

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.

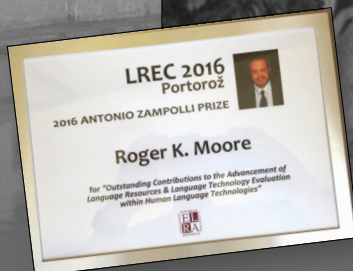


From Isolated Digits to Social Agents

How good is good enough?

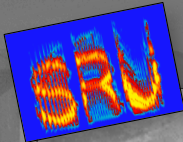


Antonio Zampolli



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

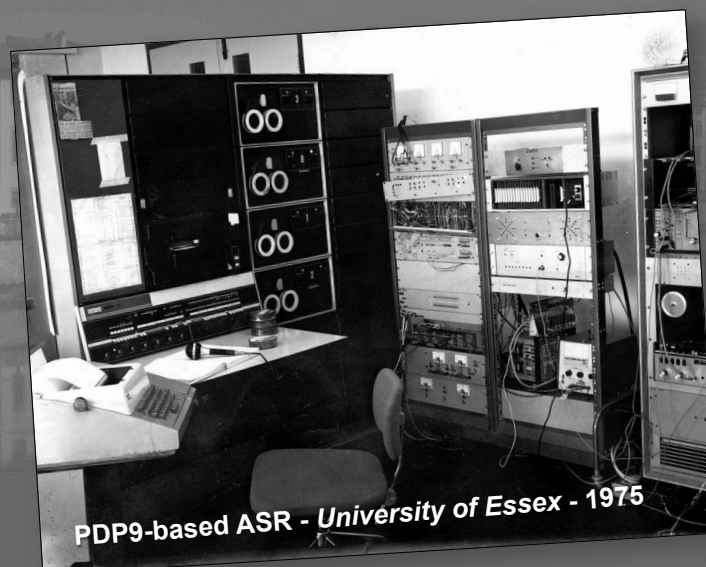
From Isolated Digits to Social Agents

How good is good enough?



From Isolated Digits to Social Agents

How good is good enough?



From Isolated Digits to Social Agents

How good is good enough?



From Isolated Digits to Social Agents

How good is good enough?



Moore, R. K. (1977). Evaluating speech recognisers. *IEEE Trans. Acoustics Speech and Signal Processing*, 25, 178-183.

From Isolated Digits to Social Agents

How good is good enough?



UK Speech Technology
Assessment Group (STAG)

Spoken Corpus Recordings in
British English (SCRIBE)

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



EU Speech Assessment
Methods (SAM) project

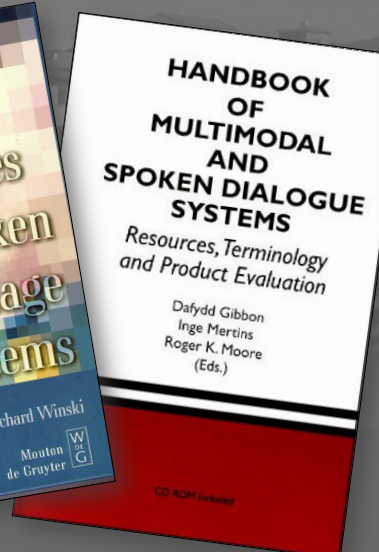


EU Expert Advisory group
in Language Engineering
Standards (EAGLES)



From Isolated Digits to Social Agents

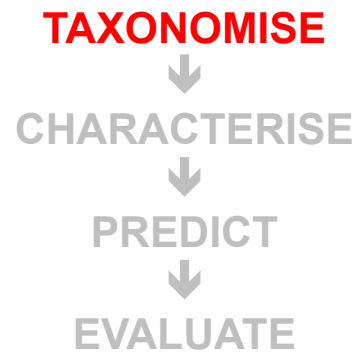
How good is good enough?



Dafydd
Gibbon

From Isolated Digits to Social Agents

How good is good enough?



The University Of Sheffield.

LREC 2016

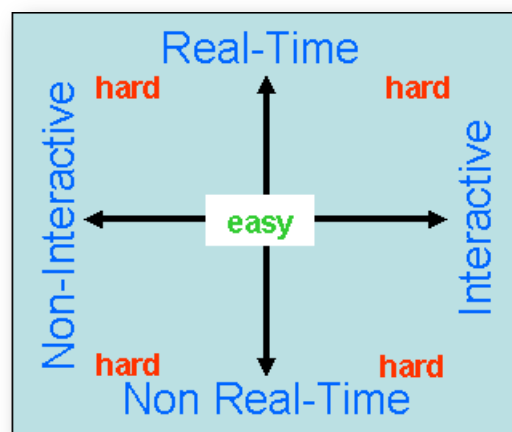
Portorož, Slovenia

27th May 2016

slide 11



Taxonomy of SLT Applications



The University Of Sheffield.

