

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/262262676>

Regional Varieties of Swedish: Models and synthesis

Chapter · September 2012

CITATION

1

READS

53

6 authors, including:



Susanne Schötz
Lund University

63 PUBLICATIONS 288 CITATIONS

[SEE PROFILE](#)



My Segerup
Lund University

7 PUBLICATIONS 11 CITATIONS

[SEE PROFILE](#)



Jonas Beskow
KTH Royal Institute of Technology

186 PUBLICATIONS 2,378 CITATIONS

[SEE PROFILE](#)



Joakim Gustafson
KTH Royal Institute of Technology

154 PUBLICATIONS 1,293 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Melody in Human–Cat Communication (Meowsic) [View project](#)



Estonian Swedish Language Structure (ESST - Estlandssvenskans språkstruktur) [View project](#)

Susanne Schötz, Gösta Bruce, My Segerup (Lund),
Jonas Beskow, Joakim Gustafson &
Björn Granström (Stockholm)

Regional varieties of Swedish: Models and synthesis

1 Introduction

The Swedish prosody model (Bruce and Gårding 1978; Bruce and Granström 1993; Bruce 2007) was developed in close interaction with intonational phonology, and has become well-known on the international research scene. In the research project SIMULEKT (Simulating Intonational Varieties of Swedish, funded by the Swedish Research Council 2007-2010, see e.g., Bruce et al. 2007) our object of study is the prosodic characteristics of different regions of the Swedish-speaking area. Various forms of speech synthesis and the Swedish prosody model play prominent roles in our work. Our main sources for analysis are three Swedish speech databases further described in section 2. The primary goal of the project is to produce more precise and thorough knowledge about phrase and utterance prosody of the major regional varieties of Swedish in global as well as local prosodic contexts. The significance of our project work is both within basic research and speech technology applications. Thus our project is expected to contribute to deepening our understanding of prosodic structuring of spoken language generally, and more specifically of dialectal variation of Swedish. It is also generating new knowledge about prosody to be used in speech synthesis systems.

The main parameters for the Swedish prosody model (Bruce and Gårding 1978; Bruce and Granström 1993; Bruce 2007) 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. Thus, the specific timing of pitch patterns of word accents in relation to stressed syllables (both primary and secondary), i.e. in simplex as well as in compound words, is considered to be a potential distinctive feature for the classification of different intonational varieties. In the SIMULEKT project major emphasis has been given to postlexical prosody, specifically phrase and utterance intonation, assuming that intonational prominence levels and tonal concatenation patterns are particularly revealing for the impression

and identification of different native accents of Swedish. We have also been emphasising patterns of tonal concatenation, both as coherence patterns between prominent words – high/low plateau, rising and falling slopes – and as boundary signals. These patterns appear to contribute to regional differentiation between distinct varieties of Swedish.

In this article we summarise our latest approaches involving modelling and synthesis of Swedish regional varieties by means of: (1) analysing and modelling intonation of compound words, (2) data-driven exploration of Swedish word accents, (3) a HMM synthesis approach for simulating dialects, where speech is generated from emphasis tagged text, and (4) experiments with a hybrid rule-based and HMM synthesiser of intonation.

2 Dialectal speech databases

Our main sources for analysis are the three Swedish speech databases *SweDia 2000*, *SpeechDat* and *NST*. *SweDia 2000* (Engstrand et al. 1997) comprises read, elicited and spontaneous Swedish dialectal speech from 12 speakers (younger and elderly men and women) each from over 100 different places in Sweden and Finland. From this database the elicited prosody material and the extensive spontaneous interviews are relevant for our project work. *SpeechDat* (Elenius 1999) contains speech recorded over the telephone from 5000 speakers and was used in our analysis and rule-based resynthesis. For our HMM synthesis approaches we selected Swedish speech material from the *NST* database (Språkbanken 2010), which includes several regional varieties of Swedish. The large speech synthesis database from a professional speaker of standard Swedish also used in the HMM approach was recorded as part of the *NST* (Nordisk Språkteknologi 'Nordic Language Technology') synthesis development approach.

3 Analysing and modelling intonation of Swedish compounds

The Swedish dialectal tonal variation has been well documented for simple words; see classical phonetic studies by, e.g., Meyer (1937) and Gårding (1977). Figure 1 shows typical pitch patterns of the five Swedish word accent types for accent I and II according to Gårding, as well as a map of the seven main intonational varieties of Swedish, corresponding to our present dialect classification scheme. Accent type 0 has the same pitch pattern for both word accents, 1A and 1B display single peaked patterns

for both accent I and II (with an earlier peak in 1A compared to 1B), while types 2A and 2B are characterised by double-peaked pitch patterns in accent II (with earlier peaks in 2A than 2B).

