

Swedish Accent Navigation

Gösta Bruce, Johan Frid and Ida Thelander

Department of Linguistics and Phonetics
Lund University, Sweden

`gosta.bruce@ling.lu.se; johan.frid@ling.lu.se; ida.thelander@ling.lu.se`

Abstract

In this paper a method for displaying dialectal, particularly intonational differences is presented. This map-based multimedia tool for simultaneous listening to speech samples and viewing acoustic records is called Swedia Prosody. It is based on Swedish speech data collected within the SweDia 2000 project from more than 100 different places in Sweden and Finland. Swedia Prosody uses both accent I and accent II words from the elicited material and short samples having a kind of prosodic equivalence from the spontaneous material. The paper describes our considerations in developing an accent navigator of this kind.

1. Introduction

The present paper gives a presentation of work conducted within the SweDia 2000 project with the aim of accounting for tonal variation among Swedish dialects. The ultimate goal of our effort is to capture the tonal identity of a dialect (or a dialect type). In the light of new data collected within the project we should also be able to develop and possibly revise the Swedish prosodic typology. Our starting point has been the intonation model devised by Bruce and Gårding [1] and developed within the research project Swedish prosody [2]. The present contribution is, however, not a report on this research itself but about the development of a method for displaying dialectal, particularly intonational differences. The idea is to be able to demonstrate accent variation in a direct and contrastive manner. Our assumption is that using fairly short speech samples should be sufficient to reveal accent identity, in particular the tonal identity of a dialect. The present paper will describe our considerations in developing an accent navigator of this kind and account for the navigator itself.

2. Background

2.1. The SweDia project

The research project SweDia 2000, The Phonetics and phonology of the Swedish dialects around the year 2000, is a cooperation between the phonetics sections at the Swedish universities of Lund, Stockholm and Umeå. It is supported by the Bank of Sweden, the Cultural Foundation between 1998 and 2004. The goals of the project are the documentation of the phonetic dialectal variation within the Swedish-speaking area, in Sweden and Finland, as well as the analysis and description of this variation around the millennium shift [3]. During the initial stage of the project recordings were made in more than 100 different communities, mainly in rural areas. For each place representatives of both the older generation (55-75 years) and the younger generation (20-30 years), both

men and women, were recorded, altogether 12 informants from each place.

The SweDia 2000 speech database relevant for intonational research consists of spontaneous interviews with informants and elicited two-word phrases from these speakers from 107 different places in Sweden and Finland, altogether more than 600 hours of speech from around 1200 speakers. The speech samples are stored in a huge digital database, where so far only the elicited material and a small fraction of the spontaneous material have been labeled for phonetic analysis. Samples of altogether around seven hours of speech from the spontaneous speech data are available on the Internet (www.swedia.nu). Around one minute of speech has been selected from each of four speakers representing two generations and both genders. These samples cover all 107 places in the SweDia database. This selection is referred to as the public database.

2.2. Intonational variation in Swedish

Our specific role within this project, which is a general phonetics and phonology project, is to describe the intonational variation among the Swedish dialect types of Sweden and Finland. The main goal of this research is to be able to account for the distinct intonational dialect types of Swedish in accordance with our basic intuition about it, i.e. to capture the tonal identity of a dialect type [4].

Unlike many of the European languages Swedish, Norwegian, and Danish are characterized by having an opposition between two word accents. The word accents appear to play a central role in Scandinavian prosody. In Swedish and Norwegian, where their main phonetic correlate is tonal, they are called accent I and accent II.

The main parameters that seem to be relevant for the description of Scandinavian and Swedish intonation are focal accentuation and phrasing within utterance prosody, and word accentuation and compounding within word prosody. Three of these four main parameters were used in an early typology by Bruce & Gårding [1] namely pitch realization of focus, pitch gestures of word accents, and pitch patterns of compounds. The specific patterns of phrase intonation that are now included were not explicitly taken care of in the Bruce & Gårding intonation model [1].

