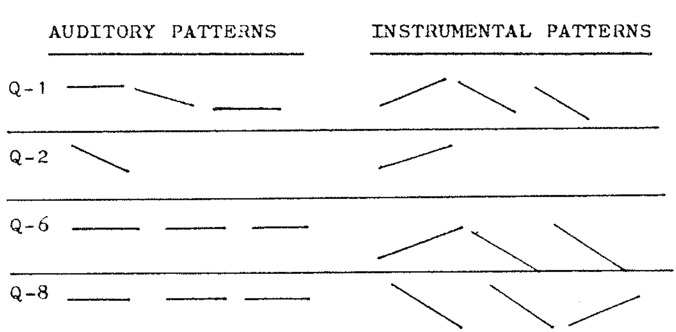
2. Auditory patterns and instrumental contours

The interval value of the rising contour in wh-questions seems to be great enough to be noticeable. All the listening subjects noted a rising pattern for wh-questions as well as for yes/no questions. There were no significant differences between the subjects regarding the F_0 contours for questions. However, statement contours differ.

As to the differences between the auditory patterns and the instrumental contours the following can be noted:

a) The contours are identical for Q-3, Q-4, Q-5 and Q-7 which indicates that the subjects have realized a rising in the fundamental frequency from 20Hz upwards.

b) In the other four questions the rising is not greater than 5-15Hz and has not been noticed. Here the auditory patterns differ from the instrumental contours. In the following the same marks are used as in the auditory test (see Appendix 2A-B-C-D).



In all the auditory patterns the pitch movements agree with those within the first three harmonics on the spectrograms.

3. Fundamental frequency values and focus

24 statement recordings and 24 question recordings of both male /MR/ and female /KVR/ have been studied.

The range for male voice /MR/:

F_o (min) in ST: 85-100 Hz, in Q : 80-100 HZ — F_o (max) in ST: 150-220 Hz, in Q : 160-220 Hz

The range for female voice /KVR/:

F_o (min) in ST: 185-210 Hz, in Q : 150-240 HZ — F_o (max) in ST: 325-390 Hz, in Q : 340-380 Hz

3.1. A different range for statements and questions has been observed only for KVR.

3.2. In all statements the focused syllable has received the highest F_o value, cf. Table 3 a) below.

	MR			KVR		
	I	II	III	I	II	III
ST-1	x-165 100-165	x-170 85-170	x-190 100-190	x-350 180-350	x-360 190-360	x-360 190-360
ST-2	x-200 100-200	x-190 85-190	x-190 90-190	x-350 200-350	x-360 200-360	x-360 200-360
ST-3	x-170 100-170	x-160 80-160	x-170 80-170	x-350 195-350	x-325 210-325	x-360 190-350
ST-4	x-200 100-200	x-200 90-200	x-190 90-190	x-370 185-370	x-360 200-360	x-350 190-350
ST-5	x-190 90-190	x-220 100-220	x-200 80-200	x-370 210-370	x-390 200-390	x-350 200-350
ST-6	x-160 90-160	x-165 85-165	x-150 90-150	x-370 195-370	x-360 210-360	x-360 200-360

Tab. 4 a) x=F_o value for focused syllable
F_o (min) and F_o (max)

3.3. For question utterances the conditions are quite different. As to yes/no questions the focused syllables have either the lowest F_o value or are close to it. As to wh-questions the sentence stressed syllables often have $F_{o\ (max)}$ values. The fundamental frequency curve falls immediately after focus to the lowest value, cf. Appendix 2A and 2B for yes/no questions, 2C and 2D for wh-questions. See also Table 4 b) below.

	x-values not always obtainable					
	MR			KVR		
	I	II	III	I	II	III
Q-1	x-140 90-180	x-110 90-190	x-100 100-180	x-160 160-340	x-240 200-360	x-290 200-360
Q-2	x-130 100-210	x-125 95-180	x-200 100-200	x-150 150-360	x-240 200-370	x-220 200-360
Q-3	x-100 100-190	x-100 95-190	x-100 80-180	x-185 185-360	x-240 190-370	x-230 200-370
Q-4	x-100 100-220	x-100 100-220	x-100 85-170	x-240 240-350	x-260 220-360	x-240 210-350
Q-5	x-220 100-220	x-200 95-200	x-190 95-200	x-350 210-350	x-360 190-370	x-330 200-330
Q-6	x-110 100-200	x-190 90-190	x-170 90-170	x-360 210-360	x-230 200-360	x-230 200-350
Q-7	x-190 90-190	x-200 100-200	x-220 90-220	x-360 220-360	x-370 195-370	---- 210-360
Q-8	x-100 100-190	x-180 100-180	x-110 90-200	---- 230-360	---- 230-380	---- 200-350

Tab. 4 b) x=F_o value for focused syllable
F_o (min) and F_o (max)

It can be stated that $F_{o\ (min)}$ and $F_{o\ (max)}$ values are related to focus in a different way for the two types of questions. An investigation of all question utterances shows that the highest F_o values occur in yes/no questions in the final, unstressed syllable regardless of the focus-position. In wh-questions the highest fundamental frequency values often occur in the focused syllable.

4. The intensity study was limited to a comparison of sentence stress and intensity peaks.

4.1. In the intensity curves the focused syllable has the highest or second highest peak. These differences are probably due to the different intensity level of different vowels.

4.2. The focus marks in the material obtained by the auditive method are positioned at the same syllable as in the test material for question utterances. In statements there is some variation in the positioning of focus marks, especially in ST-3, /statement after focus/, and in ST-6 where focus was not marked in advance, cf. TEST MATERIAL.

