

Speech Technology

Past, Present & Tech Transfer

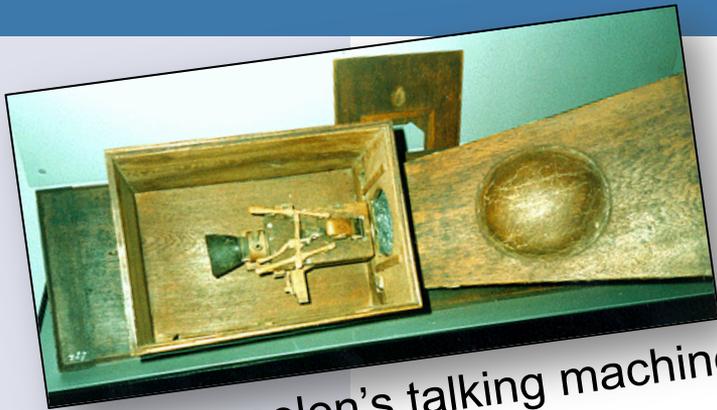
Prof. Roger K. Moore

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

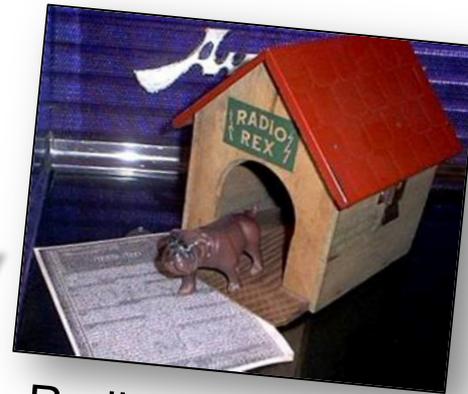


scentro
Sheffield Centre for Robotics

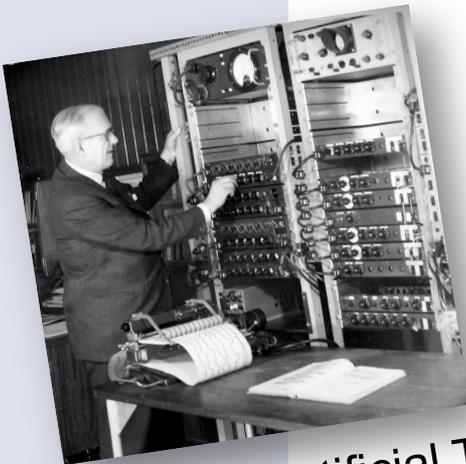
Rich History of Technological Development



Von Kempelen's talking machine (1791)



Radio Rex (1922)



Parametric Artificial Talker (1953)



Interactive Talking Doll (1987)



Speak'n'Spell (1983)



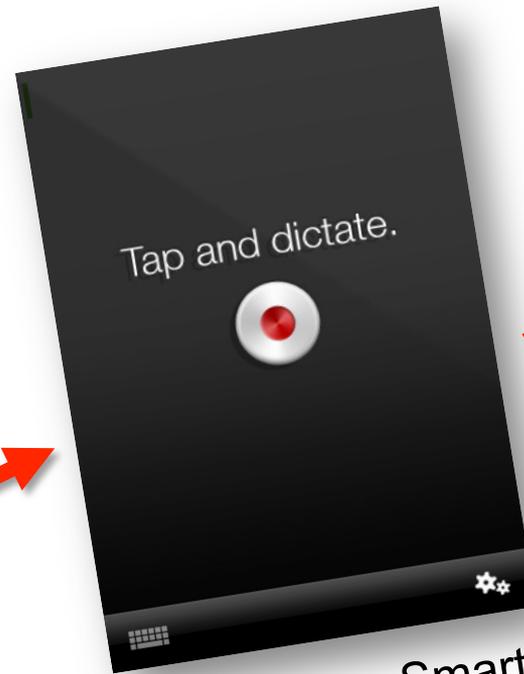
Rich History of Technological Development



Marconi 'SR128'
(1982)



Dragon 'Naturally Speaking'
(1997)

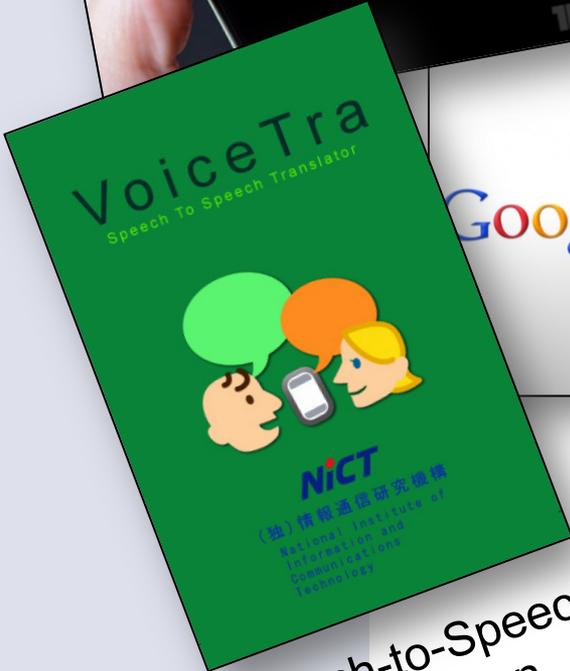
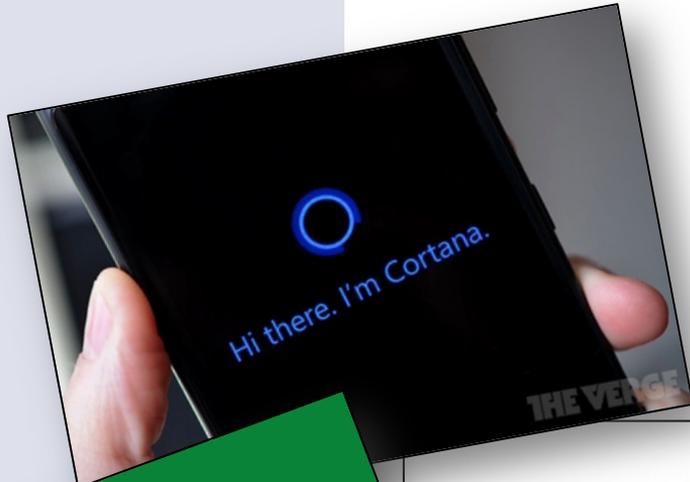


Voice dictation on SmartPhone
(2007)



Apple's "Siri"
(2011)

Rich History of Technological Development



Speech-to-Speech Translation



Apple's "Siri"
(2011)

