Automatic speech recognition for people with disordered speech: results from online and offline experiments

Mauro Nicolao, Heidi Christensen, Salil Deena, Stuart Cunningham, Phil Green, Thomas Hain

**homeService project**

Project Goal: recognising disordered speech in people’s homes using state-of-the-art techniques developed.

**Experimental data**

- UAspeech: largest, English database of dysarthric speech, 16 speakers, 11hrs.
- Enrolment data (ER), speaker dependant data: offline recordings of the keywords the participant uses to control the system. Several repetitions of a list of commands in the vocabulary (~ 30 items).
- Interaction Data (ID), speaker dependant data recorded by the participant while functioning the system. Several not uniformly distributed repetitions of given list of commands.

**Online Experiments**

- word loop grammar restricted to one-word recognition per utterance, using a menu-based dial manager, recognition dependent on dial manager state, acoustic models trained with MAP adaptation on top of UAspeech.
- 2 models used in the online experiments: mapER01 and mapER01+ID01.
- M02-ID03 and M02-ID04 recorded interleafing the two acoustic models, M02-ID06 recorded with mapER01+ID01 but different vocabulary to test model overfitting.

**Future work**

- complete the recruitment of more participants,
- change towards a keyword activation system instead of push-to-speak,
- move from command word recognition to a more natural phrase recognition,
- add Deep Neural Network (DNN) models to the atLab recogniser,
- always deliver to participants the best available system,
- involve participants in the research as much as possible, following their feedback.

**Annotation of Interaction Data**

- new recorded audio from participant needs annotation to be used as adaptation data,
- supervised annotations: humans transcribe what is said using a web interface,
- unsupervised annotations: ASR transcription from the online system, annotation tool developed in collaboration with Jonathan Kilgour

**Offline Experiments**

- Offline experiments have no dial manager,
- more lexical confusion than online experiments, every word same probability,
- more flexibility to try different configurations to eventually deploy to users.

**Performance test on the real use of homeService**

- word loop grammar restricted to one-word recognition per utterance, using a menu-based dial manager, recognition dependent on dial manager state, acoustic models trained with MAP adaptation on top of UAspeech.

**Acoustic model performance on M02-ID02**

- UAs: UAspeech SI models
- mapER01: UAspeech SI models + MAP adaptation with M02-ER01 data, tested on M02-ID02,
- mapER01+ID01: UAspeech SI models + MAP adaptation with M02-ER01 and M02-ID02 data, tested on M02-ID02,
- MAP+MAP ER01+ID01: MAP adaptation with ID01 data on top of mapER01,
- MAP+MAP ER01-ID01: MAP adaptation with ID01 data on top of mapER01

- UAs: UAspeech SI models
- mapER01+ID01: UAspeech SI models + MAP adaptation with M02-ER01 and M02-ID02 data, tested on M02-ID02.

**Dependency on the amount of supervised data in MAP adaptation**

- Adapting with varying amounts of data from M02-ER01 and M02-ID01 and testing on M02-ID02,