© 2014 The University of Sheffield

Speech Technology

Past, Present & Tech Transfer

Prof. Roger K. Moore

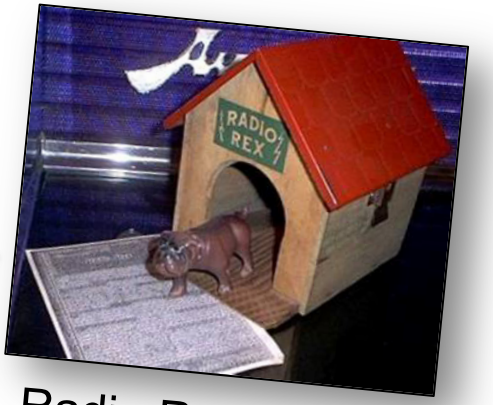
Chair of Spoken Language Processing
Dept. Computer Science, University of Sheffield
(Visiting Prof., Dept. Phonetics, University College London)
(Visiting Prof., Bristol Robotics Lab.)

scentro
Sheffield Centre for Robotics

© 2014 The University of Sheffield

Rich History of Technological Development

Von Kempelen's talking machine (*1791*)



Radio Rex (*1922*)

Parametric Artificial Talker (*1953*)

Interactive Talking Doll (*1987*)

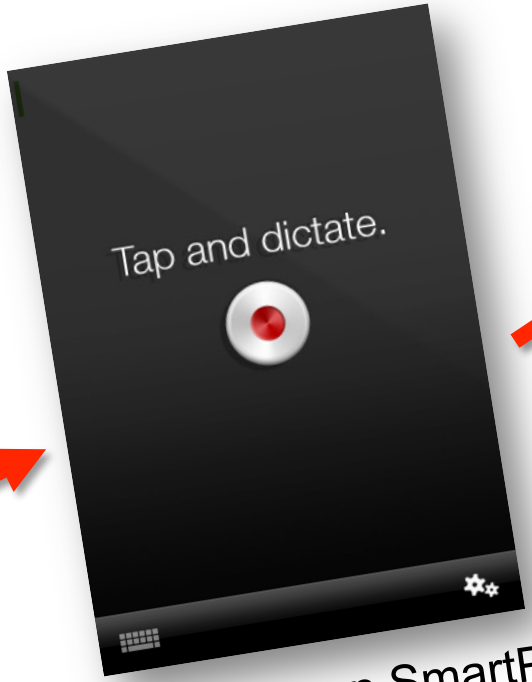
Speak'n'Spell (*1983*)

© 2014 The University of Sheffield

Rich History of Technological Development

Marconi 'SR128'
(1982)

Dragon 'Naturally Speaking'
(1997)

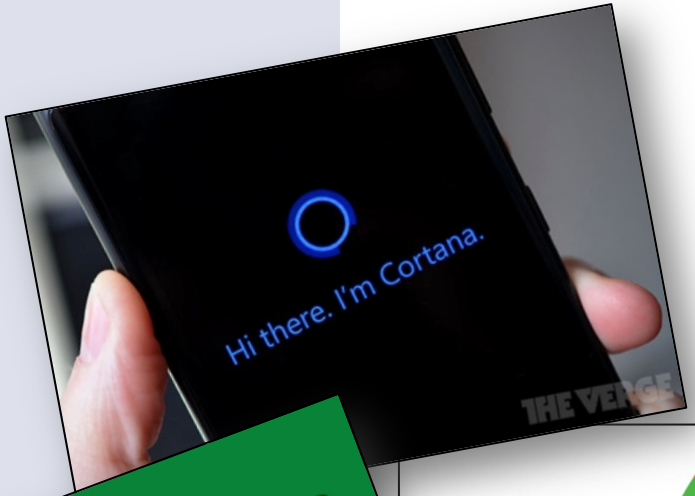


Voice dictation on SmartPhone
(2007)

Apple's "Siri"
(2011)

© 2014 The University of Sheffield

Rich History of Technological Development



Speech-to-Speech Translation

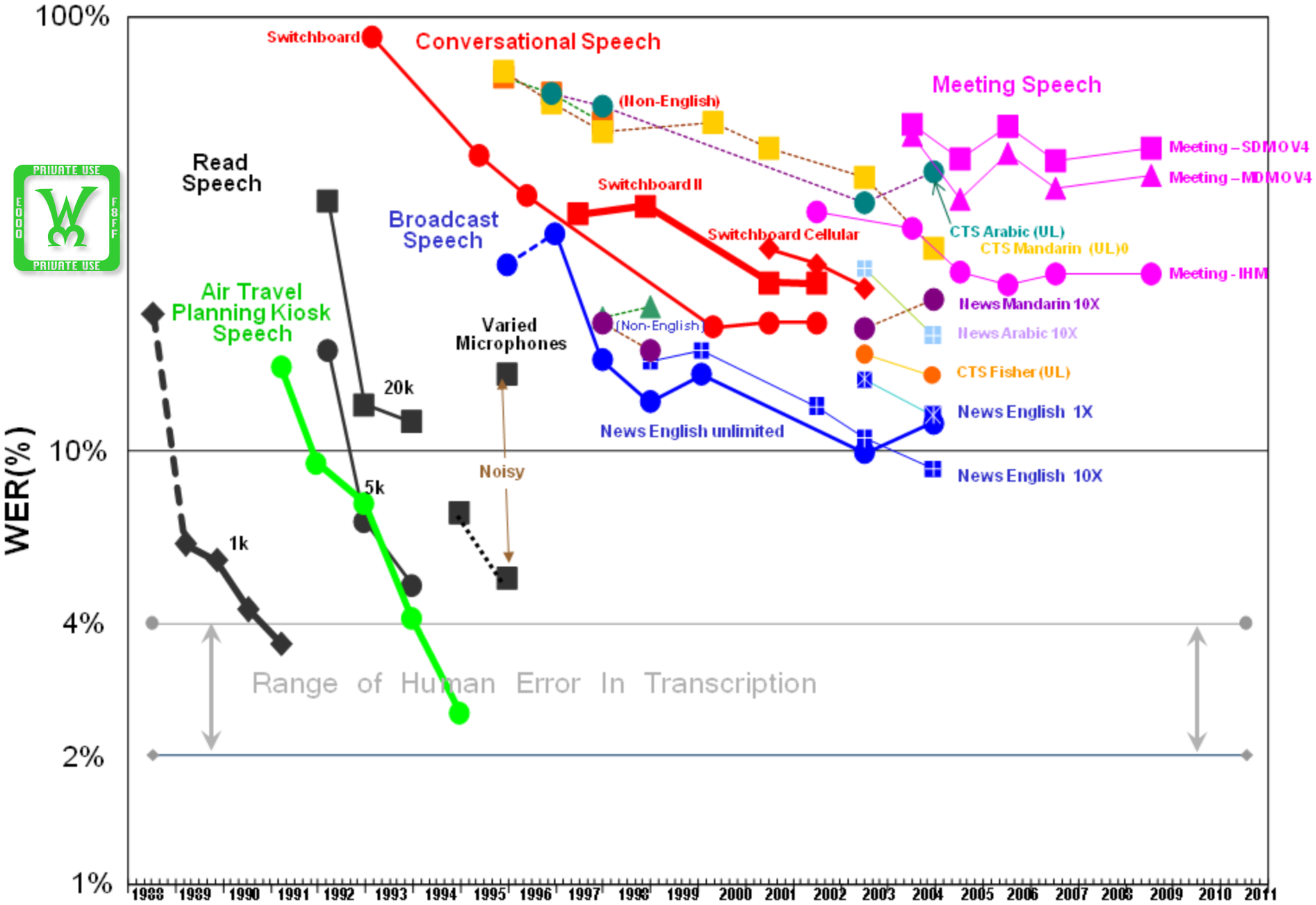
Apple's "Siri" (2011)