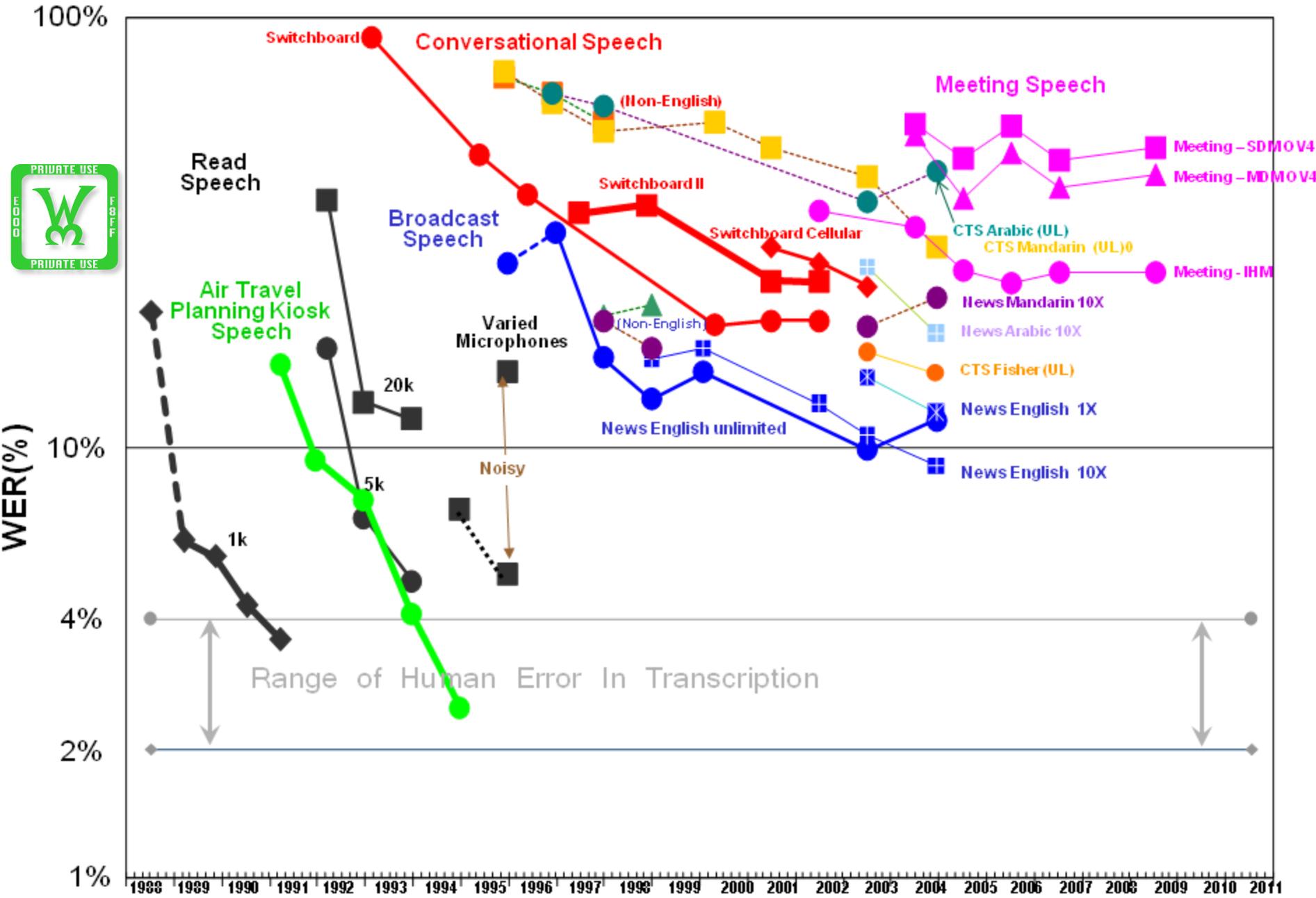


The 'State-of-the-Art'

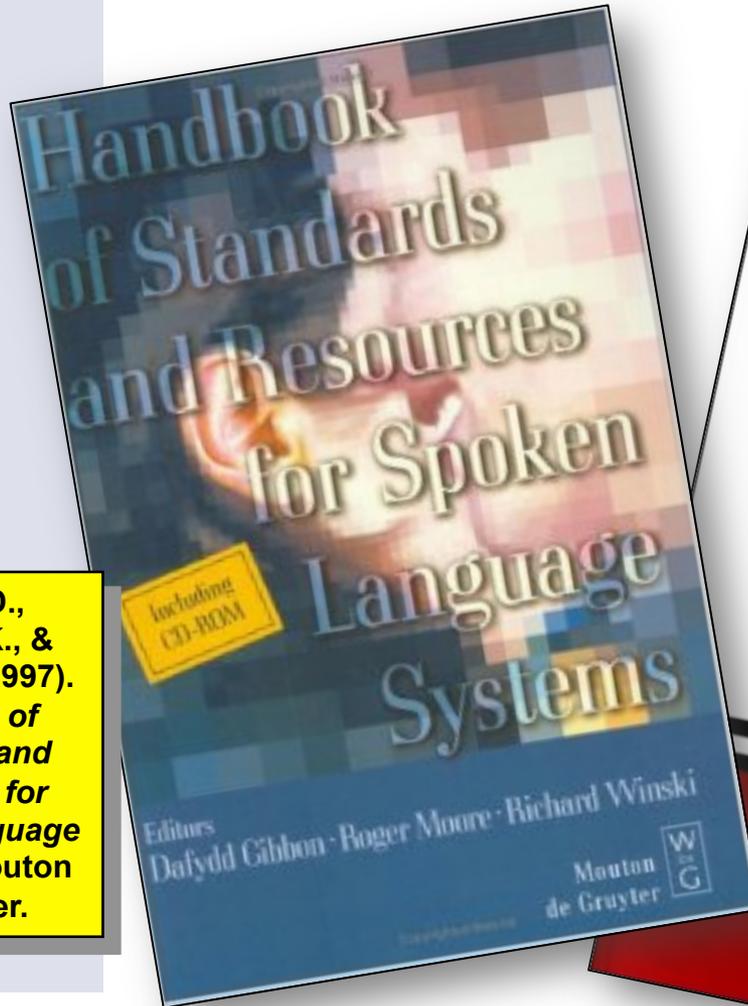


- There is steady year-on-year progress
- Improvement are coming from:
 - increase in available computer power
 - corpus-driven **statistical modelling** approaches
 - public benchmark testing (+ *standards & resources*)