Our general idea about tonal variation in Swedish is that basically the same intonational type covers a fairly large region with relatively minor internal, local variation. Using the framework for intonational typology discussed in [6] it becomes clear that more dialect types than those distinguished in the Bruce & Gårding typology can be identified. According to this typology five distinct accent types are recognized for Swedish. A basic division is into single-peaked (Far East, South, Central) and double-peaked dialect types (East, West). However, when we look beyond word intonation in natural, spontaneous Swedish, East Swedish and Far East Swedish

have to be further subdivided. A North Swedish type has been proposed earlier [4], [7] and a division of Far East Swedish into a Northern and a Southern subtype also appears called for. Table 1 shows how seven different dialect types of Swedish are characterized and distinguished by the tonal parameters discussed in Bruce [6]. The dialect types are arranged according to an estimation of their approximate geographical (latitudinal) location. The tonal features are arranged in the order of general to more specific.

Table 1. *A tentative taxonomy of Swedish intonational dialect types.*

Tonal feature	Utterance intonation				Word intonation		
	focal accentuation		phrasing		word accents		compound
<i>Dialect type</i>	distinct focal accent	extra focal peak	post-focal accent	tonal concatenation	word accent timing	word accent contrast	2ary stress = pitch synch
<i>South</i>	no	no	yes/no	low plateau	late	yes	no
<i>West</i>	yes	yes/no	yes	upslope	early	yes	no
<i>Central</i>	yes	no	yes/no	downslope	very late	yes	yes
<i>East</i>	yes	yes	yes	high plateau	very early	yes	yes
<i>Far East S</i>	no	no	yes/no	low plateau	late	no	no
<i>Far East N</i>	no	no	yes	high plateau	early	no	no
<i>North</i>	yes	yes	yes	downslope	very early	yes	yes/no

3. Swedia Prosody, elicited speech

3.1. The prosody material

Our initial approach has been to analyze and account for the elicited prosody material. This material consists of two-word phrases, where the variables are accent I / accent II and placement of focus. The phrases have been elicited by having the informants access two cards in combination with symbols for amount (numeral) and currency (noun) respectively, such as *100 \$* (in that order). Through a successive exchange of cards for either amount or currency, a natural focus location was elicited in either non-final (amount) or final position (currency). Also tonal word accent was varied through this substitution, so that the Swedish words for *100* and *\$* both have accent I, while the words for *10* and *Kr* (kronor, Swedish currency) have accent II. An example of the beginning of an elicited series of such phrases is the following using symbols: *100 \$, 10 \$, 10 Kr., 100 Kr.*, or in Swedish orthography *hundra dollar, tio dollar, tio kronor, hundra kronor*. This procedure resulted in several productions of each of the words contained, both focused and unfocused.

For the analysis of the elicited prosody material a special computer program, Swedia Prosody, has been designed by one of the authors (Johan Frid). It is a map-based multimedia tool for simultaneous listening to samples from the labeled speech database and observing acoustic records of this, specifically waveforms and pitch contours. The program is written in Tcl/Tk and uses WaveSurfer components [8] for viewing and playing audio data.

For the purpose of illustrating accent differences the noun (currency), i.e. the word *dollar* (accent I) or *kronor* (accent II)

in final, focal position has been extracted. This is not exactly the same as but by and large equivalent to the citation form of an accent I or accent II word. For each place or dialect in the speech database, one representative token was chosen as a typical production. In our labeling of the elicited speech material priority had been given particularly to elderly men as representing the most pronounced and genuine accent. Thus one example out of a number of repetitions of *dollar* and *kronor* as uttered by one elderly man (out of three informants in this category) was chosen to represent each of 100 (out of 107) dialects and places. For the remaining seven dialects the relevant elicited speech data was not complete in one way or another and had to be left out.