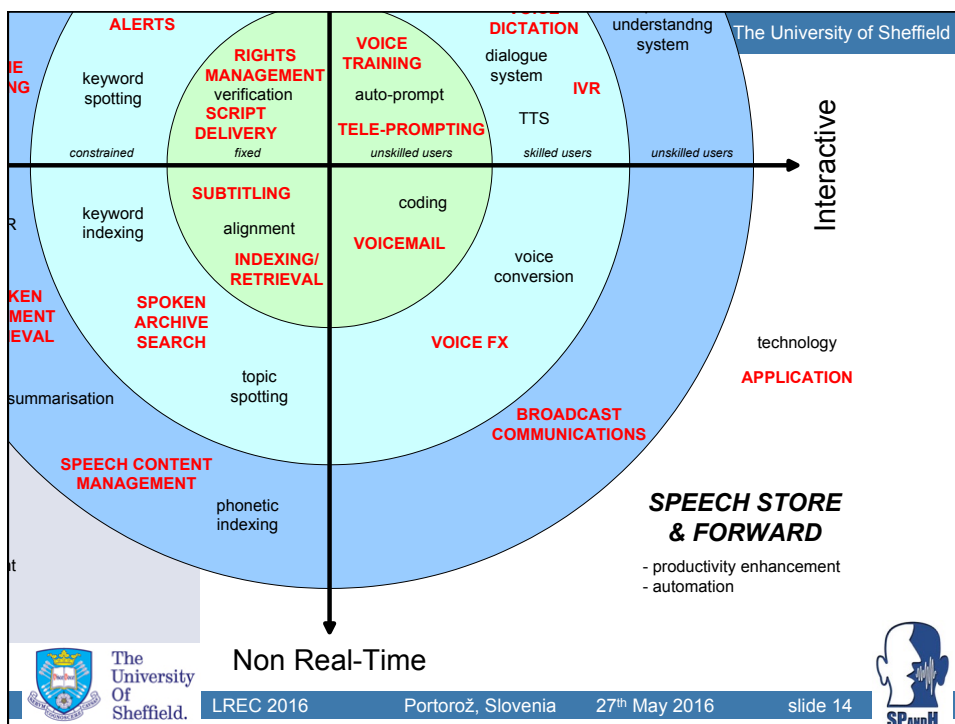
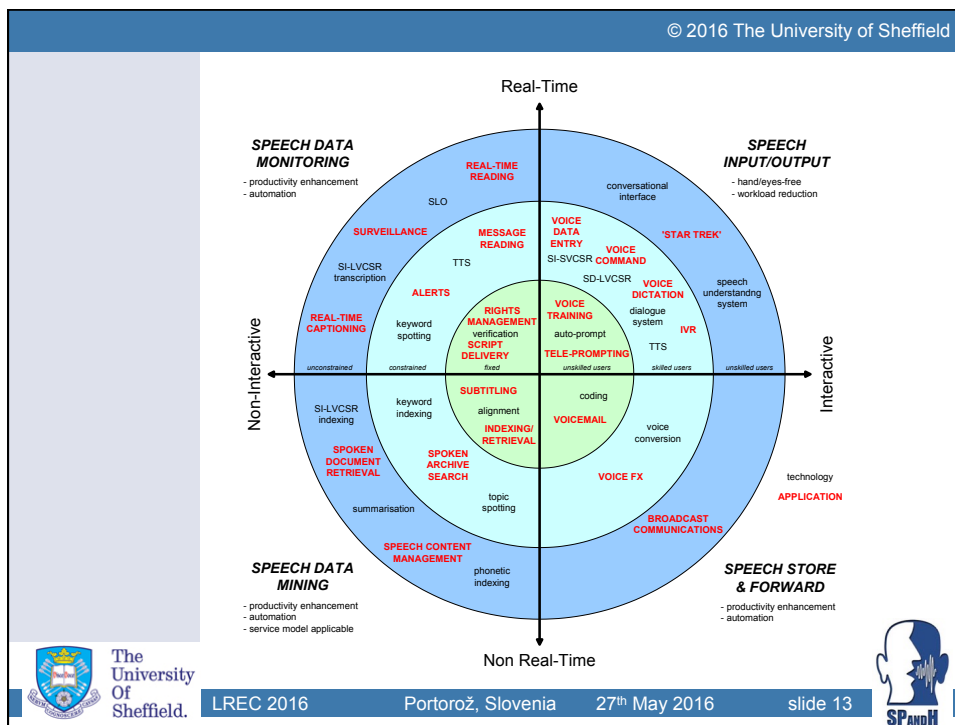
LREC 2016

Portorož, Slovenia

27th May 2016

slide 12





From Isolated Digits to Social Agents

How good is good enough?



The University Of Sheffield.

LREC 2016

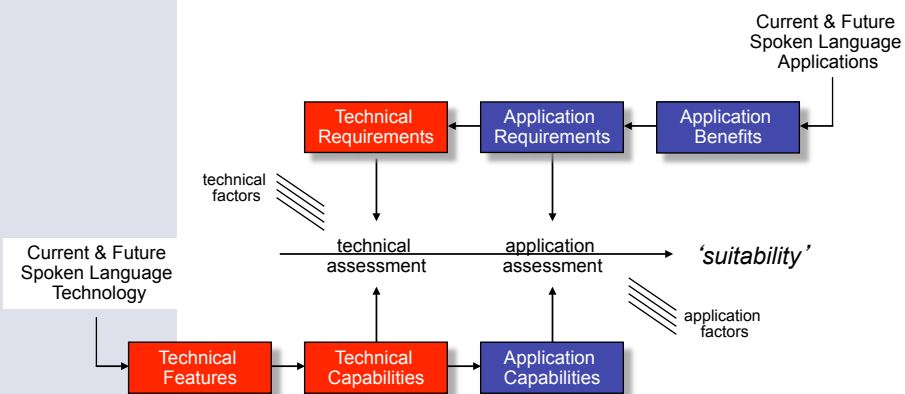
Portorož, Slovenia

27th May 2016

slide 19



Capabilities & Requirements



'Users Guide', R. K. Moore, *Eagles Handbook of Standards and Resources for Spoken Language Systems*, D. Gibbon, R. K. Moore and R. Winsky (eds.), Mouton de Gruyter, pp 1-28, 1997.