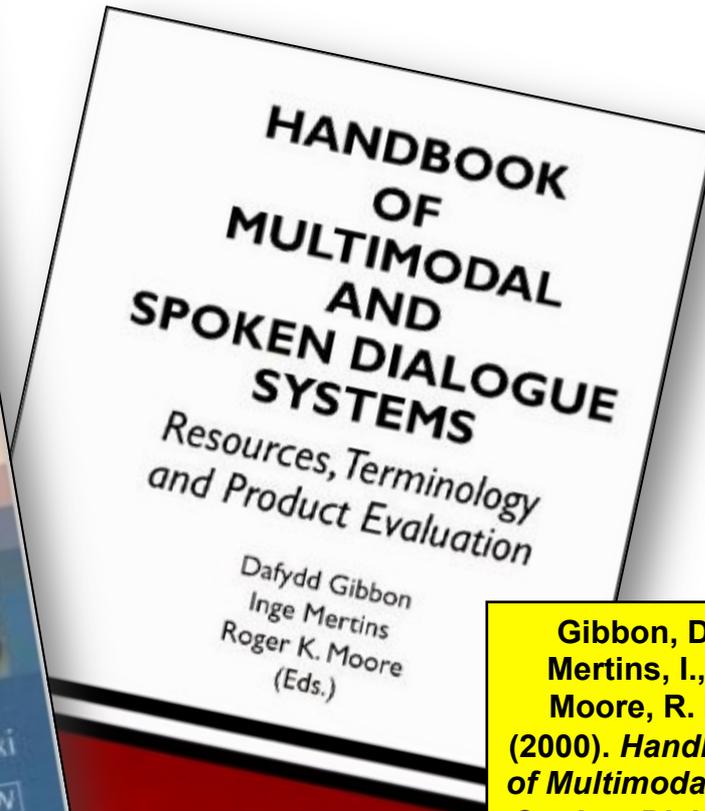
NIST STT Benchmark Test History – May. '09



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.

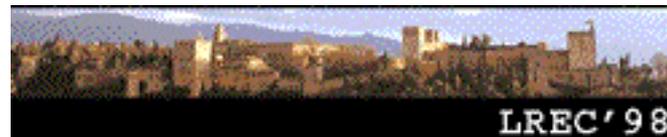


Standards & Resources



LREC 2014
Reykjavik

Harpa Conference Centre
26 - 31 May, 2014



Standards & Resources

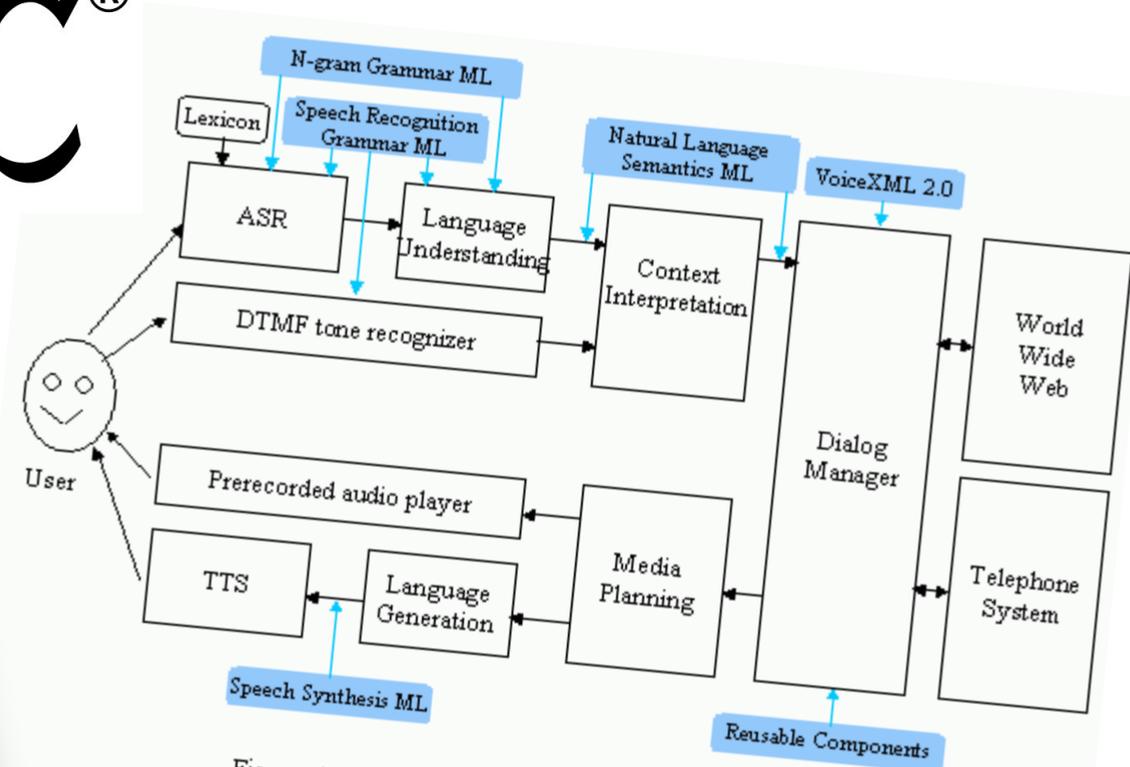
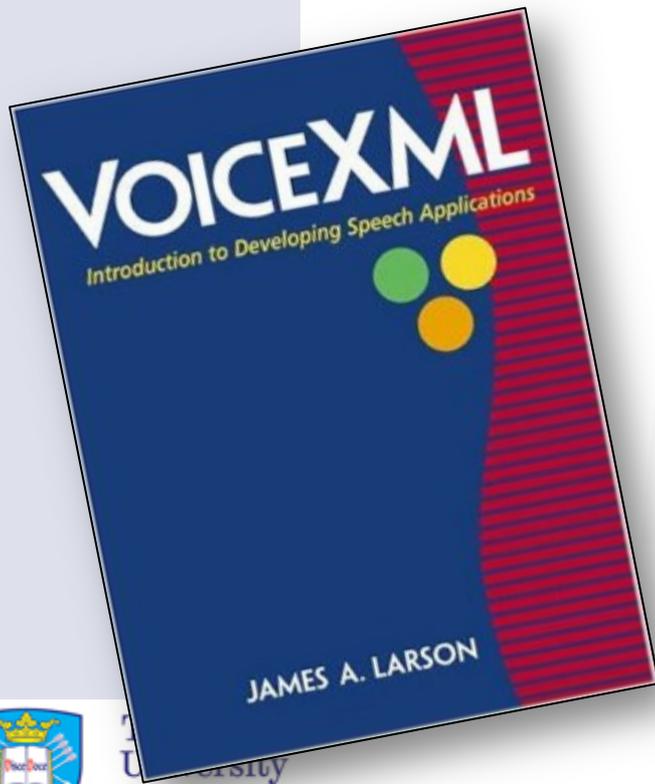


Figure 1. W3C Speech Interface Framework



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

Still Some Way to Go?

BBC News - Brummie a x
www.bbc.co.uk/news/uk-england-birmingham-20204516

Cookies on the BBC website We use cookies to ensure that we give you the best experience on our website. If you continue without changing your settings, we'll assume that you are happy to receive all cookies on the BBC website. However, if you would like to, you can **change your cookie settings** at any time.