© 2014 The University of Sheffield

The 'State-of-the-Art'

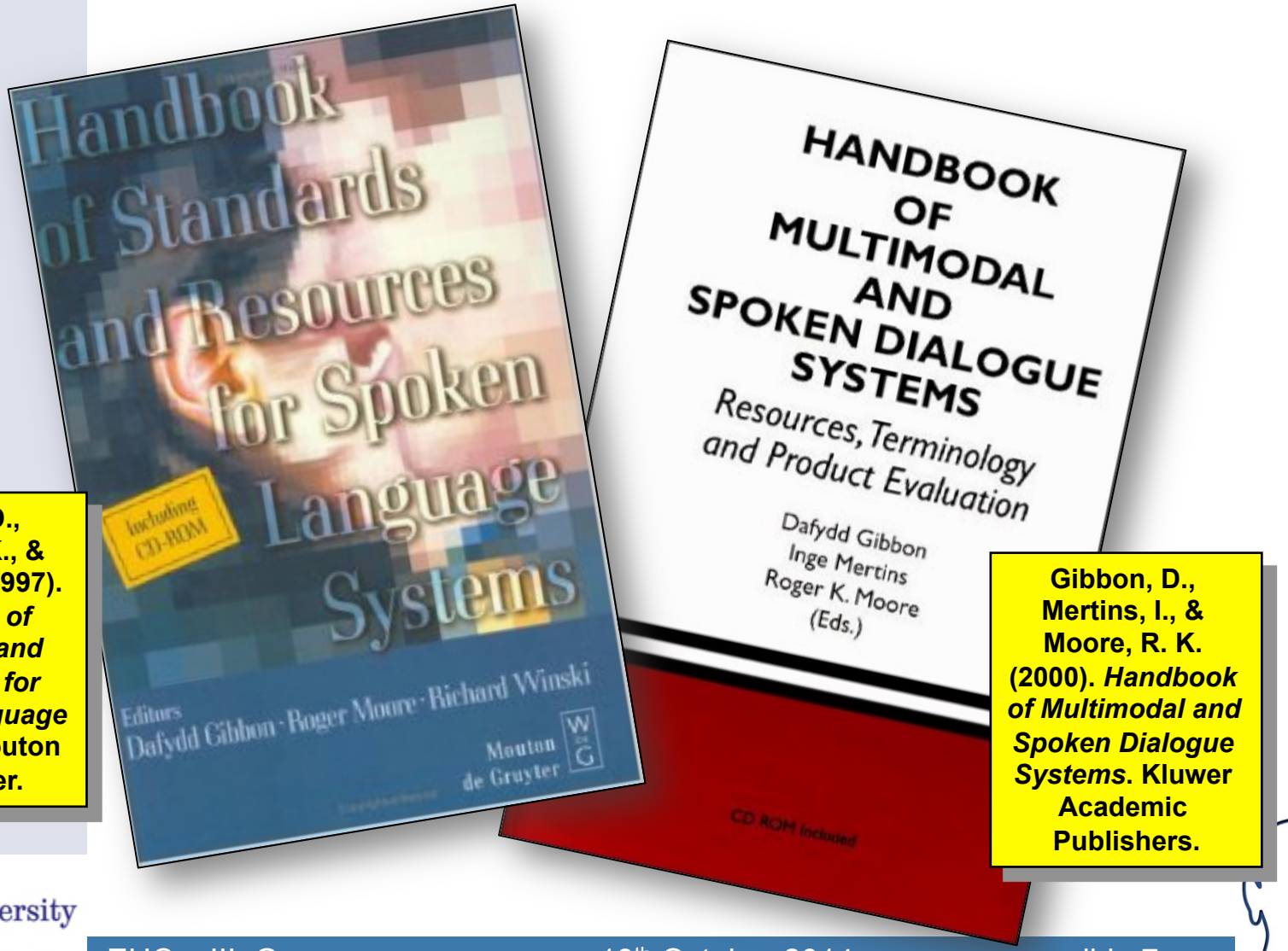
- There is steady year-on-year progress
- Improvement are coming from:
  - increase in available computer power
  - corpus-driven statistical modelling approaches
  - public benchmark testing (+ *standards & resources*)

NIST STT Benchmark Test History – May. '09



© 2014 The University of Sheffield

Standards & Resources



Gibbon, D., Moore, R. K., & Winsky, R. (1997). *Handbook of Standards and Resources for Spoken Language Systems*. Mouton de Gruyter.

Gibbon, D., Mertins, I., & Moore, R. K. (2000). *Handbook of Multimodal and Spoken Dialogue Systems*. Kluwer Academic Publishers.

