

# Speech recognition and understanding acoustic scenes

Steve Renals

<http://homepages.inf.ed.ac.uk/srenals/>

# The Cocktail Party Problem

- “How do we recognise what one person is saying when others are speaking at the same time?” – Colin Cherry, 1953
- “A few of the factors might be:
  - ...different directions
  - lip reading, gestures ...
  - different voices
  - accents differing
  - transition probabilities”

# Research Challenges

- Analysis
  - Source separation – Identify & separate talkers and other acoustic sources
  - Speech recognition – Who spoke what, and when, and how?
  - Attention
  - Speech understanding
  - Conversational understanding
- For any language, acoustic environment and task domain

# Research Challenges

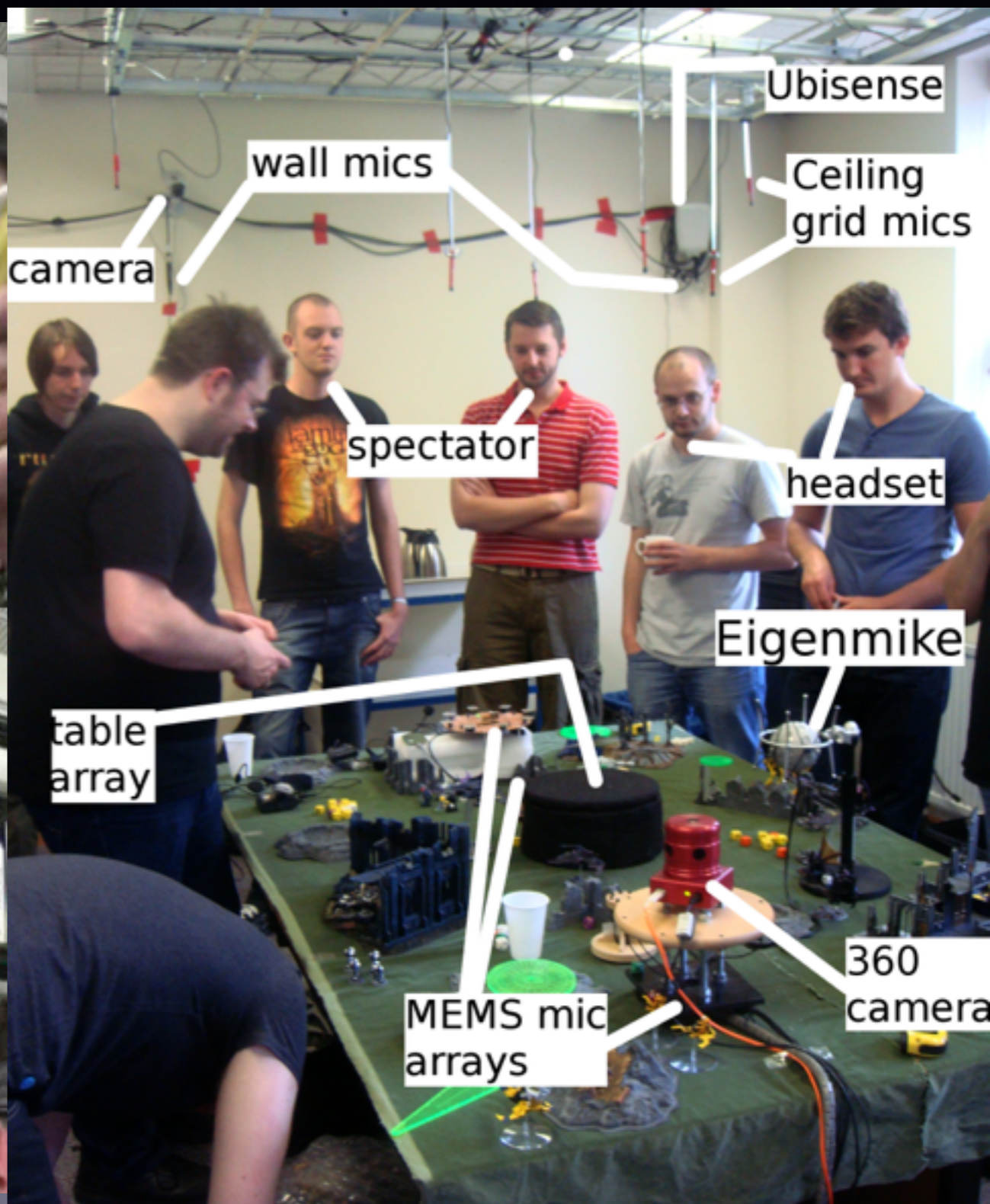
- Analysis
  - Source separation – Identify & separate talkers and other acoustic sources
  - Speech recognition – Who spoke what, and when, and how?
  - Attention
  - Speech understanding
  - Conversational understanding
- *Spoken language generation and conversational participation*

# Distant speech recognition

# Distant speech recognition



# Distant speech recognition



# Distant speech recognition





# Speech Recognition Research challenges

- Robust to noise, reverberation, overlapping talkers.....
- Semi-supervised and unsupervised training, cross-lingual models
- Models which factor different causes of variability
- Multimodal recognition and understanding

# Multi-Genre Broadcast Speech Recognition



# Deep Learning

- Coherently optimising many sub-tasks and modules
- End-to-end systems are exciting, but...
  - components require different data at different scales
  - changing one component leads to global re-optimisation
- Task-specific neural networks jointly trained
  - different modelling and optimisation approaches
  - reuse individual components in speech, NLP, ...

# Speech recognition and understanding acoustic scenes

Steve Renals

<http://homepages.inf.ed.ac.uk/srenals/>