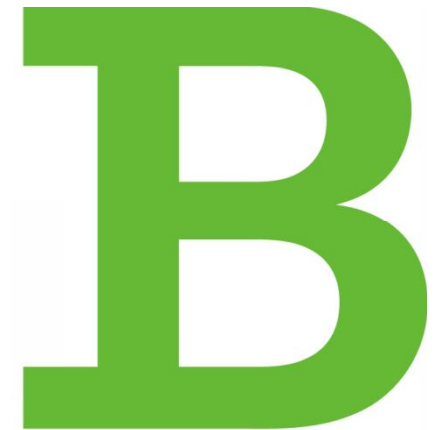




Speech and Language Technologies and Beyond

Martin Russell

School of Electronic Electrical and Computer
Engineering



Content

- Current research issues in automatic speech recognition (ASR)
- Automatic recognition of children's speech
- Paralinguistic speech technologies
- Applying algorithms from ASR to other problems
- The Speech Ark Limited

Automatic Speech Recognition

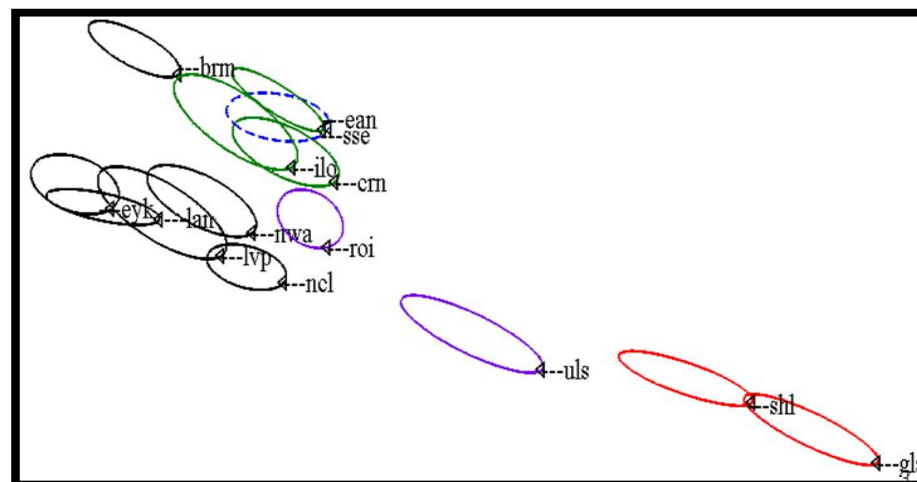
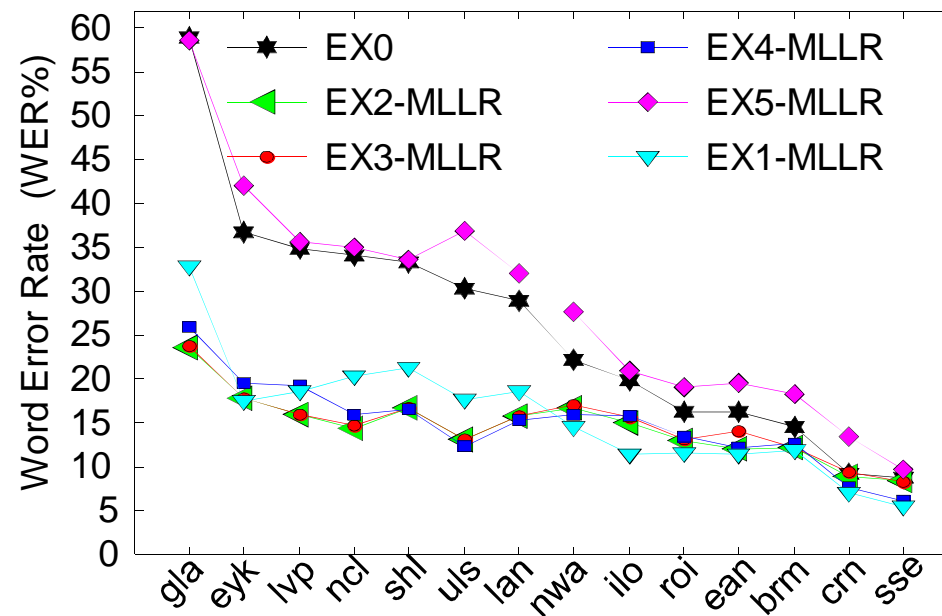
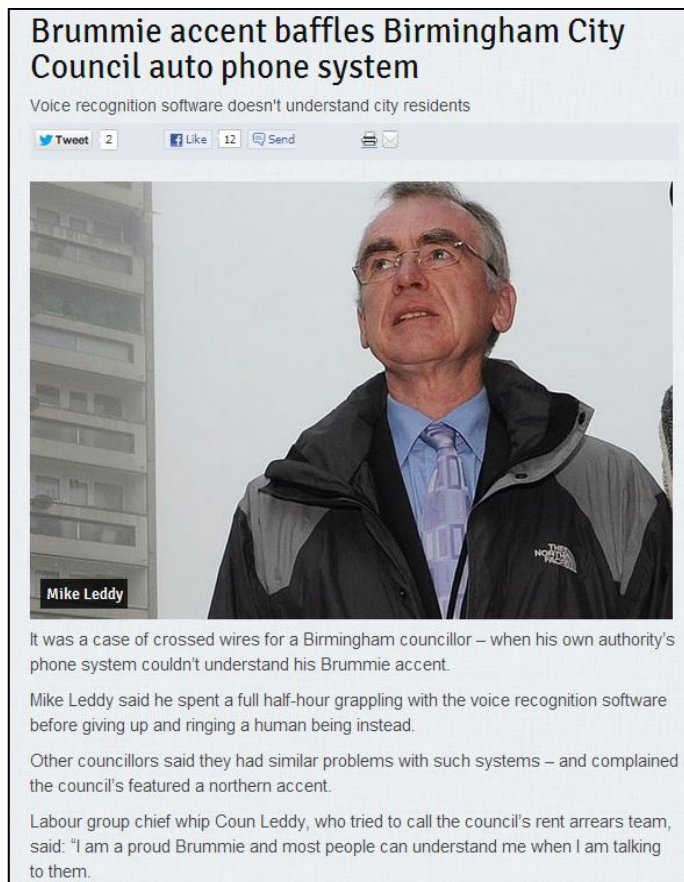
- Progress over past three decades
 - Huge progress in terms of performance (e.g. dictation systems)
 - More modest progress in terms of underlying methods – continued reliance on large statistical models
- What's new?
 - Concerns over performance and flexibility – under-resourced applications
 - New paradigms – e.g. Deep Belief/Neural Networks (DBNs / DNNs)
- “Speech Recognition by Synthesis” project
 - 3 EECE/industrial staff, 1 Research Fellow, 2 PhDs
 - Towards more compact “parsimonious” speech models that really model speech



