

EPSRC

Engineering and Physical Sciences
Research Council



Edinburgh – Cambridge – Sheffield

Natural Speech Technology Programme Overview

Steve Renals
Centre for Speech Technology Research
University of Edinburgh

23 May 2013

<http://www.natural-speech-technology.org>

Overall aim

Significantly advancing the state-of-the-art in speech technology

- making it more natural
- applied to speech recognition and speech synthesis
- approaching human levels of
 - reliability
 - adaptability
 - fluency

NST facts and figures

- EPSRC Programme Grant, May 2011–April 2016
 - Three university partners: Edinburgh, Cambridge, Sheffield
 - User group (15 organisations)
 - Scientific advisory board (4 members)
- The Team - 28 people!
 - 9 investigators
 - 14 research staff
 - 5 PhD students
- Strong links with other projects:
EPSRC, EU, DARPA, IARPA, NHS, JST, etc.

NST team

- **CSTR, University of Edinburgh:**
 - Steve Renals, Simon King, Junichi Yamagishi
 - Peter Bell, Arnab Ghoshal, Jonathan Kilgour, Heng Lu, Liang Lu, Tom Merritt, Pawel Swietojanski, Christophe Veaux, Mirjam Wester
- **Speech Research Group, University of Cambridge:**
 - Phil Woodland, Mark Gales, Bill Byrne
 - Pierre Lanchantin, Xunying Liu, Matt Shannon, Marcus Tomalin
- **Speech and Hearing Research Group, University of Sheffield:**
 - Thomas Hain, Phil Green, Stuart Cunningham
 - Heidi Christensen, Mortaza Doulaty, Charles Fox, Yulan Liu, Oscar Saz