3.2. Accent navigation

This arrangement of the elicited prosody material allows us to contrast different accents and dialects in a flexible manner and to do some accent navigation in the database. By means of the Swedia Prosody program it is possible to click on the map on one place at a time, thereby activating one man's production of either the accent I word *dollar* or the accent II word *kronor*. Simultaneously with a playback of the activated word its pitch contour (with a possible segmentation into vowels and consonants) and waveform will appear on the screen together with the place name in letters. This arrangement allows an audio-visual accent journey in Sweden and parts of Finland. Thus, for example, it is easily shown how the distinction between the two tonal word accents for one and the same dialect is manifested for most of the dialects, or for the Finland Swedish (Far East) dialects that there is no difference between accent I and accent II. Another possibility is to activate a sequence of dialects for one of the two words (accent I or accent II) and thus make an accent journey across some part of the Swedish-speaking area, i.e. for example jumping in small steps between neighboring places, or as an alternative, jumping in larger steps to contrast distinct dialect types.

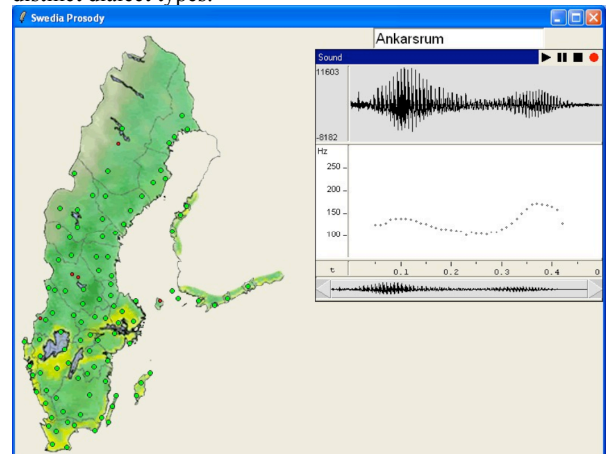


Figure 1. *Swedia Prosody; map-based multi-media tool for simultaneous listening to speech samples and viewing acoustic records. Sample production of the word 'kronor' (accent II) by an elderly male speaker from Ankarsrum (Göta Swedish) illustrating a double-peaked pitch contour.*

Thus assuming that the sample productions are representative, it is fairly easy to show how the Swedish

dialect types are tonally distinct. It is also apparent how quite abrupt some transitions between different dialect types can be. The intonational similarity between different dialects (places) within one dialect type is often also readily apparent. Figure 1 shows a map of Sweden and Finland with the places that can be activated in this way and a sample pitch contour and waveform of one speaker's production of the accent II word *kronor*.

A problem with our method has been that it is not easy to select samples from each of the speakers that are equivalent in terms of degree of involvement. Even if most of the subjects sound reasonably involved, we were sometimes left with samples that appear to be quite detached. The reason may be that from some of the speakers it has been hard to evoke sufficient interest in the specific elicitation task, which may lead to monotony. We should remember that all informants are what could be characterized as naïve subjects.

3.3. Automatic classification of dialect types

We have also attempted intonation-based automatic classification of Swedish dialects. We used the prosody material in order to build statistical models for the prediction of dialect type. F0 contour parameters such as timing and location of turning points were used. All in all there were around 4500 words. After splitting the data into a training set and a test set, we built CARTs from the training data in order to predict dialect type. We used five types: the South, West, and Central as in Table 1, whereas East and North were combined into one group and similarly, Far East S and Far East N, were combined into one group. The full procedure is described in [9].

The results were 59.1 % correct predictions. Better results were obtained when using focused words only than when using unfocused words, and, similarly, Accent II words were classified correctly more often than Accent I words.

These results show the accuracy with which dialect type may be predicted by using intonation only. Dialect type, on the whole, is rather difficult, but as dialects also show a large variation in their segmental properties there is much additional phonetic information that could be used for dialect detection.

4. Swedia Prosody, spontaneous speech

4.1. The short samples

The limitation of using samples from the elicited prosody material is that as only single words are contrasted, albeit the same word (either *dollar* or *kronor*), it is likely that a one-word utterance of this kind may reveal only part of the characteristic prosodic features of a particular dialect. The alternative method is to extract samples from our spontaneous speech database. The one-minute samples in the public database referred to above typically contain a good deal of prosodically varied speech. However, for the purpose of contrasting different dialects these samples turn out to be far too long. Therefore what would be needed are spontaneous speech samples that are sufficiently short but still prosodically varied to display characteristic features of each dialect.