Automatic Recognition of Children's Speech

- Automatic speech recognition error rates much higher for children than for adults' (even for 'matched' ASR) – *Wilpon and Jacobsen, 1996*
- Characteristics of children's speech
 - Increased variability relative to adults' speech
 - Shorter vocal tract (shifted frequency structure)
 - Higher fundamental frequency
 - Developing phonetic and linguistic skills
- Focus of majority of research to-date: adapting children's speech to work with 'adult' speech recognisers
- New project with Disney Research, USA
- Speech recognition for children – development of computationally useful models of language development in children

Accent-robust Speech Recognition



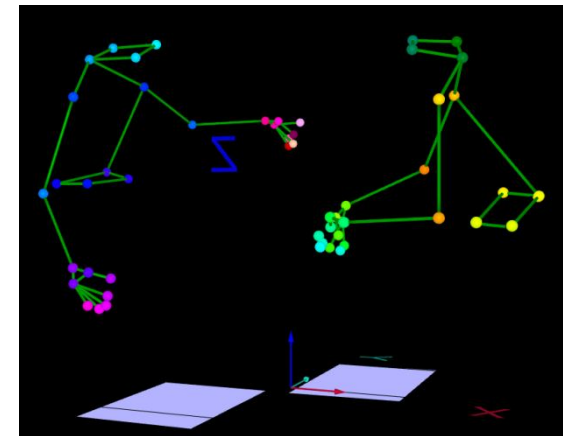
Paralinguistic speech technologies

- Paralinguistic – properties of a speech signal over and above its linguistic content
 - Examples: Language identification, speaker identification

Problem	%Acc. - Computer	% Acc. – Human
Accent ID (14 accents, British English) – unknown text	86.4%	-
Accent ID (14 accents, British English) – known text	95.2%	58.2%
Ethnic group ID (Birmingham ‘Asian’ and ‘White’)	94.3%	93.3%
Gender ID (Children)	79.2%	67%
Speaker ID (Child in a ‘class’)	90%-99%	-
Speaker ID (Child in a ‘school’)	81%	-

Beyond speech and language technology

- Speech and language recognition technologies are generic:
 - Speech recognition = Generic statistical sequential pattern recognition
 - Dialogue modelling = statistical task modelling
- Other applications
 - Gaze: eye-tracker data
 - integration of speech and gaze
 - 3D body motion data
 - Biometrics, integration of speech and gesture
 - Birdsong
 - Environment monitoring 'biocensus'
 - Human activity recognition
 - CogWatch



