

# Predicting the Future of Speech Technology

*speculation & trends*

Roger K. Moore  
20/20 Speech Ltd.

## Overview

How good is the technology now ?

What's in store for the future ?

What are the R&D challenges ?

But ... how good does it need to be ?

... and when will it be good enough ?

# Speech Technology Applications

**“Radiology report number 5 6 3 dated 19 November 1998”**



Medicine



REPORT TRANSMITTED

... Recce Report Alpha Bravo One Ends

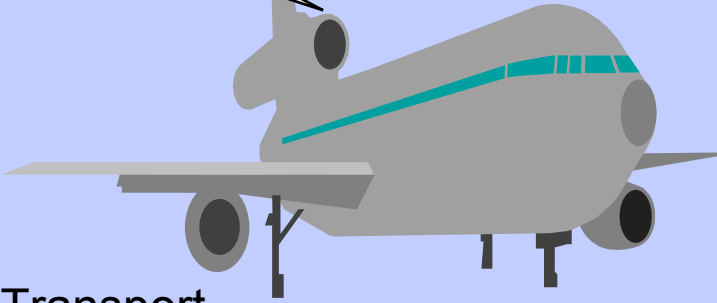
Classify Track 1234 as Hostile

ACCEPTED

Show Route to GPS Waypoint X-Ray

Speech Activation on the Battlefield


**“Engage ILS”**



**“ILS engaged”**

Transport

**“... yours sincerely etc. End memo. Please mail by tonight.”**



Office

# Progress

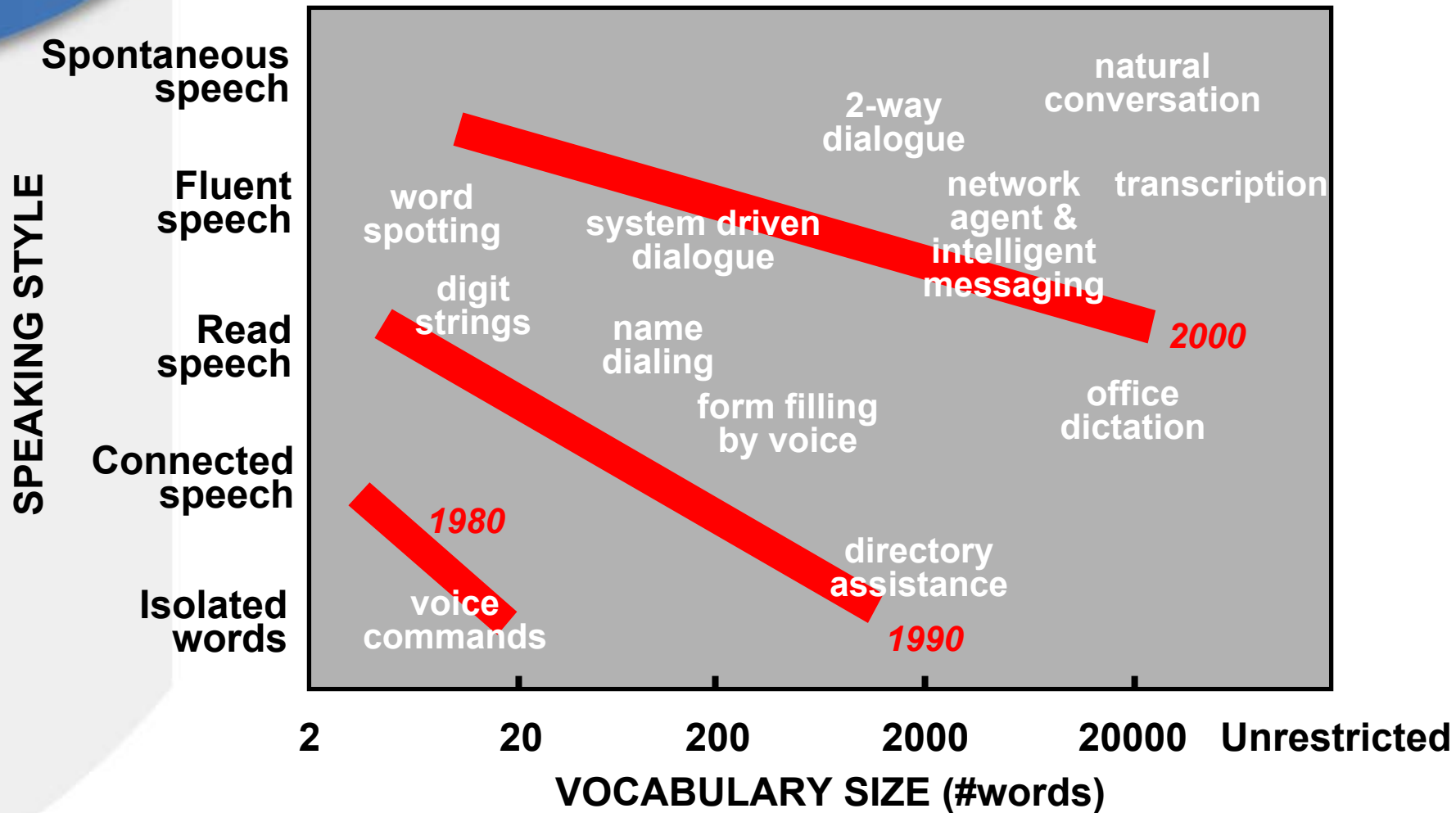


Figure by permission of Prof. Sadaoki Furui, TIT, Japan.

