Modelling Intonation in Varieties of Swedish

Gösta Bruce[1], Susanne Schötz[1], Björn Granström[2], Laura Enflo[2]

[1]Dept. of Linguistics and Phonetics, Centre for Languages and Literature, Lund University
[2]Dept. of Speech, Music and Hearing, School of Computer Science and Communication, KTH
`gosta.bruce@ling.lu.se, susanne.schotz@ling.lu.se, bjorn@speech.kth.se, lenflo@kth.se`

Abstract

The research project Simulating intonational varieties of Swedish (SIMULEKT) aims to gain more precise and thorough knowledge about some major regional varieties of Swedish: South, Göta, Svea, Gotland, Dala, North, and Finland Swedish. In this research effort, the Swedish prosody model and different forms of speech synthesis play a prominent role. The two speech databases SweDia 2000 and SpeechDat constitute our main material for analysis. As a first test case for our prosody model, we compared Svea and North Swedish intonation in a pilot production-oriented perception test. Naïve Swedish listeners were asked to identify the most Svea and North sounding stimuli. Results showed that listeners can differentiate between the two varieties from intonation only. They also provided information on how intonational parameters affect listeners' impression of Swedish varieties. All this indicates that our experimental method can be used to test perception of different regional varieties of Swedish.

1. Introduction

1.1. Background

Simulating Intonational Varieties of Swedish (SIMULEKT), funded by the Swedish Research Council, is a cooperation project between Phonetics at Lund University and Speech Communication at KTH, Stockholm. Our object of study is the prosodic variation characteristic of different regions of the Swedish-speaking area. It is apparent that prosody is a fundamental constituent of the different, native accents characterizing the regional varieties or dialect groups of Swedish [2, 5]: South, Göta, Svea, with Dala as a distinct subgroup, Gotland, North, and Finland Swedish (see Figure 1).

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

We have a fair degree of knowledge about the prosody of some of these regional varieties, like Svea and South Swedish, while other varieties like North and Finland Swedish have been much less studied. Moreover, we know more about the prosodic variation at the word level, such as pitch patterns of word accents, and markedly less about the phrase and utterance prosody of these varieties. Thus we regard it as an important research task to study in more detail the prosody and intonation of the main regional varieties of Swedish. In this study, the Swedish prosody model has a prominent role to play. A basic idea of our project work is to employ simulation of prosody by means of speech synthesis in different forms in order to be able to understand better the prosodic variation of Swedish. More information about the project work is to be found in [4].

1.2. The Swedish prosody model

The Swedish prosody model [2, 3] has been influential in the development of intonational phonology [11, 9, 10]. Originally developed as a prosodic typology for Swedish dialects, the current version of the model is described in [5]. The main parameters are for word prosody 1) word accent timing, i.e. the 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.

In our project work we start from the following broad assumptions about how the parameter values of the model are distributed among intonational varieties. Varieties of Swedish divide into two distinct groups depending on whether or not they make a clear distinction between focal and non-focal accentuation. One group (Svea, Göta) makes such a clear distinction (two levels of intonational prominence), while another group (South, Gotland-Dala, North, Finland Swedish) has more of equal prominence between successive accented words of a phrase (one level of intonational prominence). The first group is also characterized by early – while the second group has relatively late – word accent timing. Pitch patterns of compound words divide the Swedish varieties differently. Some varieties (Svea, Gotland-Dala, North) make an intonational distinction between compound and simplex words, while the other varieties (South, Göta, Finland Swedish) do not make such a distinction. Finally we suggest that concatenation patterns between prominent accents of a phrase display systematic differences between intonational varieties of Swedish. We can identify four principally distinct ways of concatenation or interpolation: high plateau, low plateau, upslope, and downslope. The major split of intonational varieties of Swedish conforms to the classical division into 'single-peaked' (South, Gotland-Dala, North, Finland Swedish) and 'double-peaked' dialect types (Svea, Göta)

[8]. It also strengthens the view that Svea and Göta varieties form a coherent geo-linguistic unit, while the other varieties fall outside the central axis.

