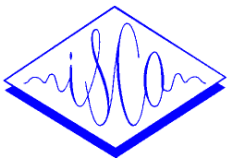

6th International Conference on Spoken Language Processing (ICSLP 2000)
Beijing, China
October 16-20, 2000

A PROMINENCE BASED MODEL OF SWEDISH INTONATION

Gunnar Fant and Anita Kruckenberg

KTH. Department of Speech, Music and Hearing

ABSTRACT

Our study is concerned with the modelling of intonation in Swedish, primarily prose reading. It is basically a superposition model with accentuation patterns added to prosodic base contours with prescribed onset, offset and declination. Boundary conditions at a juncture are related to sentence type and pause duration. There are several novel features of our approach. It is based on analysis of F0 data on a semitone scale of each syllable of a text read by five subjects of which two females. A sampling convention inspired by the canonical accent 1 and accent 2 typology of Bruce [1] is adopted. Inter-speaker variations are reduced by frequency and time normalisation, thereby bringing out common elements as well as individual global traits. The extent of F0 modulation of accent patterns are modelled as a function of the continuously scaled syllable and word prominence parameter Rs which we have introduced in earlier studies of stress and accentuation. Additional modifications with respect to the relative location of a word in a clause are introduced. Differences between speaker average and predicted data within a sentence are usually but not always small and will be systematically studied in order to derive rules for assimilation and syntactically motivated grouping. Our study contributes to detailed insights in the realisation of accent 1 and accent 2 patterns.

1. INTRODUCTION

The prominence parameter Rs was introduced by Fant and Kruckenberg [6]. The first major report on the relation of Rs to acoustic parameters including F0 was in Fant and Kruckenberg [7]. More recent studies also incorporating continuous records of a single subjects true sub- and supralottal pressure have been reported in [19-12]. Results from a syllable prominence grading test employing a jury of 15 listeners were added to the data, and displayed in synchrony with the acoustic and aerodynamic data.

The Rs parameter is. continuously scaled from 0 to 30. Fant and Kruckenberg [6] found that word prominence closely followed the most prominent syllable in the word. Typical values for stressed content words were Rs=20 and for unstressed function words Rs =10. Numerals, nouns and adjectives received somewhat greater scores than verbs and adverbs, and pronouns somewhat greater scores than prepositions and auxiliary verbs.

Focal accentuations are generally associated with a prominence of Rs>22. Contrastive and emphatic accentuation occupy the region of Rs = 25-30.

2.EXPERIMENTAL TECHNIQUES

A preliminary attempt of modelling of F0 patterns as a function of accent type, prominence Rs and location of an accented word was performed by Fant and Kruckenberg [8]. It was based on the single speaker data referred to above. The major findings obtained are still valid. The present study, first reported in [9] employed five subjects, of which two females, reading a one minute long paragraph of a novel. The prominence analysis was limited to accented syllables and was performed by two trained persons.

F0 traces on a log scale were printed out in synchrony with oscillogram, spectrogram and intensity curves [10-11]. Our calibration standard was 2 mm per semitone (st), i.e. 24 mm per octave. Measurements were made within 0.5 semitones. All F0 values were initially expressed in an absolute scale of semitones (st) relative 100 Hz.

A normalisation based on each speaker's average F0 in unstressed syllables was introduced. Accordingly, a correction of -7 st respectively -9.5 st were applied to the female data and -1 st respectively +1 st and 0 st for the male data. As a result the female data were effectively lowered to match the male data.

3. DATA SAMPLING PROCEDURE

Our F0 data sampling and labelling of accent parameters has been inspired by the canonical description and notations of Bruce [1]. A few minor additions and some specific interpretations of data labels have been introduced.

The domain of an accentuation is generally not confined to a single word. Bruce describes accent 1 as being initiated by an HL* fall from a high position H in the preceding syllable (if present) to a low position L*in the early part of the main syllable of the accented word, whereas accent 2 has an H*L fall within the primary stressed syllable. H* located close to the left boundary of the stressed vowel.

A prototype feature of accent 2 is that the H*L fall is followed by a rise of F0 to a high level in the next or a following syllable. The height of the secondary peak thus created, labelled Hg, increases with prominence. In accent 1, increasing prominence is associated with an F0 increase from L* to a high position Ha in the main syllable. These conditions of high prominence, typical of contrasting lab sentences, are traditionally referred to as sentence or focal accent.

