Some insights into talker-listener-environment coupling, energetics and the contrastive particulate structure of spoken language

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 Laboratory)

Local Timescale Variability Analysis

H&H Theory

"The speaker ... dynamically tunes the production ... for either output-oriented control (hyperspeech) or system-oriented control (hypospeech)."

"Speakers can, and typically do, tune their performance according to communicative and situational demands."

"What he/she needs to control is ... that their signal attributes possess sufficient contrast ... discriminative power that is sufficient for lexical access."

"... the lack of invariance that speech signals commonly exhibit is a direct consequence of this adaptive organization ..."

Evidence for H&H Behaviour

- People naturally tend to speak louder/differently in noise (Lombard, 1911)
- Caregivers talk differently to children (Fernald, 1985)
- Users talk differently to machines (Moore & Morris, 1992)
- Being able to hear your own voice has a profound effect on speaking (as evidenced by the need for sidetone on a telephone)
- Hearing-impaired individuals can have great difficulty maintaining clear pronunciations (or level control)
- Delayed auditory feedback causes stuttering-like behaviour
- Altered auditory feedback evokes compensations (Munhall et al, 2009; MacDonald et al, 2011)
- People with speaking difficulties (e.g. caused by cerebral palsy) report that it takes immense effort to produce even the simplest utterance
Speech Energetics


‘AnTon’ – Animatronic Tongue
What Principle is Operating Here?

- The effort/energy involved in production is being traded against the clarity/discriminability of the end product.
- Why? Because living systems have to move in order to signal, and movement costs energy.
- Large movements create clear signals but use more energy (and vice versa).
- The efficient management of energy is a major evolutionary factor in the behaviour of living systems (not just for signalling).
- Sometimes large movements are necessary due to the need to overcome an obstacle in the environment.

Analogy with Walking

Which situations are ‘natural’?

- Flat predictable surface
- Rough unpredictable surface
- Predictable obstacle
- Unpredictable obstacle
Behaviour

“feedback … is the central and determining factor in all observed behavior”


“… principal determinant of a system is not a stimulus, an event in the past, but a future result of the behaviour …”


These are examples of ‘regulatory’ processes …

A Regulatory Mechanism: Feedback Control

Figure 4. Some components of a feedback system.

Perceptual Control Theory (PCT)


Communicative Intentions?

- The ‘target’ is a perception *not* a signal
- So optimisation is over competing perceptions *not* competing signals *(e.g. not signal-to-noise ratio)*
- The intention is sufficient *contrast* at the pragmatic level *(which will lead to suitable compensations at the semantic, syntactic, lexical, phonemic, phonetic and acoustic levels)*
- The obstacles are …
  - alternative interpretations *(internal)*
  - competing signals *(external)*

An Affective Model of Behaviour

Intention

Attention

Consequence

Emotion (valence)

Motivation/ Effort (arousal) $K_p$

Motor Command

Disturbance

Appraisal

Reference Input

Actuating (Error) Signal

Feedforward (Control) Elements $g_1$

Control Signal or Manipulated Variable $u$ or $m$

Plant or Process $g_2$

Feedback Elements $h$

FORWARD PATH

Controlled Output

Attention


Reactive Speech Synthesis


This is ‘synthesis-by-recognition’
Emulation Mechanisms

- Negative feedback control works OK as long as the loop delays are not too high, but neural delays can be quite significant in living systems
- Delays can be overcome using an ‘emulation’ mechanism (pseudo-closed-loop control)


Inhibition of motor commands facilitates mental imagery and planning

Forward model facilitates prediction

Emulate ⇒ Predict ⇒ Understand

- Simulations (forward models) generate top-down expectations/predictions
- The emulation of an organism’s abilities can be recruited to emulate the behaviour of others for:
  - predicting their behaviour
  - mimicry
  - learning
- The (emergent) consequences would be:
  - an ability to interpret the behaviour of other(s) by referencing self
  - an ability to estimate the ‘beliefs, desires and intentions’ of other(s)
  - i.e. recognition and understanding


This is ‘recognition-by-synthesis’
PreSenCE
Predictive Sensorimotor Control and Emulation


PreSenCE

S: n

feeling

intention

S: m

action

S: E(U:m)

S: E(U:E(S:m))

S: E(U:i)

S: E(U:n)

motivation

feeling

sensitivity

attention

interpretation

action

needs

emulation

noise, distortion, reaction, disturbance

Emulation of System’s possible motor actions

Primary route for System’s motor behaviour
PreSenCE

System's emulation of User's emulation of System

System's interpretation of User behaviour

noise, distortion, reaction

© 2012 The University of Sheffield

LISTA Workshop, Edinburgh, 2-3 May 2012
PreSenCE

- **Energy** management
  - effort
  - attention

- **Time** management
  - synchrony
  - coordination

- **Entropy** management
  - complexity
  - combinatorics

- **Emotion** management
  - needs/drives
  - personality/mood
Predict → Optimise → Survive

- Predictive models facilitate planning and optimisation of behaviour (along the dimensions of energy, time, entropy and emotion)
- Predictive models of an organism’s interaction with the environment facilitate efficient manipulation
- Predictive models of an organism’s interaction with another organism facilitate efficient communications
- The fidelity of the predictors governs the degree of efficiencies that are possible
- Matched predictors in two organisms facilitates high levels of coupling with low information rates

Enactivism

A ‘cognitive unity’ (self-regulating self-generating closure)

Enactivism


Emulation

Empathy

Predicting other's affective state(s)

Similar to Dennett's "design stance"


Theory of Mind

Predicting other's beliefs & intentions

Similar to Dennett's "intentional stance"

Languaging

Modelling the ‘recursive particulate coupling’ between one agent and another


Predictive Coding

If a sender knows that the receiver has a predictive model of the talker, ...

… then only the difference between the intentions and the predictions need to be sent (thereby minimising the entropy)
The Listening Talker

- **declarative**
  - change state of other

- **imperative**
  - change state of world

- **interrogative**
  - change state of self

- **dialogue**
  - conversational interaction

These are being optimised by the talker

And Finally …

Would we study walking by suspending someone in the air and asking them to walk?

No? So why do we put people in a recording booth and ask them to speak?

In both cases the subject is obliged to imagine a crucial conditioning aspect of their behaviour.

An appropriate experimental methodology is the key to future progress.
Thanks for Listening to this Talker

http://www.dcs.shef.ac.uk/~roger

Universals in Vocal Behaviour?

http://www.youtube.com/watch?v=E4aMHK7AH5M