A Framework for Collecting Realistic Recordings of Dysarthric Speech - the *homeService* Corpus

Mauro Nicolao, Heidi Christensen, Stuart Cunningham, Phil Green, and Thomas Hain

University of Sheffield, United Kingdom

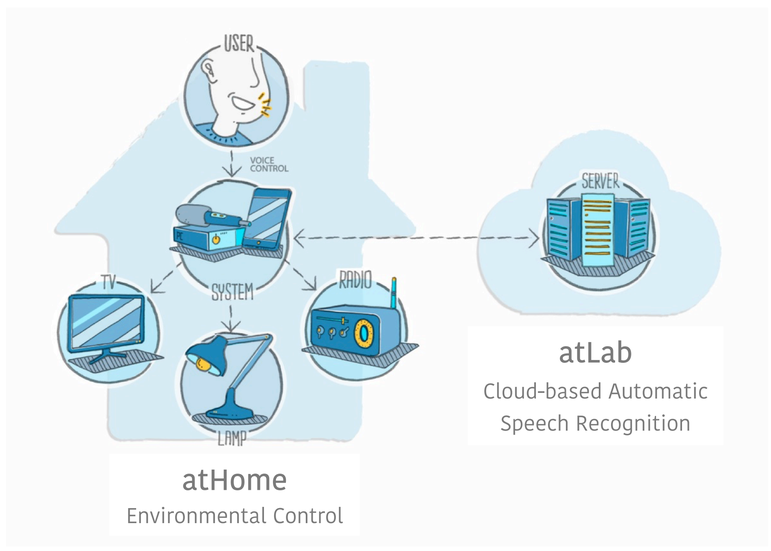
Abstract

This paper introduces a new British English speech database, named **the homeService corpus**, which has been gathered as part of the homeService project. This project aims to help users with speech and motor disabilities to operate their home appliances using voice commands. The audio recorded during such interactions consists of **realistic data of speakers with severe dysarthria**. The majority of the homeService corpus is recorded in real home environments where voice control is often the normal means by which users interact with their devices. The collection of the corpus is motivated by the shortage of realistic dysarthric speech corpora available to the scientific community. Along with the details on how the data is organised and how it can be accessed, a brief description of the framework used to make the recordings is provided. Finally, the performance of the homeService automatic recogniser for dysarthric speech trained with single-speaker data from the corpus is provided as an initial baseline. The collection process is still ongoing. Access to the homeService corpus is provided through the dedicated web page at

http://mini.dcs.shef.ac.uk/resources/homeservice-corpus/

The homeService project

The homeService project is an impact showcase for the UK EPSRC Programme Grant Project, Natural Speech Technology (NST) [1], a collaboration between the Universities of Edinburgh, Cambridge and Sheffield.



- Studying how speech technology can be of use for people with **speech disorders** and **restricted upper-limb mobility**.
- Developing an **environmental control system** to operate electronic devices such as TVs, radios, lamps, etc. through voice-commands [2].
- The **atHome** component: deployed in a user's home: software and "off-the-shelf" hardware to control participant's devices.
- The **atLab** component: cloud-based ASR system with synchronisation functions with all the atHome systems.
- Remote ASR server enables **simultaneous speech recognition** from many participants and **continuous adaptation** to user's new data.

Participants

- Having participants involved in the corpus collection has been one of the big challenges of the project.
- **Seven users** have been contacted to be part of the project in this part of the data collection.
- Two participants withdrew, three are recording, and two are starting to use the system.

User	Gender	Age	Native Language	Condition	State of recordings
M01	male	47	GBEng	cerebral palsy	interrupted
M02	male	75	GBEng	motor-neuron disease	ongoing
M03	male	22	GBEng	cerebral palsy	ongoing
M04	male	NA	GBEng	NA	recruiting
F01	female	55	GBEng	cerebral palsy	interrupted
F02	female	54	GBEng	NA	ongoing
F03	female	NA	GBEng	NA	recruiting

Online ASR experiments

- A standard state-of-the-art real-time ASR engine is used.
- Recognition performed on the 16kHz, single channel, beam-formed version of the audio.
- Perceptual linear prediction (**PLP**) acoustic features are extracted from the speech signal.
- A dysarthric background model is trained on the **UA-Speech Database** [3].
- UA-Speech model is tailored to the specific speaker's speech style with **MAP adaptation**.
- A **hierarchical grammar** is used as language model.
- M02 acoustic model continuously adapted to available data.
- F02 and M03 models are adapted with ER01test data only.

