The MGB Challenge at IEEE ASRU-2015

Natural Speech Technology

Edinburgh – Cambridge – Sheffield

Peter Bell, Pierre Lanchantin,
Oscar Saz, Jonathan Kilgour,
Phil Woodland, Mark Gales,
Thomas Hain, Steve Renals

28 May 2015

What is the challenge?

- The **Multi-Genre Broadcast** challenge
- An official challenge at this year's IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)
- Proposed and organised jointly by the three NST sites in collaboration with the BBC
- Four tasks related to speech recognition and speaker diarization of wide-domain TV output

ASRU 2015
2015 IEEE Automatic Speech Recognition and Understanding Workshop
December 13-17, 2015 • Scottsdale, Arizona, USA
IEEE
IEEE Signal Processing Society

Motivations (1)

- Create a challenge in core speech recognition research that is open to all

Motivations (1)

- Create a challenge in core speech recognition research that is open to all
- Establish common datasets and performance benchmarks on broadcast data

Motivations (1)

- Create a challenge in core speech recognition research that is open to all
- Establish common datasets and performance benchmarks on broadcast data
- Encourage researchers to work on the challenging but highly applicable broadcast media domain

Motivations (1)

- Create a challenge in core speech recognition research that is open to all
- Establish common datasets and performance benchmarks on broadcast data
- Encourage researchers to work on the challenging but highly applicable broadcast media domain
- Promote wide take-up of NST research themes and outputs

Motivations (2)

- Binds together several different strands of the NST project in a defined task

Motivations (2)

- Binds together several different strands of the NST project in a defined task
- Design a challenge where we can fully exploit the work we have done on broadcast media since the start of the project

Motivations (2)

- Binds together several different strands of the NST project in a defined task
- Design a challenge where we can fully exploit the work we have done on broadcast media since the start of the project
- Allow us to benchmark our research outputs against competitive systems from other institutions

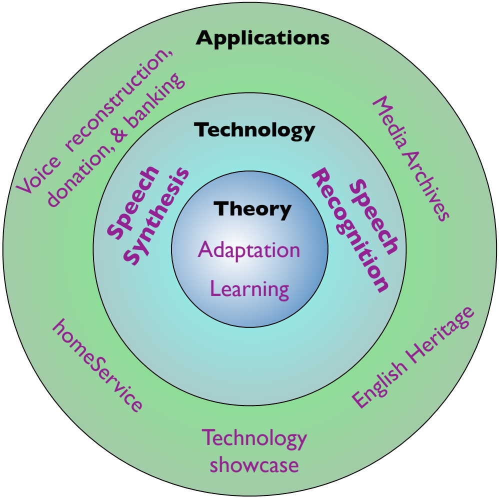
Motivations (2)

- Binds together several different strands of the NST project in a defined task
- Design a challenge where we can fully exploit the work we have done on broadcast media since the start of the project
- Allow us to benchmark our research outputs against competitive systems from other institutions
- Showcase our work on lightly supervised alignment

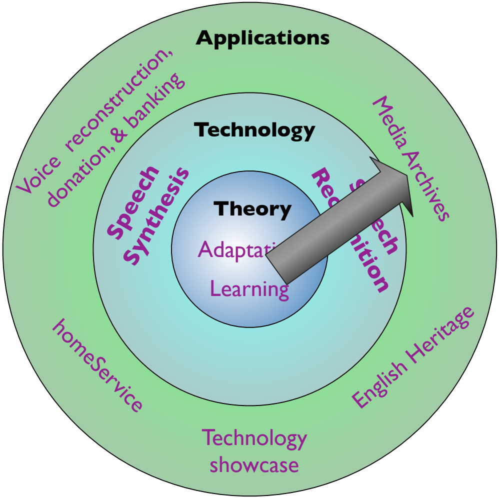
Motivations (2)

- Binds together several different strands of the NST project in a defined task
- Design a challenge where we can fully exploit the work we have done on broadcast media since the start of the project
- Allow us to benchmark our research outputs against competitive systems from other institutions
- Showcase our work on lightly supervised alignment
- Advance the state of the art in broadcast transcription

In NST


Applications
Voice reconstruction, donation, & banking
Media Archives
Technology
Speech Synthesis
Speech Recognition
Theory
Adaptation
Learning
homeService
English Heritage
Technology showcase

In NST



Applications
Voice reconstruction, donation, & banking
Media Archives
Technology
Speech Synthesis
Speech Recognition
Theory
Adaptation
Learning
homeService
English Heritage
Technology showcase

Fit with core NST themes

Applications

- Broadcast media

Natural transcription

- Wide domain coverage
- Use of rich contexts
- Longitudinal learning

Learning and adaptation

- Canonical acoustic and language models
- Structuring diverse data
- Lightly-supervised training

Principles of task design

- Standardise all training data so that differences are due to difference in algorithms and techniques, not resources

Principles of task design

- Standardise all training data so that differences are due to difference in algorithms and techniques, not resources
- Supply data pre-processing and baseline system-building recipes enabling participants to focus their time on novel research

MGB CHALLENGE

Principles of task design

- Standardise all training data so that differences are due to difference in algorithms and techniques, not resources
- Supply data pre-processing and baseline system-building recipes enabling participants to focus their time on novel research
- Low barriers to entry, even for sites who have not worked on English ASR

Principles of task design

- Standardise all training data so that differences are due to difference in algorithms and techniques, not resources
- Supply data pre-processing and baseline system-building recipes enabling participants to focus their time on novel research
- Low barriers to entry, even for sites who have not worked on English ASR
- Enable building state-of-the-art systems with immediate practical input through use of very large amounts of training data

Principles of task design

- Standardise all training data so that differences are due to difference in algorithms and techniques, not resources
- Supply data pre-processing and baseline system-building recipes enabling participants to focus their time on novel research
- Low barriers to entry, even for sites who have not worked on English ASR
- Enable building state-of-the-art systems with immediate practical input through use of very large amounts of training data
- Allow participants to choose which research areas to focus on: improving training data alignment, diarization/segmentation, genre adaptation, core algorithms...