BBC News Sport Weather iPlayer TV Radio More

NEWS BIRMINGHAM & BLACK COUNTRY

Home World UK England N. Ireland Scotland Wales Business Politics Health Education Sci/Environment Te

6 November 2012 Last updated at 12:58 482 Share

Brummie accent baffles automated phone service

An automated phone system has been taken out of service by a council because it did not understand a councillor's Brummie accent.

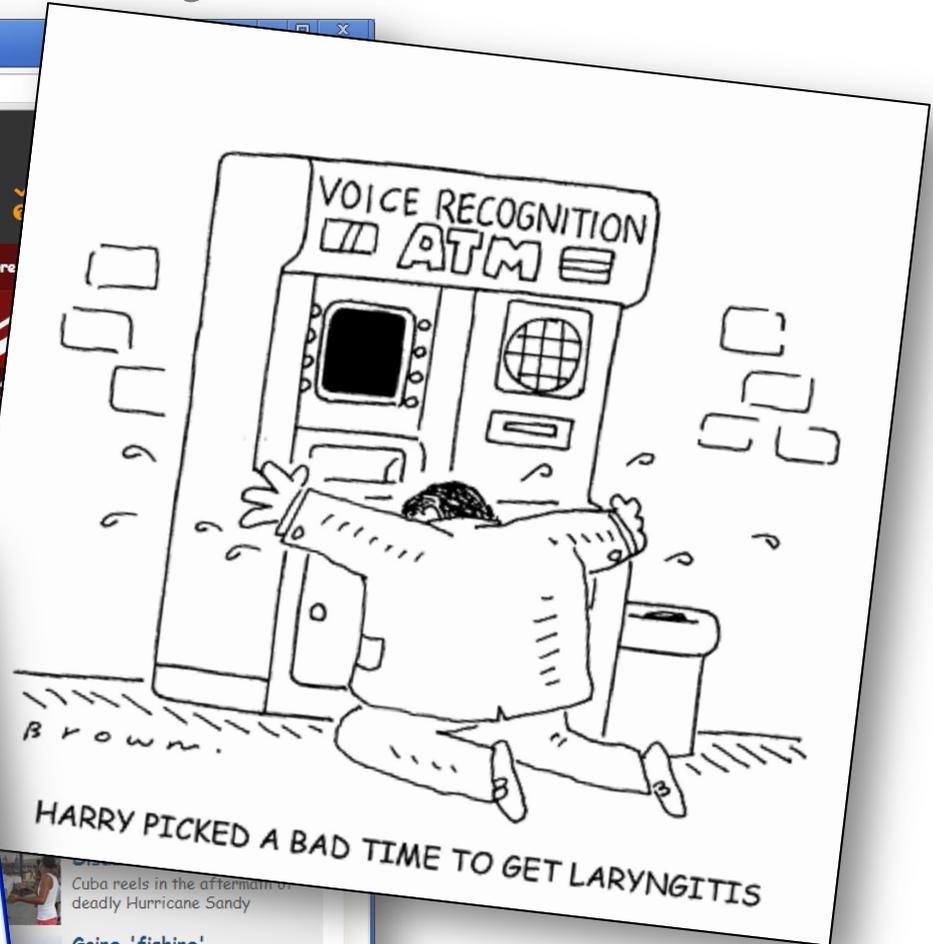
Mike Leddy said he was frustrated because his voice was not recognized by the automated system. The council said it had struggled to account for the local dialect. He said the system was taken out of service in October.

Mr Leddy said he was frustrated because he could not get through to the council. He said he had complained about the issue.

Phone Numbers, Shortcuts, Reviews & Customer Tips - Companies Worldwide - GetHuman.com

Need a company phone number? Other tips or information?

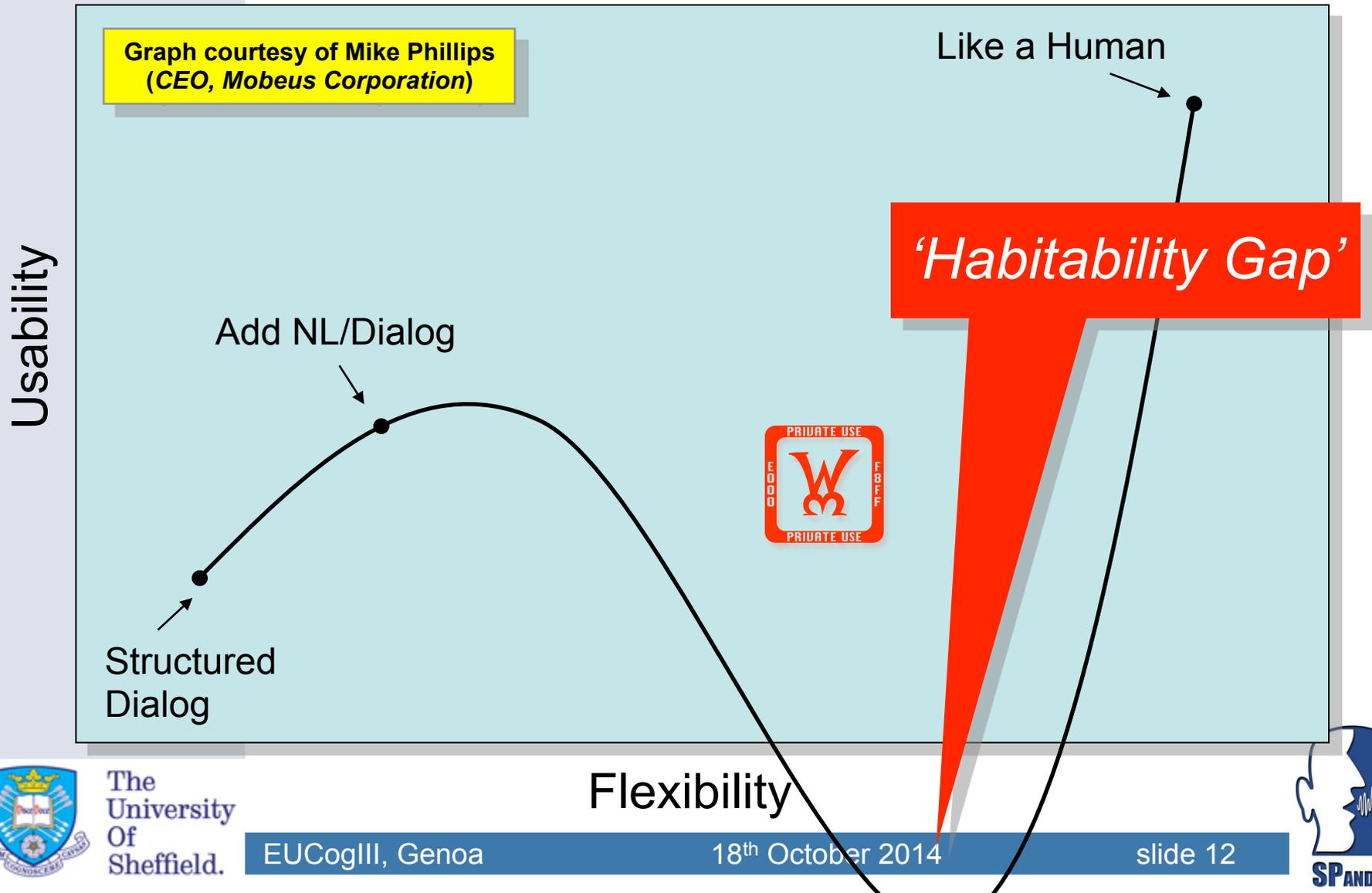
Most searched companies	phone number	steps to get a live person	Avg wait	user rating
AT&T	888-387-5270	This number connects directly with a A...	12.8 min	Average
Facebook	650-543-4800	Click for full details. Email only tec...	62.4 min	Horrible
Bank of America	877-231-9372	Keep dialing 0 at the prompts, and the...	9.5 min	Average
Comcast (Xfinity)	800-266-2278	Press *# at each prompt, ignoring mess...	15.1 min	Good
Verizon Wireless	800-922-0204	Press # at prompt, then press 0	13.7 min	Good
AT&T U-Verse	888-722-9337	Direct to human	9.5 min	Average
United Airlines	800-864-8331	Ignore the talking voice and press 0 a...	9.4 min	Average
Extrafax	800-848-5279	Direct to human	13.7 min	Average
Delta Airlines	800-492-6980	Press 0	15.0 min	Average
DirectTV	800-531-5000	Press 00 quickly in succession		