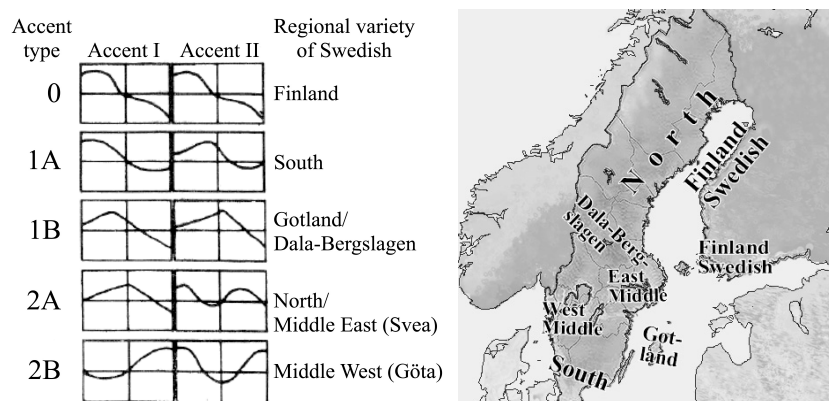


Figure 1: The five Swedish word accent types according to Gårding (1977) and approximate geographical distribution of the seven main regional varieties of Swedish.

There is reason to believe that the intonation of compound words is particularly revealing for the impression of Swedish dialectal identity. Words in all varieties of Swedish have a rhythmic structure where the word stress is not fixed to certain syllable positions. In addition to stress, Swedish has a word accent distinction (accent I - accent II) with a primarily tonal phonetic realisation. Simplex words have one stress, compounds two; primary stress on the first element, and secondary stress on the final one. Compounds are by default realised with accent II, except in South Swedish, where the word accent of the first element may be retained (see Bruce 1998) and in Finland Swedish where no accent distinction is maintained. In some regional varieties of Swedish (East Middle, Gotland, Dala-Bergslagen, North) there is an additional difference between simplex and compound words in the intonational pattern, related mainly to the accent II contour. The pitch gesture is synchronised with both primary and secondary stress. In longer compounds this creates some distance between the first and second parts of the gesture. In other varieties of Swedish (South, West Middle) there is no tonal synchronisation with the secondary stress. Riad (1998) and Bruce (2001, 2007) suggest typologies of the tonal patterns of Swedish compounds. A study of compound intonation in regional varieties of Swedish is described in Schötz et al. (2009).

3.1 Material and method

The SpeechDat database contained one token per speaker of two particularly interesting read sentences to our project: *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.) and *Flyget, tåget och bilbranschen tävlar om lönsamhet och folkets gunst.* (Airlines, train companies and the automobile industry are competing for profitability and people's appreciation.). Our object of study here were two compounds which typically appear in focal phrase-final position. *'Bil, branschen* (the car industry) represents compounds with adjacent primary and secondary stress, and *mo' biltel, fonen* (the mobile phone) compounds with longer distance between the stresses. To get a more detailed picture of some of the varieties, we further divided South, Dala-Bergslagen and West Middle Swedish into sub-varieties, making a total of 11 sub-varieties.

SpeechDat is unbalanced for dialect type. Our ambition to select 25 representative speakers for the two compounds from each regional variety was not reached. In addition, auditive analysis guided by spectrograms revealed samples containing creaky voice, which also were excluded from our analysis. We are currently recording additional speakers to get a more regionally balanced material. Table 1 shows the number of speakers (one token per speaker) for the two compounds in each sub-variety.

Regional variety (sub-variety) of Swedish	<i>bilbranschen</i>			<i>mobiletelefonen</i>		
	♀	♂	total	♀	♂	total
Finland (Nyland)	7	2	9	6	1	7
South (Southern)	11	13	24	8	14	22
South (North-eastern)	12	8	20	12	9	21
South (Småland)	8	12	20	8	13	21
Gotland	10	13	23	11	14	25
Dala-Bergslagen (Lower)	6	13	19	6	14	20
Dala-Bergslagen (Upper)	5	6	11	5	4	9
North (West Bothnia)	9	14	23	10	12	22
East Middle (Stockholm)	11	14	25	9	16	15
West Middle (Gothenburg)	9	14	23	12	13	25
West Middle (Värmland)	13	8	21	7	6	13

Table 1: Number of speakers for the two analysed words *bilbranschen* and *mobiletelefonen* in the eleven analysed sub-varieties of Swedish.

