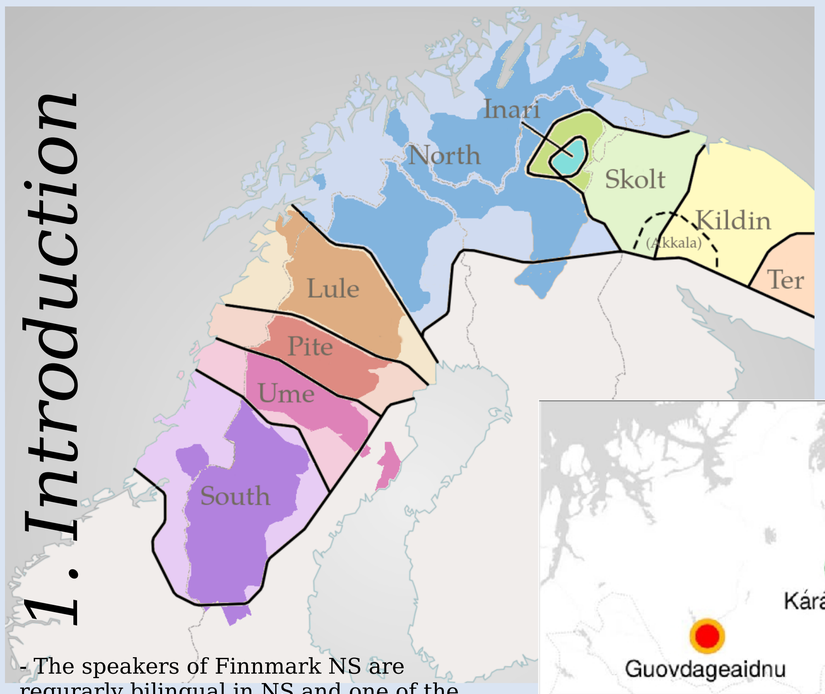
HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI
HUMANISTINEN TIEDEKUNTA
HUMANISTISKA FAKULTETEN
FACULTY OF ARTS

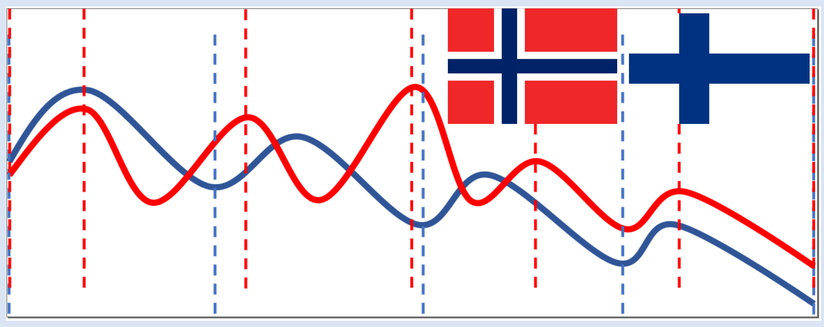
**Katri Hiovain, Antti Suni, Sofoklis Kakouros & Juraj Šimko**
**katri.hiovain@helsinki.fi**
Department of Digital Humanities
Phonetics and speech synthesis research group

# COMPARATIVE ANALYSIS OF MAJORITY LANGUAGE INFLUENCE ON NORTH SÁMI PROSODY USING WAVENET-BASED MODELING

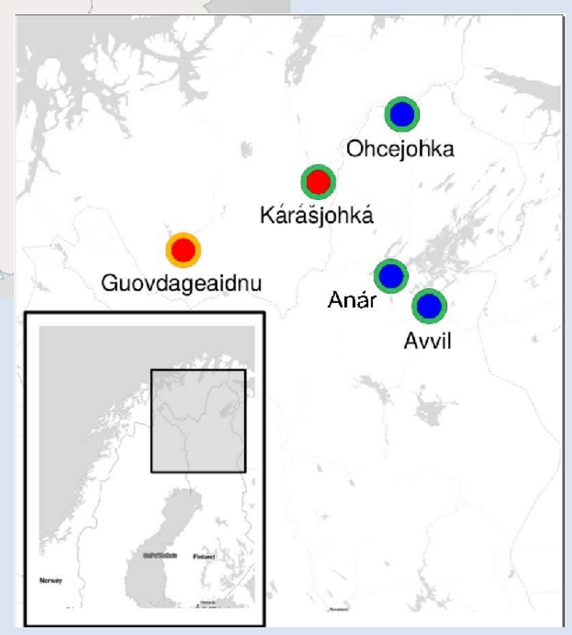
*7 joulukuuta 2020, (In press) julkaisussa: Language and Speech. ETAP Special issue: Experimental and Theoretical Advances in Prosody.*