2. Speech material

Our main sources for analysis in the project are the speech database collected in the research project SweDia 2000 and the telephone database SpeechDat.

Within the project The phonetics and phonology of the Swedish dialects around the year 2000 (SweDia 2000), supported by the Bank of Sweden for the period 1998-2003, a fairly large speech material was collected intended mainly for phonetic research [7]. The database contains recordings from more than 100 different places in Sweden and Swedish-speaking parts of Finland with minimally 12 speakers from each place (younger and elderly men and women). From this database a minor, elicited prosody material and primarily the extensive spontaneous speech material is relevant for our project work.

The SpeechDat database [6] comprises 5000 speakers of Swedish recorded over the telephone network. It was collected and annotated during the years 1997 and 1998 at the Department of Speech, Music and Hearing, KTH as a part of a European project sponsored by the EU. SpeechDat contains a wide range of utterances, from numbers, other mono- or polysyllabic words to complete sentences. The primary aim of this database is to serve as reference data for speech recognition, but all speakers also recorded two extra utterances, which were specifically designed for phonetic and prosodic analysis. One special feature of these sentences is that they contain a number of prosodically varied compound words, whose pitch patterns are assumed to be particularly revealing of intonational variety and thus dialect identity in Swedish. The two sentences are 1) *Flyget, tåget och bilbranschen tävlar om lönsamhet och folkets gunst.* 'Airlines, train companies and the car business are competing for profitability and the people's support.' and 2) *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.'. As an illustration of the dialectal variation of pitch pattterns in compounds, Figure 2 shows pitch contours of the word *företagare* /ˈfœːreˌtɑːɡare/ 'business people' representing four regional varieties.
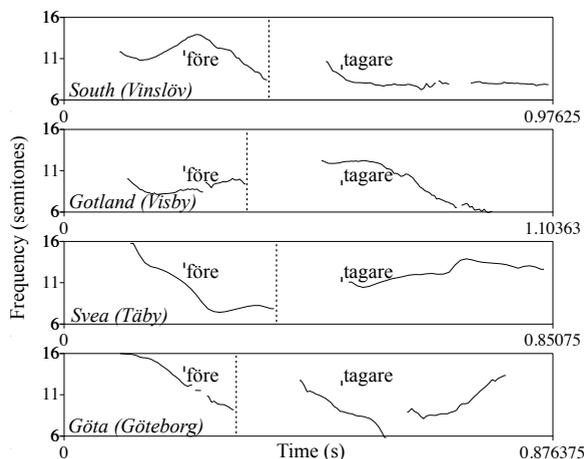

Figure 2: *Pitch contours of the word* företagare *'business people' (extracted from test sentence no. 2) of four regional varieties (female speakers, age 35-58, mean age 45).*

Compound words (and certain types of derivatives) in Swedish typically have two stressed syllables with main stress occurring in the first element and secondary stress in the final element of the word. The sample word *företagare* can be divided into the two elements *före* and *tagare*. Main stress is on the first syllable of the first element, and the first syllable of the second element receives secondary stress. The word has accent II, which is true of most Swedish compounds.

The traditional classification into single-peaked and double-peaked dialect types can be observed also in a compound word. Thus the 'single-peaked' South and Gotland varieties both have one tonal culmination point, albeit with a difference in alignment, while the 'double-peaked' Svea and Göta varieties instead have two culmination points with a corresponding difference between the these two varieties.

