SWING: A tool for modelling intonational varieties of Swedish

Beskow, Jonas; Bruce, Gösta; Enflo, Laura; Granström, Björn; Schötz, Susanne

Published in:
Proceedings of Fonetik 2008

2008

SWING:
A tool for modelling intonational varieties of Swedish

Jonas Beskow[2], Gösta Bruce[1], Laura Enflo[2], Björn Granström[2], Susanne Schötz[1] (alphabetical order)
[1]Dept. of Linguistics & Phonetics, Centre for Languages & Literature, Lund University
[2]Dept. of Speech, Music & Hearing, School of Computer Science & Communication, KTH

Abstract

SWING (SWedish INtonation Generator) is a new tool for analysis and modelling of Swedish intonation by resynthesis. It was developed in order to facilitate analysis of regional varieties, particularly related to the Swedish prosody model. Annotated speech samples are resynthesized with rule based intonation and audiovisually analysed with regard to the major intonational varieties of Swedish. We find the tool useful in our work with testing and further developing the Swedish prosody model.

Introduction and background

Our object of study in the research project SIMULEKT (Simulating Intonational Varieties of Swedish, supported by the Swedish Research Council) (Bruce et. al, 2007) is the prosodic variation characteristic of different regions of the Swedish-speaking area, shown in Figure 1.

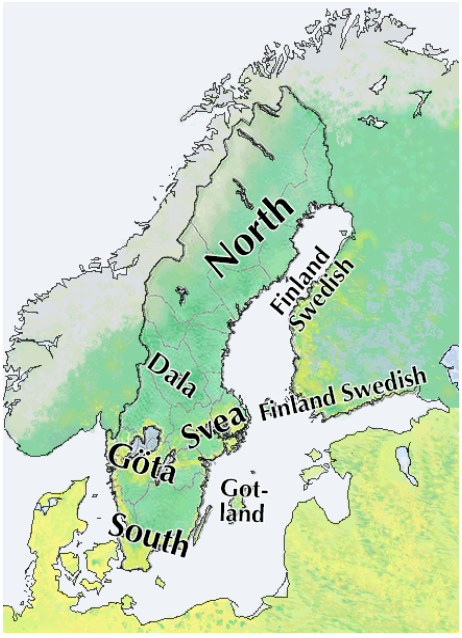

Figure 1. Approximate geographical distribution of the seven main regional varieties of Swedish.

The seven regions correspond to our present dialect classification scheme. In our work, the Swedish prosody model (Bruce & Gårding, 1978; Bruce & Granström, 1993; Bruce, 2007) and different forms of speech synthesis play prominent roles. Our main sources for analysis are the two Swedish speech databases SpeechDat (Elenius, 1999) and SweDia 2000 (Engstrand et.al, 1997). SpeechDat contains speech recorded over the telephone from 5000 speakers, registered by age, gender, current location and self-labelled dialect type. The research project SweDia 2000 collected a word list, an elicited prosody material and extensive spontaneous monologues from 12 speakers (younger and elderly men and women) each from more than 100 different places in Sweden and Swedish-speaking parts of Finland, selected for dialectal speech.

The Swedish prosody model

The main parameters for the Swedish prosody model are for word prosody 1) word accent timing, i.e. timing characteristics of pitch gestures of word accents (accent I/accent II) relative to a stressed syllable, and 2) pitch patterns of compounds, and for utterance prosody 3) intonational prominence levels (focal/non-focal accentuation), and 4) patterns of concatenation between pitch gestures of prominent words.

Background

An important part of our project work concerns auditive and acoustic analysis of dialectal speech samples available from our two extensive speech databases described in Section 1. This work includes collecting empirical evidence of prosodic patterns for the regional varieties of Swedish described in the Swedish prosody model, as well as identifying intonational patterns not yet included in the model. To facilitate our work with testing and further developing the model, we needed a tool for generating rule-based intonation.

Design

SWING comprises a number of parts, which are joined by the speech analysis software Praat (Boersma & Weenink, 2007), also serving as the graphical interface. Annotated speech samples and rules for generating intonation are used as input to the tool. The tool generates and plays resynthesis – with rule-based and speaker-normalised intonation – of the input speech sample. Additional features include visual display of the output on the screen, and options for printing various kinds of information to the Praat console (Info window), e.g. rule names and values, or the time and F_0 of generated pitch points. Figure 2 shows a schematic view of the tool design.

Table 1. Labels used for prosodic annotation of the speech samples to be analysed by the tool.