The speech material was segmented and analysed using PRAAT (Boersma and Weenink 2010). Segment boundaries were placed at the onset of every stressed vowel, and individual F0 contours were generated for every speaker and word. In order to facilitate comparison of speakers, the con-

tours were normalised for F0 by setting the minimum F0 for every speaker to 0 semitones. Mean F0 contours were calculated for each variety by calculating the mean F0 in 100 evenly distributed points per word – the number of points assigned to each segment based on the duration. The contours were plotted on a semitone scale, and compared with respect to tonal similarities and differences as well as geographic vicinity and accent type. Based on our results, we adjusted the parameters for compounds of the Swedish prosody model. Finally, we tested our parameters using rule-based resynthesis with our analysis tool SWING (Swedish INtonation Generator, see Schötz et al. 2010).

3.2 Results

Finland Swedish (type 0) and the three South sub-varieties (type 1A) all display single-peaked F0 patterns, but with different timing of the peak in relation to the primary stressed syllable. The earliest peak is found in the typical type 0 accent pattern, i.e. no word accent distinction, of Finland Swedish. Southern South Swedish has a later accent II peak and following fall compared to North-eastern and Småland South Swedish, which have similar accent II patterns. Gotland and Lower Dala-Bergslagen Swedish show high plateaux pitch patterns, while the contour of Upper Dala-Bergslagen resembles North-eastern and Småland South Swedish, though with lower pitch range. Furthermore, in addition to the high plateau, Gotland seems to have an extra tonal upglide in the secondary stressed syllable, while Lower Dala-Bergslagen Swedish has a more evenly shaped high plateau. North and East Middle Swedish (type 2A) display similar double-peaked pitch patterns, with an early peak in the primary stressed syllable and a later peak associated with the secondary stressed syllable. East Middle Swedish seems, however, to have a slightly larger pitch range than North Swedish. The two West Middle Swedish sub-varieties (type 2B) show almost identical double-peaked patterns: the first peak is aligned with the primary stressed syllable, but has later timing than in type 2A; the second peak appears in word final position.

Figure 3 shows mean F0 contours for the eleven analysed sub-varieties in the compound *mobiltelenen*. Finland Swedish (type 0) and Southern South (type 1A) Swedish display single-peaked patterns with early tonal peaks close to the primary stressed syllable. The Finland Swedish pattern is the typical type 0 pattern, while Southern South Swedish shows the typical type 1A pattern for accent I. In comparison, both North-eastern and Småland South Swedish have typical type 1A accent II patterns with later peaks. Gotland and lower Dala-Bergslagen Swedish

(type 1B) show high plateau patterns. In Gotland Swedish, there is an additional final upstep aligned with the secondary stressed syllable, while the plateau is more level in lower Dala-Bergslagen. Upper Dala-Bergslagen, however, displays a single-peaked pattern similar to North-eastern and Småland South Swedish. Similar double-peaked patterns with the first peak aligned with the primary stress and the second with the secondary stress are found in North and East Middle Swedish (type 2A). The difference between these two sub-varieties seems to be a slightly larger pitch range in East Middle Swedish. The two West Middle Swedish sub-varieties (type 2B) display close to identical double-peaked F₀ patterns with one peak aligned with the primary stressed syllable, and another peak at the end of the word. Mean F₀ contours for *bilbranschen* are shown in Figure 2.

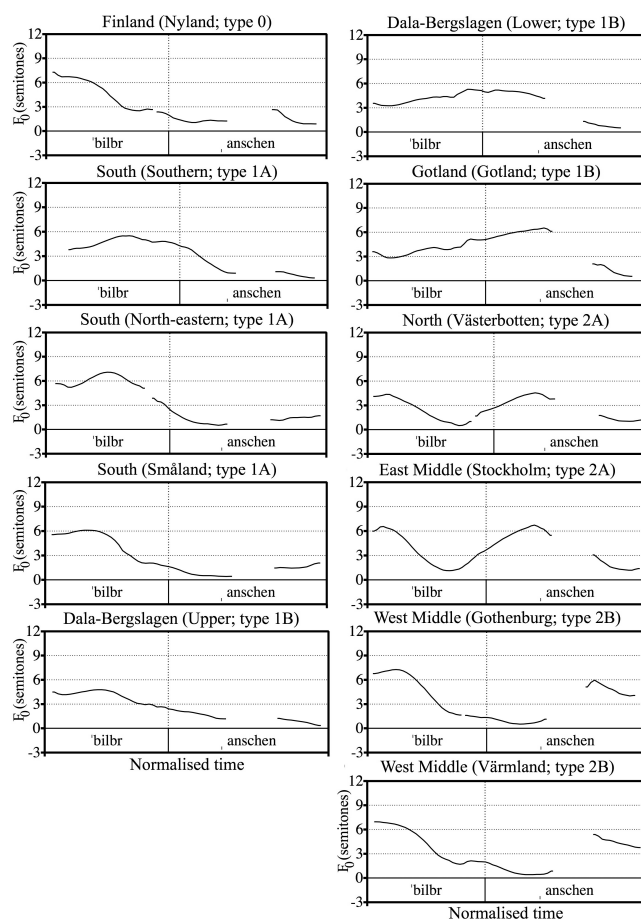


Figure 2: Mean F₀ contours for the compound *bilbranschen* in eleven varieties of Swedish (see Table 1 for information about the sample size of each variety).