3. Comparing Svea and North Swedish intonation

An interesting test case for our prosody model is the modeling of intonation in the two regional varieties of Svea and North Swedish. In terms of the citation forms of the word accents (accent I and accent II) the intonation of the two varieties of Swedish appears to be similar. Examples of such pitch patterns of disyllabic words with initial stress occurring in a focal, phrase-final position (basically equivalent to citation forms) are shown in Figure 3. In Gårding's early accent typology [8], which was based on pitch contours of citation forms from Meyer's original pioneer work [12], these pitch patterns would be classified as belonging to the same accent type (2A) with accent I having a single pitch peak in the middle and accent II having two pitch peaks, one in each syllable. In our subsequent prosody modeling [2] the Svea pitch patterns could be decomposed into features of word prosody and utterance prosody. Thus the double-peaked contour of accent II and the single-peaked contour of accent I represent the focal realization of the word accents, while in a non-focal position for accent II only the first pitch peak and for accent I a mere low pitch (preceded by a high in a non-final position) are left.
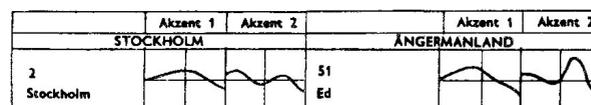

Figure 3: *Schematic pitch patterns of the citation forms of word accents in Svea (left) and North (right) Swedish (from [13]).*

Our general impression is, however, that intonation in the regional varieties of Svea and North Swedish is not at all the same and should be modeled differently. We assume that the composition of the pitch contours of accented words in the two varieties of Swedish is fundamentally distinct in a principally interesting way, something which is not immediately apparent from pitch patterns of word accents in the citation forms.

Our account of North Swedish intonation indicates that, unlike Svea, North Swedish does not exploit a regular distinction between focal and non-focal accentuation. In addition, the typical concatenation pattern between two prominent words (at some distance from each other in a phrase) in North is not a high plateau as in Svea but instead a downslope. This means that when we take aspects of utterance prosody into account, the pitch contours in Figure 3, which are quite similar, should have different interpretations for the two varieties.

In North Swedish, the basic pitch pattern of accentuation appears to be a single pitch peak with a difference in alignment for the two word accents. For accent I the pitch peak is aligned with the stressed syllable, and for accent II the peak occurs in the post-stress syllable. The first part of the accent II contour is variable between a gentle fall (as in Figure 3) and a mere low pitch level through the stressed syllable. There is reason to believe that the pitch course here is contextually determined and is an effect of the concatenation pattern exploited. The pitch contours of the two word accents seem to be fairly stable across variation in prominence and phrase position. For more details on the interpretation of North Swedish intonation, see [5].

4. Perceptual pilot experiment

In order to test our ideas about differences in intonation between Svea and North Swedish, we devised a perceptual pilot experiment. For this experiment we used variation of the two parameters of our prosody model related to utterance prosody, namely intonational prominence levels and concatenation patterns.

The test utterance was of the following form: *De' e' på kvällarna som vi dansar* 'It's in the evenings that we are dancing'. It had a syntactic structure (cleft sentence, topic fronting) eliciting the word *kvällarna* in semantic focus. The prosodically critical features of the utterance were two prominent accent II words (*kvällarna*, *dansar*) belonging to the same intonational phrase with a number of unstressed syllables both before the first stress and between the two stressed syllables. The segmental composition was also controlled to avoid consonant and vowel qualities of the test words that would be particularly revealing of regional variety. In order to make sure that the segmental make-up of the test utterance would still not bias the perceptual testing we also used a reiterant version of the test utterance with only *la* syllables.

A young male native speaker representing the Svea variety recorded the test material, both the natural version with real words and the reiterant version. In our modeling of Svea and North Swedish intonation, the pitch contours of the test utterance appear as in Figure 4. In terms of our model parameters, the Svea version has two distinct intonational prominence levels, expressed as focal accentuation on the first accented word (fall-rise = two pitch peaks) and non-focal accentuation on the second accented word (mere fall = one peak), as well as high plateau concatenation between these two words. The North Swedish version has more of equal prominence between the two accented words with basically the same accentual gesture (one pitch peak, late timing), and a downslope concatenation between these words.
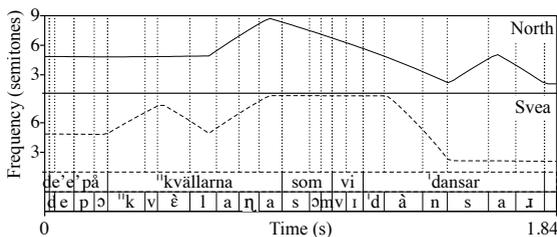

