Recognizing and Modelling Regional Varieties of Swedish

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

Published in:
Proceedings of Interspeech 2008

2008

Recognizing and Modelling Regional 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, Sweden
[2]Dept. of Speech, Music & Hearing, School of Computer Science & Communication, KTH, Sweden
{gosta.bruce, susanne.schotz}@ling.lu.se, bjorn@speech.kth.se, {lenflo, beskow}@kth.se

Abstract

Our recent work within the research project SIMULEKT (Simulating Intonational Varieties of Swedish) includes two approaches. The first involves a pilot perception test, used for detecting tendencies in human clustering of Swedish dialects. 30 Swedish listeners were asked to identify the geographical origin of Swedish native speakers by clicking on a map of Sweden. Results indicate for example that listeners from the south of Sweden are better at recognizing some major Swedish dialects than listeners from the central part of Sweden, which includes the capital area. The second approach concerns a method for modelling intonation using the newly developed SWING (SWedish INtonation Generator) tool, where annotated speech samples are resynthesized with rule based intonation and audiovisually analysed with regards to the major intonational varieties of Swedish. We consider both approaches important in our aim to test and further develop the Swedish prosody model.

Index Terms: perception (of Swedish) dialects, prosody modelling, analysis tool, resynthesis

1. Introduction

Our object of study in the research project SIMULEKT (Simulating Intonational Varieties of Swedish) [1] is the prosodic variation characteristic of different regions of the Swedish-speaking area, shown in Figure 1. The seven regions correspond to our present dialect classification scheme. In our work, the Swedish prosody model [2, 3, 4] and different forms of speech synthesis play prominent roles. Our main sources for analysis are the two Swedish speech databases SpeechDat [6] and SweDia 2000 [5]. SpeechDat contains speech recorded over the telephone from 5000 speakers, registered by age, gender, current location and self-labeled dialect type, according to Elert's suggested Swedish dialect groups [7] that is a more fine-grained classification with 18 regions in Sweden. 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.

1.1. The Swedish prosody model

The main parameters for the Swedish prosody model [2, 3, 4] 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.

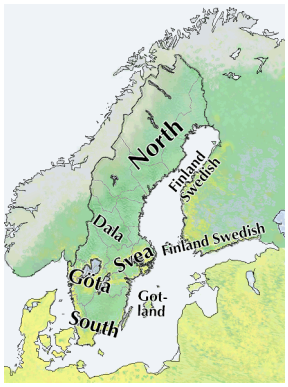

Figure 1: *Approximate geographical distribution of the seven main regional varieties of Swedish.*

1.2. Outline of the paper

This paper exemplifies two recent approaches involving recognition and modelling of Swedish regional varieties. The first involves a perception experiment where listeners were asked to recognize the geographical location of dialectal speech stimuli by clicking on a map of Sweden, while the second concerns the development of an analysis tool for testing and further developing our prosody model using rule-based intonation resynthesis.

2. Testing perception of Swedish dialects

2.1. Background

Prosody, vowels and some consonant allophones are likely to be important when trying to decide from where a person originates. The aim of this work is to develop a method which could be of help in finding out how well Swedish subjects can identify the geographical origin of other Swedish native speakers. By determining the dialect identification ability of Swedish listeners, a foundation could be made for further research involving dialectal clusters of speech. In order to evaluate the importance of the factors stated above for dialect recognition, a pilot test was put together using recordings of identical utterances from 72 speakers.

2.2. Speech material

In the Swedish SpeechDat database, two sentences read by all speakers were added for their prosodically interesting properties. One of them was used in this experiment: *Mobiltelefonen är nittiotalets stora fluga, både bland företagare och privatpersoner.* 'The mobile phone is the big hit of the nineties both among business people and private persons.'

For this test, each of Elert's 18 dialect groups in Sweden were represented by four speakers, two female and two male, with an age span as wide as possible.

2.3. Subjects

30 subjects participated in the experiment, 12 female and 18 male, with an average age of 32 and 33 years, respectively. Subjects were placed in two groups depending on where the majority of the childhood and adolescence (0-18 years) had been spent. Seven females and eleven males grew up in the central part (Svealand, Svea on the map in Figure 1) whereas five female and seven male subjects were raised in the southern part (Götaland, Göta and South on the map in Figure 1) of Sweden.

2.4. Experiment

The test was made with the scripting language Tcl/Tk and carried out in Stockholm (Svealand) and Lund (Götaland). The experiment comprised a dialect-test part, a geography test and a questionnaire. In the dialect test, the SpeechDat stimuli were played in random order over headphones and could be repeated as many times as desired before answering by clicking on a map of Sweden. The geography test included 18 Swedish towns presented one by one in written form, which were placed on the map in the same manner as for the dialect test. These towns are the most populated in each of Elert's dialect group areas. Lastly, a questionnaire was filled out by all subjects, so as to provide information about e.g. age, gender and dialectal background.