The University Of Sheffield.

LREC 2016

Portorož, Slovenia

27th May 2016

slide 20

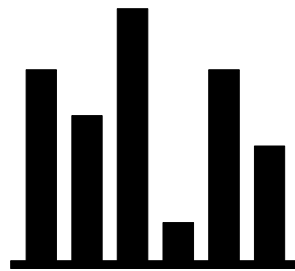


The 'Capability Profile'

Application Requirements (*Pull*)



'GOODNESS'



Performance Envelope

FEATURES



Technology Capabilities (*Push*)



The University Of Sheffield.

LREC 2016

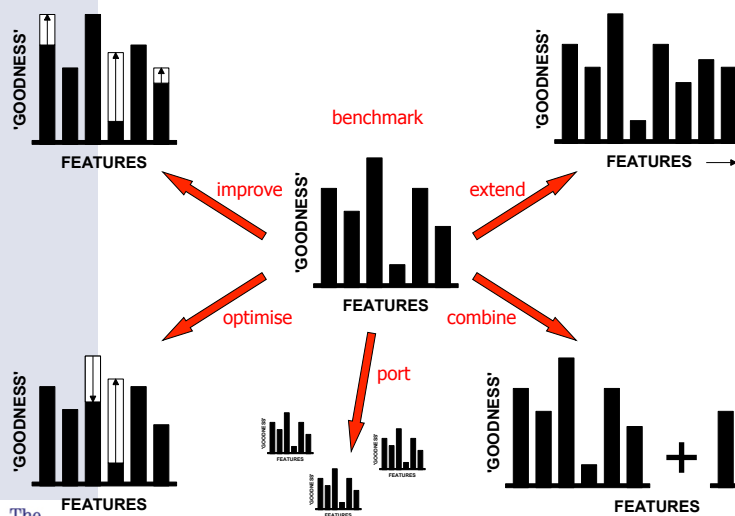
Portorož, Slovenia

27th May 2016

slide 21



Technology Migration



The University Of Sheffield.

LREC 2016

Portorož, Slovenia

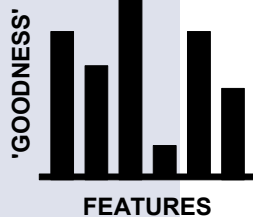
27th May 2016

slide 22



Success

- A technology doesn't have to be 'good' to be successful
- It has to be 'good enough',
(i.e. better than the alternatives)



The
University
Of
Sheffield.

LREC 2016

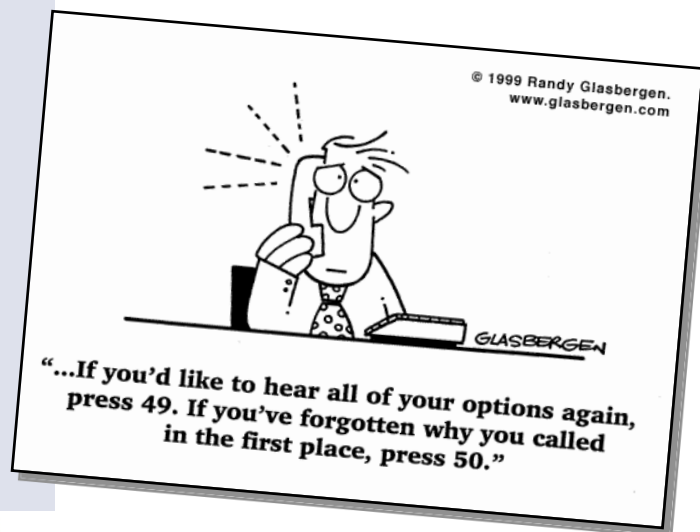
Portorož, Slovenia

27th May 2016

slide 23



One Alternative ...



The
University
Of
Sheffield.

LREC 2016

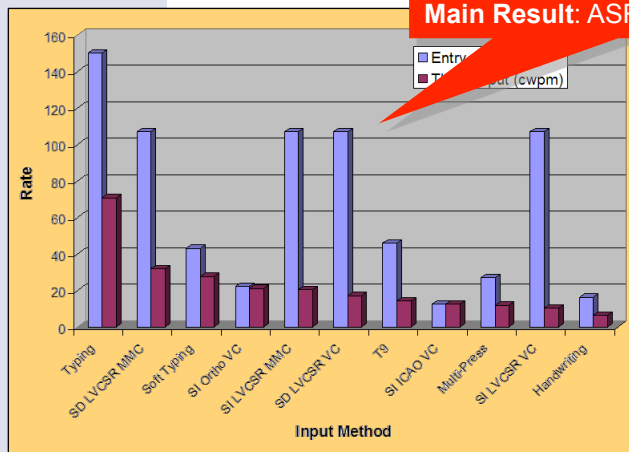
Portorož, Slovenia

27th May 2016

slide 24



Characterising Alternative Input Methods



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

Moore, R. K. (2004). Modelling data entry rates for ASR and alternative input methods, *INTERSPEECH 2004 ICSLP*. Jeju, Korea.