Exploiting spontaneous instead of elicited speech means abandoning recourse to the same, recurrent words for our speech samples. Our specific requirement would instead be to seek prosodic equivalence, i.e. to identify samples containing

sufficiently varied prosody and intonation. Our considerations here have been to select samples containing at least one accent I word, one accent II word and a compound word, each of these occurring in a prosodically prominent position of an utterance. We estimated that in order to find such samples we needed to come across passages of around 10 seconds of speech. The idea was to try to find short monologue sections containing these features of word intonation and in addition features of utterance intonation as well, such as typical pitch patterns of focal accentuation and phrasing. The samples that have been extracted in this way show some variation in length, as they have to be coherent and complete. In most cases, however, it was possible to find samples that met the specific requirement of prosodic variation that we had set.

Also for the short, spontaneous samples we chose elderly men as the main representatives of each dialect. In addition we also decided to select a second informant representing the dialect. At this point we did not choose to have a complete set of speech samples from the four representatives of each dialect for particular technical and practical reasons. This second representative of each dialect could therefore be any of the three remaining informants: elderly women, young men or women. There was a preference for choosing an informant from the younger generation, however. Only if an elderly woman turned out to be prosodically more interesting and superior to the younger informants, she would be chosen as the second informant.

4.2. Pitch patterns of compounds

According to our specific requirement at least one example of a compound should be part of each short sample. As mentioned above compound words are among the constituents that are assumed to be particularly revealing for accent identity. Even if the pitch patterns of simplex words also show clear dialectal differences, compound words typically appear to be even more tonally diversified.

In Swedish as in other Germanic languages as well, compound words are distinct from simplex words by a difference in metrical structure. While simplex words have only one stress, compound words typically have two stresses: a primary one in the first element and a secondary stress in the final element of the compound [6]. Generally, however, a compound (in a prominent position of an utterance) has only one word accent, tied to the primary stress, just like the simplex word.

Thus while compound words are typically metrically distinct from simplex words, they may or may not be distinct intonationally. In some dialect types the secondary stress may play an important role as a pitch synchronization point. In East (Svea) Swedish the secondary stress is the trigger for the focal accent rise, while in Central (Dalarna) Swedish it is the word accent fall (focal or non-focal), which is triggered by the secondary stress. For these dialect types in corresponding non-compound words the pitch gestures (rise and fall respectively) occur in immediate succession of the preceding word accent gesture. In other dialect types like South and West (Göta) Swedish, compounds and simplex words are not differentiated intonationally. Thus it is clear that the intonational structure of compound words can be quite revealing for dialect identity [10].

Figure 2 shows pitch patterns of some accent II compounds selected from the spontaneous material (the short samples) to illustrate tonal differences between prosodic

dialect types. These sample compounds are metrically similar in that they have initial primary stress. They also have secondary stress, however, not immediately adjacent but still on a non-final syllable of the word. They also occur in a focal position of an utterance. Although they all have accent II, their pitch patterns are clearly distinct.

5. Conclusion

We like to think of our Swedia Prosody accent navigator primarily as a kind of prosodic calibration instrument but also as a tool for demonstration purposes. Our scientific account of prosodic and tonal parameters for each dialect should ideally correspond to our auditory impression of the dialect as presented in the extracted speech samples. Therefore if our intonation model is capable of generating synthetic pitch patterns that match those of the sample utterances, we take this as an indication of the relative success of our modeling.

6. Acknowledgments

The SweDia 2000 project is funded by the Bank of Sweden, the Cultural Foundation between 1998 and 2004.