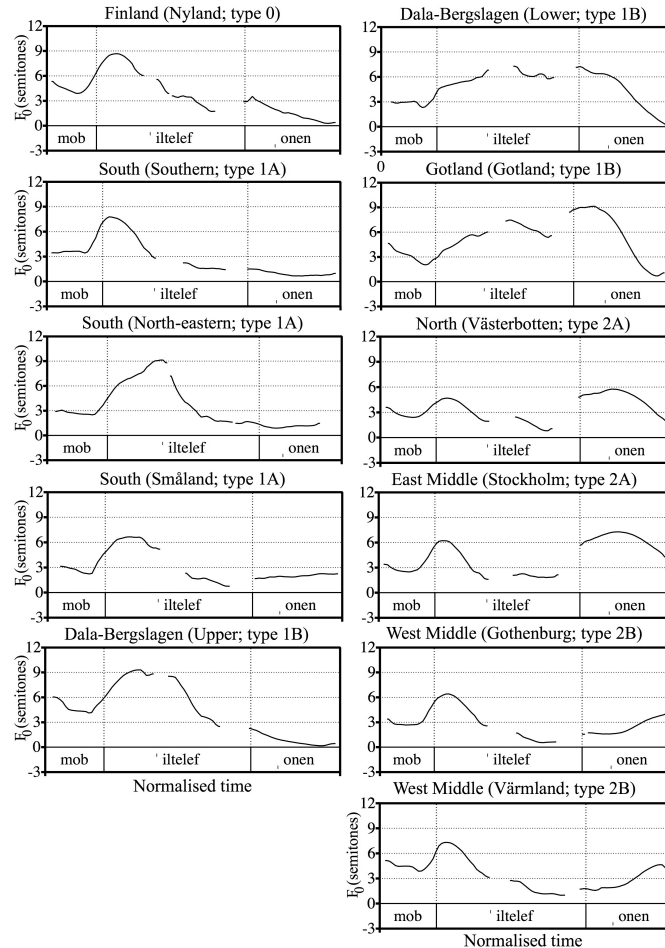


Figure 3: Mean F0 contours for the compound *mobiltelefonen* in eleven varieties of Swedish (see Table 1 for information about the sample size of each variety)

3.3 Testing the intonational compound parameters with SWING

After having adjusted the Swedish Prosody model parameters for compounds according to the results of our analyses, we tested them using SWING (Schötz et al. 2010), an analysis tool developed within the project. It comprises several parts joined by the speech analysis software PRAAT (Boersma and Weenink, 2010), which also serves as graphical interface. Using an input annotated speech sample and an input rule file, SWING generates and plays PSOLA resynthesis – with rule-based and speaker-

normalised intonation – of the input speech sample. We then used SWING to compare the model and original natural intonation in the regional varieties. If the model works adequately, there should be a close match between the original and the rule-based (SWING-generated) F0 contours. Figure 4 shows six examples of original and resynthesised intonation for the word *mobilteléfono* with SWING.

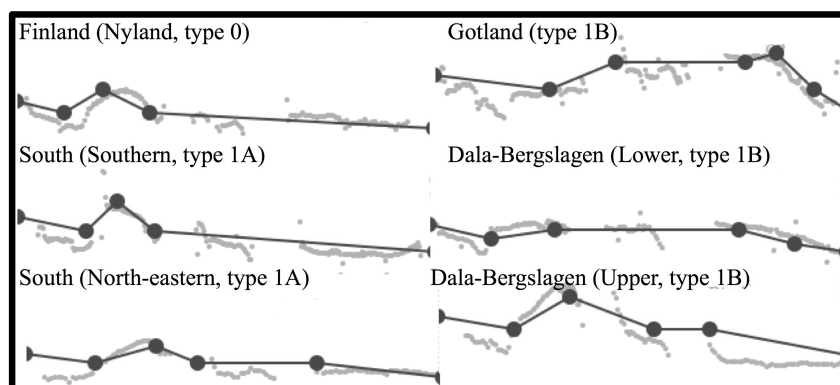


Figure 4: Comparison of modelled and original pitch of the compound *mobilteléfono* 'the mobile phone' with SWING. PRAAT Manipulation displays of the six dialect regions Finland, Southern and North-eastern South, Gotland, Lower and Upper Dala-Bergslagen Swedish (modelled rule-based resynthesised pitch: circles connected by solid line; original pitch: dotted line).