The University Of Sheffield.

LREC 2016

Portorož, Slovenia

27th May 2016

slide 25

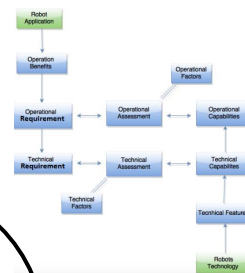
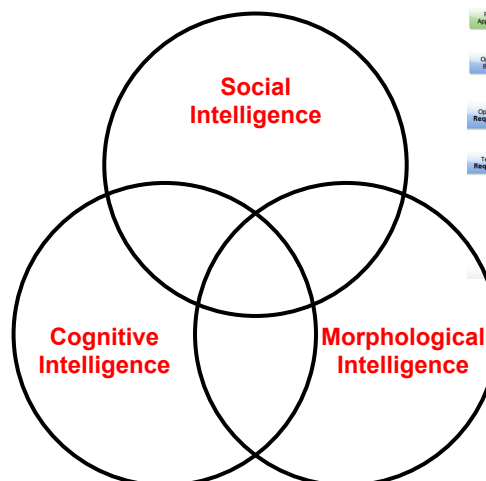


Characterising Autonomous Agents

Manal Linjawi



"What is a Robot?"



The University Of Sheffield.

LREC 2016

Portorož, Slovenia

27th May 2016

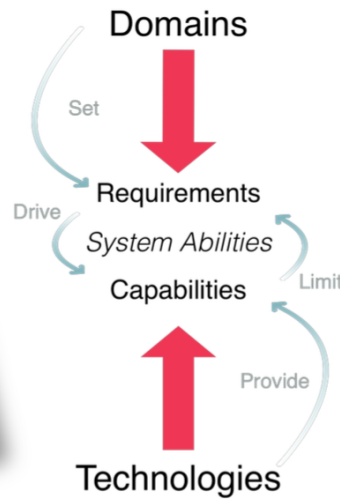
slide 26



Robotics 2020 Multi-Annual Roadmap (MAR)



SPARC. (2015). *Robotics 2020 Multi-Annual Roadmap*.



The University Of Sheffield.

LREC 2016

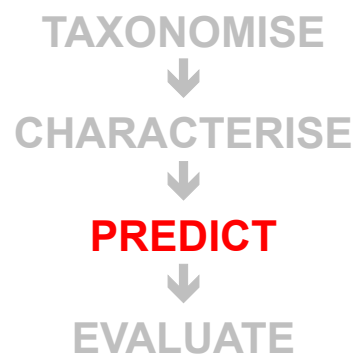
Portorož, Slovenia

27th May 2016

slide 27



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



The University Of Sheffield.

LREC 2016

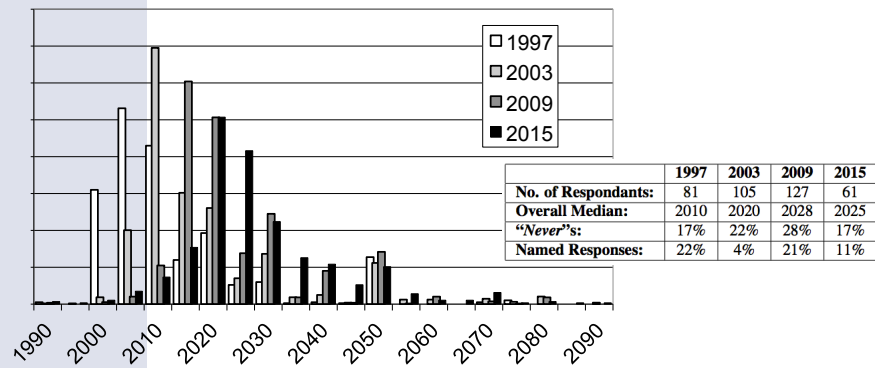
Portorož, Slovenia

27th May 2016

slide 28



Survey(s) of Community Opinion



Moore, R. & Marxer, R. (submitted). Progress and prospects for spoken language technology: results from four sexennial surveys. *INTERSPEECH-2016*.



The University of Sheffield.

LREC 2016

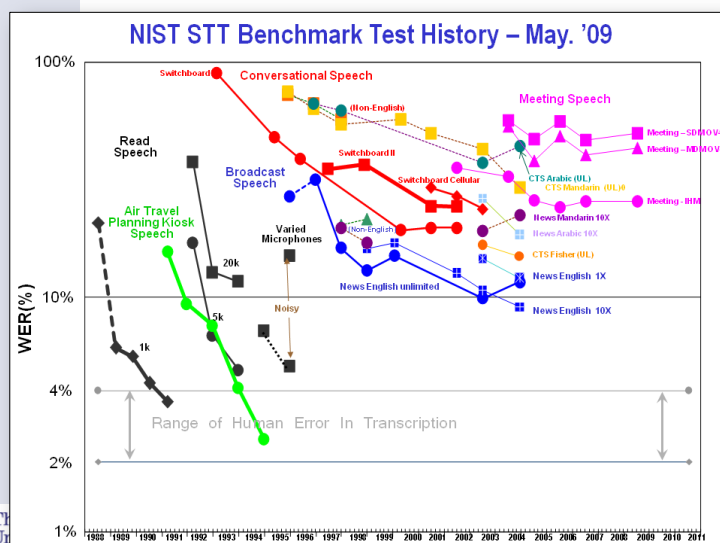
Portorož, Slovenia

27th May 2016

slide 29



Performance Trends



The University of Sheffield.

