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

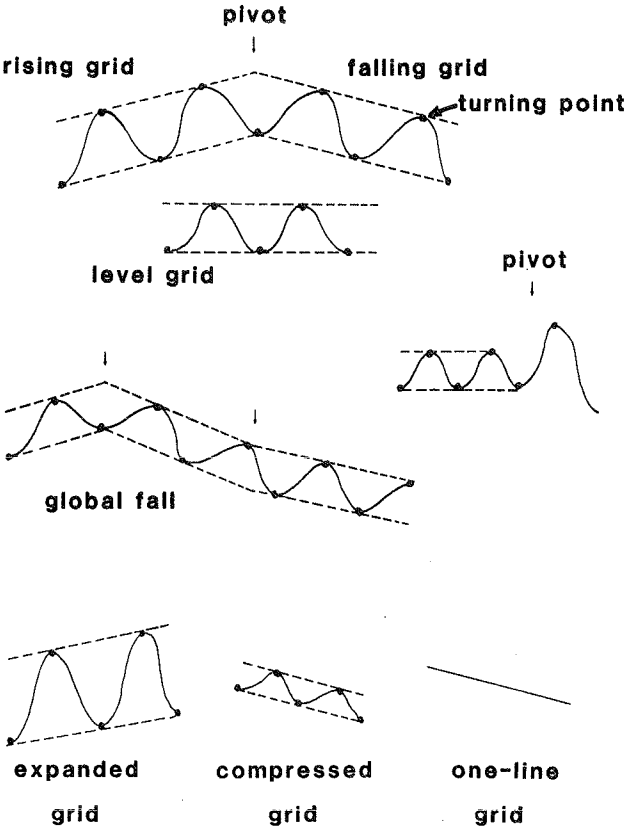
language it comes from, is that is has an undulating shape and like all undulating curves it can be efficiently described by interpolation between local maxima and minima which I call turning points. A very important observation is that some of these turning points have a rather fixed position relative 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 Yán 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.