The sub-tasks (1)

1. **Transcription of multi-genre TV shows**
   - we supply around 16 TV shows to be completely transcribed
   - show names and genre labels are provided
   - some shows are from series appearing in the training data; some are not

2. **Subtitle alignment**
   - for the same shows as Task 1, the subtitle text as originally broadcast are provided (with some automatic tokenisation applied)
   - these may differ from the verbatim audio content for a range of reasons
   - participants must produce time stamps for all words in the subtitles

The sub-tasks (2)

1. **Longitudinal transcription**
   - aim to evaluate ASR in a realistic longitudinal setting
   - participants will transcribe complete TV series, where the output from shows broadcast earlier may be used to adapt and enhance the performance of later shows
   - evaluation data will consist of a two complete TV series
2. **Longitudinal diarization and speaker linking**
   - participants aim to label speakers uniquely across a complete series
   - realistic longitudinal setting again: participants must process shows sequentially in date order

Training data

- 1,600 hours of TV, taken from 7 complete weeks of BBC output over four channels, with accompanying subtitle text
- 600M words of subtitle text from 1988 onwards
- XML metadata for all shows, generated in a standard format developed earlier in the NST project
- Data supplied to each participant subject to a license agreement between them and the BBC, for the purpose of participation in the challenge

Additional resources

To reduce the workload of system building, we also supply:

- Lightly supervised alignments of the original subtitles to the audio, produced in Cambridge (see Pierre for details)
- The Combilex British English pronunciation dictionary, previously developed at Edinburgh, plus 10,000+ automatically-generated extra pronunciations for words in the subtitle text
- Baseline speech segmentations and speaker clustering from Sheffield
- Automatic tokenisations of subtitle text from Cambridge and Edinburgh

We have also carried out careful verbatim transcription of test data, around 40 hours in total

Kaldi recipe

- Kaldi is a widely-used open-source toolkit for ASR

Kaldi recipe

- Kaldi is a widely-used open-source toolkit for ASR
- It includes easy-to-use recipes for quickly building systems on standard datasets used by the community (NST has previously contributed to these)

Kaldi recipe

- Kaldi is a widely-used open-source toolkit for ASR
- It includes easy-to-use recipes for quickly building systems on standard datasets used by the community (NST has previously contributed to these)
- We have made available a recipe to enable participants to quickly build a basic (GMM+SAT) system without needing a detailed knowledge of the data

Kaldi recipe

- Kaldi is a widely-used open-source toolkit for ASR
- It includes easy-to-use recipes for quickly building systems on standard datasets used by the community (NST has previously contributed to these)
- We have made available a recipe to enable participants to quickly build a basic (GMM+SAT) system without needing a detailed knowledge of the data
  - ○ scores 49% WER on the official development set – there is room for improvement (but this is a naturally very challenging domain)

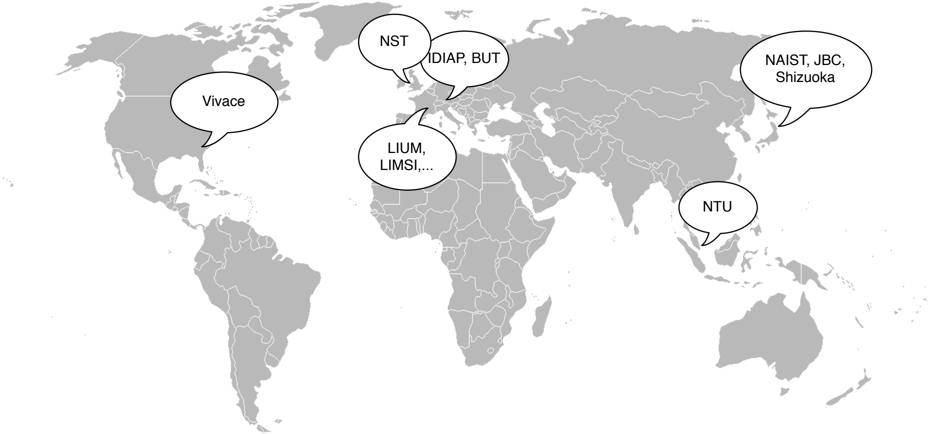
Kaldi recipe

- Kaldi is a widely-used open-source toolkit for ASR
- It includes easy-to-use recipes for quickly building systems on standard datasets used by the community (NST has previously contributed to these)
- We have made available a recipe to enable participants to quickly build a basic (GMM+SAT) system without needing a detailed knowledge of the data
  - scores 49% WER on the official development set – there is room for improvement (but this is a naturally very challenging domain)
- A full state-of-the-art recipe will be available in due course

Potential areas for research

- method to improve use of unclean training data labels
- adaptation to genre and structuring multi-genre data (see Mortaza's poster)
- processing data with diverse noise sources
- factorised approaches to learning and adaptation
- investigating scalability of training algorithms to large data
- how well do current cutting-edge techniques work on realistic data?

Around 20 participants...



Thank you

...and special thanks to the Andrew McParland and the BBC!

Find out more at http://mgb-challenge.org