5. Duration time for all utterances was examined. The results are presented below.

Tab. 5.

ST/Q No.	NUMBER of syll.	MR I	MR II	MR III	KVR I	KVR II	KVR III
Q-5	6	125	120	130	115	120	120
Q-7	6	125	120	130	115	110	120
ST-1	8	155	195	180	180	160	160
Q-1	8	140	140	150	150	150	140
Q-6	8	155	150	170	145	140	150
Q-8	8	150	140	170	125	140	140
Q-3	9	155	155	175	160	140	150
ST-2	10	190	160	160	220	190	190
ST-3	10	170	170	170	200	175	180
Q-2	10	180	165	180	190	190	170
ST-4	12	225	200	205	235	230	225
Q-4	13	215	208	208	220	210	215
ST-5	14	240	225	250	310	285	345 x)
ST-6	16	260	250	248	370	340	330

x) repeated phrase

On the whole the duration time increases with the number of syllables, but the more syllables the shorter time each unfocused syllable receives. In ST-1, where three words are focused in order to get introductory intonation, the duration time is naturally longer.

When comparing the recordings for MR it can be observed that during recording II MR the material was read at the highest speed. This could possibly be seen to confirm the expectations that recording II was correctly chosen to obtain the most natural intonation in questions /see Experimental procedures/.

For KVR the statements tend to take more time than for MR. However, questions tend to be produced faster.

6. Syllable length in focused and unfocused position

The values presented below were obtained from recording II MR discussed above. In order to check these values further, four statements and three questions were examined from recording III MR. The only difference is that the highest value for ma-, 28 cs, is not present there.

6.1. ma- occurs 19 times in statements and 13 times in questions. The duration time varies

- from 16 cs to 28 cs for focused syllable
- from 16 cs to 22 cs for unfocused

The length of the syllable ma does not seem to be influenced by whether the subject NP has sentence stress or not. Higher values can be found more often in the VP. The longest duration time for ma occurs in Q-4 which has the largest number of syllables.

For -li- similar values were found for stressed and unstressed syllables positioned in VP and outside VP. lu- is the first syllable of the verb. It occurs once as focused and lasts then 19 cs. It lasts from 10 cs. to 14 cs. as unfocused.

It has been found that the syllables discussed above tend to have a longer duration time in rhematic position. The longest duration time of all is found in the final and unstressed syllable in all question utterances: from 26 cs. to 30 cs.

SUMMARY

1. The results from the auditive investigation agree well in themselves and also with the results of the instrumental investigation regarding question utterances. All the listening subjects have defined the intonation in wh-questions as neutral.

2. Statement and question utterances have a clearly different fundamental frequency contour. This indicates that it is possible to recognize whether the produced utterance is a question or a statement only with the help of the intonation contours in the final segment of the utterance. This is valid for both types of questions.

3. However, the characteristic rising F_0 contour shows some differences for yes/no questions and for wh-questions. The differences are found in the F_0 interval values. In yes/no questions higher values are found for the final syllable of the utterance than for the same segment in wh-questions.

4. The maximal values for pitch occur in the same position as focus only in wh-questions. Regarding yes/no questions no changes of the F_0 curve could be observed when the position of sentence stress was altered.

5. The duration time for focused syllables tends to be longer than for other syllables. The longest duration time of all was found in the final unstressed syllable in questions.

The results in point 1, point 2 and point 5 agree with results arrived at by other researchers. The results in point 3 and point 4, however, throw light on the differences between the rising intonation contours in yes/no questions and in wh-questions.

One cannot draw general conclusions since the investigated material is too limited. However, in both types of question utterances a tendency towards rising F_0 contours seems to be quite clear.

REFERENCES

Benni, T. 1916. O akcencie polskim - spostrzeżenia i pomiary, Warszawa.

Bruce, G. och Gårding, E. 1978 . A prosodic typology for Swedish dialects.In:Gårding,E., Bruce,G.och Bannert, R./eds./ Nordic Prosody II: 219-228. Department of Linguistics,Lund University

Dłuska,M. 1947, II 1976. Prozodia języka polskiego. Warszawa.

Dobrogowska,K. 1978. Badania zależnosci między akcentem a intonacją w mowie nieemocjonalnej. Lingua Posnaniensis XXI 1978 s 65-76

Dukiewicz, L. 1977 Prozodyczne wyznaczniki komunikatywnej struktury wypowiedzi.In:Polonica III 1979, 5-15.

Dukiewicz,L. 1978. Intonacja wypowiedzi polskich. Wrocław.

Dukiewicz, L. 1979. Instrumental, Auditory and Funktional Analysis of Polish Intonation.In: Slavica Lundensia 7 / 1979.

Dukiewicz, L. 1982. Stosunki intensywnosciowo-iloczasowe między elementami fraz oznajmujących i pytajnych o różnej lokalizacji szczytu fonicznego.In: Polonica VIII 1982.

Gårding,E. 1974. Kontrastiv prosodi.Lund.

Gårding,E. 1979. Sentence Intonation in Swedish .Phonetica 36

Gårding,E. 1981 Contrastive Prosody; A model and its application. Studia Lingvistica 35, 1-2, 146-166.Lund.

Gårding, E. and Bruce,G.1981. A presentation of the Lund Model for Swedish Intonation. Nordic Prosody II, 33-39,Trondheim.

Gårding,E. 1984.Comparing Intonation.Working Papers 27. DEp. of Linguistics, Lund University.

Jassem,W. 1959. The phonology of Polish stress, Word 15,

Jassem,W. 1962 Akcent języka polskiego. Wrocław.