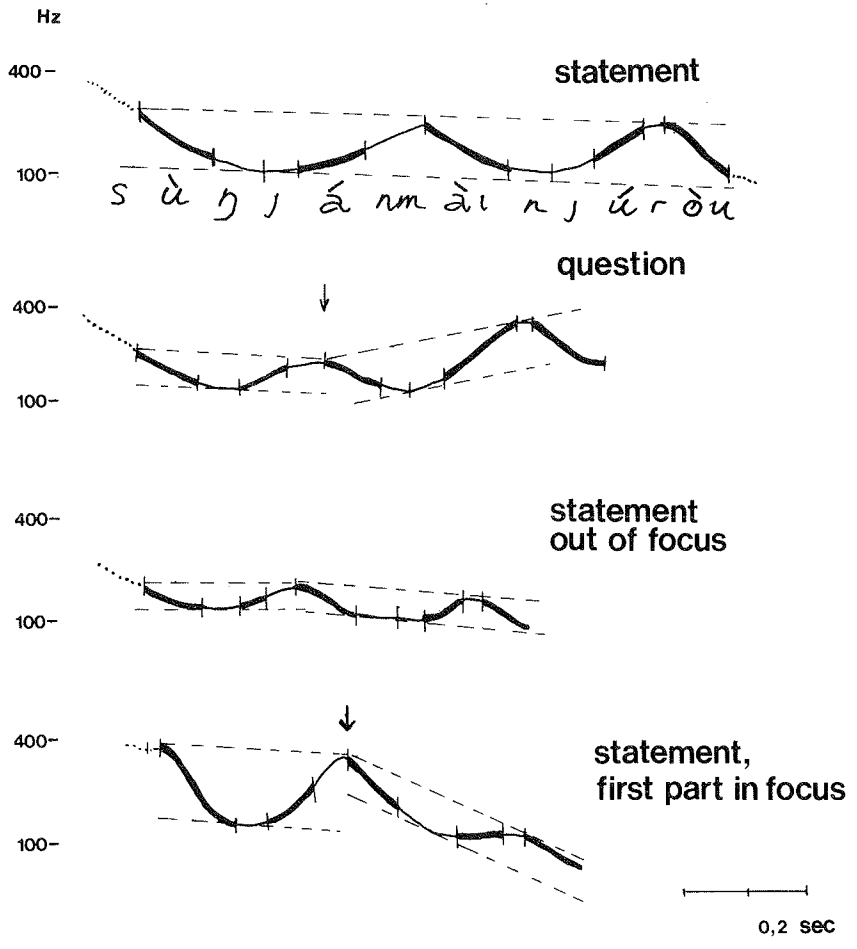


Fig.2 F0 curves from Standard Chinese Sòng Yán mài niúròu 'Song Yan säljer oxkött' in different prosodic patterns. A thick line marks the F0 curve over the vowels, a thin line marks the consonants. Broken lines are grids. An arrow marks pivot. From Gårding, Zhang & Svantesson, 1983.

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 niúrðu, 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).

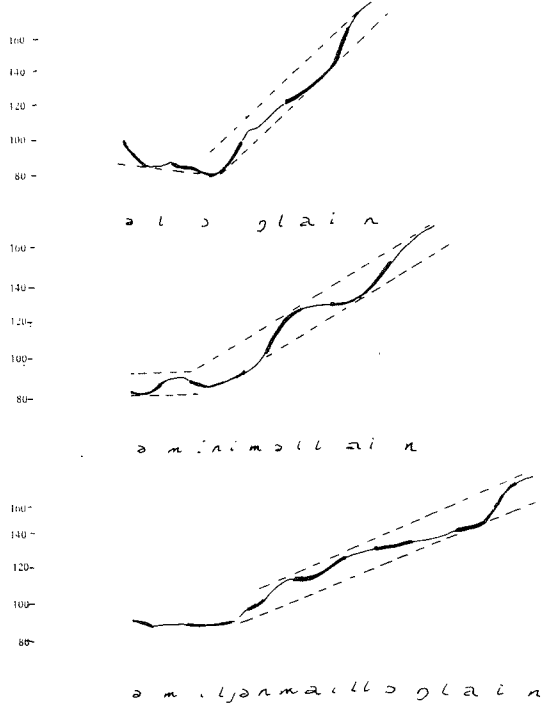


Fig. 3a F0 curves from American English yes-no questions of varying length with focus on the adjective.

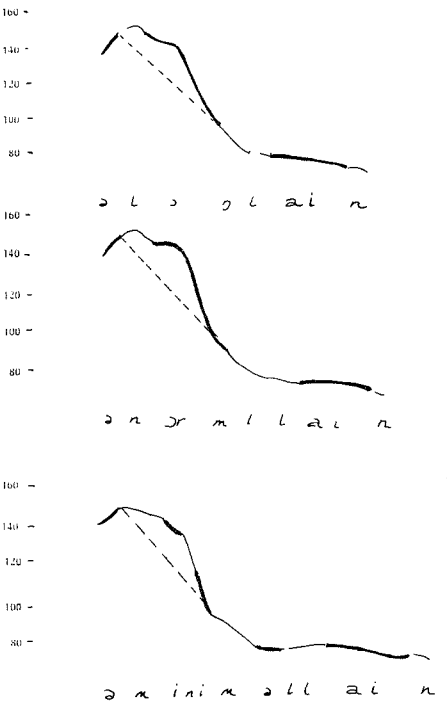


Fig. 3b Statements of varying length with focus on the adjective

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.