In our study, on the other hand, we are confronted with a continuity of accentual realisations from low to high prominence levels. We therefore have to adopt labels for F0 measures and specific sampling routines that are independent of phonological

classification. Accordingly we have labelled the core parts of accent 1 as HL*Ha and of accent 2 as H*LHg. One of our findings, to be seen in figure 4, is that Ha and Hg, potentially carrying "sentence accent", display an approximately equal rate of increase or decrease with prominence which suggests that they reflect one and the same underlying physiological mechanism. Unaccented syllables are denoted Lu.

Our present routine is confined to the sampling at two positions within the major syllable of an accented word, L* and Ha for accent 1 and H* and L for accent 2. All other syllables, i. e. those denoted H, Hg or Lu are given one sample point only. Thus the word "margarinlåda" would be denoted: Lu Lu H* L Hg Lu.

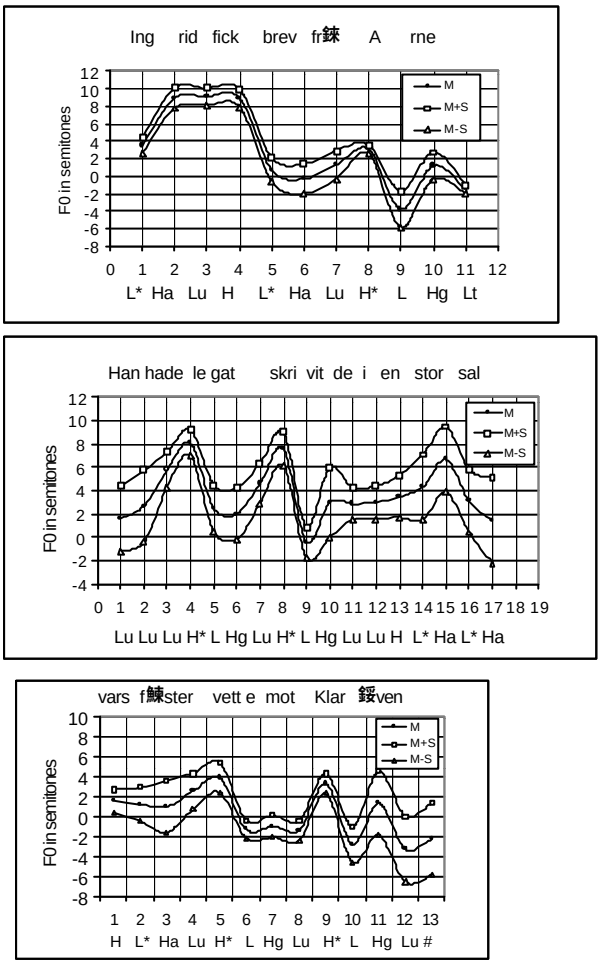

Figure 1: Five subjects, mean and standard deviation.

Lu measurements are referred to the middle of the vowel. In a weak accent 1 syllable the F0 contour may show a continuing fall instead of an L* Ha rise. As a consequence, our routines in these instances of sampling L* at the beginning and Ha at the end of the vowel produces a negative (Ha-L*). At higher degrees of prominence Ha refers to the peak of an F0 maximum in the middle of the vowel or the voiced part of the syllable.

4. F0 DATA DISPLAY

In order to visualise a connected F0 contour from the sampled data of a sentence we performed a smoothed continuous record of successive sample points based on Excel routines. The result is a time and frequency normalised intonation contour in which unvoiced portions are overbridged. The time scale is substituted by a sequence of data slots and the frequency scale is in semitones relative 100 Hz. Individual variations in timing and tempo are thus excluded and the data for any speaker is corrected for his or her deviation from the male average F0.

This system has obvious advantages for studies of the intonation of several subjects, and it provides efficient means of calculating average contours and individual variations.