2.5. Results

Subjects vary very much in their ability to locate speakers. In Figure 2, results for two listeners are displayed on the Swedish map. Dark-red dots mark the correct dialect locations and light-green dots the answers provided by the subjects. Figure 3 displays the results from the geography test in the same way. The two subjects were chosen as typical representatives of Svealand and Götaland. Both were males aged 25, but with different backgrounds. Subject 1 from Svealand was born and raised in Stockholm with parents from Stockholm and had been exposed to regional accents to a small extent. Subject 2 from Götaland was born and raised in Jönköping by parents from the same area.

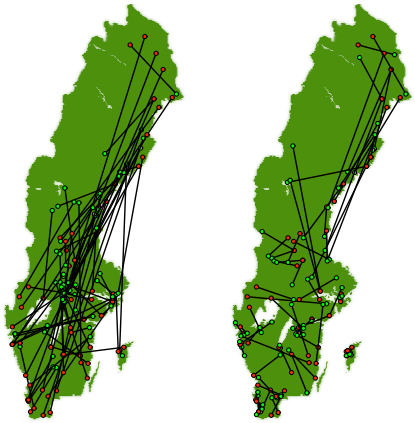

Figure 2: *Dialect test results for listener 1 from Svealand (left) and listener 2 from Götaland (right). Dark-red dots for correct locations are connected by lines with light-green dots for answers given by subject.*

Figure 3: *Geography test results for listener 1 from Svealand (left) and listener 2 from Götaland (right). Dark-red dots for correct locations are connected by lines with light-green dots for answers given by subject.*

Speakers in the test vary considerably as to how consistently they are identified. An example is displayed in Figure 4, which shows where all subjects have placed speaker no. 1, a 51-year-old female from Täby, Svealand and speaker no. 2, a 55-year-old female from Kiruna, Norrland.

The average errors in dialect placement were computed as an arbitrary unit distance on the map. Table 1 shows this mean for four different Elert dialect areas (4 speakers in each area). The subjects are divided into Svealand and Götaland listeners.

Table 1: *Götaland and Svealand listeners' average dialect location errors for speakers from four dialect areas.*

Elert	*Region*	*Svealand*	*Götaland*
18	Skåne, far south	54	40
14	Göteborg area, south west	94	58
8	Stockholm, capital area	35	40
1	Norrland, far north	330	279

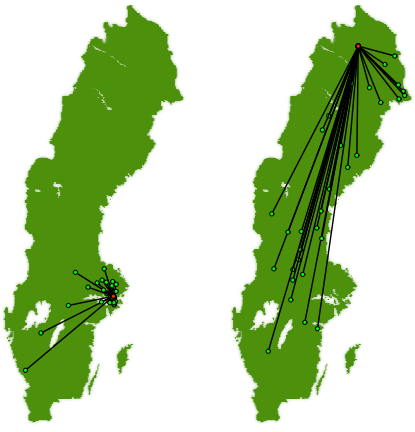

Figure 4: *Dialect test results for speaker 1 from Svealand (left) and speaker 2 from Norrland (right). Dark-red dot for correct location is connected by lines with light-green dots for answers given by all subjects.*

2.6. Discussion and future work

Our data suggests that Svealand listeners are less able to locate dialects, except their own. It is not unreasonable that it is easier to identify and locate dialects closer to you. The Götaland listeners were also good at locating Svealand speakers, possibly due to the great exposure to these dialects in media. The high error values for the northern-most part of Norrland is possibly because of long distances between towns in the northern part of Sweden, but also in part because of the subjects' less exposure to these accents. These are only some examples of results of the dialect location test. Further analysis of the data is planned in the near future, particularly using full statistical analysis. A possible extension is to use segmentally neutralized stimuli, to focus on the prosodic features of Swedish regional varieties. We also wish to use listener clustering as a tool in deciding which factors play the most important roles for distinguishing the different Swedish dialect types, which might lead to modified dialect taxonomy.

3. The SWING intonation analysis tool

SWING (SWedish INtonation Generator) is a new tool for analysis and modelling of Swedish intonation by resynthesis. The tool was developed in order to facilitate analysis of intonational varieties, particularly related to the Swedish prosody model.

3.1. Background

An important part of our project work is 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 intonational 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.

3.2. Design

SWING consists of several parts joined by the speech analysis software Praat [8], which also serves 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 5 shows a schematic view of the tool design.

