

# The Althingi ASR System

Inga R. Helgadóttir, Anna B. Nikulásdóttir, Michal Borský,  
Judy Y. Fong, Róbert Kjaran, Jón Guðnason

Reykjavik University  
Language and Voice Lab



## Abstract

- Reykjavik University developed an automatic transcription system for the Icelandic Parliament. The system entered a test phase in October 2018, and is in production since January 2019.
- A WER of 7.91% was obtained using a TDNN + rescoring with RNN-LM (in-lab test set). In-lab F-score for the punctuation model was 80.6 and 61.6 for the paragraph model.
- Including punctuation and other post-processing, a WER of  $15.0 \pm 6.0\%$  (in the wild test). The transcribers of Althingi graded 77% of the speech transcripts as Good.

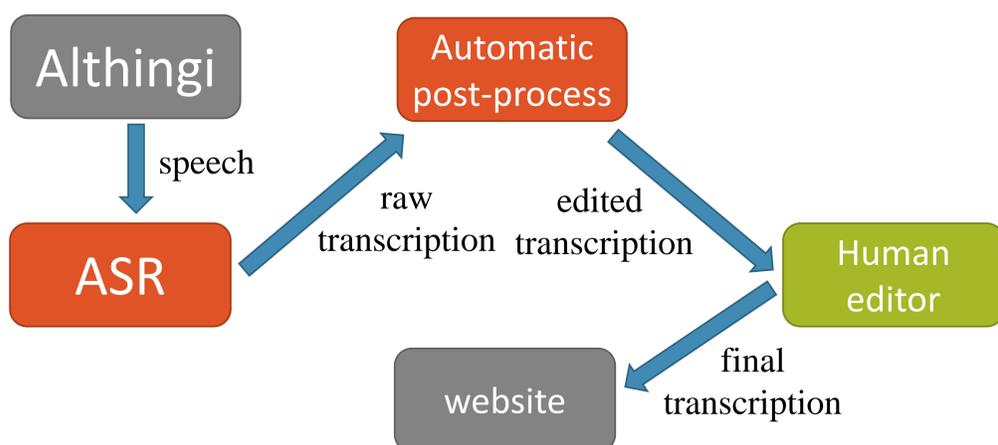
## Introduction

- The Icelandic parliament, Althingi, is dedicated to publishing speech recordings and transcripts from every meeting. A human transcriber transcribes the speech and an editor makes the transcription fit for publication
- The objective is to replace the manual transcription stage with an ASR system.
- The output text needs to be correct and readable to not slow down the editors.

## System Integration and Workflow

- The transcription system connects with Althingi servers to retrieve the templated speech XML document and an audio segment.
- The ASR process starts when the API is sent the ending time of the speech.
- The ASR transcript is created and automatic post-processing is performed.
- The resulting file is imported into Documentum.
- The manual post-processing begins and the edited speech is posted to website.
- The ASR is a virtual server and interacts with Althingi servers through the RESTful API and DFC.

## The Transcription System



## Results and Feedback from Transcribers/Editors

AM	Train set [hrs]	WER [%]	RTF
LSTM-TDNN (LF-MMI) w/sp	1500	9.07	1.03
f-TDNN w/sp	1500	9.17	0.18
f-TDNN	6200	9.05	0.21
f-TDNN w/sp	1000, re-cleaned	8.52	0.15
f-TDNN w/sp w/RNN-LM	1000, re-cleaned	8.17	0.15*
f-TDNN w/sp w/bi-RNN-LM	1000, re-cleaned	7.91	0.15*

\* RNN-LM and bi-RNN-LM takes approx. 20-times/40-times longer to rescore

## The ASR

### Acoustic Model :

- A sequence trained ANN based on a factorized TDNN.
- 11 layers with 1280 neurons + linear 256-dimension bottleneck.
- Takes 40-dim. LDA feature vectors and 100-dim. i-vectors.
- The model has a decoding RTF of 0.18.

### Language Model:

- A pruned 3-gram LM was used for decoding.
- A 5-gram and two RNN LMs were tried for re-scoring.
- RNN-LMs gave slightly better results, but we rather use the 5-gram (20x faster).
- The lexicon was based on the pronunciation dictionary from the Hjal project.

Punctuation	Precision	Recall	F-score
Comma	76.9	53.5	63.1
Period	91.0	88.4	89.7
Question mark	89.8	81.5	85.5
Colon	83.3	80.5	81.9
Overall	86.7	75.3	80.6

Err: 2.37%  
SER: 32.0%

- The editors noted that the transcripts are now available sooner than before.
- 475 speech graded as Good, 111 as Medium and 30 as Bad

## Automatic Post-processing Example

Original ASR output -> Apply grammar to rewrite numbers and abbreviate -> Apply punctuation model -> Final text after last fixes => Reference text.

- 1) að verðbólga fari ekki yfir tvö komma níu prósent ákvæði sextugustu og níundu grein laga um almannatryggingar var lögfest með tíundu grein laganna númer hundrað þrjátíu nítján hundruð níutíu og sjö í greinargerð
- 2) að verðbólga fari ekki yfir 2,9% ákvæði 69. gr laga um almannatryggingar var lögfest með 10. gr laganna nr 130/1997 í greinargerð
- 3) að verðbólga fari ekki yfir 2,9%. Ákvæði 69. gr laga um almannatryggingar var lögfest með 10. gr laganna nr 130/1997. Í greinargerð
- 4) að verðbólga fari ekki yfir 2,9%. Ákvæði 69. gr. laga um almannatryggingar var lögfest með 10. gr. laganna nr. 130/1997. Í greinargerð
- 5) að verðbólga fari ekki yfir 2,9%. <EOP> Ákvæði 69. gr. laga um almannatryggingar var lögfest með 10. gr. laganna nr. 130/1997. Í greinargerð

## Conclusion

Our system is already in use and speeding up the transcription process at Althingi. Using a clean 1000 hour subset gives better ASR transcripts than training on a 6200 hour set that only goes through basic cleaning steps. We have shown that a good text post-processing can hide some linguistic errors prone to Icelandic ASR. The responses from the staff at Althingi show that they are happy with the transcripts they receive from the ASR and it is considered useful, even for the speakers that are most often badly transcribed.

## Acknowledgements

This project was made possible through the support of Althingi's information and publications departments. The authors would like to thank Solveig K. Jónsdóttir and Þorbjörg Árnadóttir for their valuable help, Ingvi Stígsson for handling the technical aspects of the test at Althingi, as well as the editors and transcribers, for reading over the ASR transcripts, correcting them, timing the process and giving valuable comments.