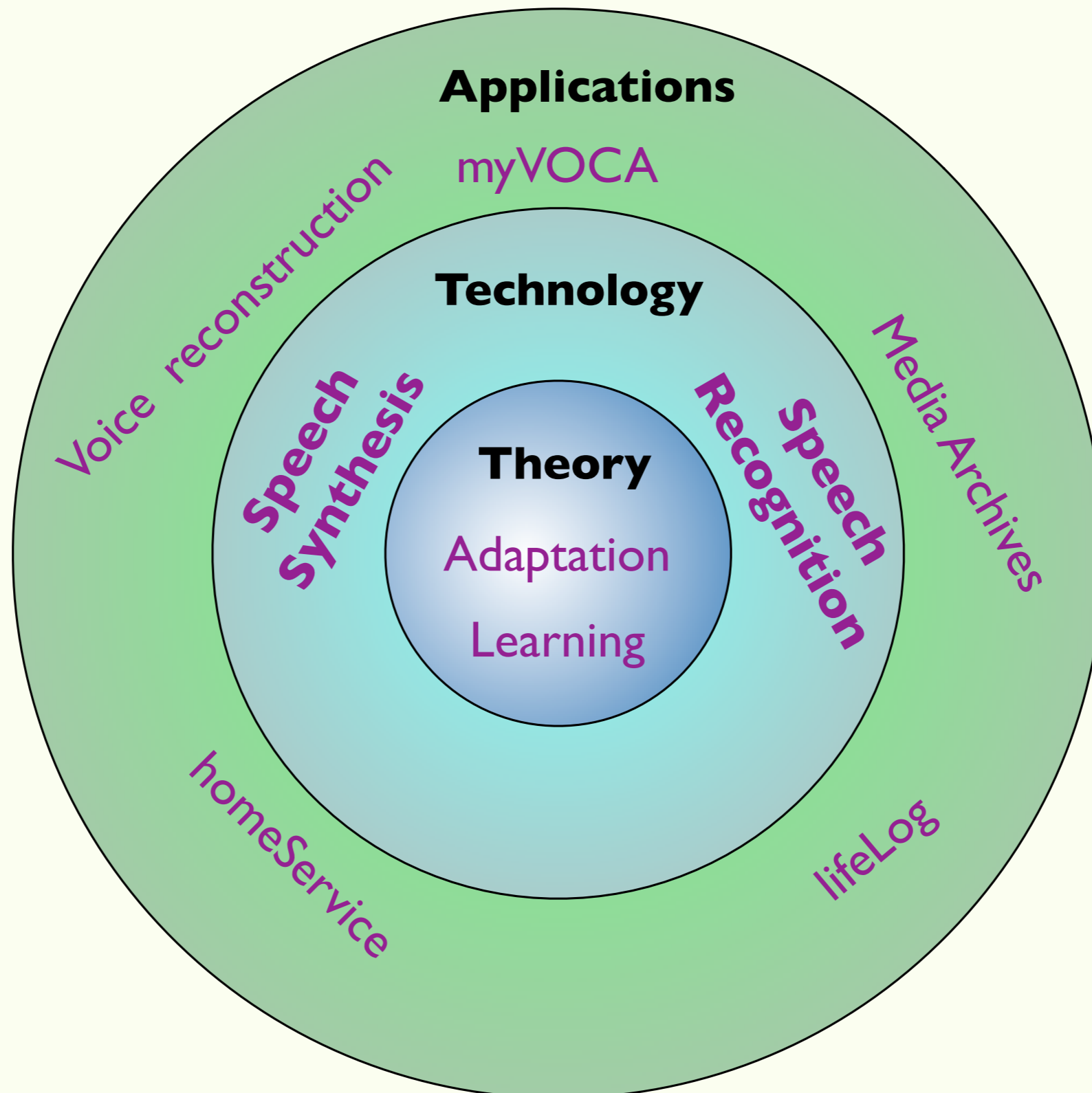
NST team

- **CSTR, University of Edinburgh:**
 - Steve Renals, Simon King, Junichi Yamagishi
 - Peter Bell, Arnab Ghoshal, Jonathan Kilgour, Heng Lu, Liang Lu, Tom Merritt, Pawel Swietojanski, Christophe Veaux, Mirjam Wester
- **Speech Research Group, University of Cambridge:**
 - Phil Woodland, Mark Gales, Bill Byrne
 - Pierre Lanchantin, Xunying Liu, Matt Shannon, Marcus Tomalin
- **Speech and Hearing Research Group, University of Sheffield:**
 - Thomas Hain, Phil Green, Stuart Cunningham
 - Heidi Christensen, Mortaza Doulaty, Charles Fox, Yulan Liu, Oscar Saz

User Group Members

- Barnsley Hospital, Medical Physics & Clinical Engineering
- Devices for Dignity
- Euan MacDonald Centre for Motor Neurone Disease Research
- NIHR CLAHRC for South Yorkshire
- Toby Churchill Ltd
- BBC Future Media & Technology
- Cereproc
- Cisco Systems
- EADS UK
- GCHQ
- Novauris
- Nuance Communications
- Quorate Technologies
- Red Bee Media
- Toshiba Research Europe