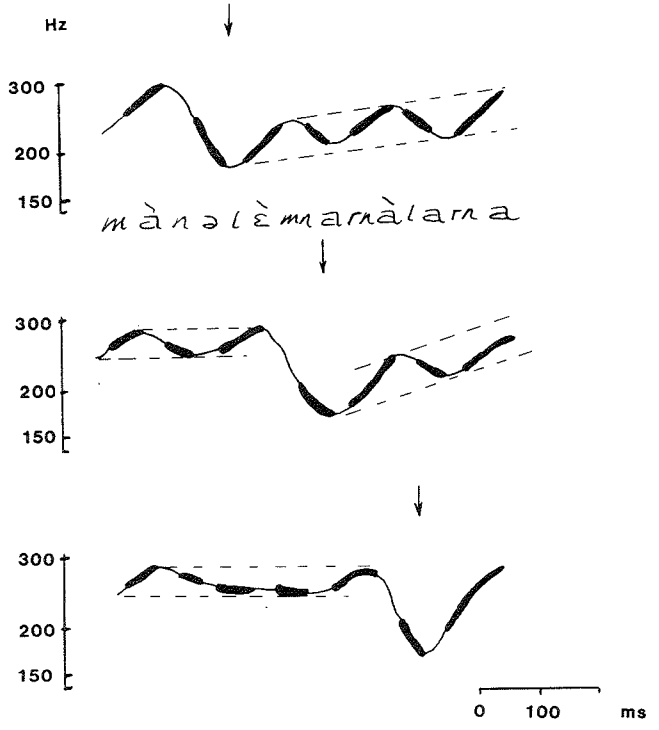


Fig. 4a F0 curves from South Swedish Manne lämnar nallarna 'Manne gives up the teddy bears'. Yes-no questions with focus in different positions.

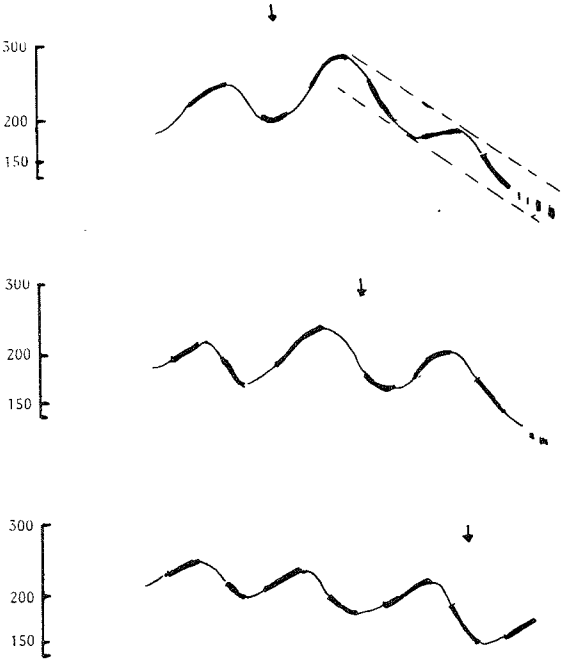


Fig. 4b Statements from South Swedish. From Anne-Christine Bredvad-Jensen (1984).