Figure 1 shows three graphs from the analysis of prose reading. First a short sentence: "Ingrid fick brev från Arne", *Ingrid received a letter from Arne,* followed by the first part of the next sentence: "Han hade legat och skrivit de i en stor sal", *He had been lying writing it in a big hall*, and finally the clause: "vars fönster vette mot Klarälven", *with windows overlooking Klarälven.*

The three curves in each graph represent the mean of the five subjects, three males and two females, and the mean plus and minus one standard deviation of inter-subject variation. The consistency is apparent, especially in the first declarative sentence, where the standard deviation does not exceed 2 semitones.

5. MODELLING AND PREDICTION

Our present stage of modelling has been limited to the following constituents.

(1) An F0 baseline of prescribed initial rise, declination and final fall is selected. We have used two alternatives: One for the initial clause of a new sentence and one for following clauses, typically a sentence final clause which ends with a relative large pause.. The latter has a steeper final fall than the intial clause., see Figure 2. In a declarative sentence the extent of the total F0 fall is approximately independent of the length. Thus our standard base curves pertain to relative positions.

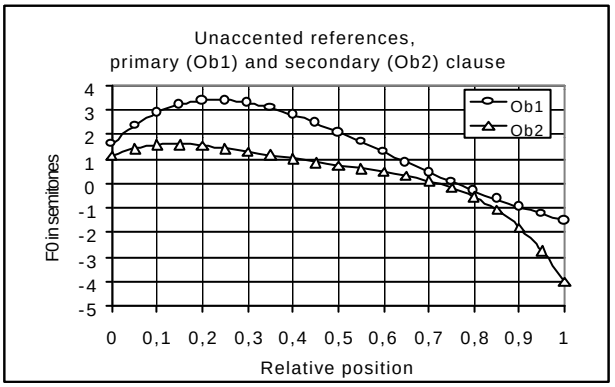

Figure 2: The two base curves for Lu slots.

We have found that clause initial and final F0 and the F0 reset at a juncture are significantly correlated to pause duration [9].

(2) The expected degree of word and syllable prominence can be predicted from word class. In the present study predictions are based on the assessed Rs data of the spoken material.

(3) A transcript of the text in terms of a sequence of accentuation and stress labels is made, and corresponding prominence values Rs are noted. Observe that all F0 slots within an accentuation domain share the same Rs, i.e. H attains the same Rs as L* and Ha, and Hg attains the same Rs as H* and L.

(4) The absolute position of a parameter slot is temporally translated to a relative position on the normalised 0 to 1 scale.

(5) F0 values are now predicted on the basis of the relative positions of accentuations and their prominences Rs. This is the central core of the prediction scheme. It is based on a detailed statistical survey of parameter values from the analyzed data. Non-linear regression equations have been derived in two steps, relating parameter values to Rs followed by a correction for position. This operation is performed separately for primary and secondary clauses. Syllables not belonging to an accent domain are given the Lu values of the particular base curve.

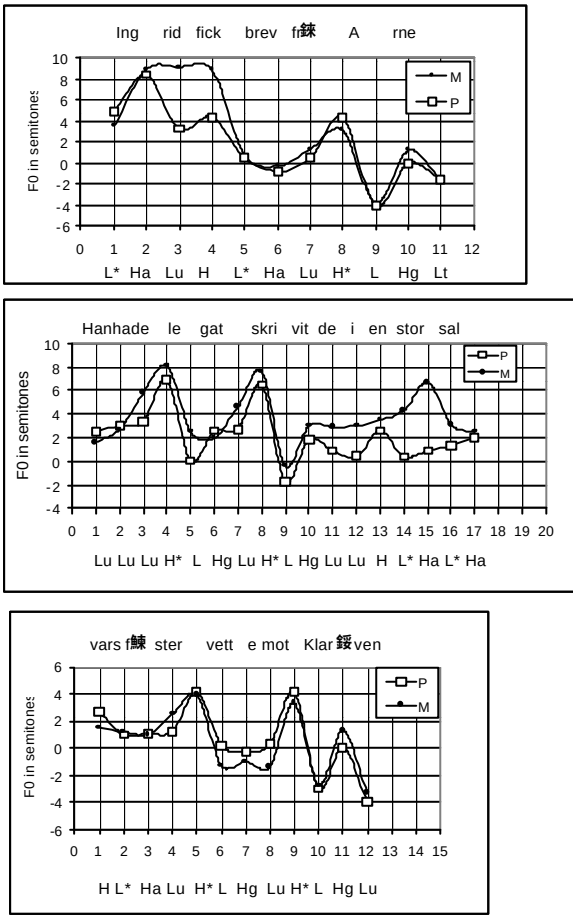

Figure 3: Measured and predicted F0 contours

(6) The results of the prediction are now displayed on the original scale of successive slots as a sequence of smoothly connected data points. In the same graph, see Figure 3, we have included the corresponding mean value contour of the 5 subjects.