As can be seen there is a fairly close match between the original and resynthesised pitch contours. An informal evaluation by two of the authors confirmed the auditory similarity between the original and resynthesis versions of each variety.

3.4 Discussion

The results from our intonation analysis of compound words show both similarities and differences within as well as between the regional sub-varieties. Finland Swedish (type 0) with no word accent distinction has a pitch pattern that is similar to that of South and Upper Dala-Bergslagen Swedish. South Swedish (1A) can be divided into two sub-varieties; Northern (with an early pitch gesture for accent II) and Southern (with retained first element accent and a later tonal gesture for accent II). Moreover, accent type 1A seems to include also Upper Dala-Bergslagen, which for simplex words is type 1B. Accent type 1B is characterised by a high plateau in both Gotland and Lower Dala-Bergslagen Swedish. However, while the plateau is fairly level in Lower Dala-Bergslagen Swedish, Gotland Swedish displays an additional final upstep in pitch in both of the

two compounds analysed here. East Middle and North Swedish retain accent type 2A for compound words. The pitch gestures in these two varieties are similar, but show a somewhat larger pitch gesture in East Middle Swedish compared to North Swedish. For the two West Middle Swedish sub-varieties Gothenburg and Värmland Swedish, the pitch patterns seem to be indistinguishable.

In several varieties of Swedish, pitch gestures vary in simplex and compound words (cf. Bruce 2010). In contrast to the simplex words, the compounds display a number of differences, and may therefore be even more revealing for the identification of different native accents of Swedish. Secondary stress as a tonal association may be critical in varieties belonging to accent type 1B and 2A, but unimportant in varieties belonging to 1A and 2B.

The adjusted Swedish prosody model parameters for compounds seem to work well when testing them with our analysis tool SWING. However, additional testing and fine-tuning of our model is still needed. Moreover, the number of regional sub-varieties needs to be increased in order for us to obtain a more detailed picture of the sub-regional variation. We have already begun recording additional speakers of the two Finland Swedish varieties Northern (Ostrobothnian) and Southern (Åland, Åboland, Nyland) Finland Swedish, and hope to add new sub-varieties to our model in the near future.

4 Data-driven exploration of Swedish word accents

As an alternative to the analysis based rule description of regional intonation patterns we wanted to try an automatic method. In a data-driven synthesis experiment we used data from the Swedish NST Database, which consists of phonetically balanced broadband recordings of 1000 speakers balanced for dialect, with 20-60 minutes of speech per speaker. We selected 15 speakers representative of the six dialect groups South, Gotland, Dala-Bergslagen, North, East Middle and West Middle Swedish. Though not prosodically balanced, the corpus still contains different types of sentences that ensure prosodic variation, e.g., statements, wh-questions, yes/no questions and enumerations. The corpus was aligned on the phonetic and word levels using the Nalign software (Sjölander and Heldner 2004) with the NST dictionary as pronunciation dictionary, which comprises more than 900.000 phonetically transcribed items with syllable boundaries marked. In addition, the text was tagged for part-of-speech using a TNT tagger trained on the SUC corpus (Megyesi 2002). We selected approximately 1000 content words, ranging from 1 to 5 syllables, with primary stress on the first syllable, from a large set of read utterances.

F0 contours were extracted, mean-normalised and temporally aligned according to vowel onset in the stressed syllable (marked with a vertical line in the figure). Pitch clouds of three accent types: accent I, accent II and compounds were generated for East Middle and South Swedish, see Figure 5.