© 2014 The University of Sheffield

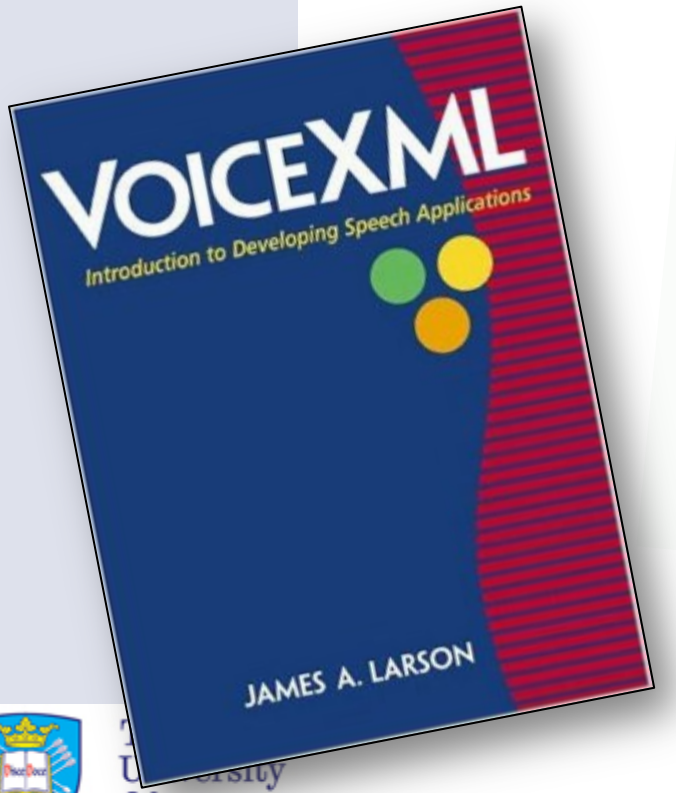
Standards & Resources

LREC 2014
Reykjavik
Harpa Conference Centre
26 - 31 May, 2014

© 2014 The University of Sheffield

Standards & Resources

W3C®



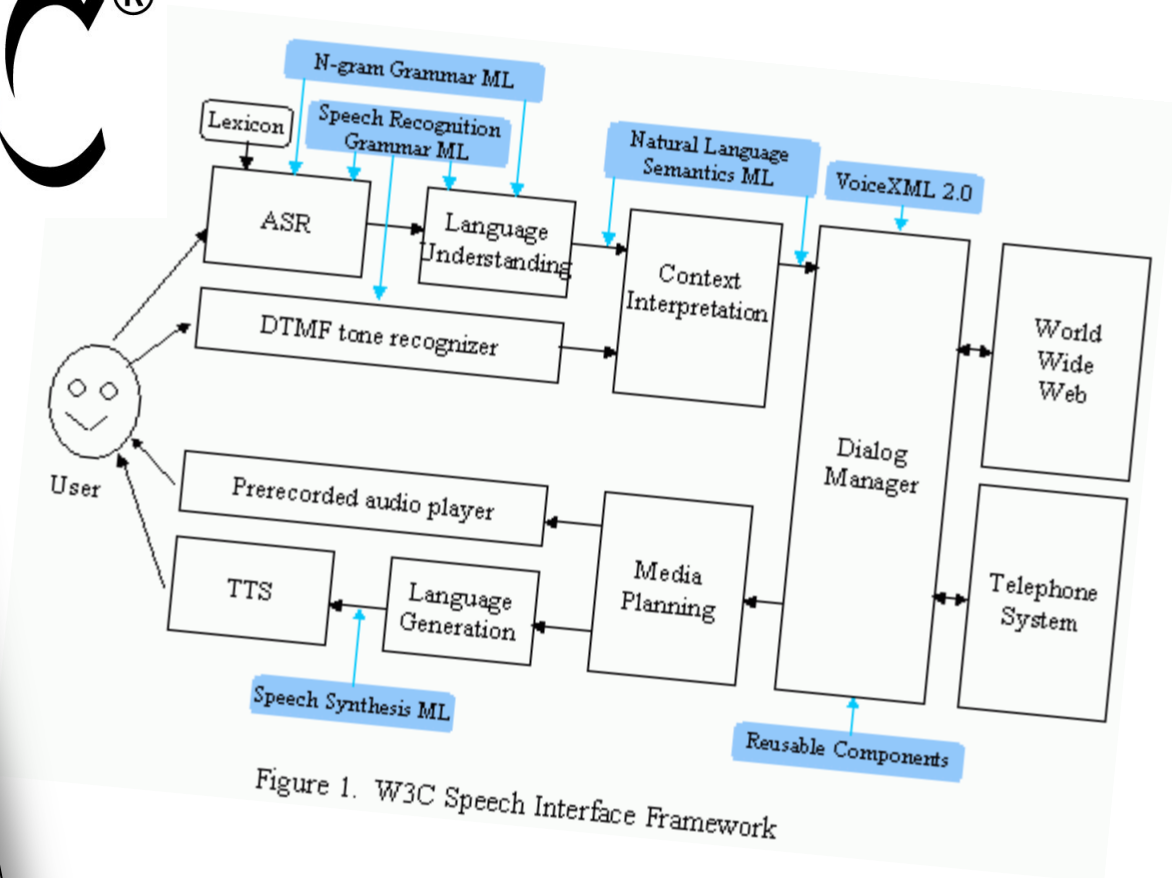


Figure 1. W3C Speech Interface Framework

© 2014 The University of Sheffield

The 'State-of-the-Art'

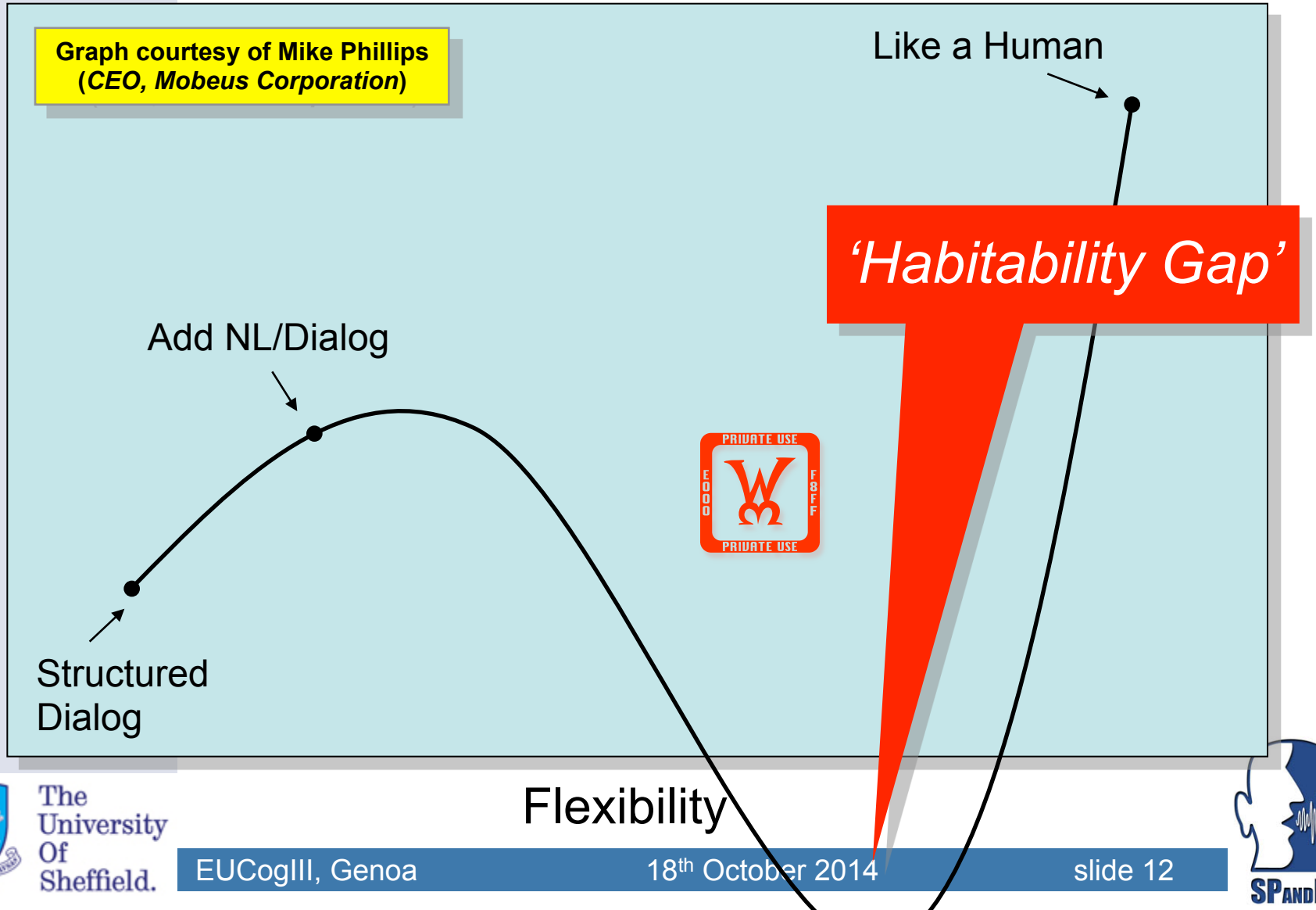
- There is steady year-on-year progress
- Improvement are coming from:
  - increase in available computer power
  - corpus-driven statistical modelling approaches
  - public benchmark testing (+ *standards & resources*)
- Progress has *not* come about as a result of deep insights into human spoken language
- Spoken language technology is
  - fragile (*in 'real' conditions*)
  - expensive (*to port to new applications / languages*)
- Speech recognition performance is reaching an *asymptote* well short of human abilities
  - 25% word error rate on conversational speech

© 2014 The University of Sheffield

Still Some Way to Go?

© 2014 The University of Sheffield

Still Some Way to Go?