(7) Observed differences between this first stage of prediction and measured values are analysed in order to be understood and corrected for in a more complete model.

6. DISCUSSION

6.1 F0 analysis and synthesis.

An overall impression from the total corpus of the five subjects text reading divided into 19 successive prosodic units, usually breath groups, is a high degree of coherence between the spoken data and predictions. Apart from some apparent shortcomings differences are of the order of 1-2 semitones only.

This first stage of modelling lacks several global features such as superimposed rising or falling intonation and syntactically motivated grouping, Gårding [13]. In the first sentence of Figure 3 we may note the lack of F0 carry over from the main syllable of the accent 1 word "Ingrid" to its second syllable and the following unstressed word "fick". However, at higher prominence levels the accent 1 F0 contour shows a peak in the main syllable followed by a minimum indicating a grouping that supports the prominence.

Our model being syllable based tends to be overdetailed in temporal fine structures. Additional rules for assimilation will be considered. On the other hand we have documented true detail patterns not considered in presently used systems [2-3]. One such is a relative early timing of the L minimum in the accent 2 H*L fall.

6.2 Comments on Swedish word accents

Figure 4 provides an overview of accent 1 and 2 parameters as a function of Rs and the relative position within a primary clause.

At high Rs levels the local F0 modulations are dominated by Ha of the main syllable of accent 1 and by H* and Hg of the secondary syllable of accent 2. These determine the upper bound of the intonation grid. The lowest part of the grid is set by accent 2 L points which usually is lower than L* and Lu.

Hg and Ha potentially carry the so called sentence accent. They are of the same magnitude and have approximately the same Rs dependency. With increasing Rs Hg grows faster than the H*. Above Rs=25 H* saturates.

At Rs>20 the accent 1 parameter H looses its relative dominance over L* which now is found in the early part of the rising branch up to Ha, i.e. (H-L*) becomes negative.

An early position enhances Ha and also reduces the depth of the accent 2 parameter L. These trends occur in addition to the initial rise of the superimposed baseline. A muscular interaction is postulated.

It has been suggested by Elert [4] and by Engstrand [5] that accent 1 could be considered as phonologically unmarked. In focal accentuation both posses sentence accent, but it is only