Speaker	Adaptation data	Type of data	Interactions	Accuracy [%]	OOV [%]
F02	ER01train	ID01train+ID01test	399	37.1	12.28
M02	ID01train.{1-5}	ID01train+ID01test	7378	74.1	2.79
M03	ER01train	ID01train+ID01test	205	38.1	10.53
Mean/Sum			**7982**	**49.8**	**8.10**

The homeService corpus v1.0

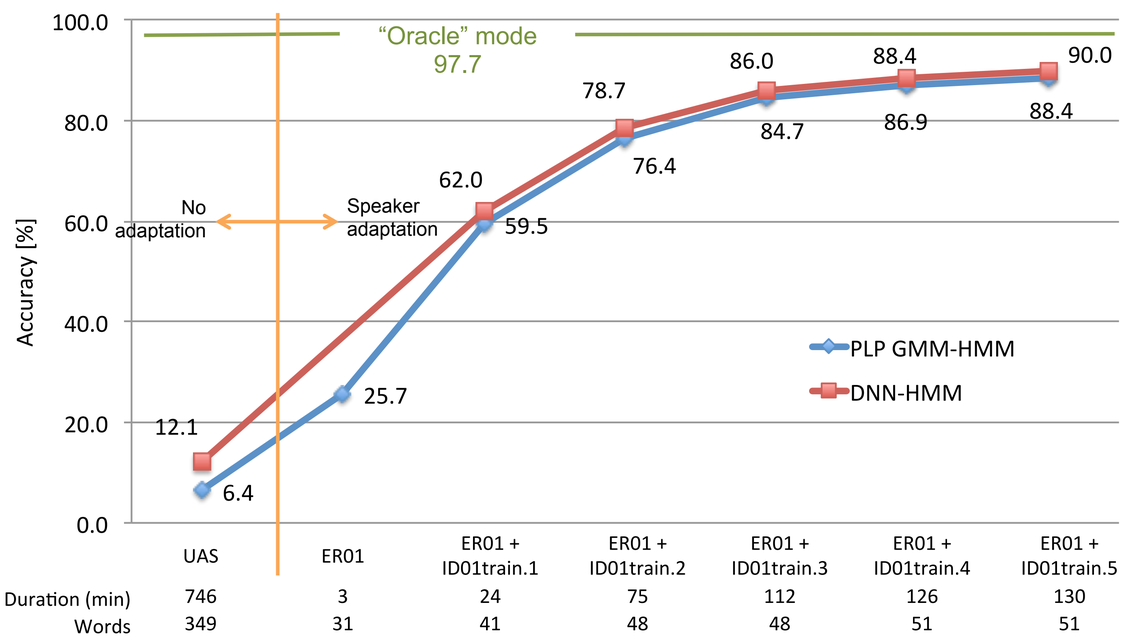
- Each participant's audio is recorded with a Microcone **microphone array** in a real home environment.
- The complete 8-channel output (6 streams from the microphone array plus a beam-formed stereo combination of those) is stored at 48kHz sampling rate and 32bit definition.
- In the current database release, a **16kHz-sampled, single channel, beam-formed** version of the audio is provided.
- Audio has been manually annotated and **transcriptions** are available with the corpus.

Speaker	Type of data	Words	Interactions	Duration
F01	ER01train	32	97	2'19"
F02	ER01train	31	314	11'58"
	ID01train	30	314	25'52"
	ID01test	16	85	5'40"
M01	ER01train	31	230	6'34"
M02	ER01train	31	130	3'16"
	ID01test	40	1571	1h44'44"
	ID01train	47	5807	6h29'40"
M03	ER01train	12	114	2'47"
	ID01train	18	169	11'26"
	ID01test	11	36	3'00"
TOTAL		**131**	**8867**	**9h27'20"**

Offline ASR experiments

- The homeService corpus v1.0 data allows us to perform several tests: e.g. optimise acoustic models for M02.
- Offline ASR use same decoder framework (**PLP GMM-HMM**) as the online one.
- A deep neural network based UA-Speech background model is also tested (**DNN-HMM**).
- In offline tests, all the entries in the grammar have the same weight.
- **"Oracle" mode** included test data in the MAP adaptation set.

Acoustic models adapted with different amount of speech data



Duration (min)	746	3	24	75	112	126	130
Words	349	31	41	48	48	51	51

Conclusions

- To the best of our knowledge, this is the first database of **spontaneous dysarthric speech** recorded in a **real-world environment**.
- Small word diversity but broad variety of **realistic acoustic recording** conditions.
- Audio collected over several months allows for **longitudinal studies** on voice variations caused by degenerative speech impairment.
- Considerable amount of more data is foreseen to be available by the end of 2016.

Acknowledgements

This work has been supported by the EPSRC Programme Grant EP/I031022/1 Natural Speech Technology (NST)

References

[1] The Natural Speech Technology (NST) Programme Grant, "homepage: http://www.natural-speech-technology.org/," 2015.

[2] H. Christensen, S. Cunningham, P. Green, and T. Hain, "homeService: Voice-enabled assistive technology in the home using cloud-based automatic speech recognition," in *4th Workshop on Speech and Language Processing (SLPAT)*, 2013.

[3] Statistical Speech Technology Group, University of Illinois, "The UA-Speech Database," *University of Illinois at http://www.isle.illinois.edu/sst/data/UASpeech/*, 2013.