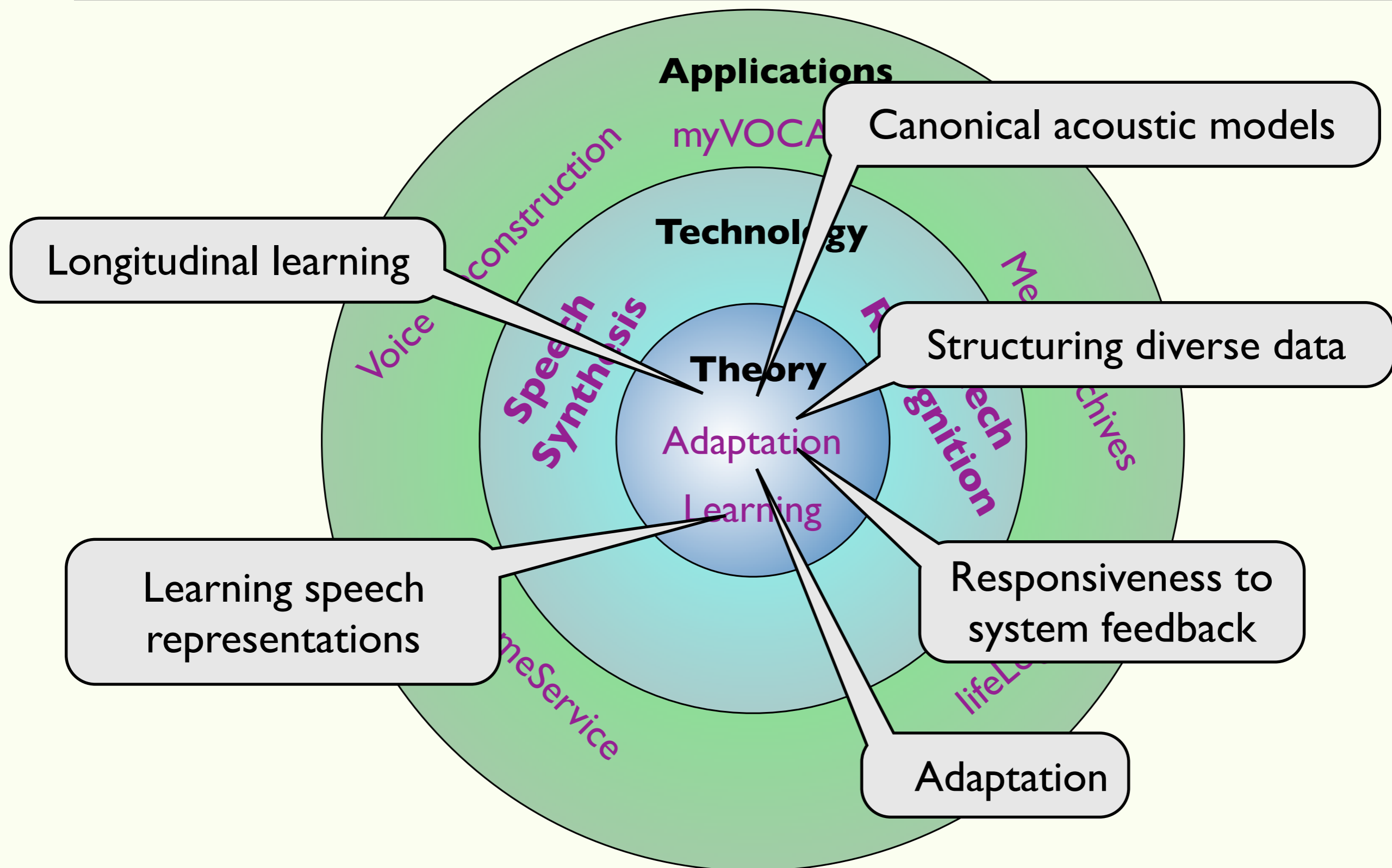
NST Research Areas



Learning and Adaptation

- Speech recognition and speech synthesis based on learning statistical models from data
- Current systems can adapt to the speaker or the domain automatically
- Challenges (for both recognition and synthesis)
 - **Factoring models** to different causes of variability
 - Almost **instantaneous adaptation**
 - **Unsupervised training** to take advantage of available data
 - **Learning not to repeat mistakes**
 - Automatically **learning representations** of speech

