



**Aalto University**  
School of Electrical  
Engineering

# Automatic Speech Recognition for Northern Sámi with comparison to other Uralic Languages

**Peter Smit, Juho Leinonen, Kristiina Jokinen, Mikko Kurimo**

*Aalto University*  
*Helsinki University*

**January 20, 2016**

# Aalto University Speech Recognition Research Group

A dialogue system  
for Northern Sámi

Speech Recognition

Under-resourced

Experiments

Video

Questions

Appendices

- Prof. Mikko Kurimo
- Broad-ranged research in Speech Recognition (ASR) and Language Modeling
  - Finnish ASR
  - Noise robust ASR
  - Subword modeling (Morfessor)
  - Conversational speech, Speech Synthesis, Speech evaluation
- Speech technology for under-resourced languages



# Contents

A dialogue system  
for Northern Sámi

Speech Recognition

Under-resourced

Experiments

Video

Questions

Appendices

## **A dialogue system for Northern Sámi**

## **Automatic Speech Recognition (ASR)**

## **ASR for Under-resourced languages**

## **Experiments**



**Aalto University**  
School of Electrical  
Engineering

**Automatic Speech Recognition for Northern Sámi with  
comparison to other Uralic Languages**  
Peter Smit  
*Signal Processing and Acoustics*

3/24

January 20, 2016

# A dialogue system for Northern Sámi

## A dialogue system for Northern Sámi

Speech Recognition

Under-resourced

Experiments

Video

Questions

Appendices

- WikiTalk for Northern Sámi
- Kids ask questions from a robot, and the robot gives information by reading Wikipedia articles to them
- Speech recognition (ASR) and speech synthesis (TTS) are needed for this application



Aalto University  
School of Electrical  
Engineering

Automatic Speech Recognition for Northern Sámi with  
comparison to other Uralic Languages  
Peter Smit  
*Signal Processing and Acoustics*

4/24

January 20, 2016

# Previous work

## A dialogue system for Northern Sámi

Speech Recognition

Under-resourced

Experiments

Video

Questions

Appendices

- Commercial Speech Synthesis system by Acapela Group
- Data and tools by University of Tromsø
- Master's thesis by Leinonen (2015)



Aalto University  
School of Electrical  
Engineering

Automatic Speech Recognition for Northern Sámi with  
comparison to other Uralic Languages  
Peter Smit  
*Signal Processing and Acoustics*

5/24

January 20, 2016

# Speech Technology for Northern Sami - Challenges

A dialogue system  
for Northern Sámi

Speech Recognition

Under-resourced

Experiments

Video

Questions

Appendices

- Only a limited amount of data and expertise is available for making ASR/TTS models



Aalto University  
School of Electrical  
Engineering

Automatic Speech Recognition for Northern Sámi with  
comparison to other Uralic Languages  
Peter Smit  
*Signal Processing and Acoustics*

6/24

January 20, 2016

# Automatic Speech Recognition

A dialogue system  
for Northern Sámi

Speech Recognition

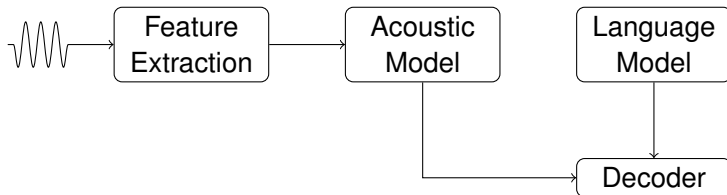
Under-resourced

Experiments

Video

Questions

Appendices



Aalto University  
School of Electrical  
Engineering

Automatic Speech Recognition for Northern Sámi with  
comparison to other Uralic Languages  
Peter Smit  
Signal Processing and Acoustics

7/24

January 20, 2016

# Automatic Speech Recognition

A dialogue system  
for Northern Sámi

Speech Recognition

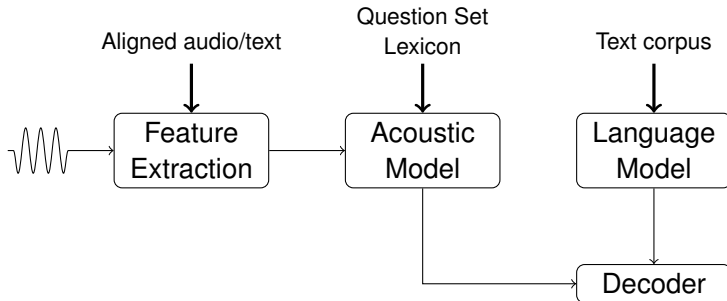
Under-resourced

Experiments

Video

Questions

Appendices





# What is needed for speech recognition?

A dialogue system  
for Northern Sámi

## Speech Recognition

Under-resourced

Experiments

Video

Questions

Appendices

- Corpus of aligned audio and text
- Corpus of text data for language modeling
- Dictionaries/lexicons (or the expertise to generate them)
- Phonetic question sets (for decision tree clustering)