Pluciński,A. 1978. Interdependencies between the Fundamental Frequency and the Intensity Level in Polish Colloquial Speech. In:Lingua Posnaniensis XXI.

Richter,L. 1978. Wpływ pozycji w zestroju akcentowym oraz długości wyrazu na czas trwania głosek w języku polskim.In:Lingua Posnaniensis XXI, 91-114.

Richter, L. 1980. Wpływ tempa mowy na czas trwania głosek w języku polskim. In:Polonica VI,1980.

Richter, L.1982. Rola poszczególnych czynników w procesie percepcji intonacji w języku polskim. In: Polonica VIII,1982.

Steffen-Batogowa,M. 1966.Versuch einer strukturellen Analyse der polnischen Aussagemelodie.Zeitschrift für Phonetik, Sprachwissenschaft u. Kommunikatinnsforschung 19,398-440.

Wierzchowska,B.1980.Fonetyka i fonologia języka polskiego. Wrocław.

Wodarz,1962.Zur Satzintonation des Polnischen. Phonetica 8/62.

APPENDIX 1

THE ORIGINAL VERSIONS IN POLISH

1. STATEMENTS /ST/

ST-1 Alina lubi maliny.

ST-2 Mama Aliny lubi maliny.

St-3....,mama Aliny lubi maliny.

ST-4 Mama Aliny lubi moje maliny.

ST-5 Mama małej Aliny lubi moje maliny.

ST-6 Mama małej Aliny lubi moje małe maliny.

2. QUESTIONS /Q/

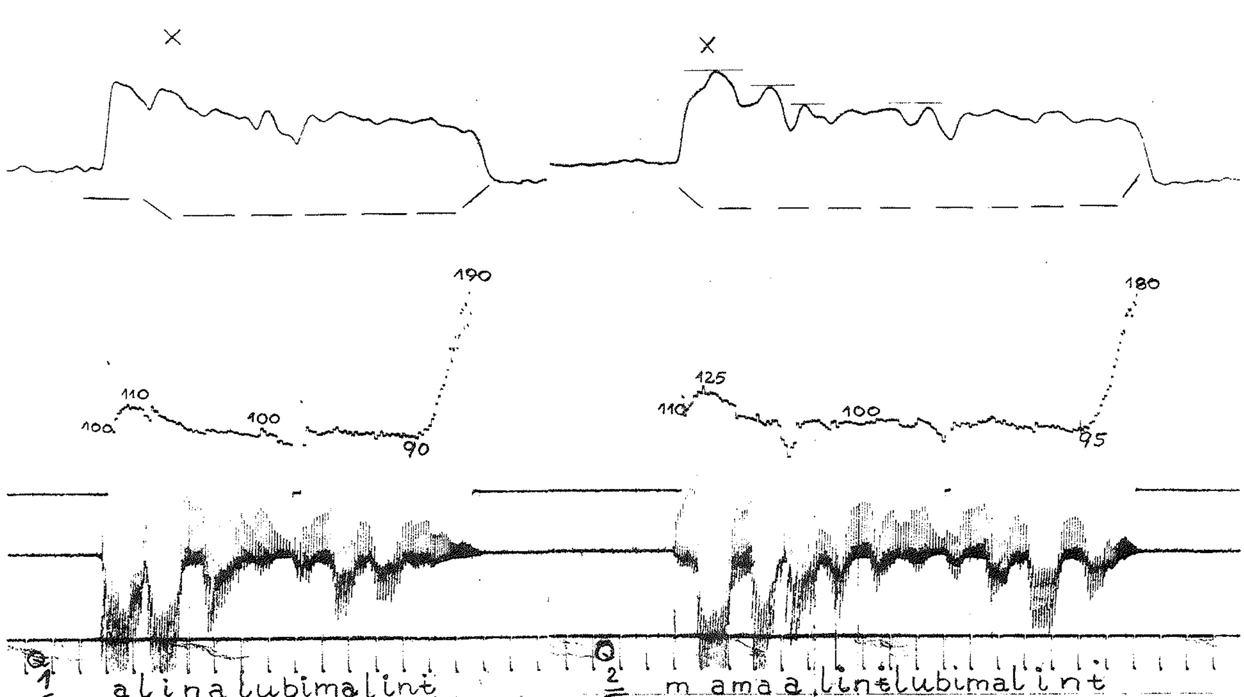
Q-1 Alina lubi maliny?	YES/NO QUESTIONS A.
Q-2 Mama Aliny lubi maliny?	
Q-3 Czy Alina lubi maliny?	YES/NO QUESTIONS B.
Q-4 Czy mama Aliny lubi małe maliny?	
Q-5 Kto lubi maliny?	WH-QUESTIONS C.
Q-6 Kto lubi małe maliny?	
Q-7 Co lubi Alina?	
Q-8 Co lubi mama Aliny?	

All words in the test material have lexical stress on the penultimate syllable.