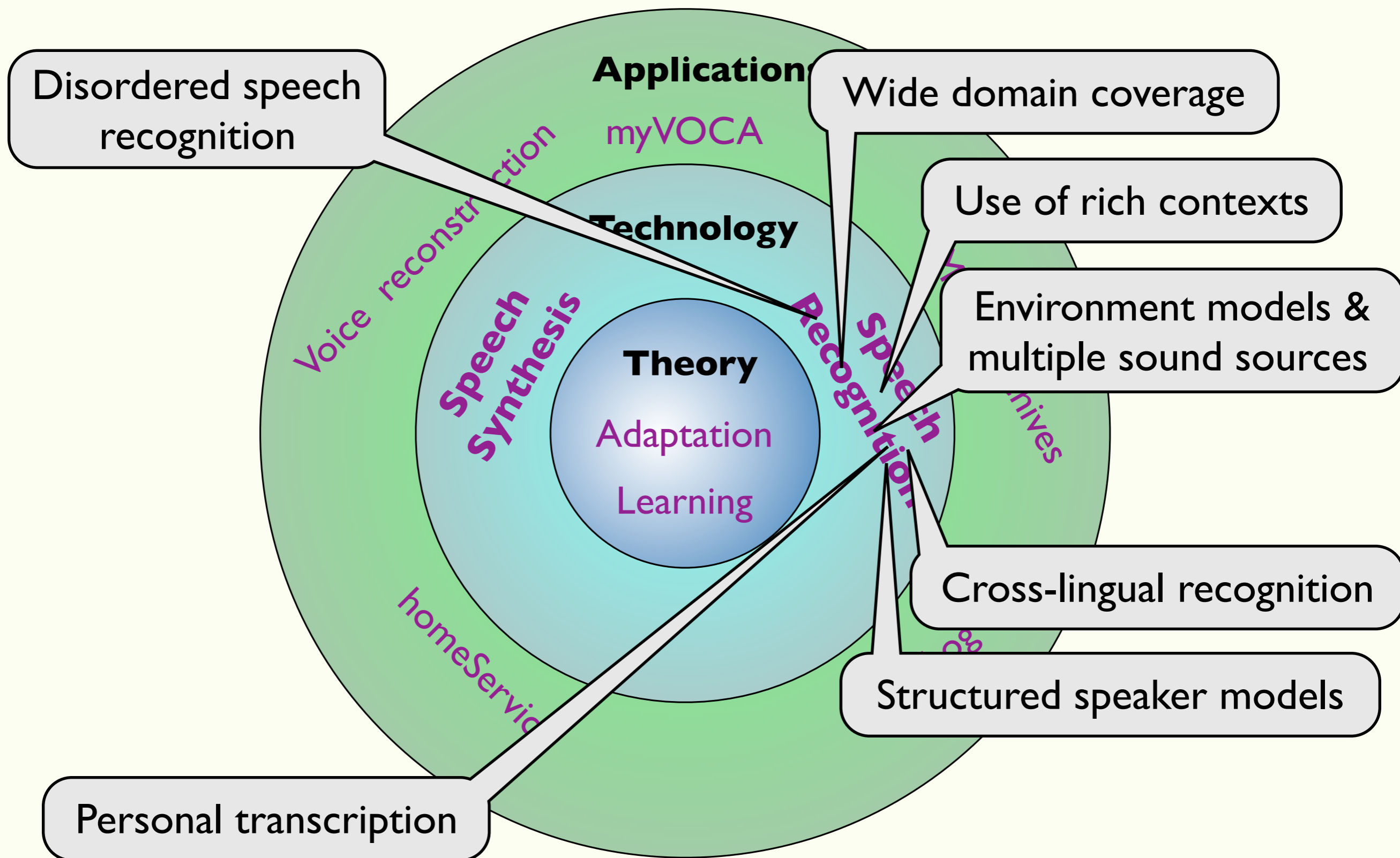
Learning and adaptation Projects



Natural Transcription

- Goal within NST – Speech recognisers that
 - output “who spoke what, when, and how”
 - give high accuracy
 - have a wide coverage of speaker, environment etc
 - are flexible and minimise in-domain training data needs
 - can be personalised
 - produce fluent output
- The “universal speech recogniser”

Speech Recognition Projects



Speech recognition Highlights

- Neural network based acoustic models
 - *Combining neural network acoustic and language models for lecture transcription (talk)*
 - *Sequence-discriminative training of deep neural networks (talk)*
 - *Multi-level adaptive networks in tandem and hybrid ASR systems (poster)*
 - *Cross-lingual knowledge transfer in DNN-based low-resource LVCSR (poster)*
- Neural network based systems give us significantly lower error rates on every task we have worked on in the past year (BBC recordings, TED talks, Globalphone, Switchboard)