Aalto University  
School of Electrical  
Engineering

Automatic Speech Recognition for Northern Sámi with  
comparison to other Uralic Languages  
Peter Smit  
*Signal Processing and Acoustics*

8/24

January 20, 2016

# Speaker-independent vs Speaker-dependent ASR

A dialogue system  
for Northern Sámi

Speech Recognition

Under-resourced

Experiments

Video

Questions

Appendices

- Speaker-independent (SI) systems works for any speaker of a language
- Speaker-dependent (SD) systems only works for the training speaker
- SI systems need audio data from many different speakers (50-100 for decent results)
- SD system only needs data from a single speaker



Aalto University  
School of Electrical  
Engineering

Automatic Speech Recognition for Northern Sámi with  
comparison to other Uralic Languages  
Peter Smit  
*Signal Processing and Acoustics*

9/24

January 20, 2016

# Speech Technology for Under-resourced Languages

A dialogue system  
for Northern Sámi

Speech Recognition

Under-resourced

Experiments

Video

Questions

Appendices

- Encourages language use
- Protects languages
- Preserves cultures



Aalto University  
School of Electrical  
Engineering

Automatic Speech Recognition for Northern Sámi with  
comparison to other Uralic Languages

Peter Smit  
*Signal Processing and Acoustics*

10/24

January 20, 2016

# Challenges for Under-resourced Languages

A dialogue system  
for Northern Sámi

Speech Recognition

Under-resourced

Experiments

Video

Questions

Appendices

- Audio Data
- Text Data
- Expertise



Aalto University  
School of Electrical  
Engineering

Automatic Speech Recognition for Northern Sámi with  
comparison to other Uralic Languages

Peter Smit  
*Signal Processing and Acoustics*

11/24

January 20, 2016

# Solutions for Under-resourced Languages

A dialogue system  
for Northern Sámi

Speech Recognition

Under-resourced

Experiments

Video

Questions

Appendices

- Audio Data
  - Audio books
- Text Data
  - Web scraping
- Expertise
  - Unsupervised methods
  - Bootstrapping on related languages



Aalto University  
School of Electrical  
Engineering

Automatic Speech Recognition for Northern Sámi with  
comparison to other Uralic Languages  
Peter Smit  
*Signal Processing and Acoustics*

12/24

January 20, 2016

# Language modeling for Uralic languages

A dialogue system  
for Northern Sámi

Speech Recognition

Under-resourced

Experiments

Video

Questions

Appendices

- High number of word forms
- Subword models are standard for ASR
- Morfessor
- $n$ -gram contexts are higher than for word models



Aalto University  
School of Electrical  
Engineering

Automatic Speech Recognition for Northern Sámi with  
comparison to other Uralic Languages  
Peter Smit  
*Signal Processing and Acoustics*

13/24

January 20, 2016

# Northern Sámi model with Wikipedia language model

A dialogue system  
for Northern Sámi

Speech Recognition

Under-resourced

Experiments

Video

Questions

Appendices

Unit	Toolkit	Speaker SF1 9-gram	Speaker SM1 9-gram
words	SRILM	52.9 / 12.7	48.7 / 11.1
morphs	SRILM	39.1 / 9.1	37.3 / 8.5
morphs	VariKN	37.6 / 8.7	34.1 / 7.9

Word error rate / Letter error rate



Aalto University  
School of Electrical  
Engineering

Automatic Speech Recognition for Northern Sámi with  
comparison to other Uralic Languages  
Peter Smit  
Signal Processing and Acoustics

14/24  
January 20, 2016

# Northern Sámi model with Big language model

A dialogue system  
for Northern Sámi

Speech Recognition

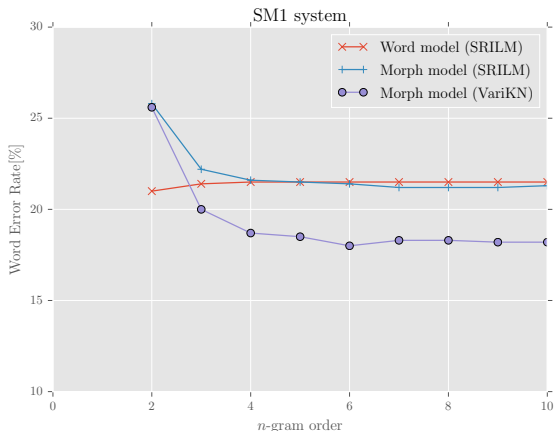
Under-resourced

Experiments

Video

Questions

Appendices





# Comparison to other Uralic Languages - datasets

A dialogue system  
for Northern Sámi

Speech Recognition

Under-resourced

Experiments

Video

Questions

Appendices

Language	Gender	Title	Hours
Estonian	Female	Nils Holgerssoni imeline teekond läbi Rootsi	16
Estonian	Male	Würst Gabriel ehk Pirita kloostri wiimsed päewad	6
Finnish	Female	Syntymättömien sukupolvien Eurooppa	12
Finnish	Male	Seitsemän veljestä	13
N-Sámi	Female	UIT-SME-TTSF	3.3
N-Sámi	Male	UIT-SME-TTSM	4.6