## 1. Introduction

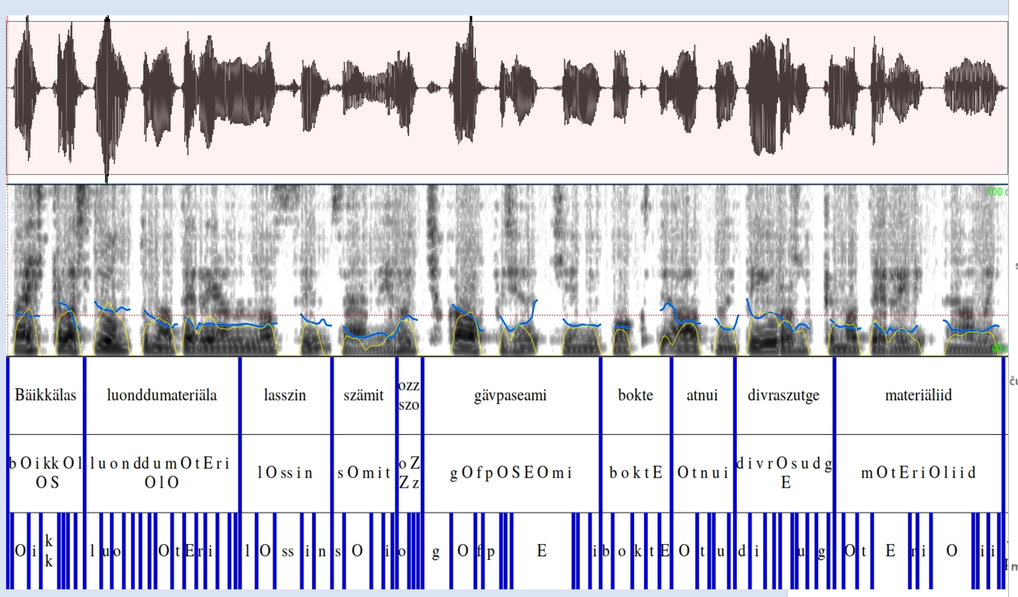


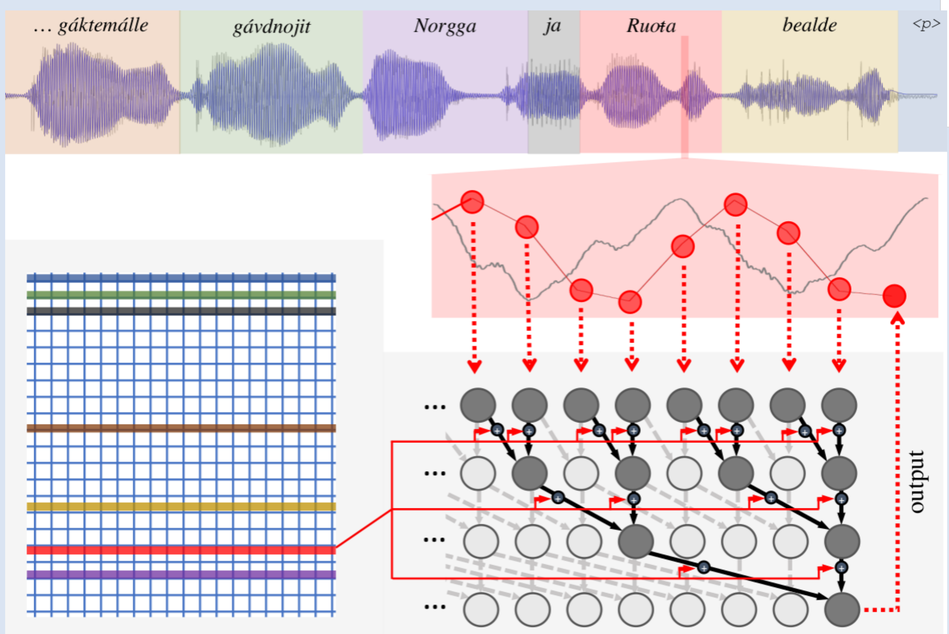
- The speakers of Finnmark NS are regurarly bilingual in NS and one of the majority languages
- We present a methodology (convolutional WaveNet synthesis models) to seek the portions of speech where prosodic differences based on maj. langs. are most robustly manifested
- Models were trained on NS speech materials (total of 214 min from 21 spkrs) from 5 NS villages, spoken in Norway and Finland.
- Data was POS-tagged, split to sentences and force-aligned using WebMAUS.
-We hypothesize that the prosodic characteristics in Norwegian NS variety reflect the features of Norwegian and respectively in Finnish varieties of NS.