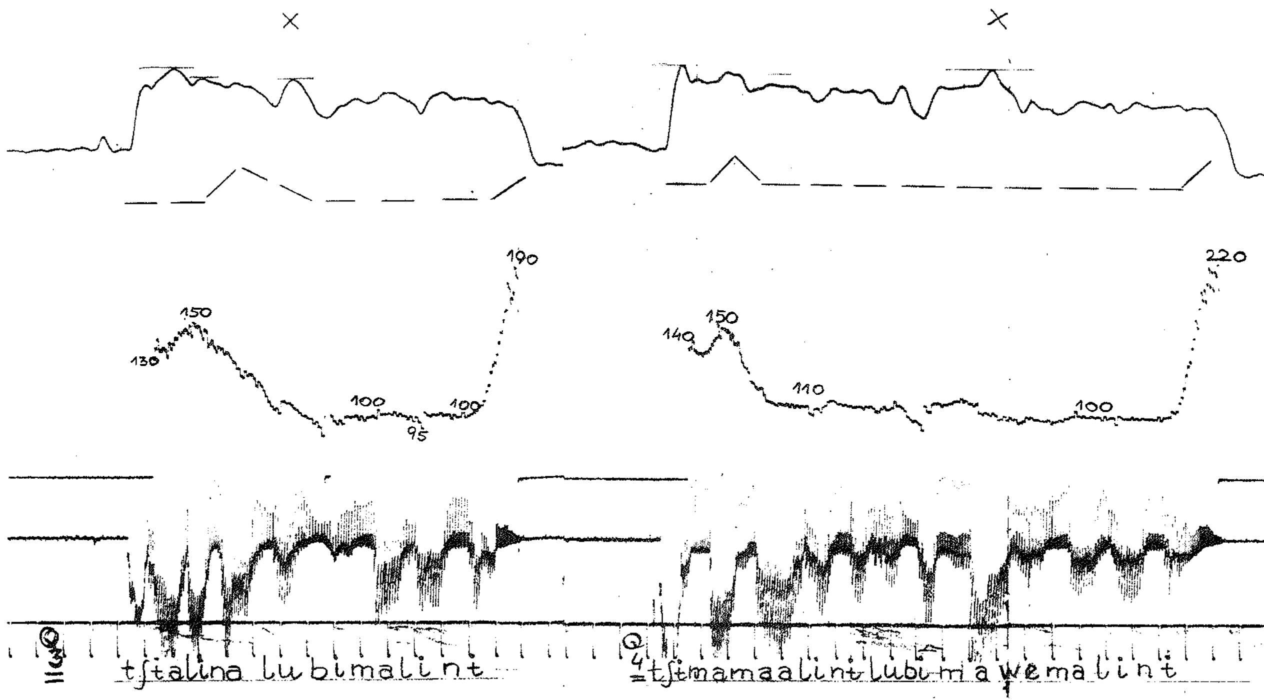
A. Unmarked yes/no-questions

B. Yes/no questions with a question particle

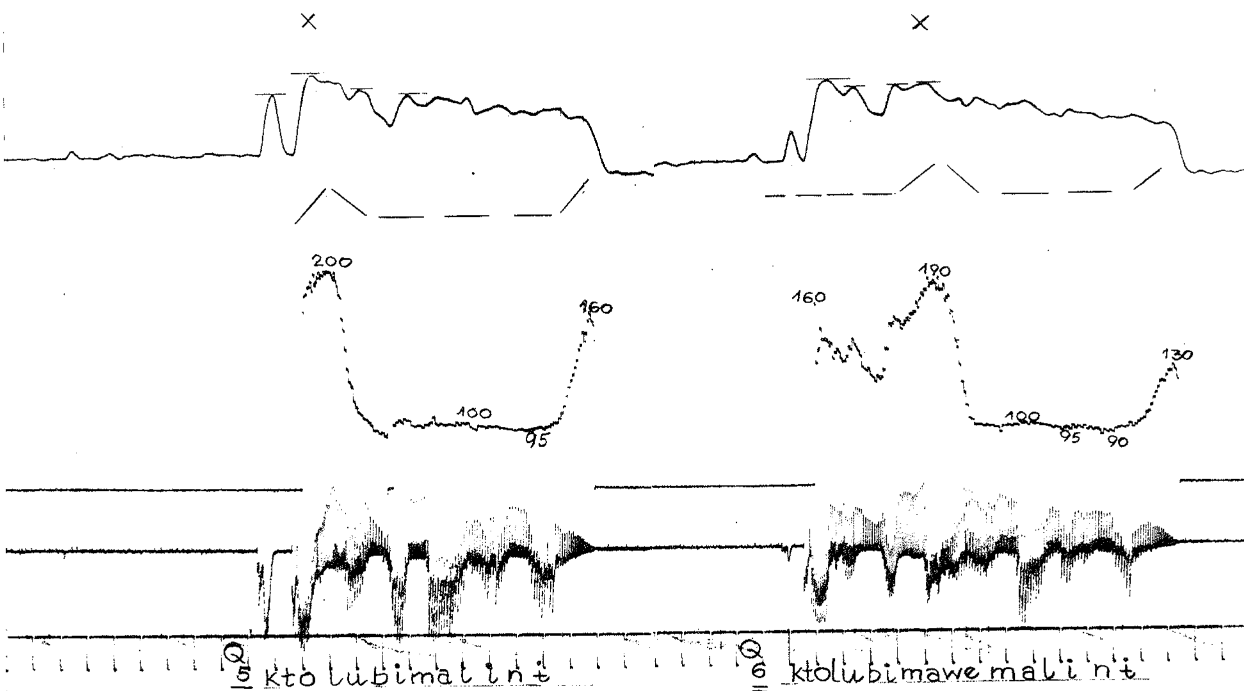
C. Wh-questions



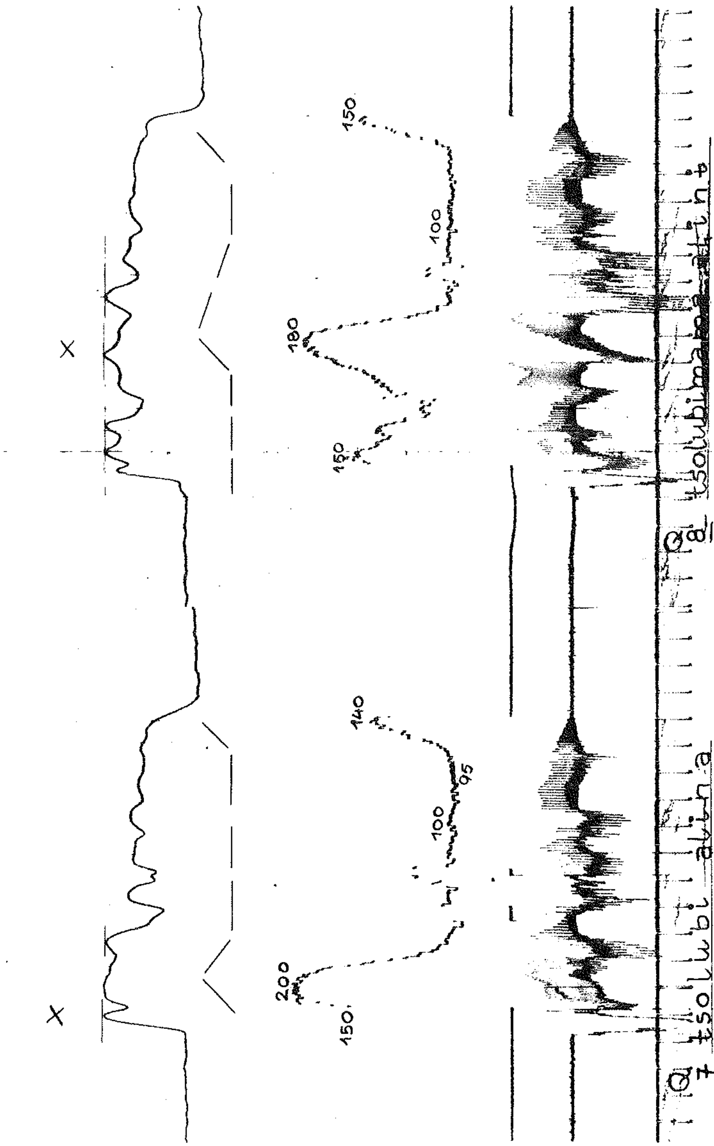
2.A. YES/NO QUESTIONS WITHOUT QUESTION PARTICLE /UNMARKED/
The auditively based F_o contours have been placed above the F_o curves. Above the intensity curve the focus marking (x) has been carried over from the test material.