Figure 4: *Stylized pitch contours of the test utterance for the Svea (dashed) and North (solid) versions (see text).*

In our testing and manipulation, the pitch values of Points 1 and 3 (see Figure 5) represented the model parameter *intonational*

prominence levels. These two points were balanced against each other, so that increasing the pitch value of Point 1 would result in a corresponding decrease of the value of Point 3 and vice versa. Thus in the one extreme case, a maximally high value of Point 1 and consequently a mere low pitch level of Point 3 in the given intonational context would represent two distinct prominence levels, typical for Svea. The other extreme case with a mere low pitch level of Point 1 and a maximally high value of Point 3 in the same context would simulate equal prominence between the two accented words of the utterance, often observed in North. The other model parameter *concatenation patterns* was controlled by changing the pitch value of Point 2 (see Figure 5) by going from a maximally high value resulting in a high plateau (Svea) to a correspondingly low value conforming to a downslope concatenation (North). We also used a wider range of the contours than in the original version of the utterance by increasing all maximum points of the contours. The effect of this manipulation was a paralinguistically more involved rendering of the utterance, and it was deemed to facilitate the task of the listeners.

4.1. Test design

The experimental method can be characterized as a production-oriented perception test. We developed a small computer program (in Praat script [1] and Java), in which subjects were asked to click with the mouse in a square on the computer screen to play stimuli with varying intonation. The task was to identify which stimulus sounded the most like the Svea and North Swedish varieties, and which sounded the most ambiguous, i.e. could belong to both varieties. Thus, the position (x and y values ranging from 1 to 100) of the clicks determined the intonation of the stimuli played. The parameter *intonational prominence levels* was controlled by the x-axis of the square, while the *concatenation patterns* parameter was controlled by the y-axis, as shown in Figure 5.
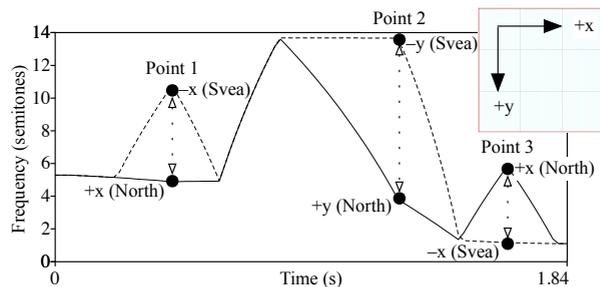

Figure 5: *Schematic pitch contours showing the variable points of the test utterance and how the position of the subjects' mouse clicks in the square (top right) of the pilot test controls the values of the pitch peaks of the stimuli; more Svea (dashed line) or more North Swedish (solid line) intonation.*

For instance, minimum values for all points (1,1) would play a typical Svea stimulus, while maximum values (100,100) would play a typical North one. Thus, our expectation was that listeners would place the cursor in the upper left part of the square for Svea and in the lower right part for North Swedish. We changed the position of origo (0,0) and the direction of the axes for the second half of the test, so that subjects who had managed to identify the positions of the two varieties in the first half would not be able to just copy their response patterns.

4.2. Procedure

Fourteen students (8 females, 6 males; mean age 24) at the Centre for Languages and Literature at Lund University took part in the experiment, which lasted about 10-15 minutes. All subjects wore headphones. First, we played one-minute samples of spontaneous Svea and North Swedish to refresh the subjects' memory of these varieties. Next, they were asked to fill out a form in the program with some information about themselves (sex, age, place of birth and childhood, dialect, familiarity with Svea and North Swedish, hearing). Written instructions were then given about the tasks of the test. In addition, we encouraged the subjects to first maximize the distance between successive clicks within the square in order to hear larger differences. The test comprised six parts with different tasks, as shown in Table 1.