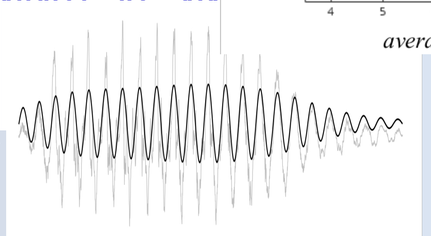

## 2. Materials




## 3. Methods




Dilated layers capturing progressively longer-term dependencies in the signal.



1) A new prosodic signal, matching the original waveform in f0, was generated for each sentence, containing no segmental information → capturing the f0 and intensity only, sampled at 800Hz

2) F0 contours extracted using customized Praat script, allowing manual correction of octave jumps and creak/noise labeling

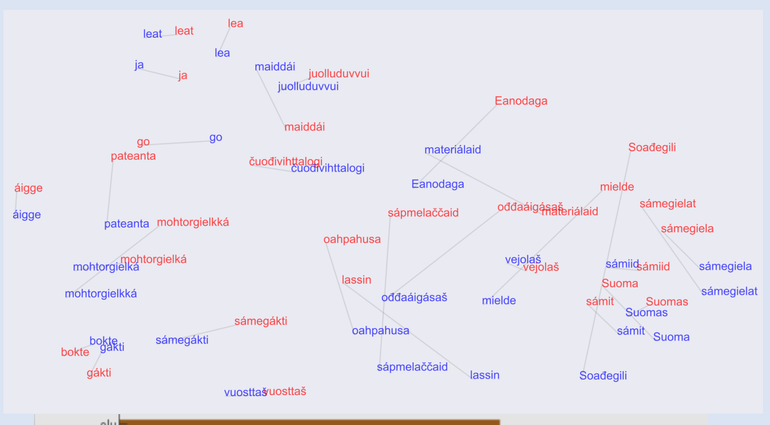
3) Speaker-normalization: the extracted F0 contours were recast to semitone scale using the median F0 for each speaker (over the entire material) as a base frequency → new F0′signals in Hz scale were recomputed using the base frequency of 100 Hz

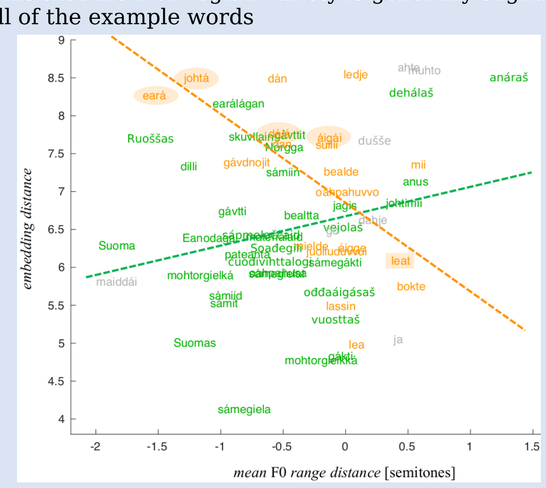
4) Training WaveNet deep network synthesis to obtain **vector representations** of individual lexical items as they are produced by, on the one hand,speakers from Finland and, on the other hand, by speakers from Norway

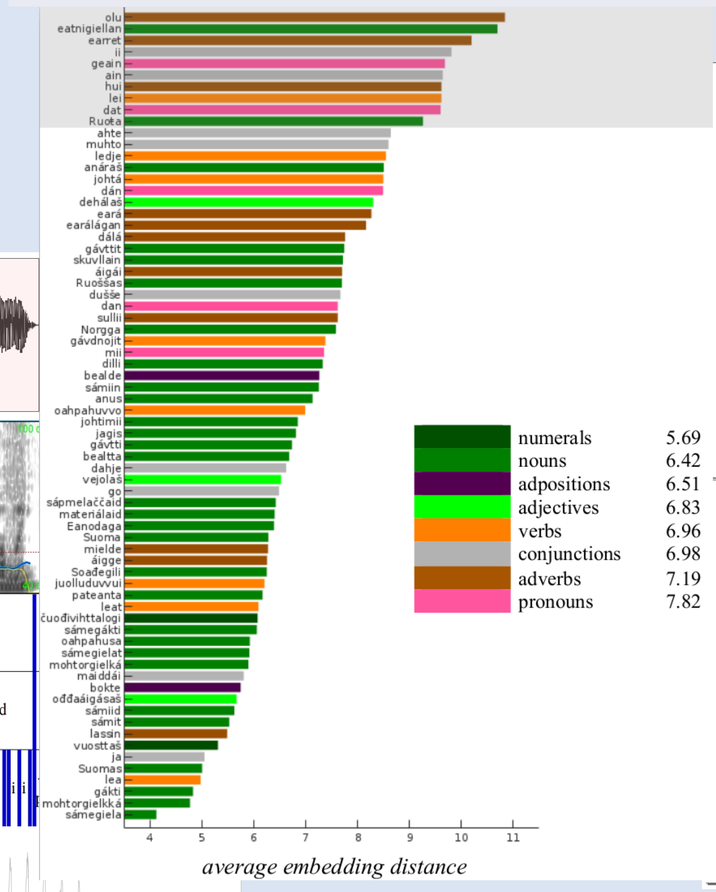
5) The network learns to predict next sample of the prosodic signal, given phrase and word identity, and the speaker's country of origin. Instead of examining the generated signals we focus on the learned word representations