LREC 2016

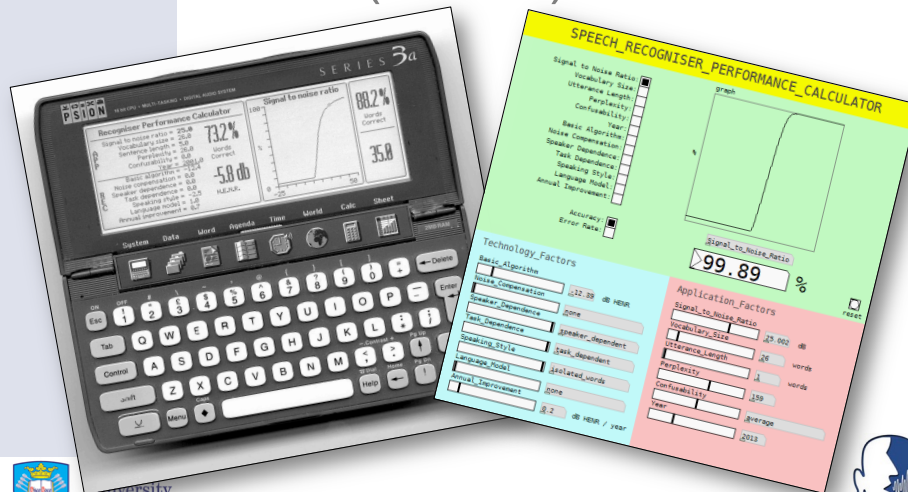
Portorož, Slovenia

27th May 2016

slide 30



Human Equivalent Noise Ratio (HENR)



University
Of
Sheffield.

LREC 2016

Portorož, Slovenia

27th May 2016

slide 31



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

Davis, K. H., Biddulph, R., & Balashek, S. (1952). Automatic recognition of spoken digits. *Journal of the Acoustical Society of America*, 24, 637–642.



The
University
Of
Sheffield.

LREC 2016

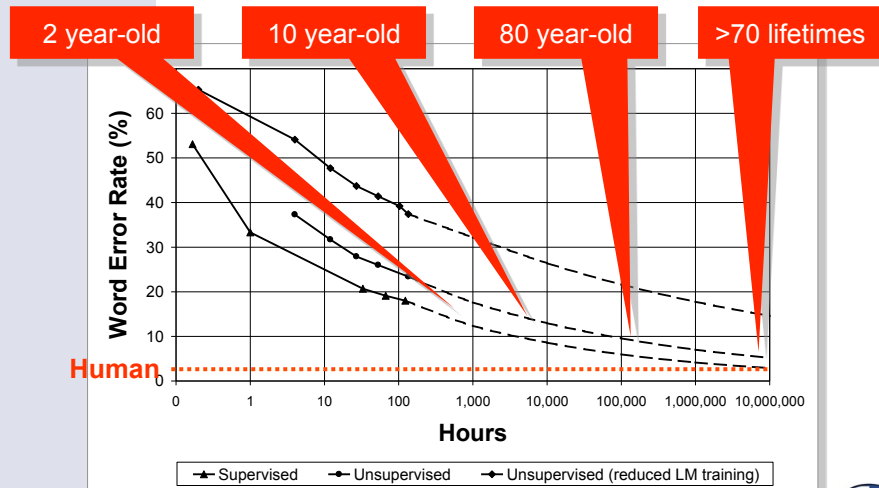
Portorož, Slovenia

27th May 2016

slide 32



How Much Data?!



The University Of Sheffield.

Moore, R. K. (2003). A comparison of the data requirements of automatic speech recognition systems and human listeners, *EUROSPEECH03*. Geneva.

LREC 2016

Portoroz, Slovenia

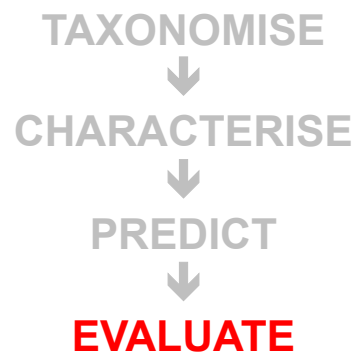
27th May 2016

slide 33



From Isolated Digits to Social Agents

How good is good enough?



The University Of Sheffield.

LREC 2016

Portoroz, Slovenia

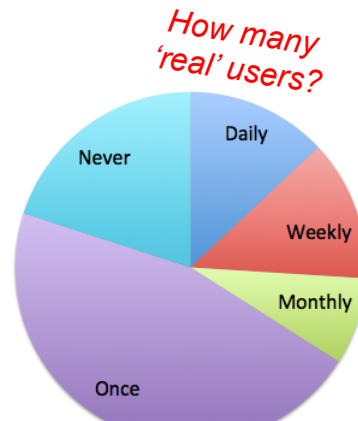
27th May 2016

slide 34



Usability Issues

accuracy
accents
noise
task familiarity
non-speech GUIs
privacy



Liao, S.-H. (2015). *Awareness and Usage of Speech Technology*. MSc Dissertation, University of Sheffield.



The University Of Sheffield.