Table 1: *The parts, stimuli and tasks of the perception test.*

Part	Stimuli	Identification task
1	real utterances	the most Svea Swedish
2	real utterances	the most North Swedish
3	real utterances	ambiguous (both Svea and North)
4	reiterant utterances	the most Svea Swedish
5	reiterant utterances	the most North Swedish
6	reiterant utterances	ambiguous (both Svea and North)

4.3. Results

We present here the preliminary results of the pilot experiment in outline. Most subjects were speakers of southern varieties of Swedish, and they had varying familiarity with North (9 poor, 5 rather good) and Svea (2 poor, 8 rather good, 4 very good) Swedish. The test results consisted of numerical pairs (x,y) of the positions which each subject had identified as the most likely candidate in each part of the test. Figure 6 displays the mean and median values of the six parts of the test. Despite a rather large dispersion of the results, the positions of the tasks to identify the most Svea sounding intonation tend to be closer to (1,1), the positions for the most North sounding variant are closer to (100,100), while the positions for the ambiguous one tend to be around (50,50). Median values show larger differences between the different tasks than mean values. Similar tendencies can be observed for both real and reiterant stimuli. A repeated-measures ANOVA (two factors: version (2 levels) and dialect (3 levels)) showed significant effects only for dialect (for x: $F(2,26)=5.591$, $p<.05$; for y: $F(2,26)=20.263$, $p<.05$). The results are basically in agreement with our general expectation.
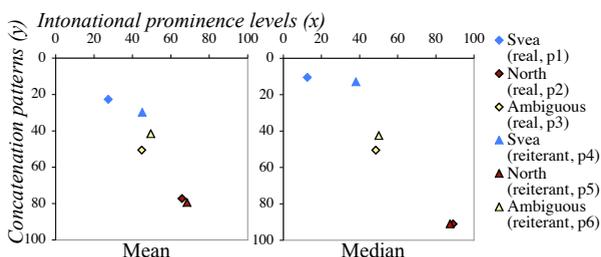

Figure 6: *Results of the perception test. Mean and median values for all subjects and all parts of the test.*

4.4. Discussion

The results of this pilot test are quite encouraging. First, they indicate that the experimental method involving a production-oriented perception test is viable for testing differences between intonational varieties of a language. We therefore intend to develop our method further and also include other combinations of model parameters in order to test listeners' ability to distinguish other intonational varieties of Swedish. Second, the pilot experiment provided information on how intonational parameters affect listeners' perception of Swedish varieties. In our interpretation, both parameters *concatenation patterns* and *intonational prominence levels* contribute to the impression of sounding either Svea or North Swedish. The fairly large dispersion of the results may be explained by the fact that most subjects were from south Sweden with varying familiarity with the North and Svea varieties. In conclusion then, the outcome of the test indicates that parameters of our prosody model can be decisive in differentiating between intonational varieties of Swedish.

5. References

[1] Boersma, P.; Weenink, D., 2007. *Praat: doing phonetics by computer (version 4.6.17)* [computer program]. http://www.praat.org/, visited 25-Aug-07.

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

[5] 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.

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

[7] 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.

[8] Gårding, E.; Lindblad, P., 1973. Constancy and variation in Swedish word accent patterns. In *Working Papers* 7, Phonetics Laboratory, Lund University, 36-110.

[9] Gussenhoven, C., 2004. *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.

[10] Jun, S-A., 2005. *Prosodic Typology*. Oxford: Oxford University Press.

[11] Ladd, D. R., 1996. *Intonational phonology*. Cambridge: Cambridge University Press.

[12] Meyer, E. A., 1937. Die Intonation im Schwedischen I: Die Sveamundarten. *Studies Scand. Philol.* 10. Stockholm University.

[13] Öhman, S., 1967. Word and sentence intonation: A quantitative model. *KTH, STL-QPSR* 2, 20-54.