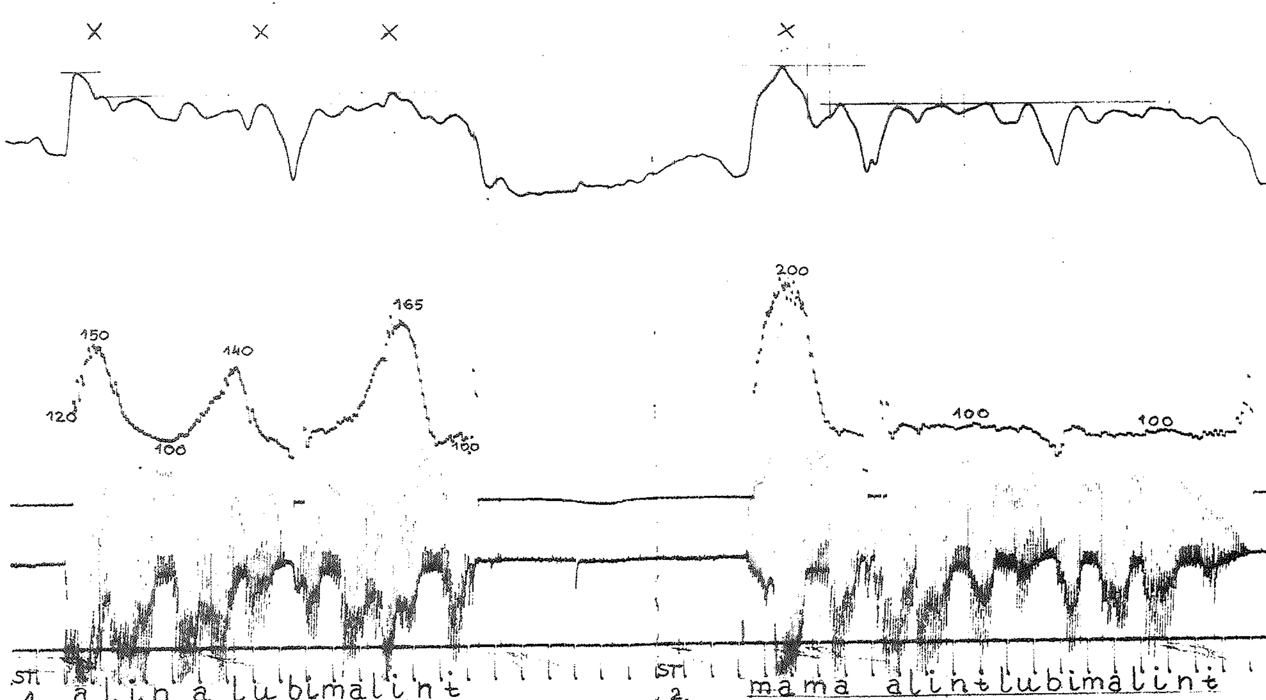
The auditively based F_o contours have been placed above the F_o curves. Above the intensity curve the focus marking (x) has been carried over from the test material.



2.C. WH-QUESTIONS. Auditively based F_0 contours have been placed above the F_0 curves. Focus marking (x) has been carried over from the test material.