The University of Sheffield



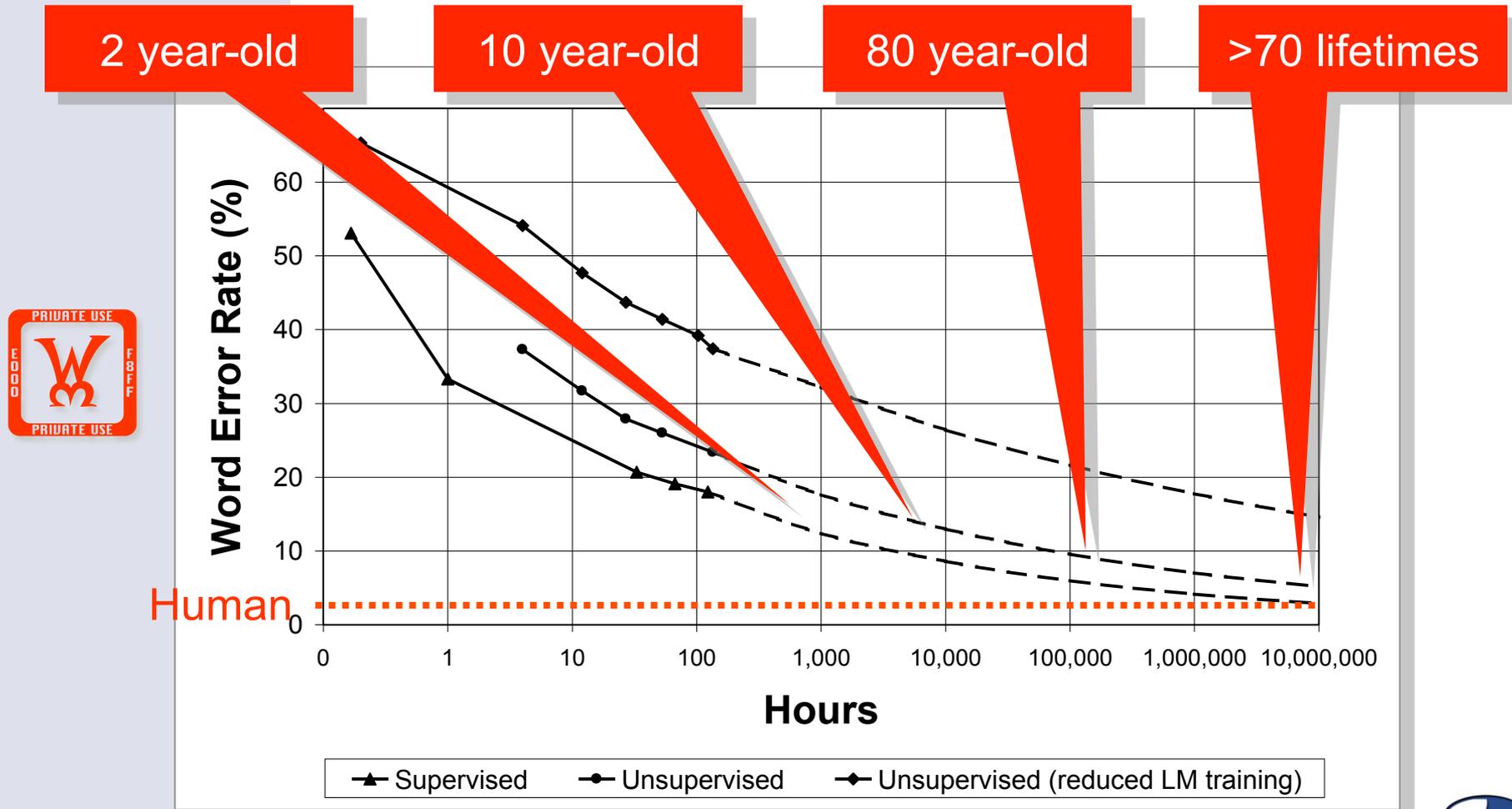
Still Some Way to Go?



Still Some Way to Go?

