HMM in speech synthesis

Project owner: Indrek Kiissel (Institute of the Estonian Language)
Team: Gerth Jaanimäe, Kaisa Saarkoppel
Goals

• Train a new Estonian synthetic speech voice using a corpus of voice recordings (smaller – 1550 recordings, larger – 4746 recordings)
• Test different training parameters
Methods – mainly HMM

• Advantages: the resulting model is quite small (in our case it was about 5.5 MB)
• Quite responsive
• HMM and other methods (decision tree, neural networks etc)
• HPC – parallel processes (larger corpus)
• Training: different parameter values and turning some so called switches on and off
The results

• Trained on both large corpus (4746 recordings) and smaller corpus (1550 recordings)
• Hard to measure (cannot use accuracy, precision, recall)
• The results were almost identical to our ears (voice models)
• Small changes from smaller corpus – smoothness/discontinuity, monotony/expressiveness
• Differences between voice models from large corpus and from smaller corpus – pace, volume, noise

The resulting model will go to EKI’s Github repository and website.
Lessons learned

• Training data plays even bigger role than we expected
• Knowledge of speech synthesis’ creation
• Linux, compiling, makefiles, debugging etc
• How SLURM works
• How to get things installed and running as non root user
Thank you all for listening!

Tänan teid kõiki kuulamast!