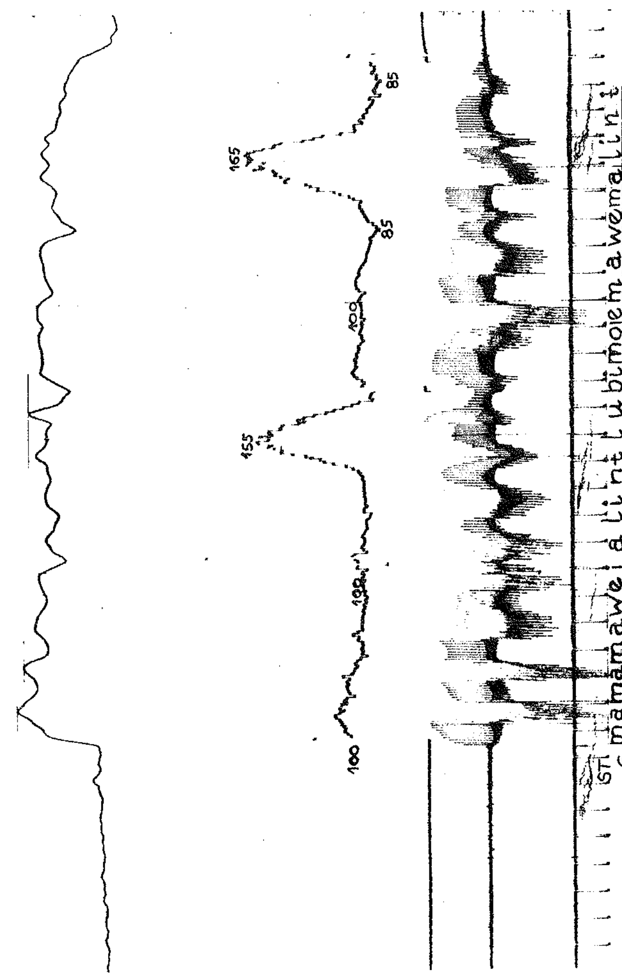


2.D. WH-QUESTIONS. Auditively based F_o contours have been placed above the F_o curves. Focus marking (x) has been carried over from the test material.



3.A. RANDOM OSCILLOGRAMS OF STATEMENTS

Focus marking (x) has been carried over from the test material.



3.B. STATEMENT OSCILLOGRAM FOR ST-6, WITHOUT FOCUS MARKED IN ADVANCE.

CURRENT RESEARCH PROJECTS

Commentator, a computer system simulating verbal behaviour
(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 non-European languages
(Swedish Council for Research in the Humanities and Social Sciences)

Eva Gårding, Mona Lindau Webb, 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

LECTURES AND PUBLICATIONS ELSEWHERE

1984

Bredvad-Jensen, A.-C. On Child Phonology. Paper presented at the Eighth Scandinavian Conference of Linguistics, 30/8 - 1/9 1984

Bruce, G. On the phonetics of rhythm: Evidence from Swedish Paper presented to the ASA meeting, Norfolk, Virginia, USA

Goodluck, H. & Håkansson, G. Children's syntax - grammar and experimental evidence. In Proceedings from Child Language Symposium, Paper 1, Child Language Research Institute, Department of Linguistics, Lund University

Gårding, E. Constancy and Variation in Standard Chinese tonal patterns, Prosody Club, Department of Linguistics, Stockholm University, November.

Gårding, E. Intonation models in general, for Chinese in particular. Department of Linguistics, Gothenburg University, October.

House, D. Transitioner i talperceptionen: implikationer för perceptionsteori och applikationer i hörselvården. Abstract in Nord & Trampe (eds): Tal Ljud Hörsel, föredrag och abstracts från det första TLH-symposiet 10-11 mars 1983 Stockholms universitet, p. 108. Institutionen för lingvistik, Stockholms universitet

Håkansson, G. Lärarens lexikonanvändning - hur lärare talar till elever på olika nivåer. In Hyltenstam, K & Maandik, K. (eds), Nordens språk som målspråk. Stockholms universitet

Lindau, M. Phonetic differences in glottalic consonants. Journal of Phonetics 12:2, 147-57.

Lindau, M. Vowels and diphthongs in Hausa. Paper presented at the 15th Conference on African Linguistics, Los Angeles, March 1984.

Magnusson, E. & Nauclér, K. Language disordered children's reading and spelling. In Proceedings from Child Language Symposium, pp. 50-55. Paper 1, Child Language Research Institute, Department of Linguistics, Lund University

Nauclér, K. Läsning och språkutveckling med särskild tonvikt på språkförsenade barn. Paper presented to 9:e Nordiska konferensen för läspedagoger, 30/7 - 2/8, Vejle, Denmark.

Nauclér, K. & Magnusson, E. Language disordered children's reading and spelling. Preliminary Results. Paper presented to the 7th World Congress of Applied Linguistics, Aug 5-10, Brussels, Belgium

Sigurd, B. Computer simulation of spontaneous speech production. Proc. COLING 84. Stanford

Sigurd, B. Om Jaså, Bra, Precis och andra returord. språkvård 1984: 3

Svantesson, J-O. Consonant harmony in child and adult phonology. Proceedings from Child Language Symposium, Lund, May 17-18, 1984, 92-102. Child Language Research Institute, Lund.

Svantesson, J-O. Sketch of the tense-mood-aspect system of Kammu. In Ö. Dahl and D. Kős-Dienes, ed., Selected working papers from the tense-mood-aspect project, Stockholm, p. 44-48.

Svantesson, J-O. Some results form the Third World Languages project. Academia Sinica, Institute of Acoustics, Beijing. Sept. 1984.

