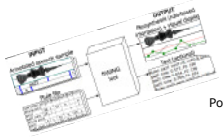
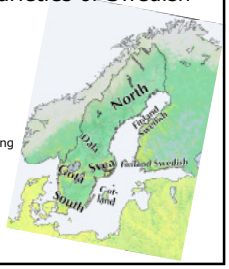
SWING:

A tool for modelling intonational varieties of Swedish

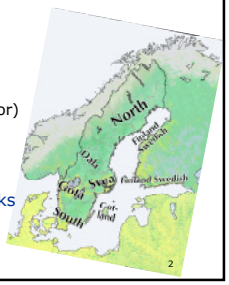


Gösta Bruce
Susanne Schötz
SIMULEKT (VR)
Potsdam-Lund Prosody Meeting
June 2008



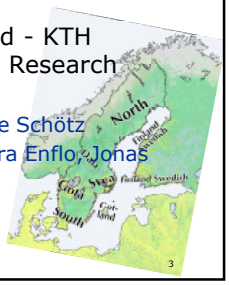
Outline

- SIMULEKT
  - Background and goals
  - The Swedish Prosody Model
  - Speech databases
  - Methods
  - Map
- SWING (SWedish INtonation Generator)
  - Background and purpose
  - Design
  - Demo
  - Discussion and concluding remarks



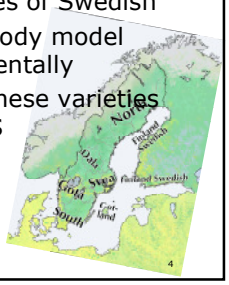
SIMULEKT

- Simulating intonational varieties of Swedish (SIMULEKT)
- Joint research project Lund - KTH supported by the Swedish Research Council (2007-2009)
  - Lund: Gösta Bruce, Susanne Schötz
  - KTH: Björn Granström, Laura Enflo, Jonas Beskow



Goals

- Gain more precise knowledge about the major intonational varieties of Swedish
- Develop the Swedish prosody model theoretically and experimentally
- Simulate the prosody of these varieties using SPEECH SYNTHESIS

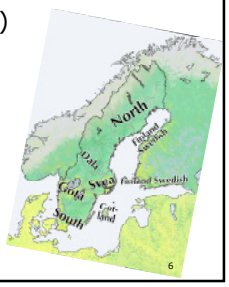


Two speech databases

- SweDia 2000 (Bank of Sweden, 1998-2003)
  - ≈ 1200 speakers of Swedish from >100 communities (Sweden & Finland)
    - elicited and spontaneous speech
- SpeechDat (EU project, reference data for ASR)
  - ≈ 5000 speakers (from all over Sweden)
    - read, telephone transmitted speech
    - two prosodically interesting sentences

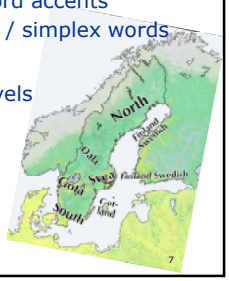
The Swedish prosody model

- Bruce & Gårding (1978)
- Bruce & Granström (1993)
- Bruce (2007)



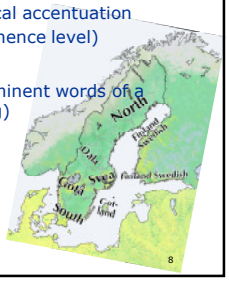
Features of the prosody model

- *Word prosody*
  - Timing characteristics of word accents
  - Pitch patterns of compound / simplex words
- *Utterance prosody*
  - Intonational prominence levels
  - Concatenation patterns



Utterance intonation

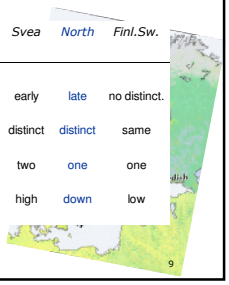
- Intonational prominence levels
  - clear distinction focal / non-focal accentuation (= 2 prominence levels)
  - no clear distinction focal / non-focal accentuation i.e. equal prominence (= 1 prominence level)
- Concatenation patterns
  - tonal concatenation between prominent words of a phrase (pitch patterns of phrasing)



The Swedish intonation model

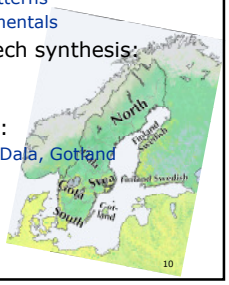
- Updated and revised (Bruce, 2007)
  - a tentative taxonomy of Swedish intonational varieties:

Intonational variety:	*South*	*Göta*	*Gotl.-Dala*	*Svea*	*North*	*Finl.Sw.*
Prosodic feature:						
Timing characteristics of word accents	late	early	late	early	late	no distinct.
Pitch patterns of compounds	same	same	distinct	distinct	distinct	same
Intonational prominence levels	one	two	one	two	one	one
Concatenation patterns	low	up	down	high	down	low



Methods

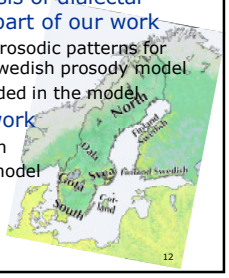
- Analysis and description of recorded speech:
  - prosodic (tonal and temporal) patterns
  - coordination of prosody and segmentals
- Simulating prosody using speech synthesis:
  - resynthesis of recorded speech
  - "real" speech synthesis
- Focus on seven main varieties:
  - South, Göta (West), Svea(East), Dala, Gotland North, and Finland Swedish