LREC 2016

Portorož, Slovenia

27th May 2016

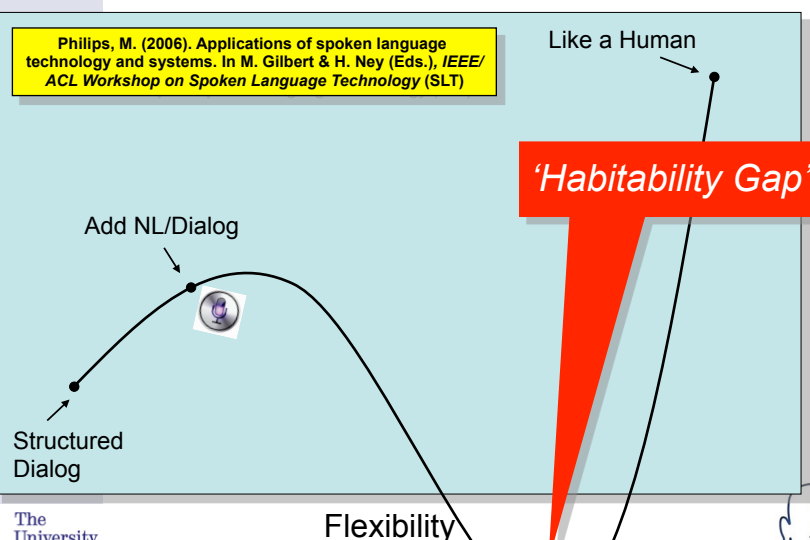
slide 35



Usability Issues

Phillips, M. (2006). Applications of spoken language technology and systems. In M. Gilbert & H. Ney (Eds.), *IEEE/ACL Workshop on Spoken Language Technology (SLT)*

Usability



The University Of Sheffield.

LREC 2016

Portorož, Slovenia

27th May 2016

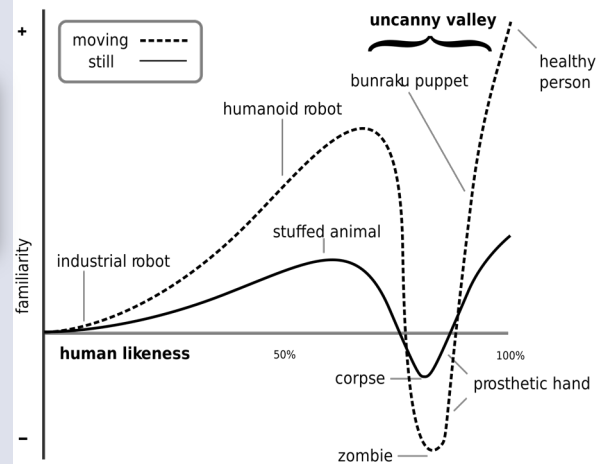
slide 36



The Uncanny Valley



Masahiro Mori



Mori, M. (1970). Bukimi no tani (the uncanny valley). *Energy*, 7, 33-35.



The University Of Sheffield.

LREC 2016

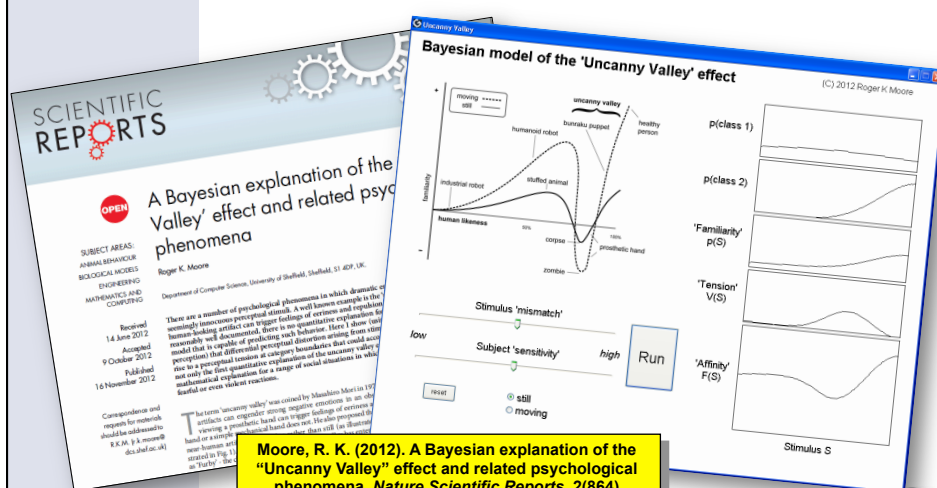
Portorož, Slovenia

27th May 2016

slide 37



A Quantitative Model



Moore, R. K. (2012). A Bayesian explanation of the "Uncanny Valley" effect and related psychological phenomena. *Nature Scientific Reports*, 2(864).



The University Of Sheffield.

LREC 2016

Portorož, Slovenia

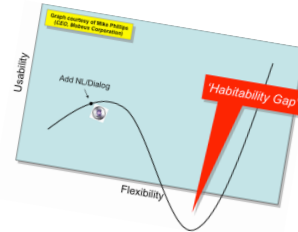
27th May 2016

slide 38



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*)



The
University
Of
Sheffield.

LREC 2016

Portorož, Slovenia

27th May 2016

slide 39



WIRED

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

[http://www.wired.com/magazine/
2011/11/pi_mori/](http://www.wired.com/magazine/2011/11/pi_mori/)



The
University
Of
Sheffield.