Söderbergh, R. (Ed.) Proceedings from Child Language Symposium, Lund, May 17-18, 1984. Paper No 1, Child Language Research Institute, Department of Linguistics, Lund University

Wood, S. Articulatory compensation in normal speech. Paper presented at the Tal-Ljud-Hörsel Summer School, Särö

1985

Bruce, G. Scandinavian pitch accent and intonation. In van der Hulst, H. & Smith, N (eds), Pitch accent volume (forthcoming)

Bruce, G. Structure and functions of prosody. Paper presented to the French-Swedish Seminar on Speech, Grenoble

Engdahl, E. Constituent questions: the Syntax and Semantics of Questions with Special Reference to Swedish. D. Reidel Publishing Co., Dordrecht

Gårding, E. Comparing intonation across languages. Phonology Symposium. Stockholm, May

Gårding, E. En intonationsmodell och dess förankringar. Plenary lecture, Symposium FRN Speech, Sound, Hearing, Göteborg, March

Gårding, E. En intonationsmodell och dess förankringar. Helsinki and Turku, May

Horne, M. English sentence stress, grammatical functions and contextual dependence. To appear in Studia Linguistica 39 (1).

Horne, M. On the predictability of English Sentence Stress (focal accent) within a discourse. Paper to be presented at the First Nordtext Symposium, Trondheim, 16-18 August 1985

House, D. Implication of spectral changes for categorizing tonal patterns. Paper presented at the second TLH-symposium, 14-15 March, Göteborg

House, D. Compensatory perceptual strategies used by hearing-impaired listeners. Paper presented at the second TLH-sympsoium, 14-15 March, Göteborg

House, D. Compensatory use of acoustic speech cues and language structure by hearing-impaired listeners. Paper presented at the Conference on Communication and Handicap - Aspects of Psychological Compensation and Technical Aids, 3-4 June, Stockholm

Lindau, M. The story of /r/. In V. Fromkin, ed,. Linguistic phonetics. Essays in honor of Peter Ladefoged. New York: Academic Press.

Lindau, M. Hausa vowels and diphthongs. UCLA Working papers in phonetics 60, 40-54.

Lindau, M., K. Norlin and J-O. Svantesson. Resultat från projektet Några viktiga språk i tredje världen. Tal-Ljud-Hörsel, Göteborg, March 1985.

Lindau, M., K. Norlin and J-O. Svantesson. Cross-linguistic differences in diphthongs. 109th Meeting of the Acoustical Society of America, Austin, Texas, April 1985. Abstract in The Journal of the Acoustical Society of America, 77, Supplement 1, p. 100.

Magnusson, E. Rim som lek och allvar. Paper presented at Audiologopaedagogisk forening, 10-13 April, Nyborg Denmark

Magnusson, E. Speech and language disorders and immigrant children. Paper presented to the Symposium on the Education of Handicapped Immigrant Children and Youth, May 28-30, Brioni, Yugoslavia

Magnusson, E. & Nauclér, K. Linguistic Awareness, cognitive level and short term memory in language disordered and normally speaking pre-school children. Paper presented at the First International Congress of Applied Psycholinguistics, June 17-20, Barcelona, Spain

Norlin, K. Faryngalisering i arabiska. Paper presented at Tal-Ljud- Hörsel meeting, Göteborg, 14-15 March

Norlin, K. Pharyngealization in Arabic. Paper presented at the Conference on Phonology and Phonetics, Department of Linguistics, Stockholm University, 10 May

Pettersson, T. & Wood, S. A Spectrographic Study of Vowel reduction in Bulgarian. In Davidsson C et al. (eds), X Nordiska Slavistmötet, 13-17 augusti 1984. Meddelanden från Stiftelsens för Åbo Akademi forskningsinstitut nr 102, pp. 173-184

Sigurd, B. CONVERSATOR. Dialog på dator. In Proc. Tal-Ljud-Hörsel, Göteborg 1985

Sigurd, B. Språkvetenskapens frontlinjer. Installationsföredrag. In Vitterhetsakademiens årsbok 1985

Svantesson, J-O. Vokalharmoni i mongoliska och andra språk. Tal-Ljud-Hörsel, Göteborg, March 1985.

Svantesson, J-O. Tradition and reform in China´s minority languages. HSFR Symposium Traditionen i moderniseringsprocessen i Öst- och Sydöstasien, Stockholm, April 1985.

Svantesson, J-O. Vowel harmony in Mongolian. Mini-conference on phonology and phonetics, Dept. of Linguistics, Stockholm, May 1985.

Söderbergh, R. Early Reading with Deaf Children. Prospects, Quarterly Review of Education, Unesco, Vol. XV, No. 1

Söderbergh, R. (forthcoming). Reading in Early Childhood - and then? (preliminary title). To appear in Andersson, T. (ed.) Cribnotes on Infants' and Toddler's First Steps Toward Reading

Söderbergh, R. (forthcoming). Acquisition of spoken and written language in early childhood. In J.H. Danks, I. Kurcz & G. Shugar (Eds.), Knowledge and Language. Amsterdam: NorthHolland

Söderbergh, R. (forthcoming). From communication as language to language as communication. In Proceedings from Interdisciplinary symposium on precursors of early speech, Wennergren Center, Stockholm, Sept. 19-22, 1984

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
WP 25 1982	ROBERT BANNERT	Some Phonetic Characteristics of a Model for German Prosody
	GÖSTA BRUCE	On Rhytmic Alternation