7. References

- [1] 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.
- [2] Gårding, E., 1982. Swedish prosody. *Phonetica* 39, 288-301.
- [3] 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, Reports in Phonetics* 4, Department of Philosophy and Linguistics, Umeå University, 97-100.
- [4] Bruce, G., 2003a. Tonal variation in Swedish. To appear in *Proceedings of the Symposium on Cross-Linguistic Studies of Tonal phenomena: Historical Development, Phonetics of Tone and Descriptive Studies*, S. Kaji (ed.). Tokyo: ILCAA, Tokyo University of Foreign Studies.
- [5] Gårding, E., 1977. *The Scandinavian word accents*. Lund: Gleerup.
- [6] Bruce, G., 2004. An intonational typology of Swedish. To appear in *Speech Prosody 2004*, Proceedings from a conference in Nara, Japan, March 23-26.
- [7] Bruce, G., 2001. Secondary stress and pitch accent synchronization in Swedish. In *Nordic Prosody VIII*, W. van Dommelen & T. Fretheim (eds.). Frankfurt am Main: Peter Lang, 33-44.
- [8] Sjölander, S.; Beskow, J., 2000. Wavesurfer – an open source speech tool. *Proceedings of the ICSLP 2000*, Beijing, vol.4, 464-467.
- [9] Frid, J., 2003. *Lexical and Acoustic Modelling of Swedish Prosody*. Lund: Department of Linguistics and Phonetics.
- [10] Riad, T., 1998. Towards a Scandinavian accent typology. In *Phonology and morphology of the Germanic languages*, W. Kehrein & R. Wiese (eds.). Tübingen: Max Niemeyer, 77-109.

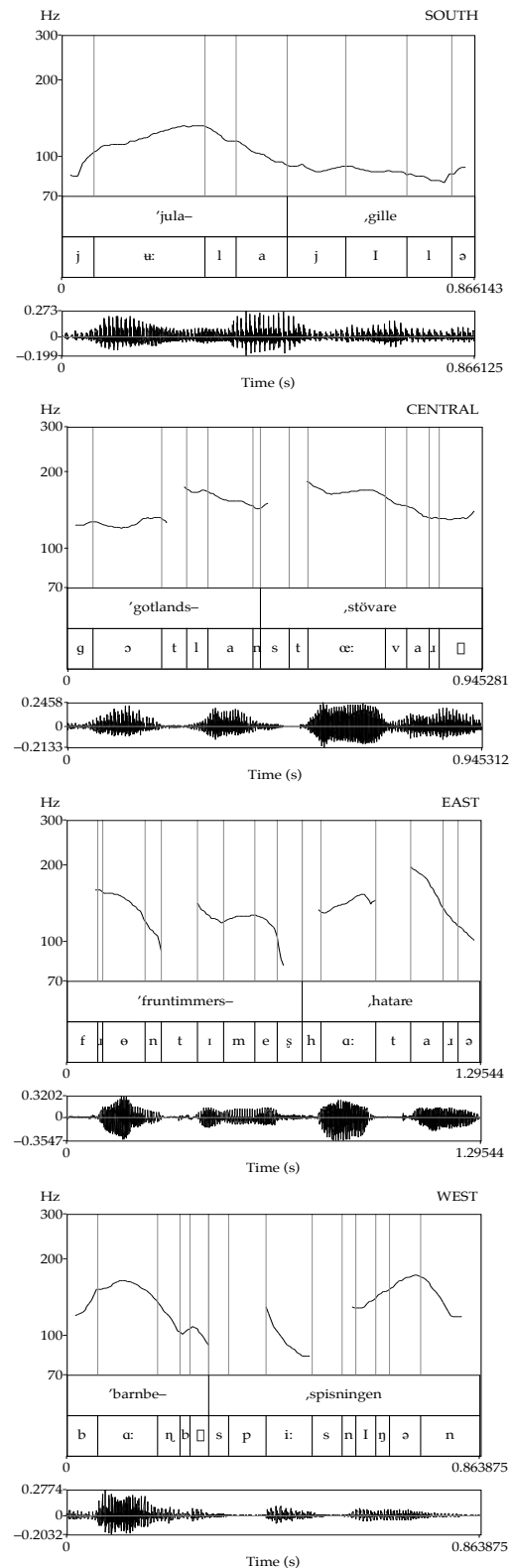


Figure 2. Tonal differences between four Swedish dialect types. F0 contours of sample accent II compounds in focal position. Examples are produced by elderly men from Våxtorp(South), Spröge (Central), Villberga (East) and Öxabäck (West) respectively.