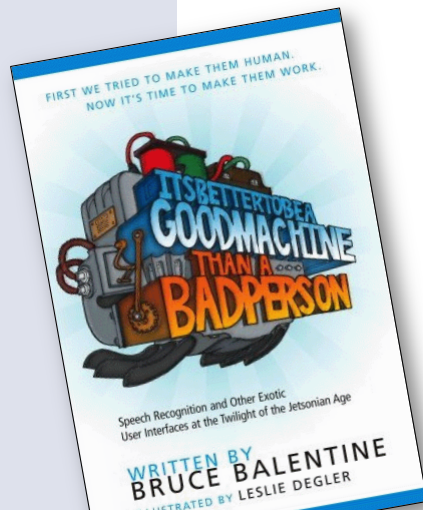
LREC 2016

Portorož, Slovenia

27th May 2016

slide 40





Balentine, B. (2007). *It's Better to Be a Good Machine Than a Bad Person: Speech Recognition and Other Exotic User Interfaces at the Twilight of the Jetsonian Age*. ICMI Press.

WIRED

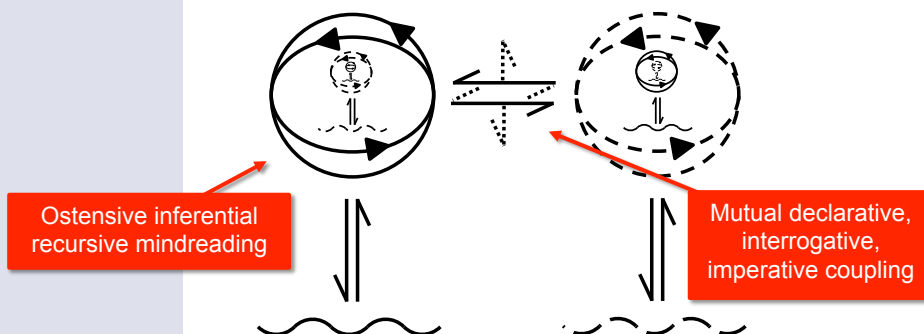
Wired: Do you think it's possible to bridge the uncanny valley?

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.

http://www.wired.com/magazine/2011/11/pl_mori/



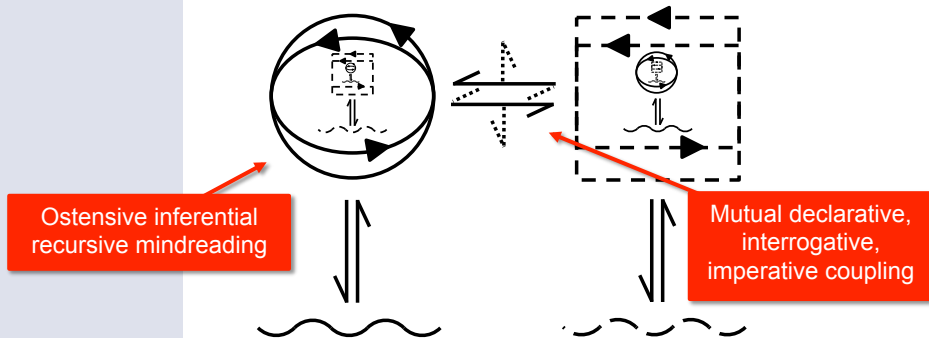
Human-Human Languageing



Moore, R. K. (2016). Is spoken language all-or-nothing? Implications for future speech-based human-machine interaction. In *International Workshop on Spoken Dialogue Systems (IWSDS)*. Saarisekä, Finland.



Human-Machine Languageing



Moore, R. K. (2016). Is spoken language all-or-nothing? Implications for future speech-based human-machine interaction. *In International Workshop on Spoken Dialogue Systems (IWSDS)*. Saarisekä, Finland.



The University Of Sheffield.

LREC 2016

Portorož, Slovenia

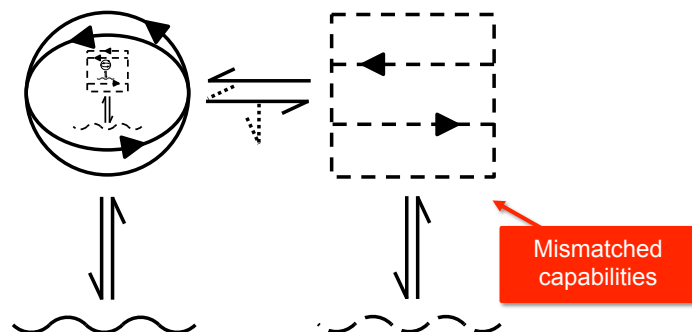
27th May 2016

slide 43



Human-Machine Interaction

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



Moore, R. K. (2016). Is spoken language all-or-nothing? Implications for future speech-based human-machine interaction. *In International Workshop on Spoken Dialogue Systems (IWSDS)*. Saarisekä, Finland.



The University Of Sheffield.

LREC 2016

Portorož, Slovenia

27th May 2016

slide 44



Resources Needed?



The University Of Sheffield.

LREC 2016

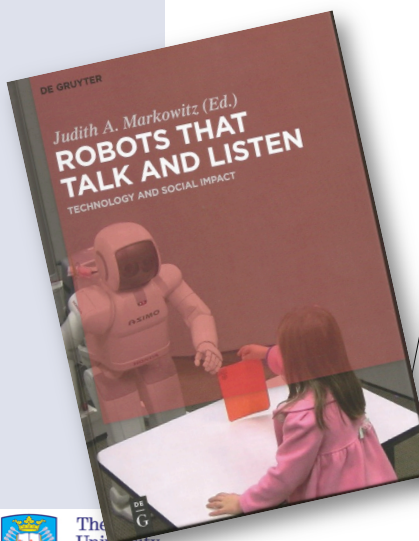
Portorož, Slovenia

27th May 2016

slide 45

SPandH

Where to find out more ...