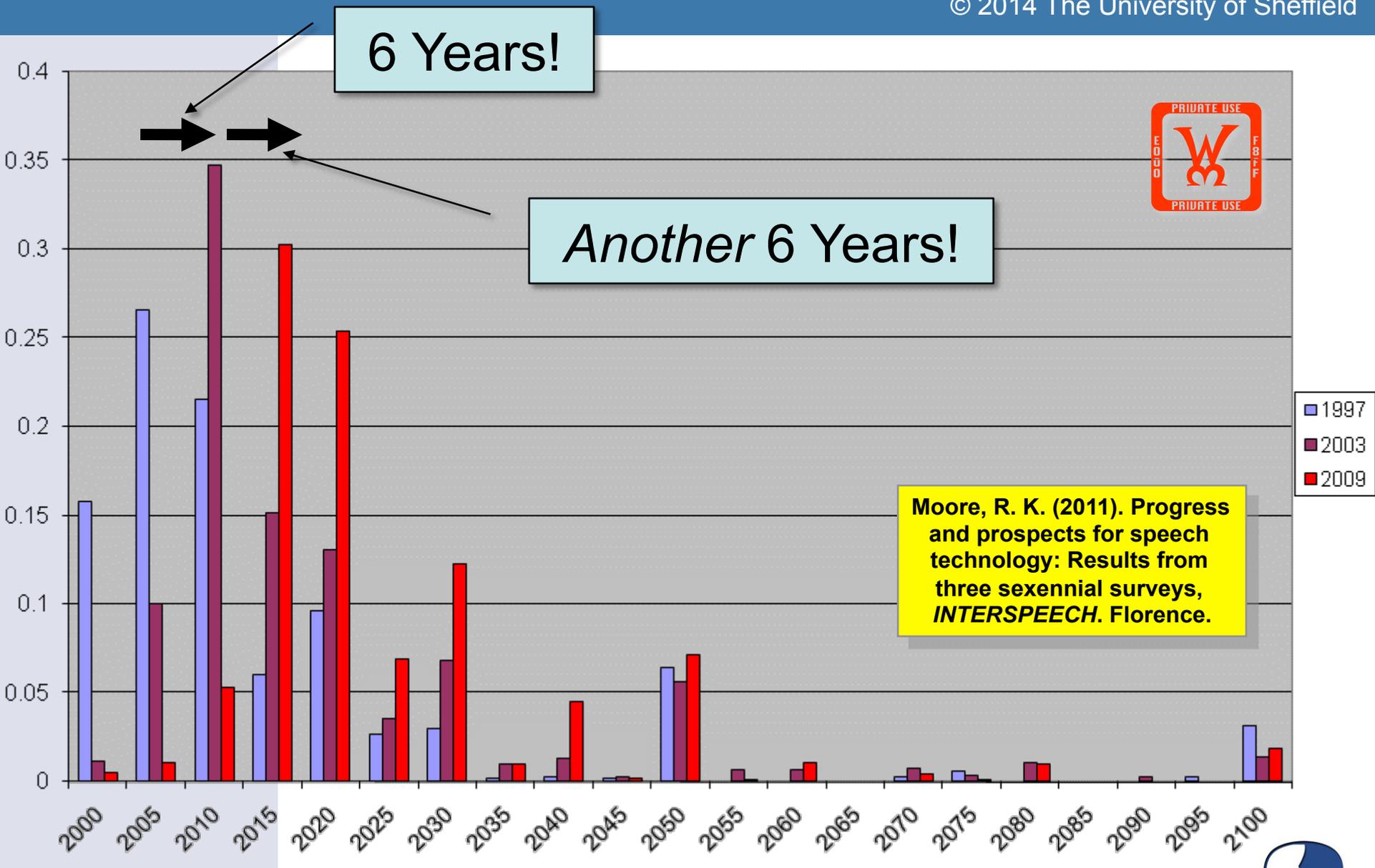


“There’s No Data Like More Data”



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





6 Years!

Another 6 Years!

Moore, R. K. (2011). Progress and prospects for speech technology: Results from three sexennial surveys, *INTERSPEECH*. Florence.



Still Some Way to Go?

Chatbot Conversation x
theinstitute.ieee.org/technology-focus/technology-topic/chatbot-conversation-goes-viral

The IEEE news source

the institute

Search The Institut Search

IEEE

About Quarterly Issue Archive Multimedia Blog Email Alert RSS

HOME PEOPLE **TECH FOCUS** BENEFITS BRIEFINGS OPINIONS REGION NEWS
CAREER & EDUCATION

HOME / TECH FOCUS / TECH TOPIC

Chatbot Conversation Goes Viral

A tale of two talking bots
By ANIA MONACO 17 November 2011

What happens when two artificial-intelligence bots talk to each other? Does their conversation resemble a human one? What do they talk about? Two Cornell University Ph.D. students led by an IEEE member decided in August to find out.

The students, Igor Labutov and Jason Yosinski, performed an experiment in which they set up two chatbots—artificial intelligence programs that learn and then mimic human conversation—to talk to each other via two laptops. They added voices to the bots, turned them into avatars with an animation program, shot a video of the interaction, and posted it on

SHARE AND PRINT
127

RELATED CONTENT

TECH FOCUS:
7 April 2008
In the Movies: Humans and Robots in Love

OPINIONS:
7 January 2008
Will Humans Marry Robots?

TECH FOCUS:
5 February 2010
Advances in Artificial Intelligence Make Games Smarter

Most Recent Most Viewed Most Commented

OPINIONS: An Indian Perspective on Cloud Computing

BRIEFINGS: Standards Association Names Managing Director

PEOPLE: Student Member Develops Device That Mimics Brain

PEOPLE: Sri Chandrasekaran: Spearheading Standards Development in India

BENEFITS: Conference Explores Consumer Electronics' Future

More recent >

ADVERTISEMENT

Stay Current and Share your Opinion.

Cleverbot



Still Some Way to Go?



Dangers of
misrepresenting
capabilities



The 'State-of-the-art'

Over-reliance on data?

Too easy to 'tweak' algorithms?



- Progress is steady year-on-year
- Improvements 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
 - **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?





Speech Communication 18 (1996) 205–231

**SPEECH
COMMUNICATION**

Towards **increasing** speech recognition error rates !

Hervé Boulard^{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 available techniques. In this case, almost any sufficiently new approach will necessarily hurt the accuracy of existing systems. However, if progress is to be made against the remaining

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



Challenges & Competitions

CHiM CHALLENGE

The Synthesizer Song
INTERSPEECH2007



SynSIG Speech Synthesis Special Interest Group

- ▀ Main Page
- ▀ SynSIG committee
- ▀ Publications
- ▀ Speech Synthesis Workshops
- ▀ Software
- ▀ Education
- ▀ Samples
- ▀ Evaluation
- ▀ Blizzard Challenge
- ▀ Pointers
- ▀ Future events
- ▀ Index
- ▀ Recent changes

