Examples of BBC data processing

Pierre Lanchantin, Yanhua Long, Andrew Liu, Mark Gales, Phil Woodland
Cambridge University Engineering Department

Natural Speech Technology
Edinburgh – Cambridge – Sheffield

UNIVERSITY OF CAMBRIDGE

- We processed a large diversity of shows from the BBC with their associated raw transcriptions
- Our goal is to automatically extract useful information from the raw data and to produce complete, accurate and reliable transcriptions
- This task is central for the design of ASR & TTS systems

Time stamp correction

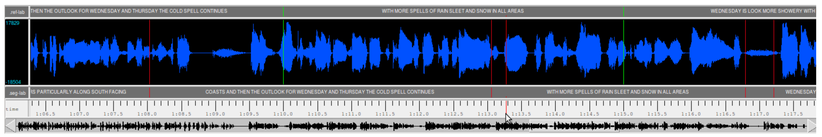


Figure: Time-stamp correction: top: original ts, bottom: corrected

- Positions of time-stamp can be inaccurate due to:
  - quantisation (1s, 5s...)
  - show-specific time lag
  - human error
- We show different ways to detect and correct these issues

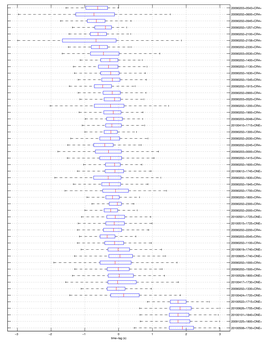


Figure: Show-specific time-lag

Decoding

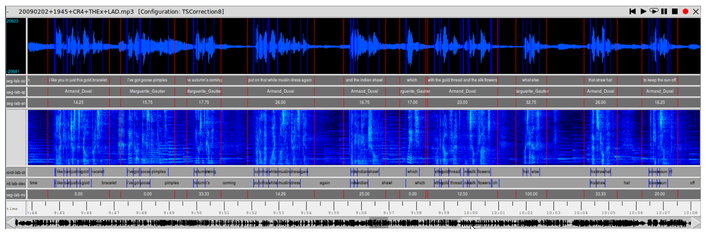


Figure: Alignment and extraction of metadata

- Performance of the ASR system depends strongly on the nature of the show. We present:
  - different decoding results and associated refined transcriptions for different nature of shows
  - different show-specific Matching Error Rate profiles

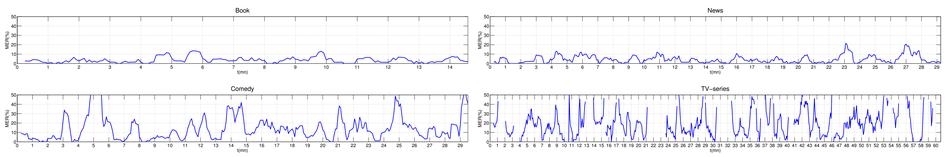
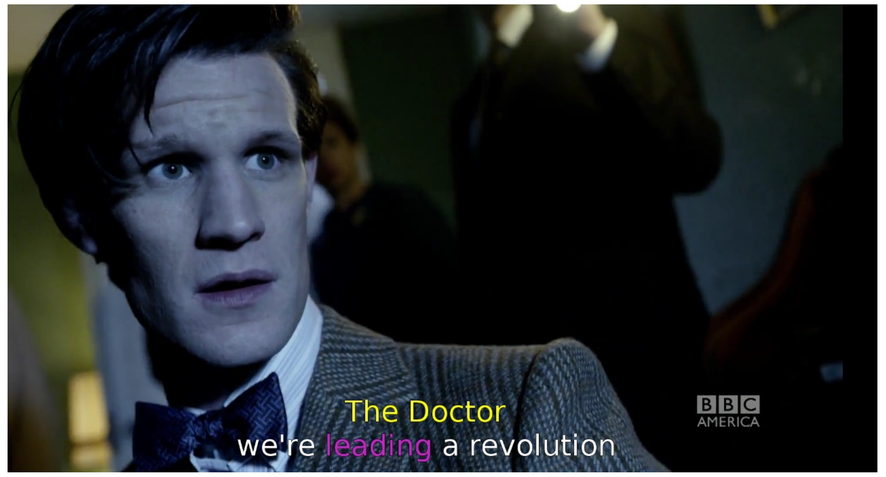


Figure: Typical MER show profiles (30s sliding window)

Trailer Decoding & Alignment



- We tested our ASR system on a TV-series trailer from the BBC youtube channel. This trailer includes expressive speech, multiple speaker, music and sound effects.
- 2 results are presented:
  - Decoding using the LM and acoustic models trained on the BBC data (Err=58.47%)
  - Alignment

TTS from unstructured data

- We selected speech from several shows based on SNR and MER
- We then trained models for HMM-based speech synthesis using the Edinburgh CSTRVoiceClone system

Database

- Speaker: Alec Broers
- Extracted for 6 Reith lectures (3:30h)

Data Selection

- SNR $>$ 35dB
- Matching Error Rate $<$ 25%
- Selection of 1:30h of speech