# Comparison to other Uralic Languages

A dialogue system  
for Northern Sámi

Speech Recognition

Under-resourced

2.5 hours of audio data

Experiments

Video

Questions

Appendices

Language	Voice	TRAIN+WIKI		BIG	
		WER	LER	WER	LER
Estonian	EF1	39.6	15.8	25.0	11.4
Estonian	EM1	39.2	13.3	25.5	9.6
Finnish	FF1	25.2	4.1	8.9	2.1
Finnish	FM1	35.8	7.7	24.9	5.6
Northern Sámi	SF1	37.5	8.5	23.7	5.5
Northern Sámi	SM1	39.5	9.4	20.9	4.9



# Comparison to other Uralic Languages - increase hours

A dialogue system  
for Northern Sámi

Speech Recognition

Under-resourced

Experiments

Video

Questions

Appendices

Voice	#hours	2.5 hours		All data	
		WER	LER	WER	LER
EF1	8	25.0	11.4	18.8	8.3
EM1	4.5	25.5	9.6	23.2	8.4
FF1	9	8.9	2.1	8.1	1.9
FM1	10	24.9	5.6	19.8	3.7
SF1	2.5	23.7	5.5	23.7	5.5
SM1	3.5	20.9	4.9	18.1	4.2



# State-of-the art Uralic Languages

A dialogue system  
for Northern Sámi  
Speech Recognition  
Under-resourced  
**Experiments**  
Video  
Questions  
Appendices

Language	Description	WER	LER
Estonian	Broadcast conversations	17.9%	1
Estonian	Oral presentations	26.3%	
Finnish	Speecon testset		2.9% <sup>2</sup>
Estonian	Telephone speech	33.1%	11.9% <sup>3</sup>
Finnish	Telephone speech	21.6%	6.8%

<sup>1</sup>Tanel Alumäe (2014). “Recent improvements in Estonian LVCSR”. In: Spoken Language Technologies for Under-Resourced Languages.

<sup>2</sup>Janne Pytkkonen and Mikko Kurimo (2012). “Analysis of Extended Baum-Welch and Constrained Optimization for Discriminative Training of HMMs”. In: Audio, Speech, and Language Processing, IEEE Transactions on 20.9, pp. 2409–2419. ISSN: 1558-7916. DOI: 10.1109/TASL.2012.2203805.

<sup>3</sup>Teemu Hirsimäki, Janne Pytkkonen, and Mikko Kurimo (2009). “Importance of high-order n-gram models in morph-based speech recognition”. In: Audio, Speech, and Language Processing, IEEE Transactions on 17.4, pp. 724–732.



# Future work

A dialogue system  
for Northern Sámi

Speech Recognition

Under-resourced

Experiments

Video

Questions

Appendices

- Expand recognizer to be Speaker Independent



Aalto University  
School of Electrical  
Engineering

Automatic Speech Recognition for Northern Sámi with  
comparison to other Uralic Languages

Peter Smit  
*Signal Processing and Acoustics*

20/24

January 20, 2016

# Video

A dialogue system  
for Northern Sámi

Speech Recognition

Under-resourced

Experiments

**Video**

Questions

Appendices

<https://youtu.be/14jYeViJ0X0>  
<https://www.youtube.com/v/14jYeViJ0X0>



**Aalto University**  
School of Electrical  
Engineering

**Automatic Speech Recognition for Northern Sámi with  
comparison to other Uralic Languages**  
Peter Smit  
*Signal Processing and Acoustics*

21/24  
January 20, 2016

# Questions

A dialogue system  
for Northern Sámi

Speech Recognition

Under-resourced

Experiments

Video

**Questions**

Appendices

## Questions?



**Aalto University**  
School of Electrical  
Engineering

**Automatic Speech Recognition for Northern Sámi with  
comparison to other Uralic Languages**  
Peter Smit  
*Signal Processing and Acoustics*

22/24

January 20, 2016

# OOV rate

A dialogue system  
for Northern Sámi

Speech Recognition

Under-resourced

Experiments

Video

Questions

Appendices

	Word	Morph
Female	22%	0%
Male	20%	0%



Aalto University  
School of Electrical  
Engineering

Automatic Speech Recognition for Northern Sámi with  
comparison to other Uralic Languages  
Peter Smit  
*Signal Processing and Acoustics*

23/24

January 20, 2016



# Text corpora

A dialogue system  
for Northern Sámi

Speech Recognition

Under-resourced

Experiments

Video

Questions

Appendices

Language	Source	#sent	#tokens	#types
Estonian	Wiki	895k	10M	778k
Estonian	newspaper	19M	229M	3.8M
	+ web + broadcast			
Finnish	Wiki	2.2M	22M	1.5M
Finnish	Kielipankki	13M	143M	4.1M
Northern Sámi	Wiki	10k	88k	20k
Northern Sámi	Den samiske tekstbanken	990k	12M	475k