Label	*Description*
pa1	primary stressed (non-focal) accent 1
pa2	primary stressed (non-focal) accent 2
pa1f	focal accent 1
pa2f	focal accent 2

Rules

The Swedish prosody model is implemented as a set of rule files – one for each regional variety in the model – with timing and F_0 values for critical points in the rules. These files are simply text files with a number of columns, where the first one contains the rule names, and the following columns contain three pairs of values,

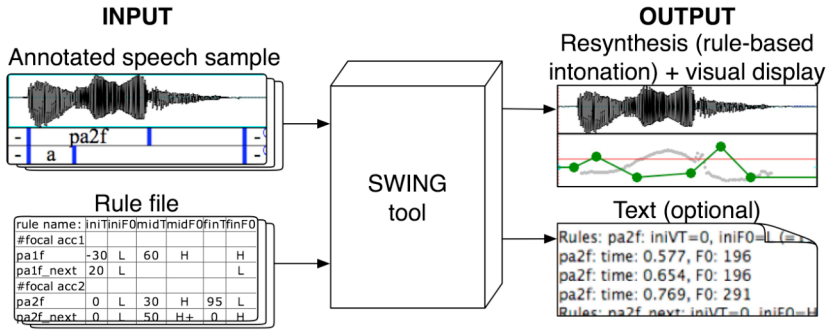

Figure 2. Schematic overview of the SWING tool.

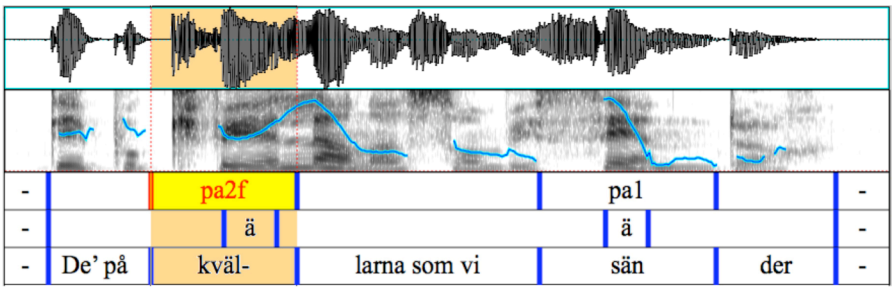

Figure 3. Example of an annotated input speech.

Speech material

The input speech samples are annotated manually. Stressed syllables are labelled prosodically and the corresponding vowels are transcribed orthographically. Table 1 shows the prosodic labels used in the current version of the tool, while Figure 3 displays an example utterance with prosodic annotation: *De' på kvällarna som vi sänder* 'It's in the evenings that we are transmitting'.

corresponding to the timing and F_0 of equally many critical pitch points of the rules. The three points are called *ini* (initial), *mid* (medial), and *fin* (final). They contain values for the timing (T) and F_0 (F0). Timing of F_0 points is expressed as a percentage into the stressed syllable, starting from the onset of the stressed vowel. If no timing value is explicitly stated in the rule, the pitch point is by default aligned with the onset of the stressed vowel. Three values are used for F_0: L (low), H (high) and H+ (extra high, used in focal accents). The mid

pitch point is optional; unless it is needed by a rule, its values may be left blanc. Existing rules are easy to adjust, and new rules can be added. Table 3 shows an example of the rules for South Swedish. Several rules contain a second part, which is used for the pitch contour of the following (unstressed) interval (segment) in the annotated input speech sample. This extra part has 'next' attached to its rule name. Examples of such rules are pa1f and pa2f in Table 2.

points with the Praat built-in Manipulation feature. Figure 4 shows a Praat Manipulation object for an example utterance. The light grey line under the waveform shows the original pitch, while the circles connected with the solid line represent the rule-generated output pitch contour. In the Praat interface, the user can easily compare the original and the resynthesized sounds and pitch contours, and further adjust or manipulate the output pitch contour (by

Table 2: Example rule file for South Swedish with timing (T) and F_0 (F0) values for initial (ini), mid (mid) and final (fin) points.

Rule name	***iniT***	***iniF0***	***midT***	***midF0***	***finT***	***finF0***
global (phrase)		L			L	
concatenation		L			L	
pa1f (focal accent 1)	-10	L	20	H+	50	L
pa1f next (extra gesture)		L			L	
pa2f (focal accent 2)		L	40	L		H+
pa2f next (extra gesture)		H+	30	L		L
pa1 (non-focal accent 1)	-30	L	10	H	40	L
pa2 (non-focal accent 2)		L	50	L		H
pa2 next (extra gesture)		H	30	L		L

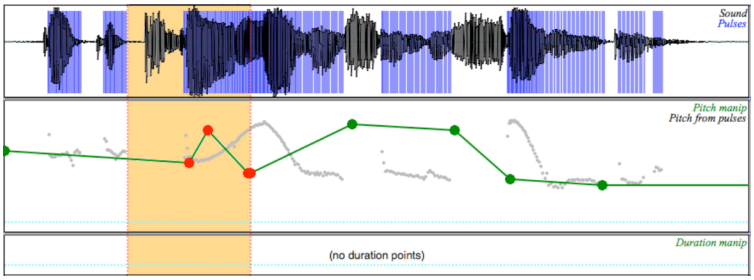

Figure 4. Praat Manipulation display of a South Swedish utterance with rule-generated Svea intonation (circles connected by solid line; original pitch: light-grey line).