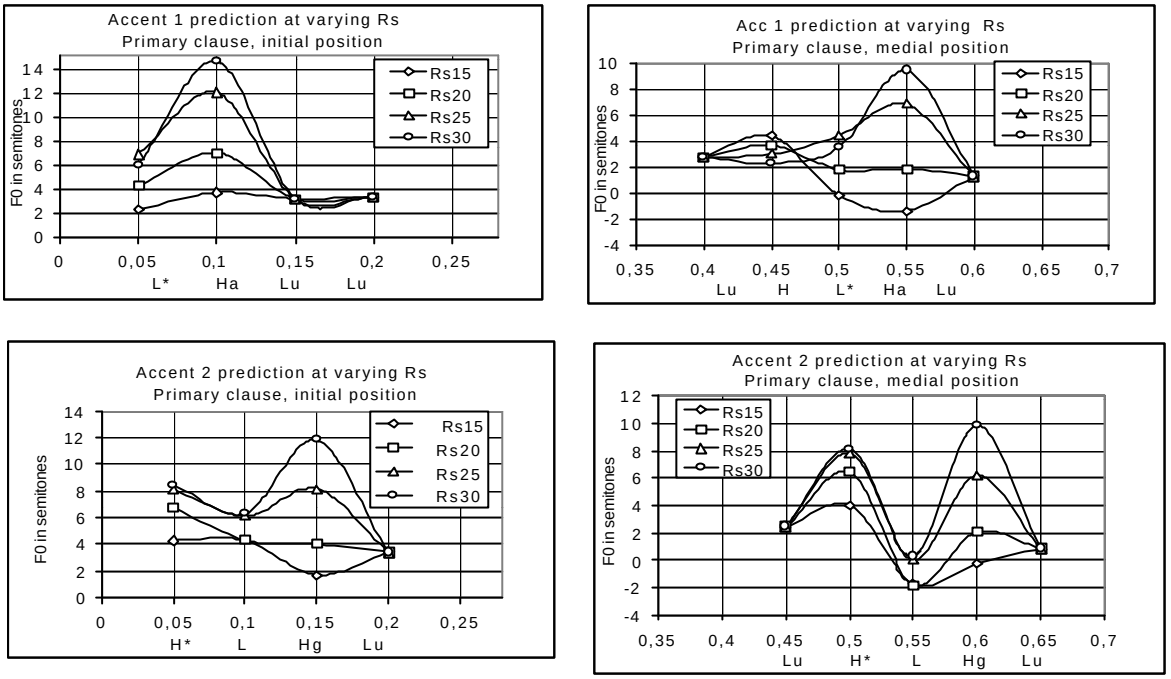

Figure 4: Accent 1 and 2 parameters as a function of Rs and relative position in an initial clause.

accent 2 that conveys an additional accent pattern. However, as pointed our by Bruce [1] the identity lies in the timing. In focal accentuation this is primarily a matter of Ha occurring earlier than Hg, whilst the similar relation of H with respect to H* loses significance.

Our study will continue with extension of the modelling and with perceptual assessments. Preliminary results point at a high degree of acceptance as long as deviations from a norm are within the limits of individual variations. Our model should also be useful for descriptive purposes.

8. REFERENCES

1. Bruce, G. *Swedish Word Accents in Sentence Perspective.* Lund: Gleerup, 1977.
2. Bruce, G and Granström, B. "Prosodic modeling in Swedish speech synthesis". *Speech Communication* 13. 63-74, 1993.
3. Carlson, R. and Granström, B. "Word accent, emphatic stress, and syntax in a synthesis-by-rule scheme for Swedish". *STL-QPSR* 2-3/1973. 31-35.
4. Elert, C.- C. *Allmän och Svensk Fonetik.* Almqvist& Wiksell. Uppl. 7**,** 1995.
5. Engstrand, O. "Phonetic Interpretation of the Word Accent Contrast in Swedish." *Phonetica* 52. 171-179, 1995.
6. Fant, G., and Kruckenberg, A. "Preliminaries to the Study of Swedish Prose Reading and Reading Style." *STL-QPSR* 2/1989. 1-83.
7. Fant, G. and Kruckenberg, A.. "Notes on stress and word accent in Swedish." *Proceedings of the International Symposium on Prosody, Sept 18 1994, Yokohama.* Also published in *STL-QPSR* 2-3/1994. 125-144.
8. Fant, G., and Kruckenberg, A.. "Prominence and accentuation. Acoustical correlates." *The Swedish Phonetics Conference, 1998, Stockholm University* ed. by P. Branderud and H. Traunmüller, 142-145. 1998.
9. Fant, G., and Kruckenberg, A. "F0-patterns in text reading." *Proc of Fonetik 1999, Gothenburg papers in theoretical linguistics*, ed. by Jens Allwood. Göteborg University, 53-56.
10. Fant, G., and Kruckenberg, A. "Prominence correlates in Swedish Prosody." *Proceedings of ICPhS-99*, San Francisco, Vol. 3. 1749-1752.
11. Fant, G., Kruckenberg, A. and Liljencrants, L. "Acoustic-phonetic analysis of prominence in Swedish." To be published in Antonis Botinis, editor, *Intonation. Analysis, Modelling and Technology*. Kluwer, Academic Publishers, 55-86. 2000.
12. Fant, G., Kruckenberg, A., Hertegård, S. and Liljencrants, J., 1997. "Accentuation and subglottal pressure in Swedish." *Proc. ESCA workshop on Intonation. Athens, Greece*, 1997, 111-114.1997.
13. Gårding, E. "Intonation in Swedish." In *Intonation Systems*, ed. by Daniel Hirst and Alberto Di Cristo. Cambridge University Press 1998. 112-130. Also in *Lund University Linguistics Department, Working papers* 35. 63-1988.