Blizzard Challenge

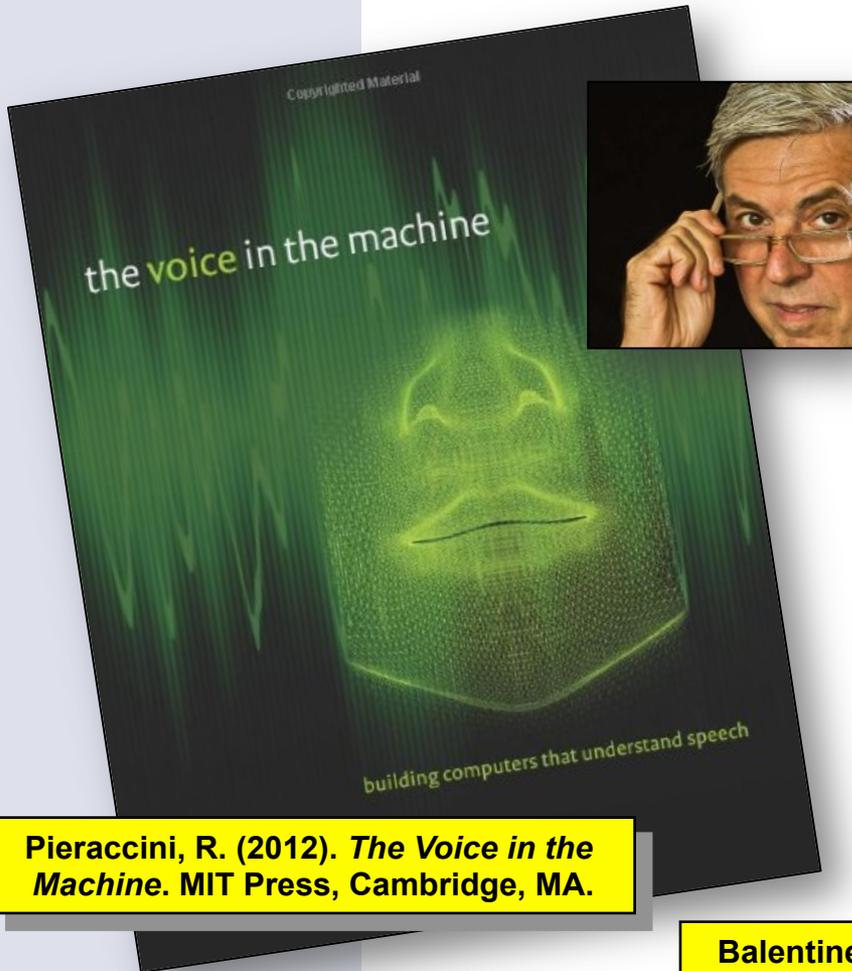
In order to better understand and compare research techniques in building corpus-based speech synthesizers on the same data, the Blizzard Challenge has been devised. The basic challenge is to take the released speech database, build a synthetic voice from the data and synthesize a prescribed set of test sentences. The sentences from each synthesizer are then evaluated through listening tests.

Contents [hide]

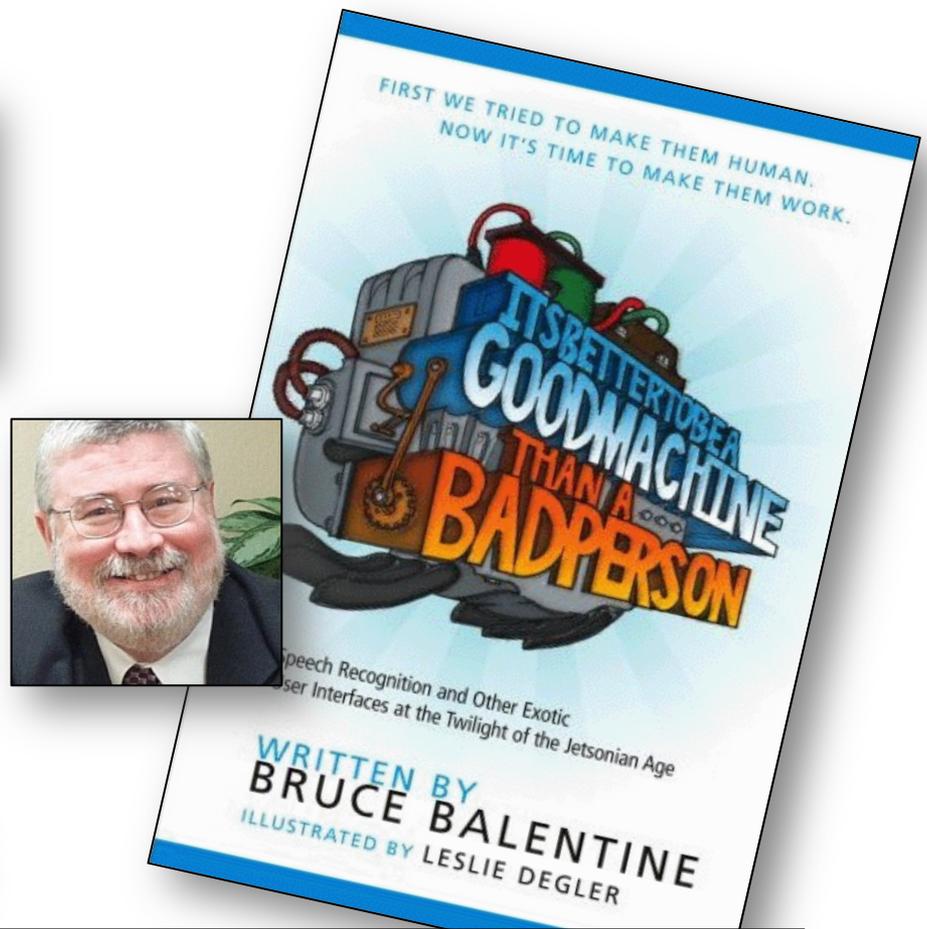
- 1 General information
 - 1.1 Mailing list
- 2 The annual challenges
- 3 Tools and data



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.

Where to Find Out More

Copyrighted Material

the voice in the machine

building

Progress and Prospects

www.dcs.shef.ac.uk/~roger/progress.html

My Bookmarks Prof. Roger K. Moore Save to Mendeley

Inbox - rogerkmoore007@gmail... Google Calendar Progress and Prospects

Home About Me Research Publications Teaching Seminars Administration

Prof Roger K Moore

BA(Hons) MSc PhD FIOA FISCA MIET MIEEE [contact](#)

[Follow @rogerkmoore](#)

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)

2014

- June:** Amadeus Capital Partners leads a £750K seed funding round for Cambridge University's latest spinout VocalIQ.
- June:** A rumour circulates that Samsung may be interested in purchasing Nuance.
- June:** A furore breaks out after Prof. Kevin Warwick announces that a chatbot (based on a 13-year-old Russian boy called Eugene Goostman) passed the Turing Test at an event hosted by the Royal Society.
- Apr:** Apple confirmed that it had acquired Novauris Technologies sometime during 2013.
- Mar:** Thad Starner at the Georgia Institute of Technology demonstrates a prototype dolphin translator called Cetacean Hearing and Telemetry (CHAT).
- Mar:** Google's Larry Page Admits "Speech Recognition is Not Very Good".
- Mar:** A report by MarketsandMarkets predicts that the speech analytics market will be worth \$1.33 Billion by 2019.

TO MAKE THEM HUMAN.
TIME TO MAKE THEM WORK.

BETTER TO BE A
MACHINE
THAN A
PERSON

otic
the Jetsonian Age

ENTINE
EGLER

Machine Than a
Exotic User
ve: ICMI Press.

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



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 been active in standards work, has been active in standards work, and has been active in standards work.

