From Isolated Digits to Social Agents

How good is good enough?

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.

From Isolated Digits to Social Agents

How good is good enough?

Antonio Zampolli
From Isolated Digits to Social Agents
How good is good enough?

UK Speech Research Unit

From Isolated Digits to Social Agents
How good is good enough?

RSG10: NATO Research Study Group on Speech Processing
From Isolated Digits to Social Agents

How good is good enough?

PDP9-based ASR - University of Essex - 1975
From Isolated Digits to Social Agents
How good is good enough?

From Isolated Digits to Social Agents

How good is good enough?

International Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques (COCOSDA)

Spoken Corpus Recordings in British English (SCRIBE)

EU Expert Advisory group in Language Engineering Standards (EAGLES)

UK Speech Technology Assessment Group (STAG)

EU Speech Assessment Methods (SAM) project

From Isolated Digits to Social Agents

How good is good enough?
From Isolated Digits to Social Agents

How good is good enough?

TAXONOMISE
CHARACTERISE
PREDICT
EVALUATE

Taxonomy of SLT Applications

Real-Time
Non-Interactive
Non Real-Time
Interactive

easy
hard
hard
hard
Command and Control Systems

Dictation Systems

Interactive Voice Response (IVR) Systems

Voice-Enabled Personal Assistants

Embodied Conversational Agents (ECAs)

Autonomous Social Agents (robots)
From Isolated Digits to Social Agents

How good is good enough?

TAXONOMISE

CHARACTERISE

PREDICT

EVALUATE

Capabilities & Requirements

Current & Future Spoken Language Technology

Technical Requirements

Application Requirements

Application Benefits

Technical Features

Technical Capabilities

Application Capabilities

‘suitability’

Current & Future Spoken Language Applications

Technical factors

application factors

The ‘Capability Profile’

Application Requirements (Pull)

‘GOODNESS’

Performance Envelope

Technology Capabilities (Push)

Technology Migration

FEATURES

improve

benchmark

extend

combine

port

optimise

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES
Success

- A technology doesn’t have to be ‘good’ to be successful
- It has to be ‘good enough’, (i.e. better than the alternatives)
Characterising Alternative Input Methods

Main Result: ASR WER < 2xHSR WER

Using Fitts’ Law to predict data entry rates on a PDA


Characterising Autonomous Agents

Manal Linjawi

“What is a Robot?”

Social Intelligence

Cognitive Intelligence

Morphological Intelligence
Robotics 2020 Multi-Annual Roadmap (MAR)


From Isolated Digits to Social Agents
How good is good enough?

TAXONOMISE
CHARACTERISE
PREDICT
EVALUATE
Survey(s) of Community Opinion


Performance Trends

NIST STT Benchmark Test History – May. '09
Human Equivalent Noise Ratio (HENR)

Machine SR = Human SR

- Extrapolation using the ‘Human Equivalent Noise Ratio’ (HENR) suggests that:
  - progress on new tasks ~0.7 dB/year
  - progress overall ~0.2 dB/year

- Predictions:
  - transcription of read newspapers … by 2010
  - transcription of freestyle speech … by 2017
  - recognition of digit strings … by 2052
  - recognition of alphabet letters ... by 2060


From Isolated Digits to Social Agents

How good is good enough?

TAXONOMISE
CHARACTERISE
PREDICT
EVALUATE
Usability Issues

- accuracy
- accents
- noise
- task familiarity
- non-speech GUIs
- privacy

How many 'real' users?


Usability Issues

- Flexibility
  - Like a Human
  - 'Habitability Gap'


Add NL/Dialog

Structured Dialog
The Uncanny Valley


A Quantitative Model

How to Avoid the Uncanny Valley (and maximise ‘usability’)

• Minimise conflicting cues

• Align an agent’s ‘affordances’:
  – visual (i.e. what it looks like)
  – vocal (i.e. what it sounds like)
  – behavioural (i.e. how it behaves)
  – cognitive (i.e. what it appears to know)
  – linguistic (i.e. how it communicates)

Wired: Do you think it’s possible to bridge the uncanny valley?

Wired: Do you think it’s possible to bridge the uncanny valley?

Mori: Yes, but why try? I think it’s better to design things like Honda’s Asimo, which stops right before it gets to be uncanny.


Human-Human Languaging

Ostensive inferential recursive mindreading


Human-Machine Languaging

Ostensive inferential recursive mindreading

Mutual declarative, interrogative, imperative coupling

Human-Machine Interaction

Communication, interaction and language are core cognitive capabilities (not peripheral sensorimotor components)

Mismatched capabilities
Resources Needed?


Where to find out more …

http://www.dcs.shef.ac.uk/~roger/MarkowitzCh12manuscript.pdf
Where to find out more …


http://www.dcs.shef.ac.uk/~roger/progprogress.html
And finally …

Ethical/standards issues for the LREC community to pioneer:

- **Privacy**
  - ‘intelligent’ agents eavesdropping on user conversations
  - ‘intelligent’ agents sharing information about users

- **Security**
  - ‘intelligent’ agents being hacked

- **Misrepresentation**
  - ‘intelligent’ agents pretending to be more capable than they really are
  - developers/manufacturers making unwarranted claims about their agents’ capabilities

Thank You

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