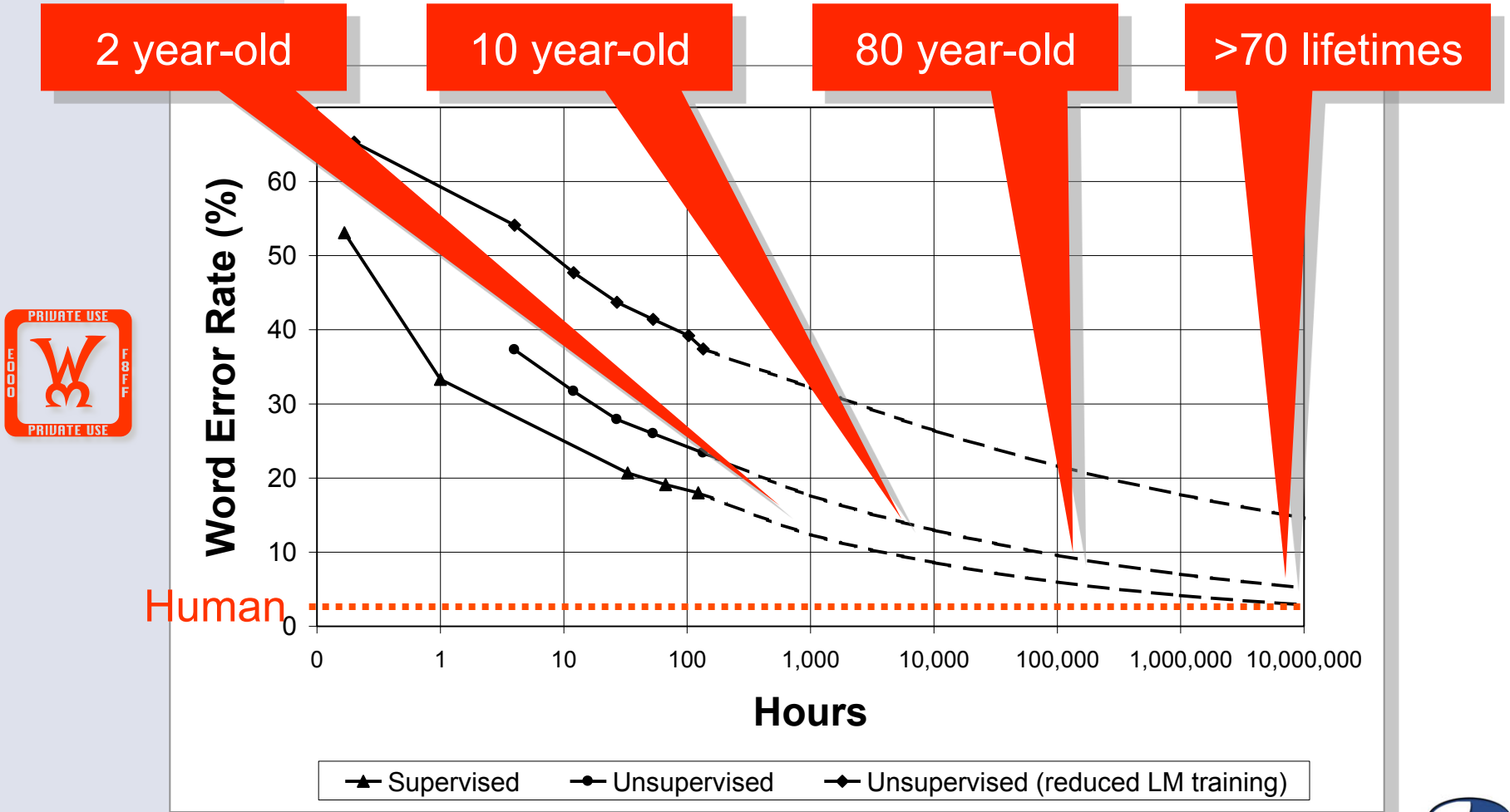


© 2014 The University of Sheffield

Still Some Way to Go?

© 2014 The University of Sheffield

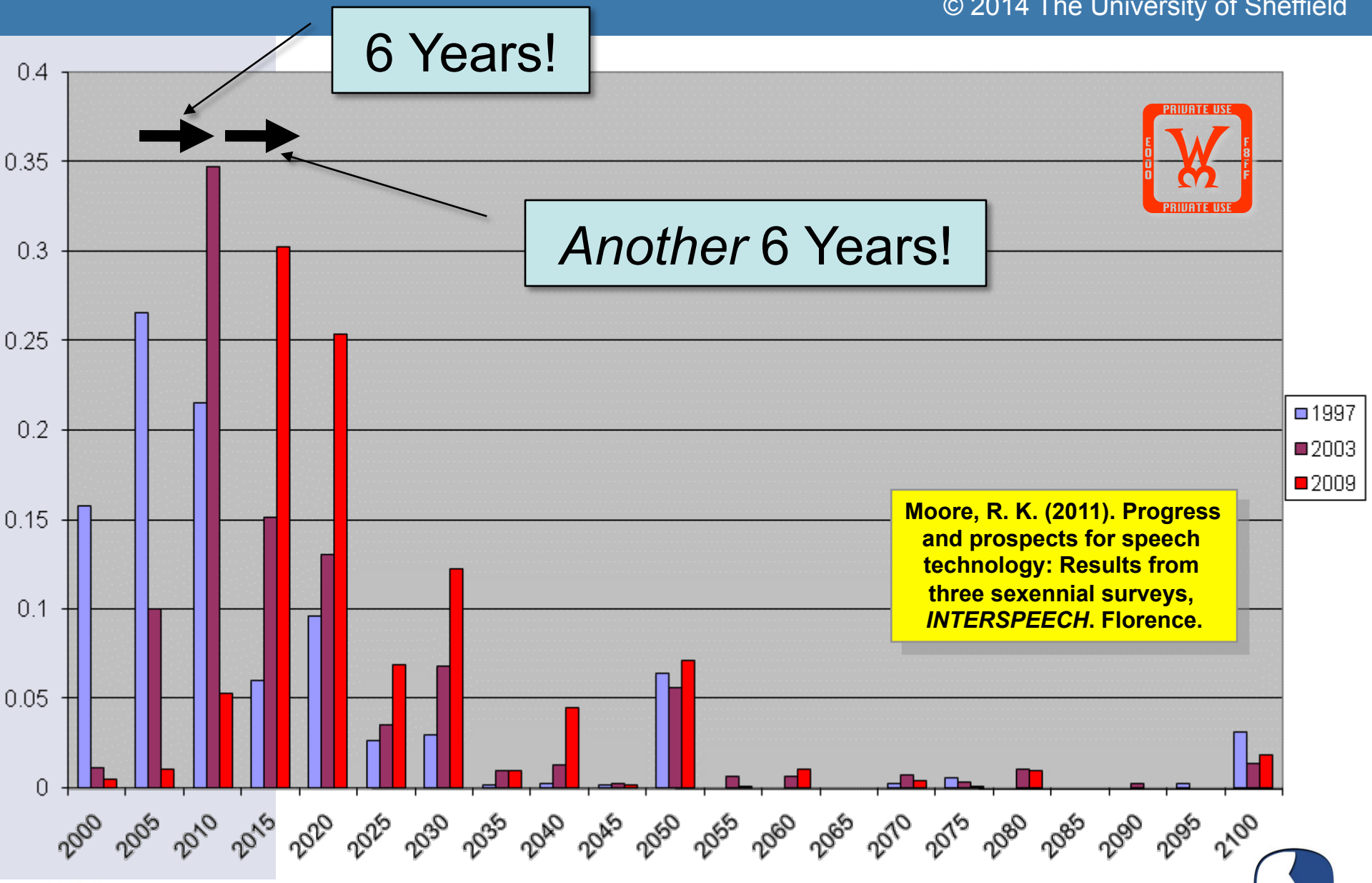
"There's No Data Like More Data"



