Dept. for Speech, Music and Hearing
Quarterly Progress and Status Report

The FK prosody system

Fant, G. and Kruckenberg, A.

journal: TMH-QPSR
volume: 48
number: 1
year: 2006
pages: 001-003

KTH Computer Science and Communication

http://www.speech.kth.se/qpsr

The FK prosody system

Gunnar Fant and Anita Kruckenberg

Introduction. Notes on prosodic research

The FK prosody system originates from a co-operation between Gunnar Fant and Anita Kruckenberg established already in 1984. It started with a project labeled ”God Svenska” (Fant & Kruckenberg, 1986) A study (Fant, Nord & Kruckenberg, 1986) was directed to individual variations in timing but also to the realization of prosodic boundaries.

These issues were followed up in Fant & Kruckenberg (1989), which is the major publication from the early period. It outlined the durational correlates of stress and introduced a technique of perceptual continuous scaling of word and syllable prominence. Later more detailed studies with 15 listeners (Fant & Kruckenberg, 1999) resulted in a measure, labelled RS (received stress) ranging from 0 to 30 which now has become an important parameter in our prosody modelling. In this article, and in more detail in Fant, Kruckenberg & Liljencrants (2000), changes in F0 and duration following shifts of focal emphasis in a sentence was illustrated.

A quantal theory of speech timing has been proposed (Fant & Kruckenberg, 1996). It is based on observations of rhythmically preferred time intervals in pausing. Other aspects of prosodic investigations to be mentioned are durational correlates of stress in French and English compared to Swedish (Fant, Kruckenberg & Nord, 1991). We have also studied prosodic patterns in poetry reading (Kruckenberg & Fant, 1993).

Developing the FK model

A strength of our text-to-speech system is that it has a solid foundation from speech analysis. Our main corpus is a study of five subjects, 3 males and 2 females, reading a text from a Swedish novel. Records were assembled of phone and syllable durations and of at least one F0 measure per syllable. Later studies included pausing.

A most important decision was to adopt a semitone scale for F0. It has allowed us to extract comparable data for males and females. We also performed a normalization of speaking rate by substituting continuous F0 contours by syllable aligned sequences of sampled data. These were defined by a phonological marking of syllables as carrying accent 1 or accent 2 or being unaccented.

As a result of the F0 normalization, we attained representative group average contours of each sentence and also measures of individual variations in accentuation and prominence. A remarkable finding was the overall success in speaker normalization. In spite of individual variations, the standard deviation of F0 within a sentence was less than 2 semitones. The spread of F0 data may further be reduced when corrections for individual variations in RS are taken into account.

We may thus claim that our analysis has revealed language specific production patterns. The prominence parameter RS has a central role. In addition to relative positions, RS is a determinant of both F0 modulation and of duration.

In synthesis, RS attains default values according to word class and is a major tool for experimental variation of stress patterns.

Phone durations are scaled according to major vowel and consonant classes and with respect to RS and position within a syllable. Our reference data derive from a specific databank, DS, of one expert reading a larger text material.

A major aspect of prosodic grouping is the pausing and occasional resetting of an underlying F0 contour. The FK model imposes the superposition of local accent patterns on one or more smoothly shaped base curves within a sentence, see Figure 1. Each of these is determined by a rise/fall contour and a pause. Such breaks usually occur in connection with breathing.

To the fine structure of prosodic grouping belong pauses that occur without an associated base curve shift. All pauses are combined with final lengthening, and final lengthening alone is often sufficient to mark a minor boundary.

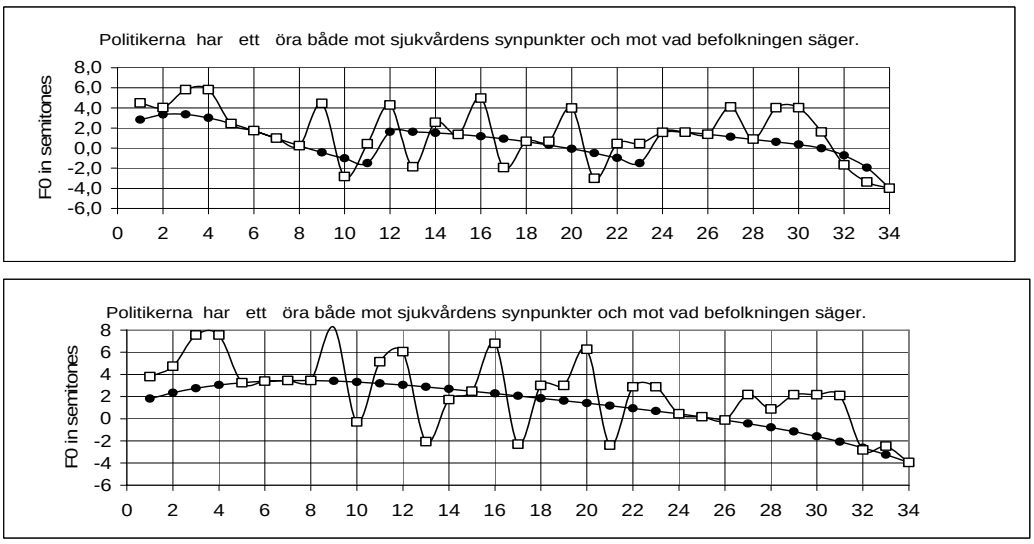

Figure 1. F0 accent modulations superimposed on three prosodic base curves above, and below on a single base curve.