The SWING tool procedure

Analysis with the SWING tool is fairly straightforward. The user selects one input speech sample and one rule file to be used with the tool, and which (if any) information about the analysis (rules, pitch points, debugging information) to be printed to the console. A Praat script generates resynthesis of the input speech sample with a rule based output pitch contour. Generation of the output pitch contour is based on 1) the pitch range of the input speech sample, which is used for speaker normalisation, 2) the annotation, which is used to find the time and prosodic gesture to generate, and 3) the rule file, which is used for the values of the pitch points in the output. The Praat graphical user interface provides immediate audio-visual feedback of how well the rules work, and also allows for easy additional manipulation of pitch

moving one or several pitch points) and the annotation files. The rule files can be adjusted in any text editor.

Testing the Swedish prosody model with SWING

SWING is now being used in our work with testing and developing the Swedish prosody model. Testing is done by selecting an input sound sample and a rule file of the same intonational variety. If the model works adequately, there should be a close match between the F_0 contour of the original version and the rule-based one generated by the tool. Figure 5 shows examples of such tests of an utterance in the Svea and South Swedish varieties. Interesting pitch patterns found in our material which have not yet been implemented in the rules are also analysed using the tool.

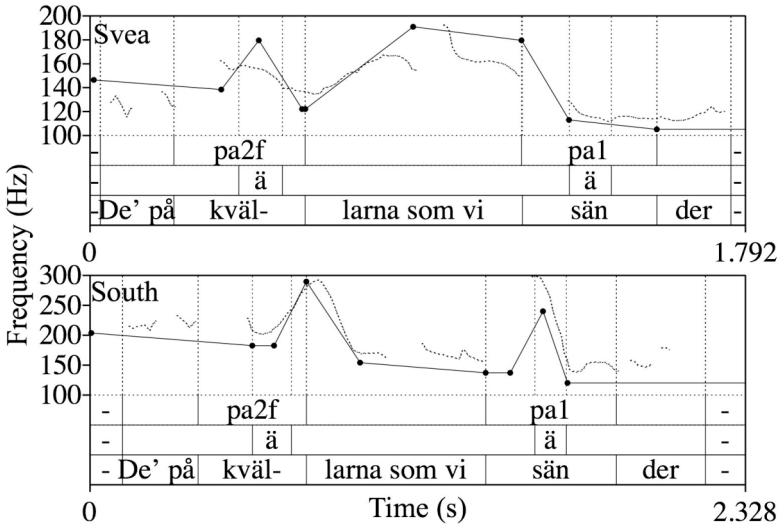

Figure 5. Original and rule-based intonation of the utterance *De' på kvällarna som vi sänder* 'It's in the evenings that we are transmitting for Svea and South Swedish (original pitch: dotted line; rule-generated pitch: circles connected with solid line).

Discussion and future work

Although SWING still needs work, we already find it useful in our project work of analysing speech material as well as testing our model. We consider the general results of our model tests to be quite encouraging. The tool has so far been used on a limited number of words, phrases and utterances and with a subset of the parameters of the Swedish prosody model, but was designed to be easily adapted to further changes and additions in rules as well as speech material. We are currently including more speech samples from our two databases, and implementing other parameters of the Swedish prosody model, such as rules for compound words. Our near future plans include evaluation of the tool by means of perception tests with natural as well as rule-generated stimuli.

References

Boersma P. and Weenink D. (2007) Praat: doing phonetics by computer (version 4.6.17) [computer program]. http://www.praat.org/, visited 12-Mar-08.

Bruce G. (2007) Components of a prosodic typology of Swedish intonation. In Riad T. and Gussenhoven C. (eds) Tones and Tunes, Volume 1, 113-146, Berlin: Mouton de Gruyter.

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

Bruce G. and Granström B. (1993) Prosodic modelling in Swedish speech synthesis. A prosodic typology for Swedish dialects. Speech Communication 13, 63-73.

Bruce G., Granström B., and Schötz S. (2007) Simulating intonational varieties of Swedish. Proc. of ICPhS XVI (Saarbrücken, Germany)

Engstrand O., Bannert R., Bruce G., Elert C-C., and Eriksson A. (1997) Phonetics and phonology of Swedish dialects around the year 2000: a research plan. Papers from FONETIK 98, PHONUM 4, Umeå: Department of Philosophy and Linguistics, 97-100.

Elenius K. (1999) Two Swedish SpeechDat databases - some experiences and results. Proc. of Eurospeech 99, 2243-2246.