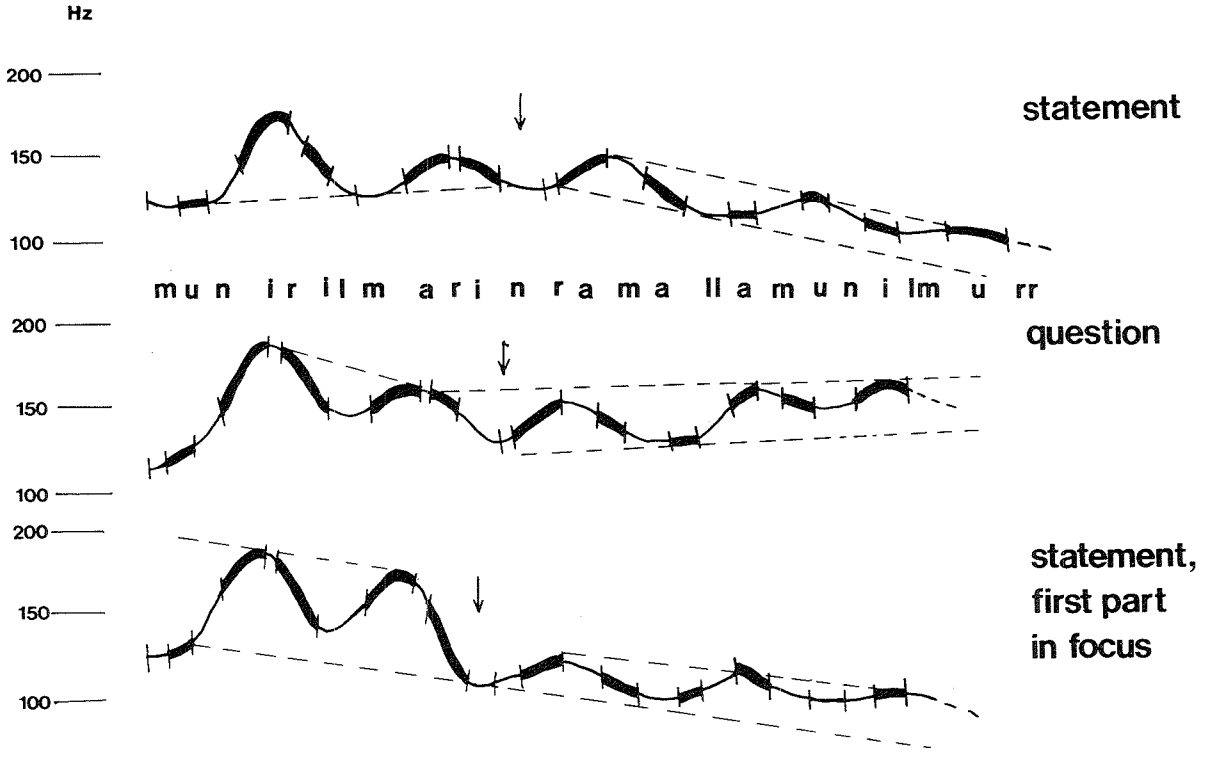


Fig. 5 F0 curves from Cairo Arabic Munír ilmárin ráma llamún ilmúrr
'The lithe Munir throw the bitter lemons away'. From Kjell Norlin

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

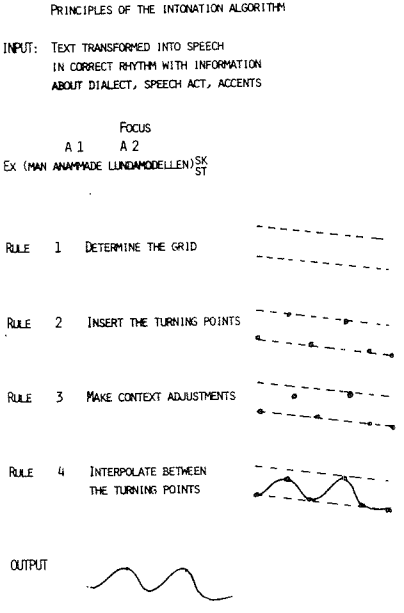


Fig. 6 The main principles of the generative scheme.

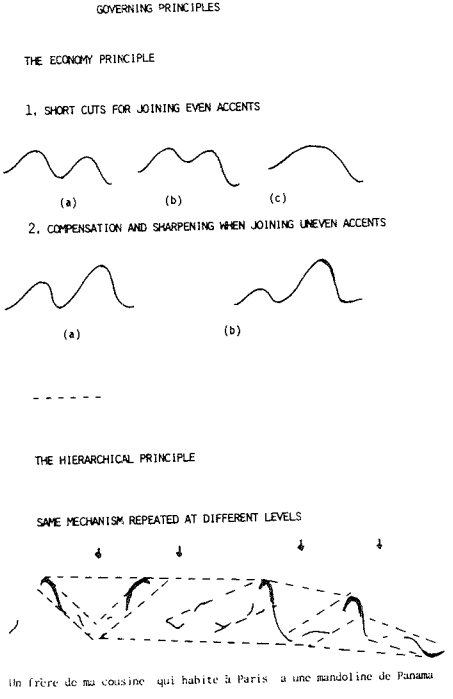


Fig.7 Some governing principles.

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 studie 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