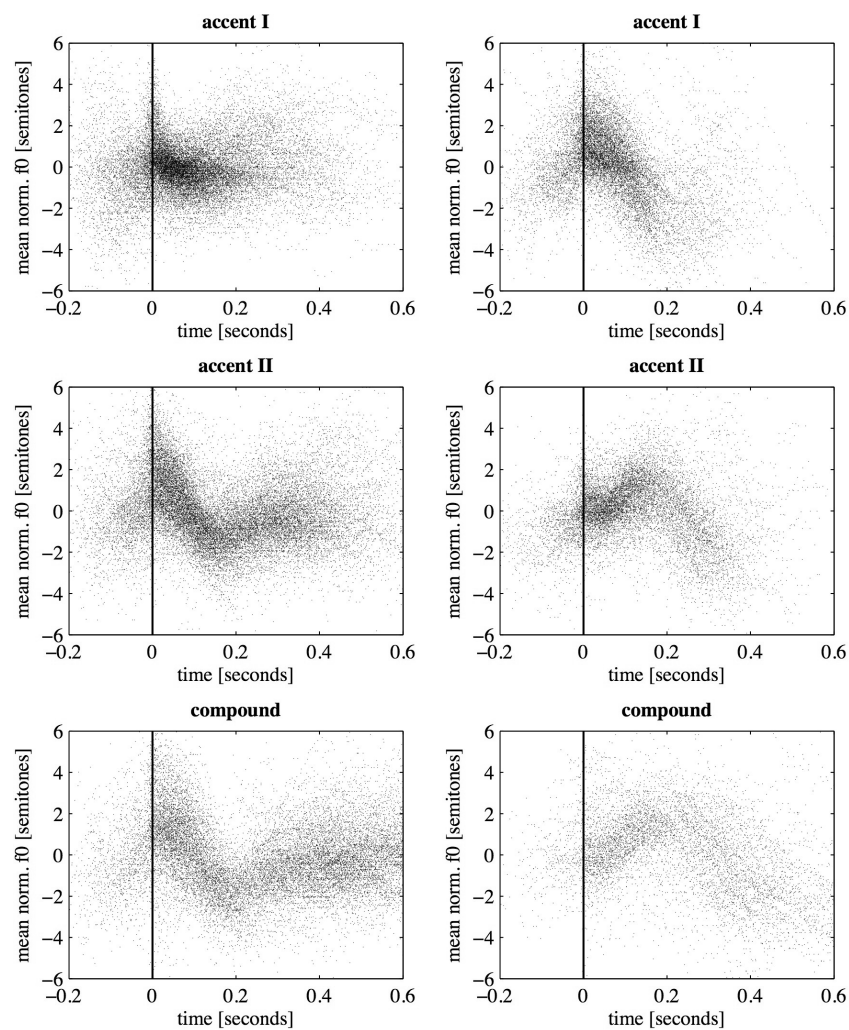


Figure 5: Pitch clouds for *East Middle* (left) and *South* (right) dialects. Vertical lines mark the alignment, i.e. the vowel onset of the stressed syllables.

Although the pitch clouds display a considerable spread, the data-driven exploration produced clouds that clearly indicate the dialect difference in accent II and compounds, with two peaks in the *East Middle* case and a

single peak for *South*. For *South* it is clear that the temporal alignment of the peak is later in accent II than in accent I. It can be noted that the words represented in the different clouds typically have different length. There here is an overrepresentation of monosyllabic accent I words, since we selected only those with stress on the first syllable. Accent II words are primarily disyllabic, while a majority of the longer words are compounds containing both accent I and accent II realisations in the South Swedish dialect. This type of analysis gives insight into features that influence prosodic realisations, which is valuable both in HMM synthesis and for fine-tuning the SWING rules.

5 HMM synthesis experiments

During the last decade, most speech synthesisers have been based on prerecorded pieces of speech resulting in improved quality, but with lack of control in modifying prosodic patterns (Taylor 2009). The research focus has been directed towards how to optimally search and combine speech units of different lengths. A synthesis approach that has gained interest in recent years is HMM based synthesis (Tokuda et al. 2000). In this solution the generation of speech is based on a parametric representation, while the grapheme-to-phoneme conversion still relies on a large pronunciation dictionary. This approach has been successfully applied to a large number of languages, including Swedish (Lundgren 2005). HMM synthesis is an entirely data-driven approach to speech synthesis. As such it gains all its knowledge about segmental, intonational and durational variation in speech from training on an annotated speech corpus. Given that the appropriate features are annotated and made available to the training process, it is possible to synthesise speech with high quality at both segmental and prosodic levels. Another important feature of the HMM synthesis making it an interesting choice for studying dialectal variation is that it is possible to adapt a voice trained on a large data set (2-10 hours of speech) to a new speaker with only 15-30 minutes of transcribed speech (Watts et al. 2008). In this study we used 20-30 minutes of dialectal speech from the Swedish part of the NST corpus.

The typical HMM synthesis model can be decomposed into a number of distinct layers. At the acoustic level, a parametric source-filter model (MLSA-vocoder) is responsible for signal generation. Context dependent HMMs, containing probability distributions for the parameters and their 1st and 2nd order derivatives, are used for generation of control parameter trajectories. In order to select context dependent HMMs, a decision tree that uses input from a large feature set to cluster the HMM models was applied. In this study, we used the standard model for acoustic and

HMM level processing, and we focussed on adapting the feature set for the decision tree for the task of modelling dialectal variation. The feature set typically used in HMM synthesis includes features on segment, syllable, word, phrase and utterance level. Segment level features include immediate context and position in the syllable; syllable features include stress and position in the word and phrase; word features include emphasis, part-of-speech tag (content or function word), number of syllables, position in the phrase etc.; phrase features include phrase length in terms of syllables and words; utterance level includes length in syllables, words and phrases.