<http://www.dcs.shef.ac.uk/~roger/MarkowitzCh12manuscript.pdf>

Roger K. Moore
From talking and listening robots to intelligent communicative machines

Abstract It is a popular view that the future will be inhabited by intelligent talking and listening robots with whom we shall converse using the full palette of linguistic meaning available to us as human beings. Of course, recent technical and engineering developments such as Siri would appear to suggest that important steps are being made in that direction – and indeed they are. However, it is argued here that we need to go far beyond our current capabilities and understanding towards a more integrated perspective, simply introducing state-of-the-art speech technology with a state-of-the-art robot is very unlikely to lead to effective human-robot interaction. We need to move from developing robots that simply talk and listen to developing intelligent communicative machines that are capable of truly understanding human behavior, and this means that we need to look beyond speech, beyond words, beyond meaning, beyond communication, beyond dialog and beyond one-off interactions.

1 Introduction

The idea that humanity's future world will be populated by intelligent robots, and that we will converse with them in exactly the same way that we interact with human beings, is a compelling image in contemporary science fiction. From Star Wars' golden android C-3PO to Pixar's rusty dust-busting WALL-E, we have a collective vision of only this world. If scientists and engineers, however, are an exception to this vision, then they are not only wrong, but also dangerous. We have a collective vision of the future that is not only wrong, but also dangerous.

Moore, R. K. (2015). From talking and listening robots to intelligent communicative machines. In J. Markowitz (Ed.), *Robots That Talk and Listen* (pp. 317–335). Boston, MA: De Gruyter.



The University Of Sheffield.

LREC 2016

Portorož, Slovenia

27th May 2016

slide 46



Where to find out more ...

International Journal of Advanced Robotic Systems

OPEN ACCESS ARTICLE

Introducing a Pictographic Language for Envisioning a Rich Variety of Enactive Systems with Different Degrees of Complexity

Invited Feature Article

Roger K. Moore¹*

¹ University of Sheffield, Sheffield, UK
*Corresponding author(s) E-mail: r.k.moore@sheffield.ac.uk

Received 17 June 2016; Accepted 11 January 2016
DOI: 10.5772/62244

© 2016 Author(s). Licensee InTech. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Moore, R. K. (2016). Introducing a pictographic language for envisioning a rich variety of enactive systems with different degrees of complexity. *Int. J. Advanced Robotic Systems*, 13(74).

University of Sheffield LREC 2016 Portorož, Slovenia 27th May 2016 SPANDH

Where to find out more ...

PROGRESS & PROSPECTS IN SPEECH TECHNOLOGY

Having been actively involved in speech technology R&D for over four decades, I'm often called upon to deliver my personal perspective on the progress that's been made in the past and the prospects we're likely to witness in the years to come. In order to inform these views, I've not only conducted a number of **surveys** of the speech technology R&D community, but I've also exploited the ability of my **Human Equivalent Noise Ratio (HENR)** model to extrapolate automatic speech recogniser performance into the future.

In addition, I maintain a personal **timeline of significant events** in our field (including some infamous quotations and notable predictions) which it is hoped will provide a useful resource for students and researchers interested in learning how the speech technology field has developed over the years.

The Past, Present and Future (?) of Speech Technology

A personal timeline history of significant historical events in speech technology R&D (and related topics)

Jan: Report published by Allied Market Research predicts the world Intelligent Virtual Assistant (IVA) market to reach \$3.6 Billion by 2020.

Jan: Xuedong Huang, Microsoft's Chief Speech Scientist, is quoted as saying "Speech recognition is really close to reaching parity with humans, in the next three years" but adds "understanding is a different story".

Jan: Marvin Minsky dies.

Jan: Microsoft moves its deep learning CNTK toolkit to GitHub.

Jan: VocalZoom signs an agreement to integrate its technology alongside iFLYTEK.

Jan: Amazon announces that Amazon Echo's Alexa can be able to read Kindle books aloud.

Jan: Baidu releases Warp-CTC - deep learning software used to build their Deep Speech 2 speech recognition system.

Jan: The 15th annual Deloitte Technology, Media & Telecommunications Predictions report the most persuasive developments of 2016 will be in cognitive technologies such as speech recognition, natural language processing and machine learning.

Jan: OnMobile divests its speech technology assets to France's Voicebox Technologies for €650,000.

Jan: SoundHound Inc. collaborates with NVIDIA to bring deep learning-based natural language understanding to cars.

2015

Dec: Semantic Machines, a startup with artificial intelligence technology and talent from Apple and Google (such as Larry Gillick and Dan Roth), raises \$12.3 million in new funding.

Dec: Baidu claims voice recognition is now competitive with humans in some settings.

Dec: Microsoft releases speech and video recognition AI

<http://www.dcs.shef.ac.uk/~roger/progress.html>

University of Sheffield LREC 2016 Portorož, Slovenia 27th May 2016 SPANDH

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



The
University
Of
Sheffield.

LREC 2016

Portorož, Slovenia

27th May 2016

slide 49