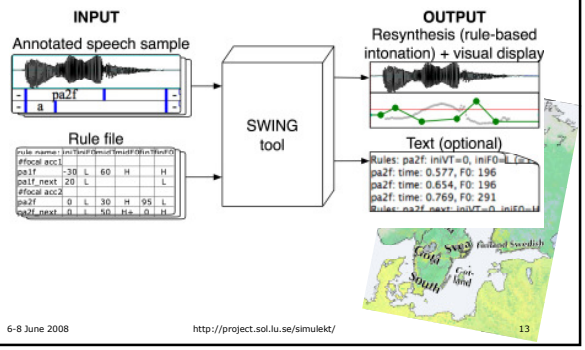
Map of Sweden and the Swedish-speaking parts of Finland

SWING: SWedish INtonation Generator

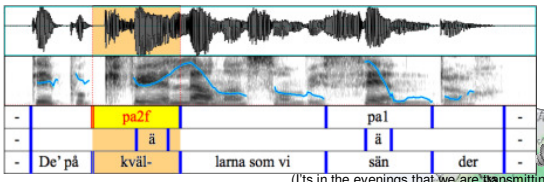
- Background
  - Auditive and acoustic analysis of dialectal speech samples important part of our work
    - collect empirical evidence of prosodic patterns for the regional varieties of the Swedish prosody model
    - identify patterns not yet included in the model
  - A tool would facilitate our work
    - generate rule-based intonation
    - test and further develop the model



SWING: Design



Input: Annotated speech sample



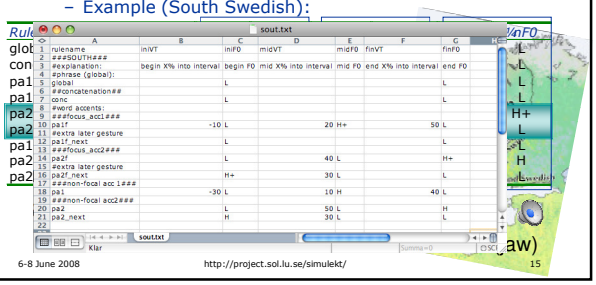
(I'ts in the evenings that we are transmitting)

- Manual segmentation in Praat (TextGrid)

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

Input: Rule file

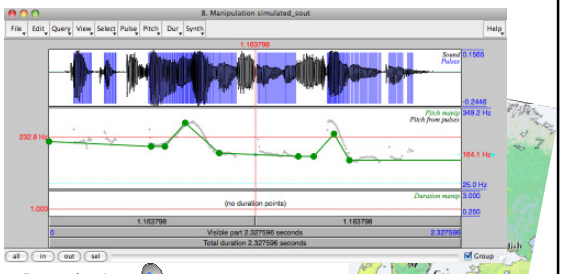
- Parameters of the Swedish prosody model
- One rule file for each regional variety
  - Example (South Swedish):



SWING tool

- Praat script (Two files: 1) input selection, 2) actual script)
  - Extract
    - pitch and duration values from input speech sample (Sound)
    - vowel and prosodic labels from input annotation (TextGrid)
  - Calculate
    - output pitch points using input rule file and extracted values
    - global (utterance, e.g. downstep, downdrift) values using factors
  - Generate
    - output pitch contour (PitchTier)
    - PSOLA resynthesis + visual feedback (Manipulation object)
    - debugging info (text)

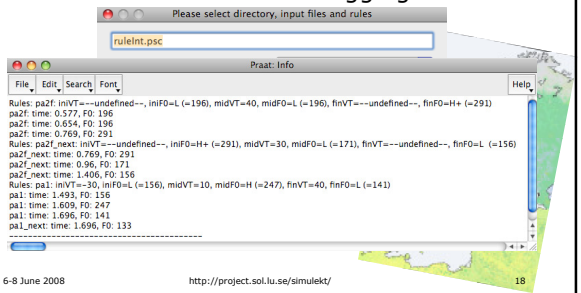
Output: Resynthesis + Manipulation object



- Resynthesis
- Original

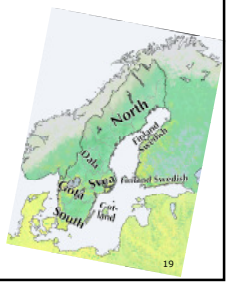
Output: Text (optional)

- For information and debugging



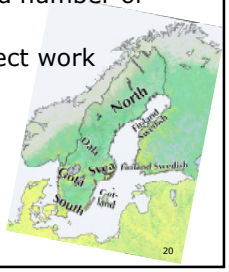
SWING Demo

SWING



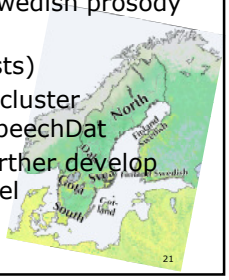
Discussion

- Still under development
- So far used only on limited number of samples
- Already useful in our project work

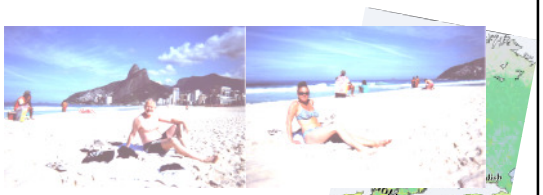


Future work

- More speech samples and dialects
- More parameters of the Swedish prosody model (e.g. compounds)
- Evaluation (perception tests)
- Use with KTH program to cluster intonational varieties in SpeechDat
- Use SWING to test and further develop the Swedish prosody model



Thank you!



Two Swedish phoneticians relaxing at IPA-nema (Rio de Janiero, Brazil) after the Speech Prosody 2008 conference