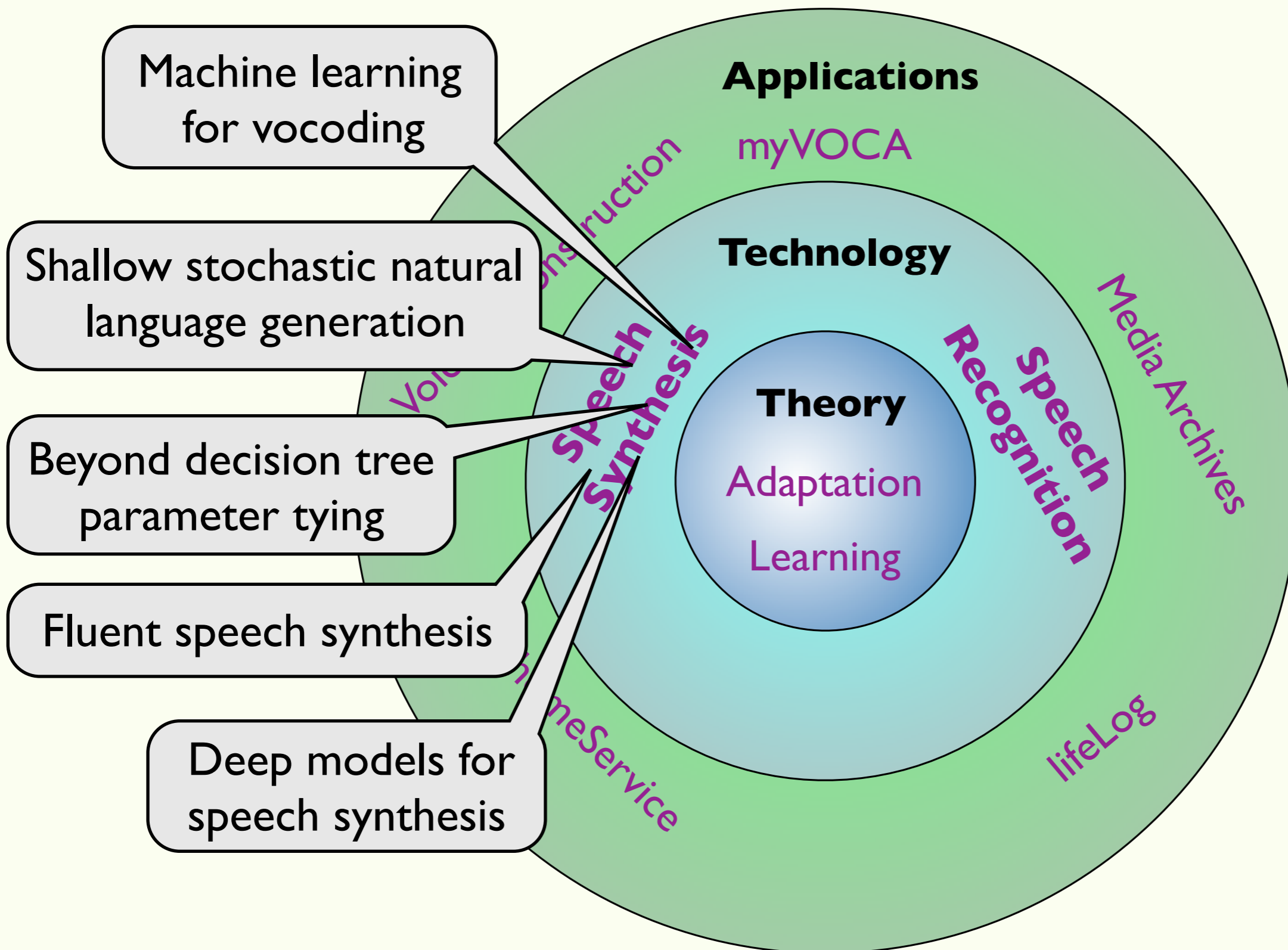
Speech recognition Highlights

- Language models
 - *Cross-domain paraphrasing for improving language modelling using out-of-domain data (talk)*
- Pronunciation models
 - *Acoustic data driven pronunciation lexicon for speech recognition (talk)*
- Acoustic factorisation
 - *Asynchronous factorisation of speaker and background in speech recognition (talk)*
- Disordered speech
 - *A comparative study of adaptive, automatic recognition of disordered speech (poster)*

Natural Synthesis

- Long term vision: Fully controllable speech synthesis, indistinguishable from a human voice, with high intelligibility in all acoustic conditions.
- Goals within NST
 - Statistical parametric synthesis,
 - controllable in terms of speech parameters,
 - adaptable without new data,
 - personalisable with minimal data,
 - high degree of expressivity if required.

Speech Synthesis Projects



Speech synthesis Highlights

- HMM-based speech synthesis
 - *Multiple-average-voice-based speech synthesis (talk)*
 - *Deep neural network for speech synthesis (talk)*
 - *Fast, low-artifact speech synthesis considering global variance (poster)*
 - *A grapheme-based method for automatic alignment of speech and text data (poster)*
- Demos
 - *Touchscreen accent interpolation*
 - *Kinect-based speech synthesis*

homeService and clinical applications

- homeService
 - *The NST homeService application: recent system and experimental developments (talk)*
 - Demo + android app
- Voice banking and reconstruction
 - *An update on voice banking and voice reconstruction (talk)*
 - Android app

Summary

- Major achievements in year 2
 - state of the art deep neural network acoustic modelling systems
 - significant effort in development of approaches and infrastructure to cross-domain ASR
 - multiple-average-voice-model-based speech synthesis to closely match target speaker characteristics
 - investigations of the dimensions of variation in the voice bank data (age, gender, accent, ...) combining meta-data and acoustics
 - homeService – demo system, and ASR of disordered speech