PRIVATE USE
PRIVATE USE

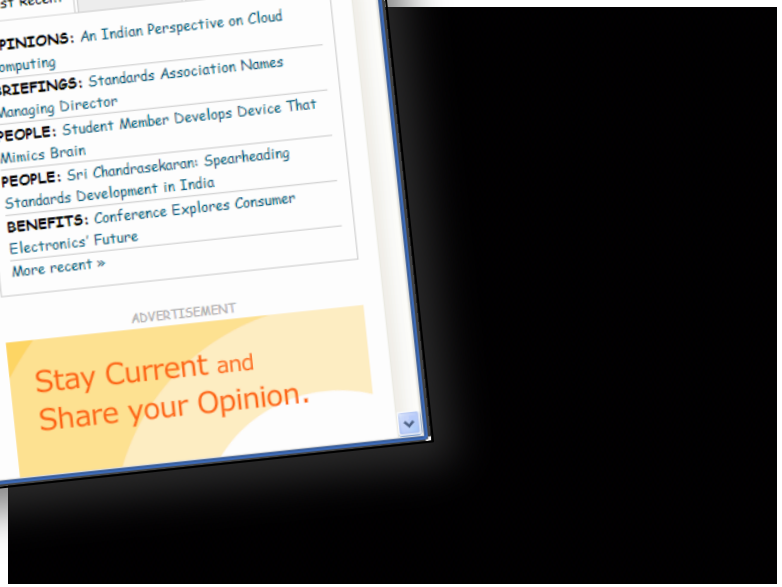
Moore, R. K. (2003). A comparison of the data requirements of automatic speech recognition systems and human listeners, *EUROSPEECH03*. Geneva.

© 2014 The University of Sheffield



© 2014 The University of Sheffield

Still Some Way to Go?



The University Of Sheffield.

© 2014 The University of Sheffield

Still Some Way to Go?

Dangers of misrepresenting capabilities

© 2014 The University of Sheffield

The 'State-of-the

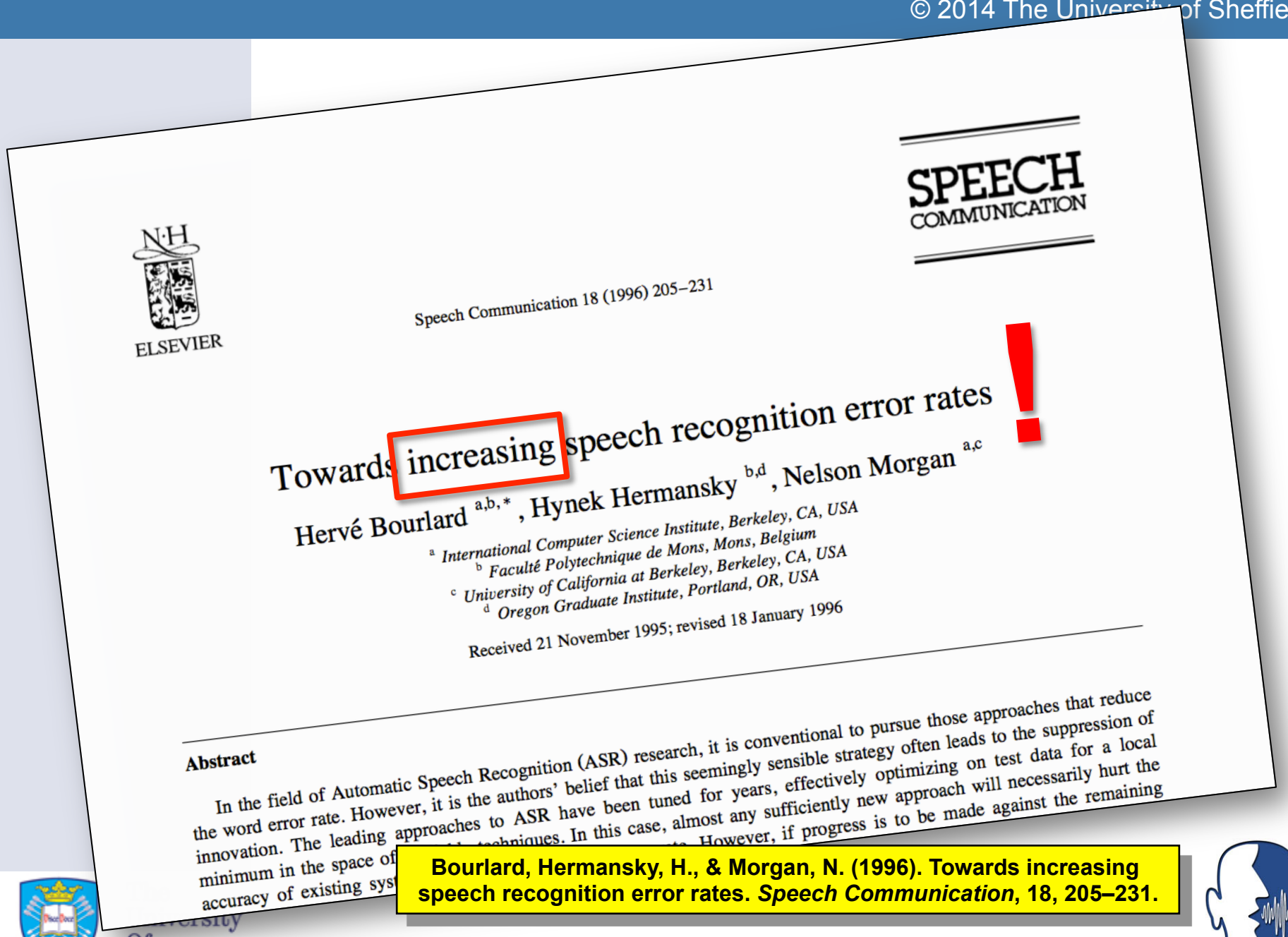
Over-reliance on data?

Too easy to 'tweak' algorithms?

- is steady year-on-year progress
- Improvement are coming from:
  - increase in available computer power
  - corpus-driven statistical modelling approaches
  - public benchmark testing (+ *standards & resources*)
- Progress has *not* come about as a result of deep insights into human spoken language
- Spoken language technology is
  - fragile (*in 'real' conditions*)
  - expensive (*to port to new applications / languages*)
- ASR performance is reaching an *asymptote* well short of human abilities
  - 25% word error rate on conversational speech

Inhibits creativity?

© 2014 The University of Sheffield



ELSEVIER

Speech Communication 18 (1996) 205–231

SPEECH COMMUNICATION

Towards increasing speech recognition error rates !

Hervé Bourlard [a,b,*], Hynek Hermansky [b,d], Nelson Morgan [a,c]

[a] International Computer Science Institute, Berkeley, CA, USA
[b] Faculté Polytechnique de Mons, Mons, Belgium
[c] University of California at Berkeley, Berkeley, CA, USA
[d] Oregon Graduate Institute, Portland, OR, USA

Received 21 November 1995; revised 18 January 1996

Abstract

In the field of Automatic Speech Recognition (ASR) research, it is conventional to pursue those approaches that reduce the word error rate. However, it is the authors' belief that this seemingly sensible strategy often leads to the suppression of innovation. The leading approaches to ASR have been tuned for years, effectively optimizing on test data for a local minimum in the space of [illegible] techniques. In this case, almost any sufficiently new approach will necessarily hurt the accuracy of existing sys[illegible] However, if progress is to be made against the remaining

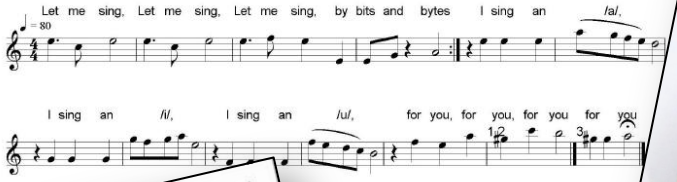
Bourlard, Hermansky, H., & Morgan, N. (1996). Towards increasing speech recognition error rates. *Speech Communication*, 18, 205–231.