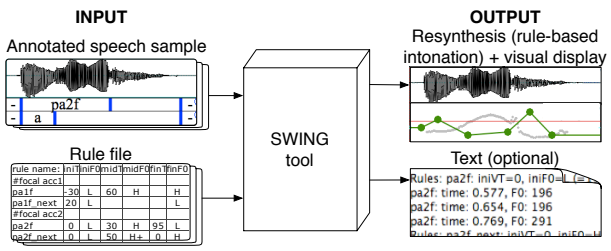

Figure 5: *Schematic overview of the* SWING *tool.*

3.2.1. Speech material

The speech samples to be used with the tool are first manually annotated. Stressed syllables are labeled prosodically and the corresponding vowels are transcribed orthographically. Table 2 shows the prosodic labels used in the current version of the tool, while Figure 6 displays an example utterance with prosodic annotation: *De' på kvällarna som vi sänder* 'It's in the evenings that we are transmitting'.

Table 2: *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 focal accent 1
pa2f	focal focal accent 2

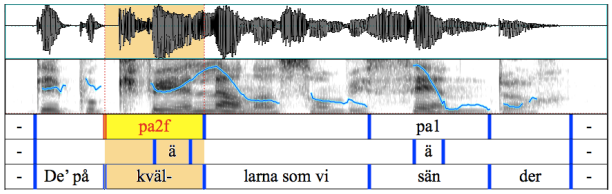

Figure 6: *Example of an annotated input speech sample.*

3.2.2. 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 contains the rule names, and the following columns contain three pairs of values, 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 is expressed as a percentage into the stressed syllable, starting from the onset of the stressed vowel. If no value is explicitly stated in the rule, the point is aligned with the beginning 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 can 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 3.

3.3. 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

Table 3: *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

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 points with the Praat built-in *Manipulation* feature. Figure 7 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 moving one or several pitch points) and the annotation files. The rule files can be adjusted in any text editor.

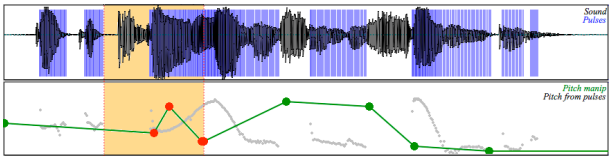

Figure 7: *Praat Manipulation display of a South Swedish utterance with rule-generated Svea intonation (circles connected by solid line; original pitch: light-grey line).*

3.4. 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 rule-based one generated by the tool. Figure 8 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.

3.5. 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.

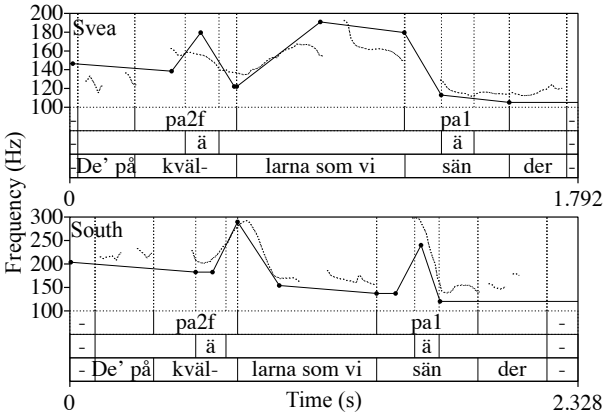

Figure 8: *Original and rule-based intonation of the utterance* De' på kvällarna som vi sänder *'It's in the evenings that we are transmitting' for Svea and South Swedish (original pitch: dotted line; rule-generated pitch: circles connected with solid line).*

4. Acknowledgment

This work is supported by a grant from the Swedish Research Council.

5. References

[1] Bruce, G.; Granström, B.; Schötz, S., 2007. Simulating Intonational Varieties of Swedish. *Proc. of ICPhS XVI*, Saarbrücken, Germany.

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

[3] Bruce, G.; Granström, B., 1993. Prosodic modelling in Swedish speech synthesis. A prosodic typology for Swedish dialects. *Speech Communication* 13, 63-73.

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

[5] Engstrand, O.; Bannert, R.; Bruce, G.; Elert, C-C.; 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.

[6] Elenius, K., 1999. Two Swedish SpeechDat databases - some experiences and results. *Proc. of Eurospeech 99*, 2243-2246.

[7] Elert, C.-C., 1994. Indelning och gränser inom området för den nu talade svenskan - en aktuell dialektografi. In *Kulturgränser - myt eller verklighet.*, Edlund, L.E. (Ed.). Umeå, Sweden: Diabas, 215-228.

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