For our present experiments, we have also added a speaker level to the feature set, since we train a voice on multiple speakers. The only feature in this category at present is dialect group, which is one of *North*, *Dala-Bergslagen*, *East Middle*, *West Middle*, *Gotland* and *South*. In addition to this, we have chosen to add a morphological feature to the word level stating whether or not the word is a compound, since compound stress pattern is often a significant dialectal feature in Swedish (Bruce et al. 2007). At the syllable level we have added explicit information about lexical accent type (accent I, accent II). The resulting voices have clearly identifiable dialectal features. However, in order to investigate the effect of different intonational realisations, more fine-grained control is needed. The next chapter describes the integration of SWING rules into the HMM synthesiser.

6 Hybrid synthesis and evaluation

The SWING tool requires information about phoneme alignment, pitch range, syllable stress and accents. These features are all automatically generated in the HMM synthesis process, which makes it possible to use SWING rules to generate pitch contours automatically from an emphasis-tagged text, which in turn can be used to replace or supplement the HMM-generated pitch contours prior to sound rendering. Current work in our project concerns using the rules obtained with SWING to generate intonation for the main regional varieties of Swedish to be used by the HMM synthesiser. We have integrated SWING into the HMM synthesis framework, so that it may be driven by the same input features as the HMM synthesis, and the durations generated by the HMMs trained on a speaker of the target dialect. This makes it possible to replace or augment the HMM-generated F0 track by the one generated by SWING before rendering the waveform. As an example of how the new hybrid SWING/HMM synthesiser works, figure 6 shows the F0 tracks generated for two sentences using the South Swedish versions of SWING and HMM. In addition, it shows a mixed track that consists of a 10-to-90 blend between the

HMM and SWING F0 tracks. The motivation for producing a mixed version is that the SWING rules tend to produce flat pitch contours without any micro-prosodic movements in unstressed positions. Informal listening tests by the authors indicate that the mixed version retains the regional characteristics of the SWING-generated prosody but adds microvariations which results in more natural-sounding speech. This shows the feasibility of blending rule-based and data-driven pitch contours in the same synthesiser. The new hybrid SWING/HMM synthesiser will also allow more careful investigation of the SWING rules, since large sets of perceptual stimuli can be automatically generated under controlled conditions.

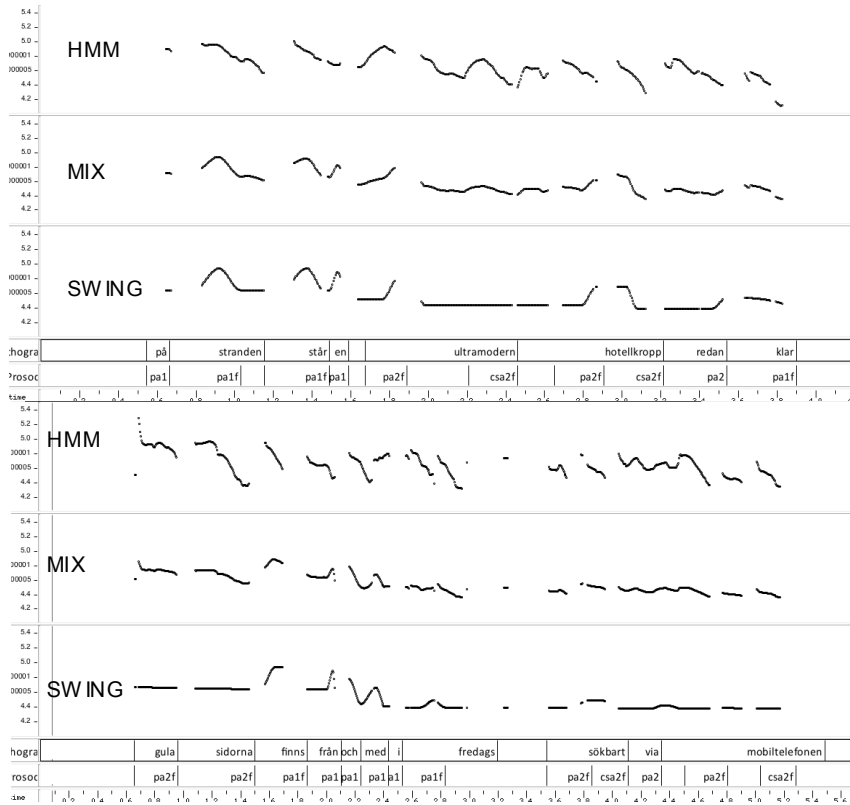


Figure 6: F0 tracks for two sentences generated using South Swedish versions of HMM, SWING and a 10-to-90 mix.