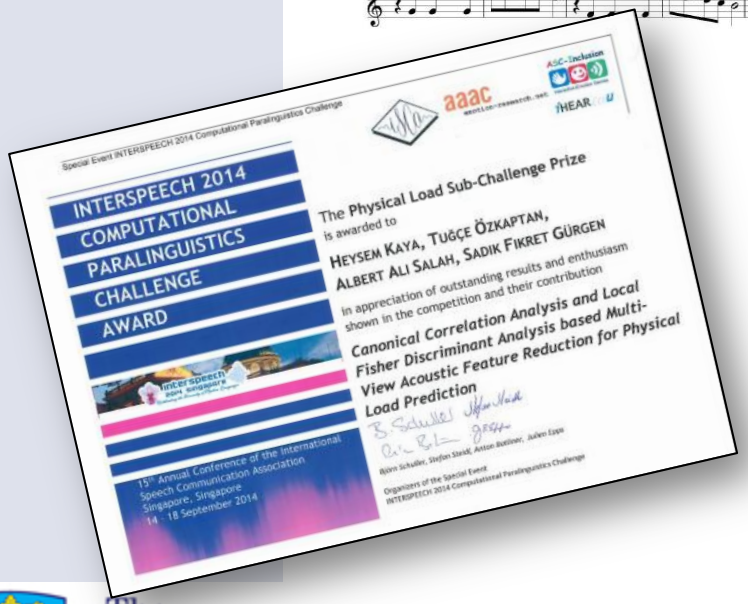
© 2014 The University of Sheffield

Challenges & Competitions

CHiME CHALLENGE

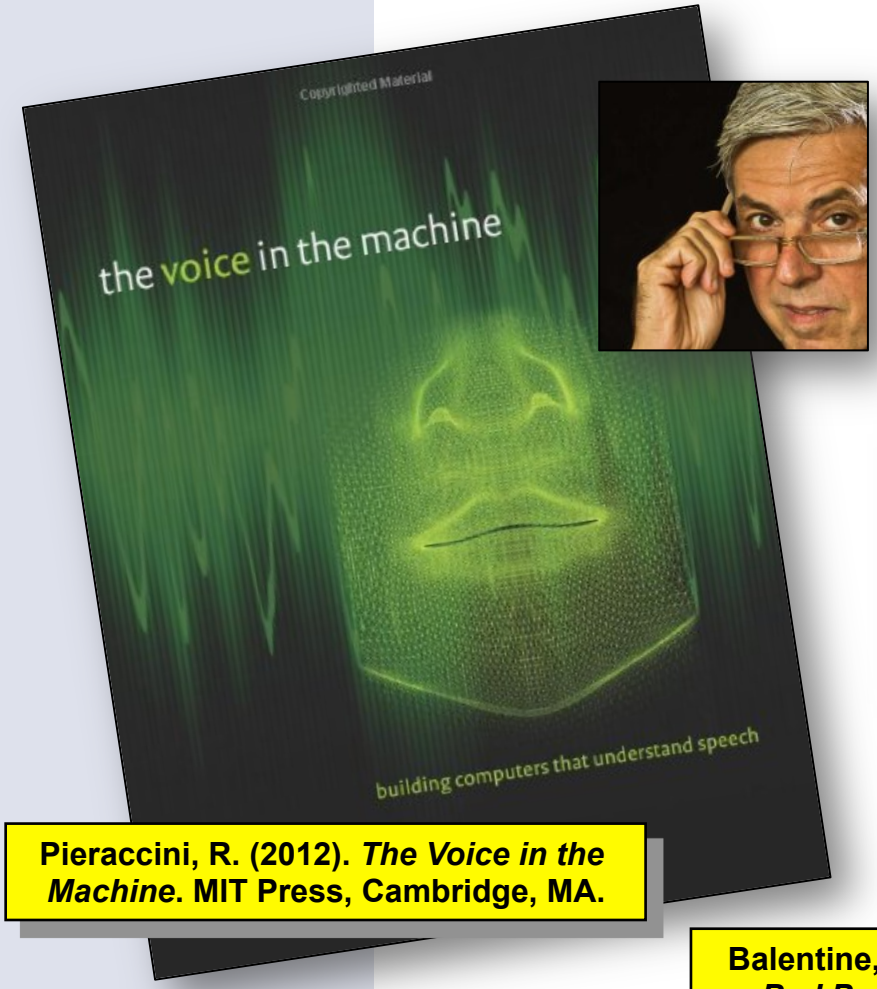
The Synthesizer Song
INTERSPEECH2007

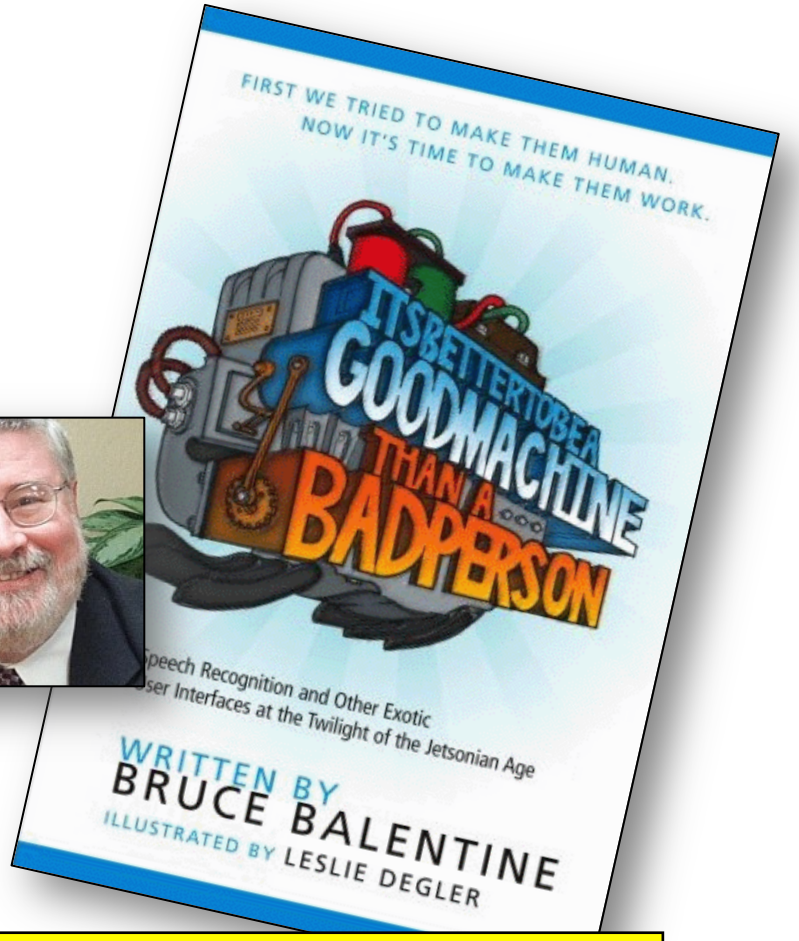




© 2014 The University of Sheffield

Where to Find Out More



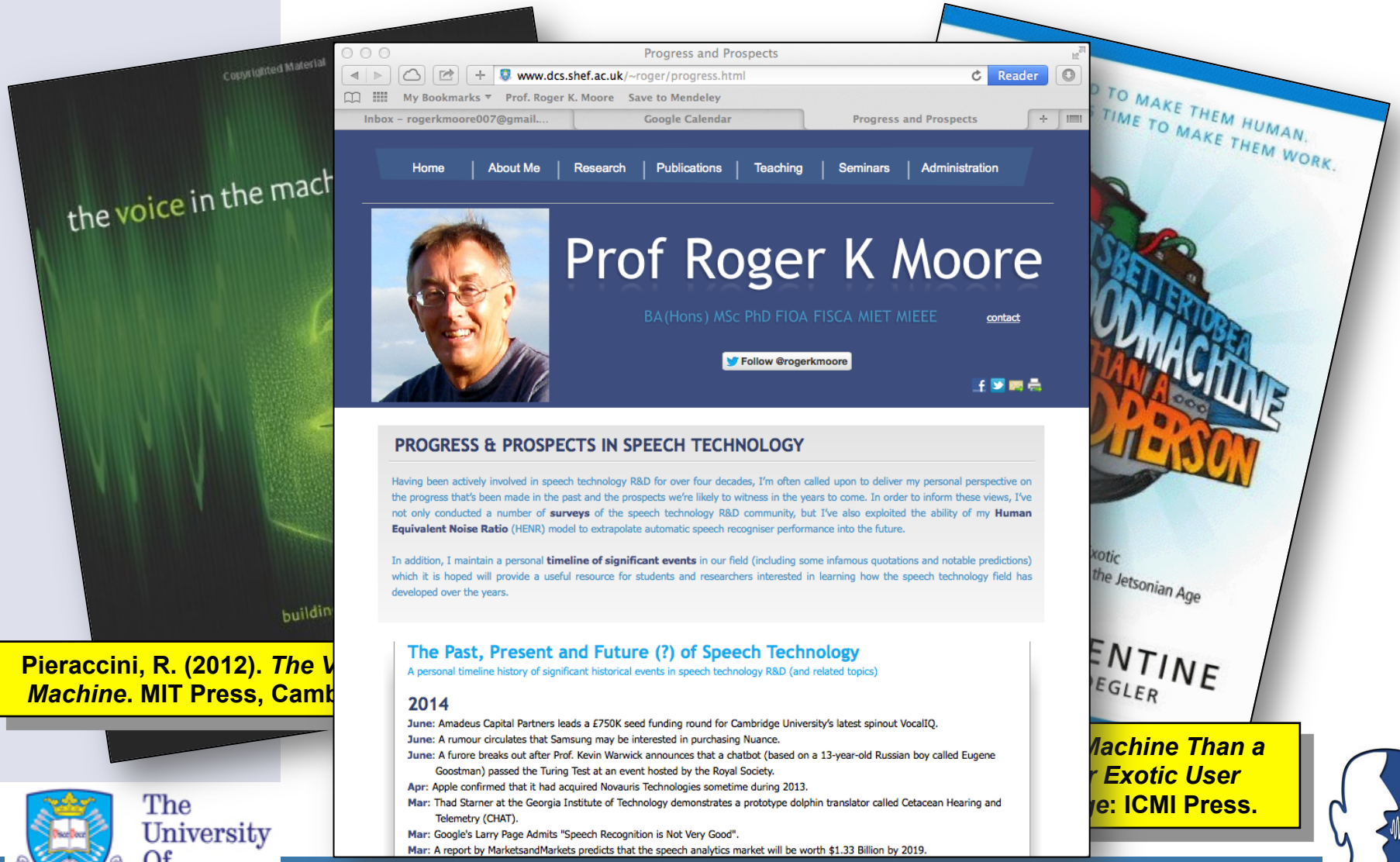


Pieraccini, R. (2012). ***The Voice in the Machine***. **MIT Press, Cambridge, MA.**

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.**

© 2014 The University of Sheffield

Where to Find Out More



Pieraccini, R. (2012). *The V*
Machine. MIT Press, Camb

Machine Than a
r Exotic User
e: ICMI Press.

© 2014 The University of Sheffield

FORTHCOMING NEW BOOK

ROBOTS THAT TALK AND LISTEN

Technology and Social Impact

Edited by: Judith Markowitz

J. Markowitz Consultants

Chicago, IL 60660

About the volume editor

Dr. Judith Markowitz is known internationally as a thought leader in speech processing. She has ... has been active in standards ... editor

Robots That Talk and Listen examination of speech and lang... functional, and social perspective...

Language is an extremely diffic... Contributors addresses the lingu... that robots need to acquire in or... prowess of fictional automata... multi-faceted ability to learn la... designed for communication ... requires structure, meaning, re... in a coherent dialogue.

Speech-enabled automata are... factory automata, teacher's... Such automata fill roles that... well as an automaton, such... children.

The social impact of the... specific roles they provide... of those issues, notably ac... robots, and developing eth...

From Talking and Listening Robots to Intelligent Communicative Machines

Roger K. Moore

University of Sheffield

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 expression 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 interfacing 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

