Computational Audio Analysis

*Interpreting Intentional Behaviour*

**Prof. Roger K. Moore**

Chair of Spoken Language Processing
Dept. Computer Science, University of Sheffield, UK
(Visiting Prof., Dept. Phonetics, University College London)
(Visiting Prof., Bristol Robotics Lab.)

---

**Motivation**

- Lots of excellent research on *(computational)* auditory scene analysis
- But we have not yet developed a sufficiently sophisticated theoretical framework for general-purpose analysis of audio scenes
- Much existing research has been focused on classifying the *surface* phenomena associated with acoustic events
- Little attention has been given to interpreting the underlying *intentional behaviour* associated with living organisms
- Intentionality is a particularly important driver of communicative behaviour *(such as speech)*
The ‘design’ stance

“Things do what they are supposed to do”

The ‘intentional’ stance

“Things do what they do on purpose”

The ‘physical’ stance

“Things obey the laws of physics and do what they do”

http://www.soundjay.com/
Interpreting Intentional Behaviour

Human beings are exceptionally good at assigning motives (agency) to anything …


Modelling Intentional Behaviour

- Could use a traditional Bayesian approach …

\[
\Pr(\text{intentions} \mid \text{behaviour}) = \frac{\Pr(\text{behaviour} \mid \text{intentions}) \Pr(\text{intentions})}{\Pr(\text{behaviour})}
\]

- We just need the usual vast quantity of appropriately annotated data, and away we go!

- Unfortunately this simple model doesn’t really capture the reality of the behaviours involved
Dependencies

- Intentional behaviour implies a ‘coupling’ between an agent and its environment (which may include other agents)
- This means that communicative behaviours (such as speech) are highly conditioned on the context
- Also, living systems constantly strive to minimise physical (and neural) effort
- Hence, interaction is an actively managed competition between effort and effectiveness
- So where can we turn for inspiration?
  - agent-based modelling
  - cognitive neuroscience

Agent-Based Modelling

Beliefs, Desires and Intentions (BDI) architecture

Agent-Based Modelling

- BDI is a proven technology for agent-based modelling, but it doesn’t specify how to recognise/interpret behaviour.
- Nevertheless, it does appear to capture some important features of ‘intelligent’ behaviour …
  - Beliefs ≡ priors (≡ memory)
  - Desires ≡ goals
  - Intentions ⇒ planning

Cognitive Neuroscience

- "recognising the intentions of others is based … on the same mechanisms underlying the formation of one’s own motor intention”
- Pr(intentions | behaviour) = \frac{Pr(behaviour | intentions) Pr(intentions)}{Pr(behaviour)}
- Possible mechanism … ‘Mirror Neurons’
Mirror Mechanisms

It is hypothesised that projecting ‘self’ onto ‘other’ facilitates …
– interpretation of intentions
– action understanding
– imitation
– learning (by imitation)
– empathy
– Theory of Mind (ToM)
– gestural communication
– evolution of speech & language


So … back to basics!

Can we agree on a primitive conceptual framework?
The evolution of physical events

Continuous cycle of cause and effect

Realise intentions behaviour
Interpret meanings observations beliefs motivations needs attention desires

Agent

Degree of effort expended depends on factors such as urgency (i.e. emotional ‘arousal’)

© 2013 The University of Sheffield
Dagstuhl Seminar: Computational Audio Analysis 3-8 Nov. 2013 slide 13

© 2013 The University of Sheffield
Dagstuhl Seminar: Computational Audio Analysis 3-8 Nov. 2013 slide 14
Another intentional agent

REALISE
- intentions
- desires
- needs

INTERPRET
- attention
- meanings
- observations

AGENT

WORLD

Another intentional agent

TRANSFORM
- causes
- effects

The accuracy with which the causes can be estimated depends on the fidelity of the inverse transform
WORLD

AGENT

causes → SEARCH → causes

TRANSFORM

WORLD

AGENT

causes → TRANFORM → effects

SEARCH

 Negative-feedback control loop matches output of a forward model with observed effects

Intended actions and/or consequences may or may not occur (depending on the fidelity of the transforms)
An agent manipulating the world

Error signal reflects the agent's appraisal of its actions (i.e. emotional valence)

Negative-feedback control loop ensures consequences match intentions (even with unpredictable disturbances)

Control loop emulates consequences of actions (when consequences cannot be observed)
An agent manipulating the world

**AGENT**

Intentions → SEARCH → TRANSFORM

actions → TRANSFORM → consequences

Predictive control loop emulates (imagines) consequences of actions prior to action selection

An agent manipulating another agent

**AGENT**

Intentions → SEARCH → TRANSFORM

Degree of search effort depends on agent-agent relationship (i.e. social 'dominance')

**AGENT**

Behaviour → TRANSFORM → Intentions

Synthesis-by-Analysis
An agent interpreting another agent

ARTIST

behaviour

SEARCH

TRANSFORM

WORLD

AGENT

intentions

TRANSFORM

One signalling, the other interpreting

ARTIST

behaviour

SEARCH

TRANSFORM

WORLD

AGENT

intentions

TRANSFORM

So could this!!

This agent's model of the 'other' agent could be constructed from a model of itself!!
One signalling, the other interpreting

AGENT

WORLD

Intentions → SEARCH → Transform

SEARCH

Analysis-by-Synthesis

Synthesis-by-Analysis

Search

An intentional agent

AGENT

WORLD

Motivations

Intentions

Beliefs

Meanings

Needs → Desires → Behaviour

Behaviour → Attention → Observations

Agents' intentions, world's interpretations.
The whole enchilada!

WORLD

Issues for Discussion

- How important is it that we acknowledge that the world contains intentional agents?
- Can we envisage a unified computational modelling approach which is capable of integrating the relevant contingencies?
- What are the implications of modelling 'self'?
  - recursion?
  - context dependency?
  - embodiment?
- Can an agent ever understand a natural scene if it is not (or has never been) part of it?
- Bottom line … What does an automated agent need to know about the world and the entities it contains in order to make sense of a general audio scene?
Thank You

If you’ve been affected by any of the issues raised in this talk

please contact me at

r.k.moore@dcs.shef.ac.uk

or have a look at this …