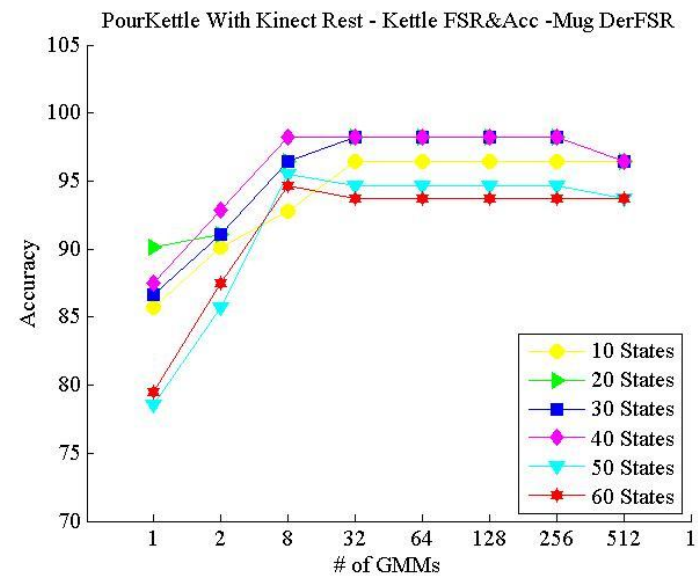
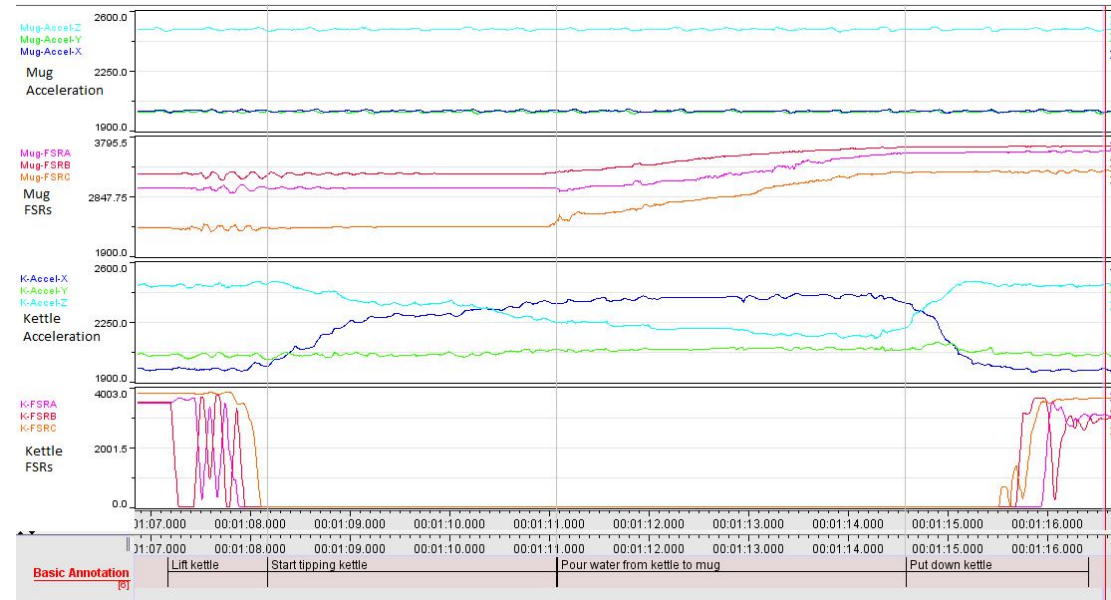


CogWatch

- European Commission (FP7) funded research project
- Rehabilitation of stroke patients:
 - Many experience long term physiological and/or cognitive disabilities.
 - A significant proportion suffer from Apraxia or Action Disorganisation Syndrome (AADS) - impairment of cognitive abilities to carry out activities of daily living (ADL).
 - Current focus is tea-making
- CogWatch objective:
 - Develop technology to automatically track patient's progress through tea making
 - Provide feedback and cues as appropriate
- Technology
 - Instrumented Objects
 - 3D motion tracking (MS kinect)
 - Pattern recognition techniques from speech and language processing



CogWatch





The Speech Ark

- Spin-out from University of Birmingham, 2008
- Creates and sells speech corpora for speech technology research and development
- Sales:
 - Academic “research purposes only” licences to universities in UK, Europe, India, Australia
 - Commercial licences to company in USA
 - Currently negotiating with research lab in Japan

Any questions?