WP 25 1982	EVA GÅRDING, JIALU ZHANG & JAN-OLOF SVANTESSON	A Generative Model for Tone and Intonation in Standard Chinese
	DAVID HOUSE	Perceptual Interaction between Fo Excursions and Spectral Cues
	GABRIELLA KOCH & ANNE-CHRISTINE BREDVAD-JENSEN	Intonation Contours in Different Registers
	MONA LINDAU	Glottalic Consonants in Hausa and some other Languages
	EVA MAGNUSSON	PhD Thesis Summary: The Phonology of Language Disordered Children
	MAGNUS NORDENHAKE & JAN-OLOF SVANTESSON	Duration of Standard Chinese Word Tones in Different Sentence Environments
	KJELL NORLIN	Acoustical Analysis of Fricatives in Cairo Arabic
	ANN-CHRISTINE OHLSSON	Voice Problems and Occupational Environment
	THORE PETTERSSON & SIDNEY WOOD	Vowel Reduction in Bulgarian
	BENGT SIGURD	How to Make a Text Production System Speak
	JAN-OLOF SVANTESSON	Acoustical Analysis of Chinese Fricatives and Affricates
	URSULA WILLSTEDT	Perceptual Criteria for Differentiating between Dialect Types
WP 26 1983	THORE PETTERSSON	Electoconvulsive Therapy and the Lateralization of Linguistic Functions
WP 27 1984	ROBERT BANNERT	Towards a Model for German Prosody
	ANNE-CHRISTINE BREDVAD-JENSEN	Tonal Geography. Geographical Variation in Declarative and Interrogative Intonation along the West Coast of Sweden
	GÖSTA BRUCE	Aspects of Fo Declination in Swedish
	CHRISTINA DRAVINS	Prosody and the Hemispheres
	EVA GÅRDING	Comparing Intonation

WP 27 1984	DAVID HOUSE	Experiments with Filtered Speech and Hearing-Impaired Listeners
	KURT JOHANSSON	Perceptual Experiments with Duration versus Spectrum in Swedish Vowels
	MONA LINDAU	Testing a Model for Hausa Intonation
	EVA MAGNUSSON, KERSTIN NAUCLÈR & EWA SÖDERPALM	Form or Substance? The Linguistic Awareness of Pre-school Children and School Children Investigated by Means of a Rhyming Test
	ULRIKA NETTELBLADT	PhD Thesis Summary: Developmental Studies of Dysphonology in Children
	KJELL NORLIN	Acoustic Analysis of Vowels and Diphthongs in Cairo Arabic
	JAN-OLOF SVANTESSON	Vowels and Diphthongs in Standard Chinese
	JIĀLǓ ZHĀNG	Perceptual Studies of Chinese Intonation Using Foreign Listeners

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	*Eva 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.

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

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:-
Nr 10 - 1985	Kerstin Nauclér, Eva Magnusson Språket i skrift - om normal och försenad utveckling av tal och skrift 60:-
Nr 11 - 1985	Bengt Sigurd, Jean-Jacques Berthout, Robin Cooper Språk på dator - Hur datorn kan konversera, förstå tolka morse och undervisa i franska 45:-

(Prices in Swedish Crowns)
Swedish Post Office Giro account 439 34 92 - 6

LINGUISTICS DEPARTMENT

STAFF

Chairman	Eva Gårding
Director of Studies	Thore Pettersson
Library	Jana Geggus
Office	Eva Körner, Britt Nordbeck

General Linguistics

Berthout, Jean-Jacques	BA; Assistant
Cooper, Robin	PhD; Docent
Dahlbäck, Hans	Fil Kand; Research Assistant, extra teacher
Engdahl, Elisabet	PhD; Research Fellow
Fornell, Jan	Research Assistant
Geggus, Jana	Fil Kand; Assistant
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
Research Students:	Jean-Jacques Berthout, Hans Dahlbäck, Sheila Dooley, 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
Eeg-Olofsson, Mats	Fil Kand; Programmer (Substitute for Bengt Mandersson)
Gårding, Eva	Fil Dr; Professor; Head of Division
Hadding, Kerstin	Fil Dr; Professor Emeritus
House, David	BA; Assistant, extra teacher
Johansson, Kurt	Fil Lic; Lecturer

Jönsson, Karl-Gustav	Technician
Körner, Eva	Secretary
Lindau Webb, Mona	PhD; Research Assistant; Phonetics Dept, UCLA
Mandersson, Bengt	Tech Dr; Civ Ing; Computer Engineer (leave of absence)
Nauclér, Kerstin	Fil Dr; Logoped; Research Assistant; extra teacher
Nettelbladt, Ulrika	Fil Dr; Logoped; extra teacher
Norlin, Kjell	Fil Kand; Research Assistant
Svantesson, Jan-Olof	Fil Kand; Research Assistant
Wood, Sidney	Fil Dr; Research Fellow
Research Students:	Antonis Botinis, Anne-Christine Bredvad-Jensen, Christina Dravins, David House, Dieter Huber, Birgitta Kuylenstierna-Nadel, Boris Larnert, Ann-Christine Ohlsson, Lili-Ann Rudberg, Gabriella Stenberg-Koch, Paul Touati, Kjell Weimer, Ursula Willstedt

<u>Voice and Speech Training</u>

Norman, Lennart	Fil Kand; Voice and Speech teacher

<u>Child Language Research</u>

Bredvad-Jensen, Anne-Christine	Fil Kand; Research Assistant
Mårtensson, Bodil	Secretary
Söderbergh, Ragnhild	Fil Dr; Professor; Head of Division

Address: Department of Linguistics and Phonetics
Lund University
Helgonabacken 12
S-223 62 LUND

Lunds universitet
Reprocentralen, 1985