J. Markowitz (Ed.), *Robots That Talk and Listen*. Boston, MA: De Gruyter.

DE GRUYTER

121 High Street, Third Floor · Boston, MA 02110 · T (857) 284-7073 · F (857) 284-7358

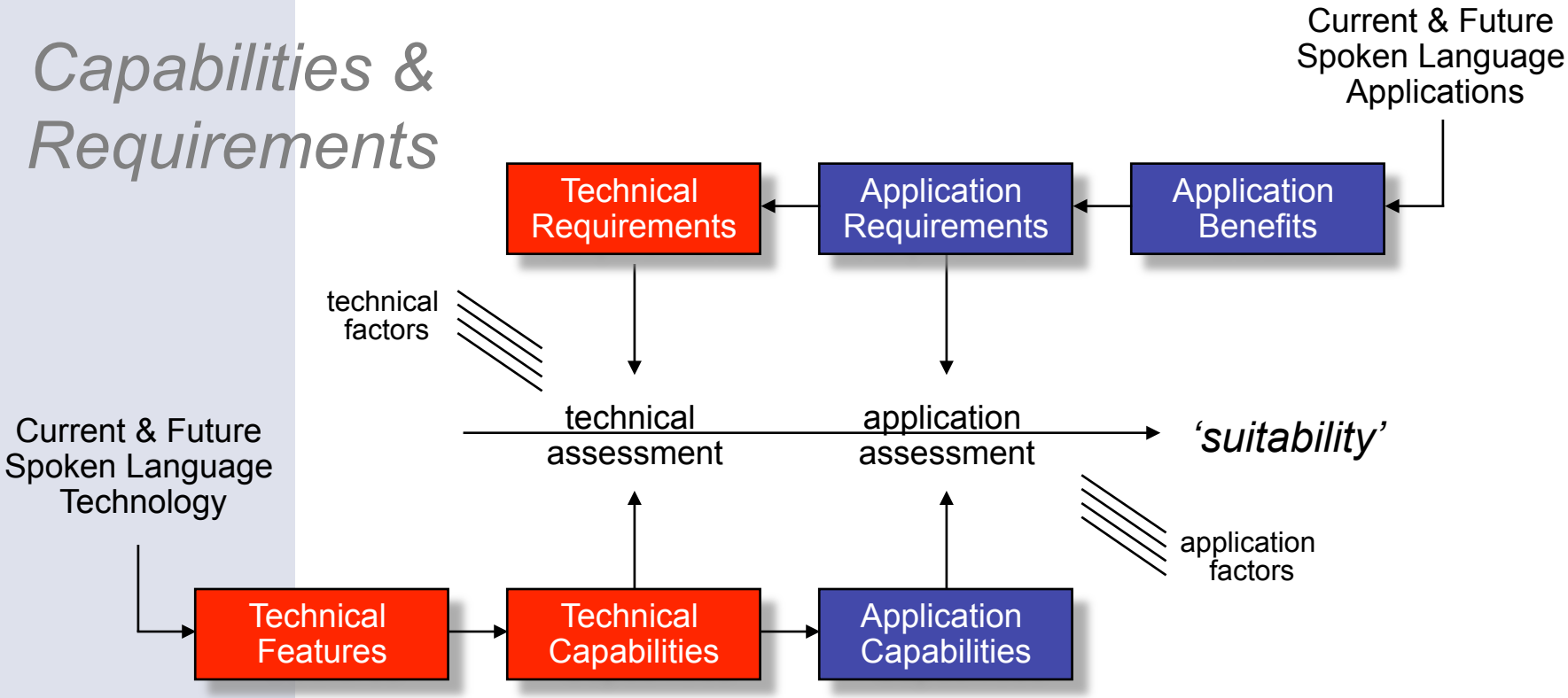
info@degruyter.com · www.degruyter.com

Thank You
Any questions?
http://www.dcs.shef.ac.uk/~roger

© 2014 The University of Sheffield

Performance Characterisation

Capabilities & Requirements

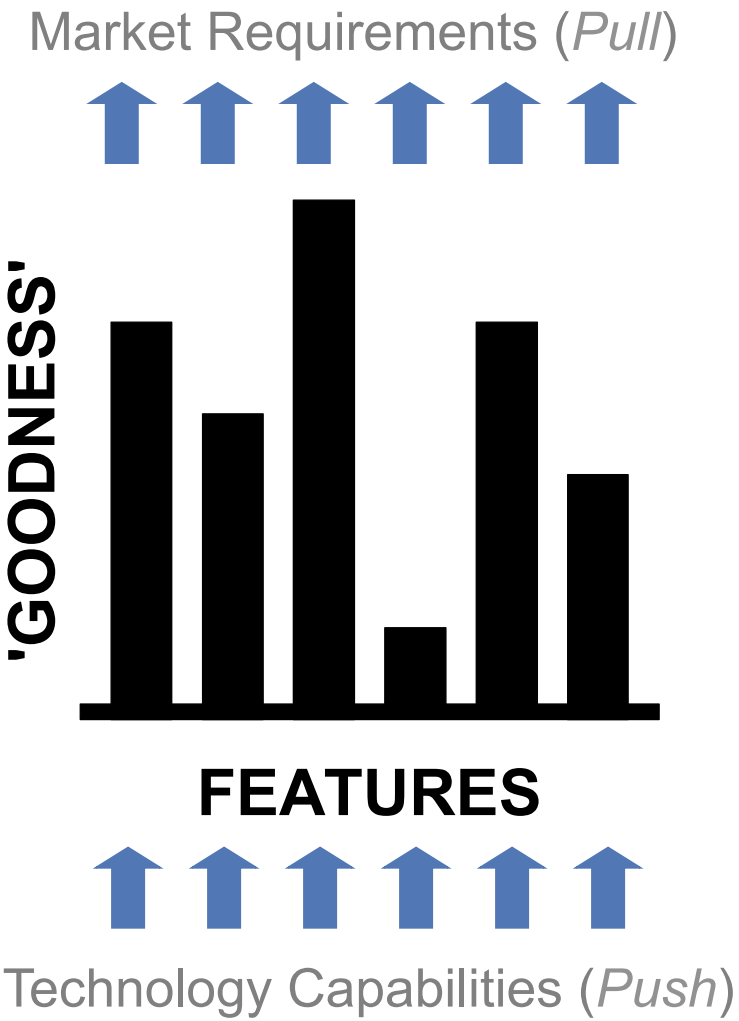


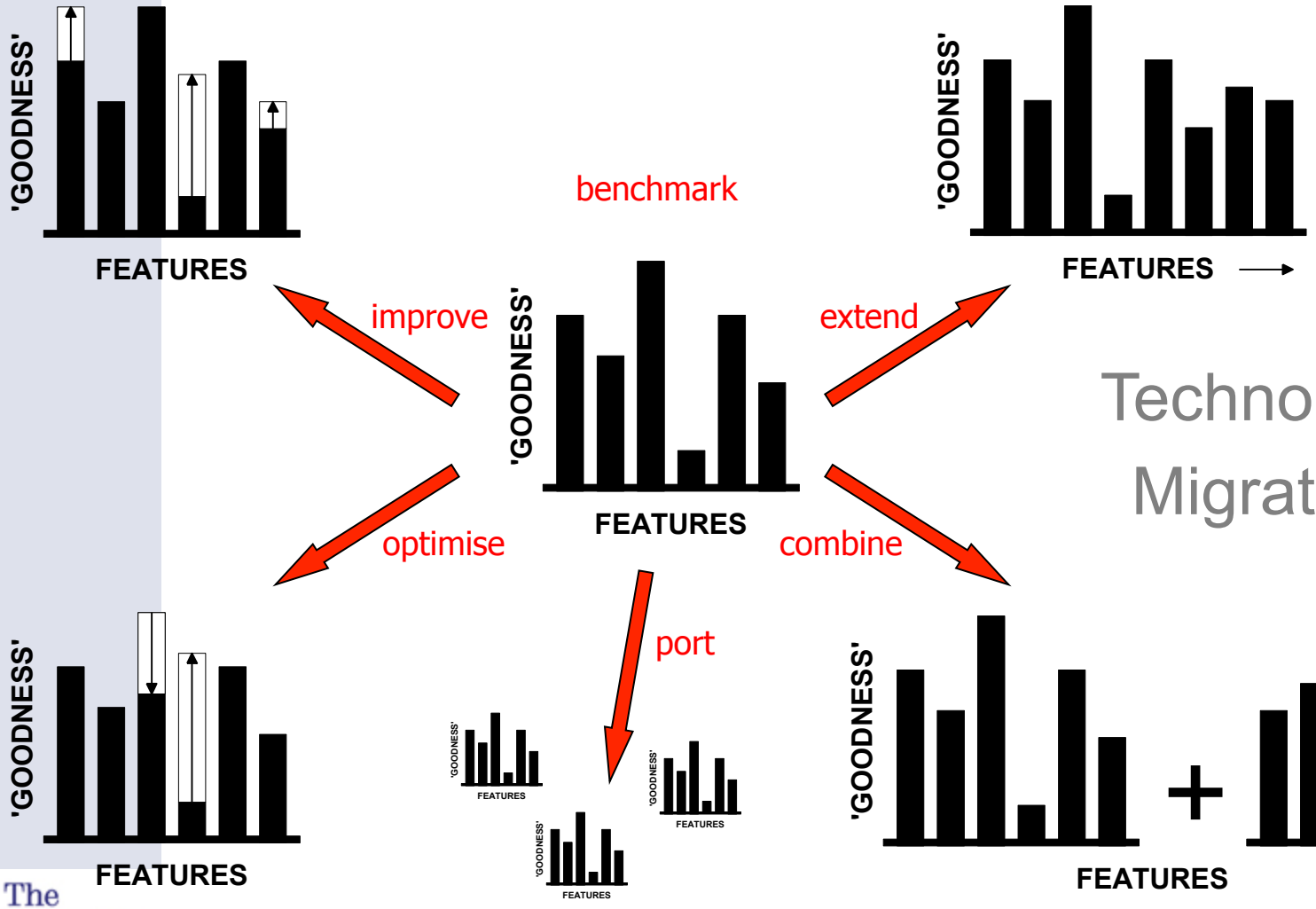
'Users Guide', R. K. Moore, *Handbook of Standards and Resources for Spoken Language Systems*, D. Gibbon, R. K. Moore and R. Winsky (eds.), Mouton de Gruyter, pp 1-28, 1997.

© 2014 The University of Sheffield

Performance Characterisation

The 'Capability Profile'
(*performance envelope*)



© 2014 The University of Sheffield

Performance Characterisation