Robots That Talk and Listen is an examination of speech and language from a functional, and social perspective.

Language is an extremely difficult task. Contributors addresses the linguistic challenges that robots need to acquire in order to process of fictional automata. This multi-faceted ability to learn language requires a multi-faceted ability to learn language designed for communication. This requires structure, meaning, and context in a coherent dialogue.

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

The social impact of the specific roles they provide of those issues, notably access to robots, and developing ethical

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.



The
University
Of
Sheffield.

DE GRUYTER

121 High Street, Third Floor - Boston, MA 02110 - T (857) 284-7073 - F (857) 284-7358
info@degruyter.com · www.degruyter.com

EUCogIII, Genoa

18th October 2014

slide 23





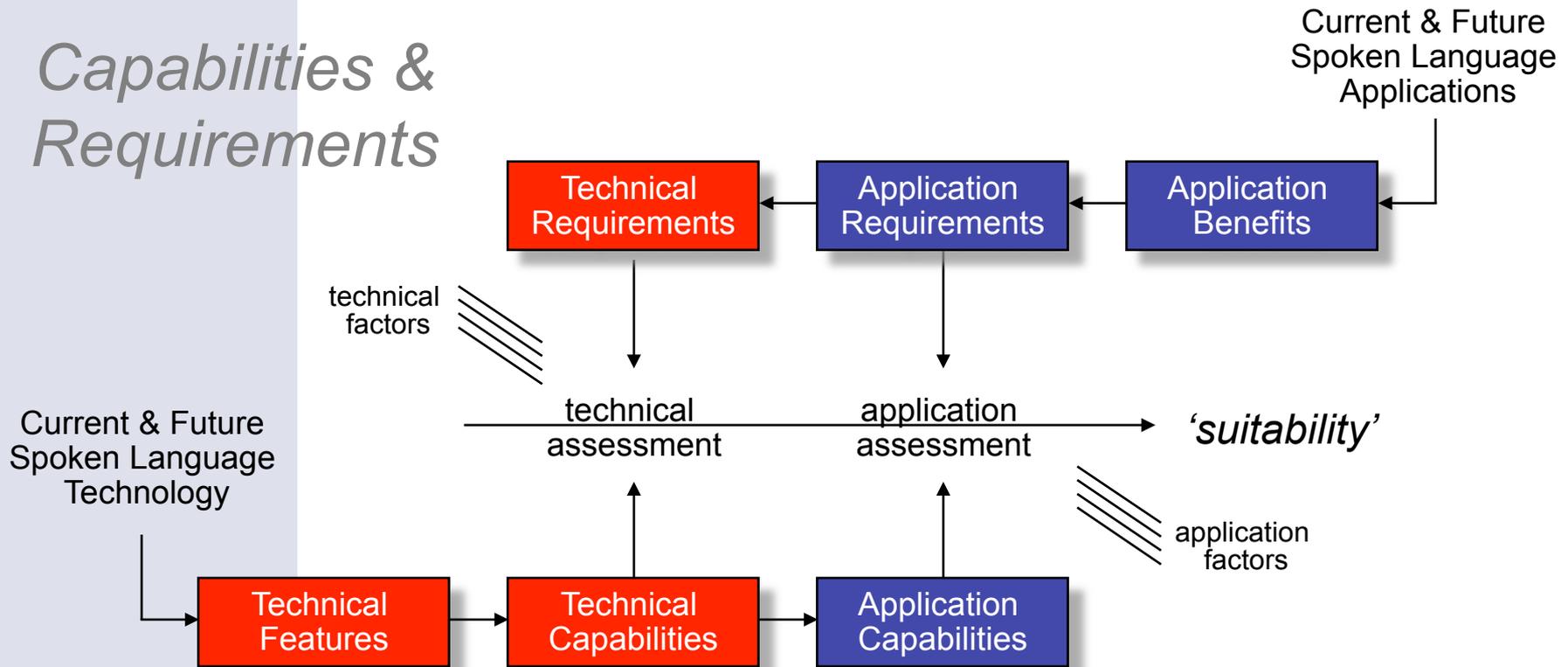
Thank You

Any questions?

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

Performance Characterisation

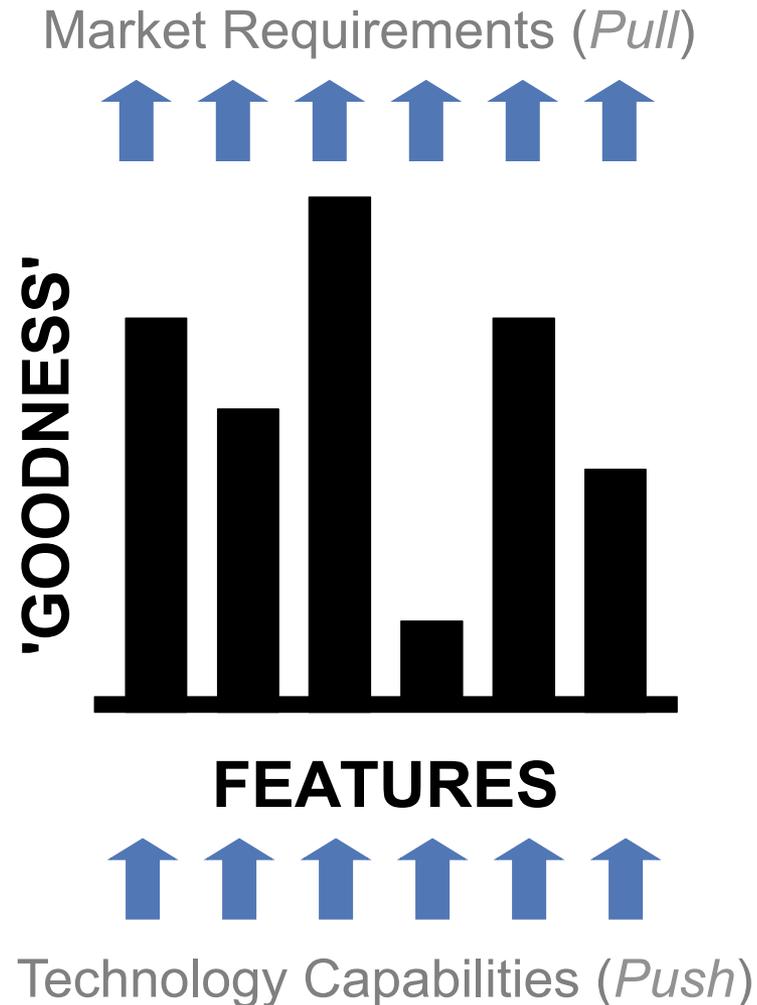
Capabilities & Requirements



'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.

Performance Characterisation

The 'Capability Profile'
(*performance envelope*)



Performance Characterisation