We have found that subjects may differ substantially in pausing within a complete sentence but much less in pausing between sentences. (Fant, Kruckenberg & Barbosa Ferreira, 2003).

The first major report on FK synthesis was from a meeting in Aix en Provence (Fant et al., 2002). It has been followed up by Fant & Kruckenberg (2004a,b,c; 2006). The most complete account of our work is to be found in the recent book of Fant (2004).

Present status

At present, the core of procedures and rules for synthesis are confined in Excel files, supplemented by a number of internal reports and memoranda. They assume as an input the access to a lexical search for word class, phonetic transcription and accent 1 and accent 2 assignments. In order to facilitate the processing, our present Excel programs for Mbrola synthesis are being translated to Java format by Rickard Carlsson, a thesis student at Uppsala University. However, there remains the overall programming for lexical access in a complete system.

A decomposition of a text into syllables is required. We have developed our own rules, which depart from standards in some details.

Of major importance is our novel set of parsing rules for determining prosodic grouping and pausing. They are stated in plain Word language and need to be transferred to rules for dictionary interface. They have been tested in synthesis with quite promising results in comparison with our accumulated analysis data.

As far as we can judge, FK modelling and FK text-to-speech synthesis has a potential of providing a more natural prosody than any other now existing system. What remains is programming for lexical interface.

Of special interest to us are applications in experimental studies of reading style. Here the FK system has the advantage of covering all levels of the production process, linguistic as well as acoustic. An exception is voice source characteristics, which are not included in Mbrola synthesis. They appear to be of secondary importance, once F0 and duration are correctly encoded. We possess an accumulated knowledge base for voice source dynamics in connected speech (Fant, 1997). There are also unpublished reports.

References

Fant G (1997). The voice source in connected speech. *Speech Communication* 22:125-139.

Fant G (2004). *Speech Acoustics and Phonetics. Selected Writings.* Kluwer/Springer.

Fant G & Kruckenberg A (1986). Projekt god svenska – databas. *Internrapport Inst f Talöverföring och Musikakustik*, KTH, Stockholm.

Fant G & Kruckenberg A (1989). Preliminaries to the study of Swedish prose reading and reading style. *STL-QPSR* 2/1989, 1-83.

Fant G & Kruckenberg A (1996). On the quantal nature of speech timing. *Proc ICSLP;* 2044-2047.

Fant G & Kruckenberg A (1999). Prominence correlates in Swedish prosody. *Proc of Int Conf of Phonetic Sciences (ICPhS*), San Francisco 3, 1749-1752.

Fant G & Kruckenberg A (2004a). The FK prosody model. Analysis and synthesis. (To be publ in Proc of Nordic Prosody IX, Lund, Aug 2004).

Fant G & Kruckenberg A (2004b). Prosody by rule in Swedish with language universal implications. *Proceedings Prosody 2004*, Nara, 405-408.

Fant G & Kruckenberg A (2004c). An integrated view of Swedish prosody. Voice production, perception and synthesis. In: Fant G: *Speech Acoustics and Phonetics, Selected Writings*, Kluwer Academic Publisher/Springer.

Fant G & Kruckenberg A (2006). Individual and contextual variations of prosodic parameters. *TMH-QPSR, KTH*, 48: 5-8.
http://www.speech.kth.se/qpsr/tmh/

Fant G, Kruckenberg A & Barbosa Ferreira J (2003). Individual variations in Pausing. A study of read speech. *Proc from Fonetik 2003, Phonum 9*, Umeå University, 193-196.

Fant G, Kruckenberg A & Nord L (1991). Durational correlates of stress in Swedish, French and English. *Journal of Phonetics* 19: 351-365.

Fant G, Kruckenberg A & Liljencrants J (2000). Acoustic-phonetic analysis of prosody in Swedish. In: Botinis A (ed) *Intonation. Analysis, Modelling and Technology*. Kluwer, Academic Publishers, 55-86.

Fant G, Kruckenberg A, Gustafson K & Liljencrants J (2002). A new approach to intonation analysis and synthesis of Swedish. *Speech Prosody* 2002, Aix en Provence. Also in *Fonetik 2002*, TMH-QPSR 2002.

Fant G, Nord L & Kruckenberg A (1986). Individual variations in text reading. A data-bank pilot study. *STL-QPSR* 4/1986, 1-17.

Kruckenberg A & Fant G (1993). Iambic versus trochaic patterns in poetry reading. *Nordic Prosody VI*, Stockholm, 123-135.

Additional references are to be found in the home page of Gunnar Fant
http://www.speech.kth.se/~gunnar/