7 Concluding remarks and future work

Our approaches with analysis, modelling, rule-based resynthesis and HMM synthesis have added significantly to our knowledge about the intonation in regional varieties of Swedish. By combining our efforts in a hybrid synthesiser, we were able to successfully simulate intonation of at least one regional variety of Swedish. In future studies we plan to test the hybrid synthesiser with more regional varieties. Moreover, we are currently analysing and modelling regional variation in phrasing using the SpeechDat material. Our pilot studies of intonational phrasing in varieties of Swedish are described in Schötz and Bruce (2010).

This article describes the last few studies that we carried out together with our dear colleague and friend Professor Gösta Bruce. It is of great value to us, since Gösta passed away in June 2010 following a short illness. As his co-workers it has been important to be able to complete our last research work together by finishing this article. To have been colleagues and project members in SIMULEKT with Gösta Bruce has been a privilege, and we hope to be able to continue our work in his spirit.

8 References

- Boersma P and D. Weenink (2010): PRAAT: Doing phonetics by computer (version 5.1.41) [computer program]. *Webpage* <http://www.praat.org>, visited 6-April 2010.
- Bruce, G. (1998): *Allmän och svensk prosodi*. Department of Linguistics, Lund University.
- Bruce, G. (2001): Secondary stress and pitch accent synchronization in Swedish. In: W. van Dommelen, T. Fretheim (eds): *Nordic Prosody VIII* (pp. 33-44), Frankfurt am Main: Peter Lang.
- Bruce, G., B. Granström and S. Schötz (2007): Simulating Intonational Varieties of Swedish. *Proceedings of the 17th International Congress of Phonetic Sciences, Saarbrücken, Germany*.
- Bruce, G. (2007): Components of a prosodic typology of Swedish intonation. In: T. Riad, C. Gussenhoven (eds): *Tones and Tunes, Volume 1* (pp. 113-146). Berlin: Mouton de Gruyter.
- Engstrand O., R. Bannert, G. Bruce, C-C. Elert and A. Eriksson (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.
- Gårding, E. (1977): *The Scandinavian word accents*. Lund: Gleerup.
- Lundgren A. (2005): *HMM-baserad talsyntes*. Master's thesis, KTH, TMH, CTF.
- Megyesi B. (2002): *Data-Driven Syntactic Analysis - Methods and Applications for Swedish*. Ph.D. thesis, Department of Speech, Music and Hearing, KTH, Stockholm.
- Meyer, E. A. (1937): Die Intonation im Schwedischen. *Die Sveamundarten. Studies Scand. Philol.* **10**. Stockholm University, Sweden.

- Riad, T. (1998): Towards a Scandinavian accent typology. In: W. Kehrein, R. Wiese (eds): *Phonology and morphology of the Germanic languages* (pp. 77-109). Tübingen: Max Niemeyer.
- Schötz, S., G. Bruce and B. Granström (2009): Modelling compound intonation in Dala and Gotland Swedish. *Proceedings of Fonetik 2009, Department of Linguistics, Stockholm University, Sweden*.
- Schötz, S., J. Beskow, G. Bruce, B. Granström and J. Gustafsson (2010): Simulating Intonation in Regional Varieties of Swedish. *Proceedings of the 5th International Conference of Speech Prosody, Chicago, USA*.
- Schötz, S. and G. Bruce (2010): Phrase-initial pitch patterns in South Swedish. *Proceedings of the 5th International Conference of Speech Prosody, Chicago, USA*.
- Sjölander K and Heldner M (2004): Word level precision of the NALIGN automatic segmentation algorithm. *Proceedings of the XVIIth Swedish Phonetics Conference, Fonetik 2004, Stockholm University, Sweden*, 116–119.
- Språkbanken (2010): The Norwegian HLT resource collection. Securing NST's language resources. *Webpage <http://www.spraakbanken.nib.no/sits.page>*, visited 30-November-10.
- Taylor P. (2009): *Text-To-Speech Synthesis*. Cambridge: Cambridge University Press.
- Tokuda K, T. Yoshimura, T. Masuko, T. Kobayashi and T. Kitamura (2000): Speech parameter generation algorithms for HMM-based speech synthesis. *Proceedings of the International Conference of Acoustics, Speech and Signal Processing, Istanbul, Turkey*, 1315-1318.
- Watts O, J. Yamagishi, K. Berkling and S. King. (2008): HMM-based synthesis of child speech. *Proceedings of the 1st Workshop on Child, Computer and Interaction, Chania, Greece*.

134 Susanne Schötz, Gösta Bruce, My Segerup, J. Beskow, J. Gustafson & Björn Granström