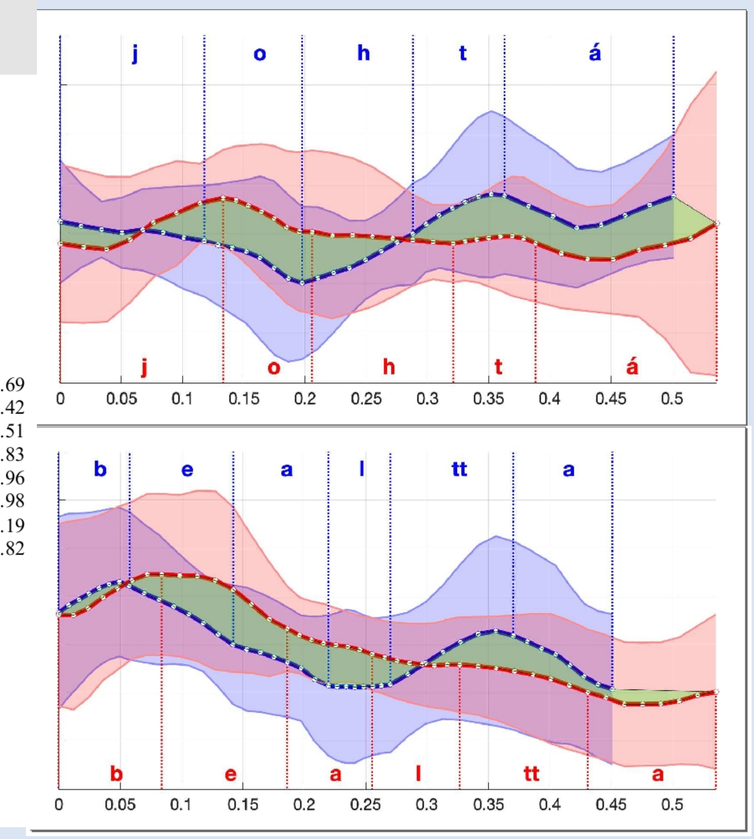
## 4. Results

- 70 most frequent words were taken into the analysis
- A measure of differences between the prosodic realization of the words: the distance between the embedding vectors of words renditions by the Finnish side and the Norwegian side NS speakers were calculated
- For selected words we calculated the mean F0 contours and standard deviationbands, separately for the Finnish and the Norwegian North Sámi speakers
→ In the examples, the timing ofthe peaks are different in the varieties, especially in the disyllabic words, different in certain parts of speech
→ When considering the duration of all example words (the x-axis), it seems that the Norwegian variety is generally slightly longer in all of the example words




## 5. Discussion

-The presented analysis provides an example of typological research where the initial hypotheses are formulated with the help of a machine learning technique that can extract a complex statistical model of speech (prosody) for the investigation of prosodic phenomena of interest.

-Our results indicate that at least for the words realized most differently between the Finnish and Norwegian bilinguals, the NS speakers from Finnish side produce the 'noun' group words with relatively greater, and the 'verb' group words with relatively smaller F0 excursions compared to their Norwegian side counterparts.

**-The Finnish–North Sámi bilingual speakers realize the prominence differences between the two word categories to a greater, more consistent extent than Sámi speakers from Norway.**

- While the Norwegian readers mark each word intonationally, the Finnish speakers reduce the intonation in the 'verb' portion. This is consistent with the findings ofless F0 variance in verbs compared to content words in Finnish language

- Potential issues: "universal read speech prosody", variable recording quality (WaveNet is very sensitive to noise in data), reading fluency...