aurix

# Speech Technology on the Move

Aurix<sup>®</sup> activator



**Wireless**  
**Delivered**

Enabling the wireless community

**Kane**  
EMAIL PLAYER

**Dixons**

**Argos**

 **Handango<sup>™</sup>**

**QinetiQ**



**JAGUAR**

THE ART of PERFORMANCE



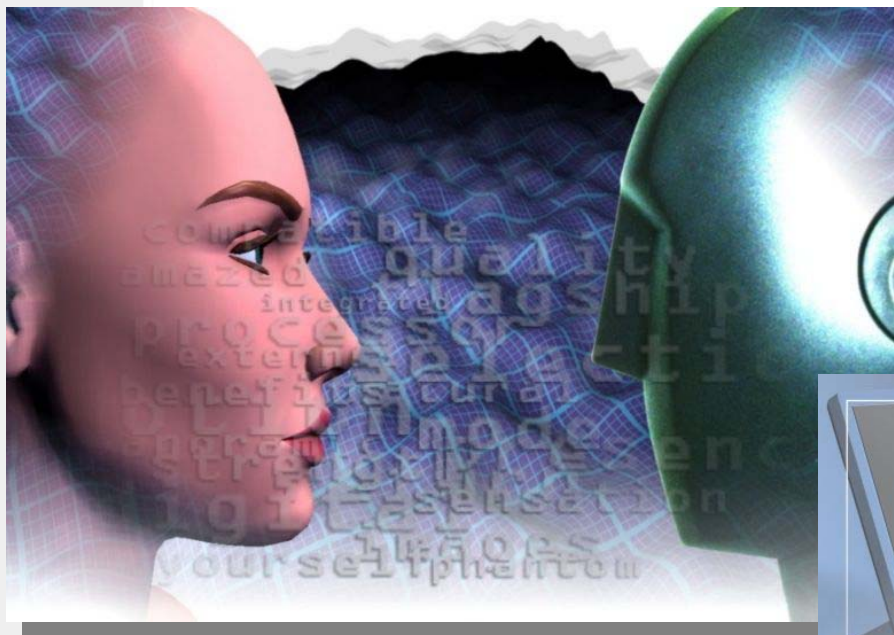
Ministry of  
**DEFENCE**

**Aurix<sup>®</sup> asr**  
**Aurix<sup>®</sup> tts**



aurix

# What's in Store for the Future ?



# What's in Store for the Future ?

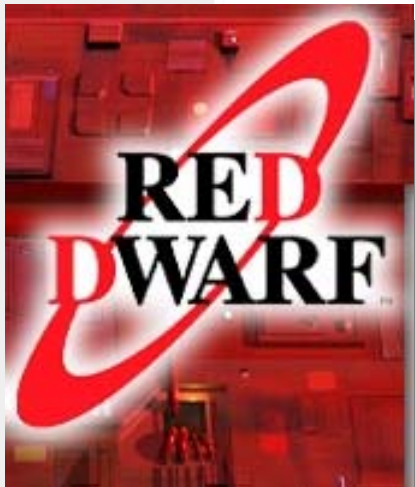


“It is hard to predict ...”

“... especially the future.”

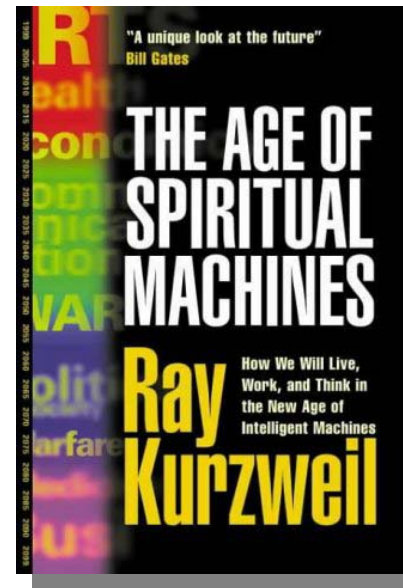
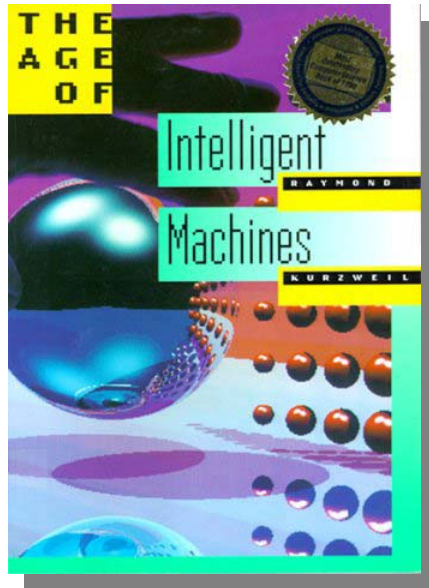
**Niels Bohr, 1922**

# aurix Science Fiction

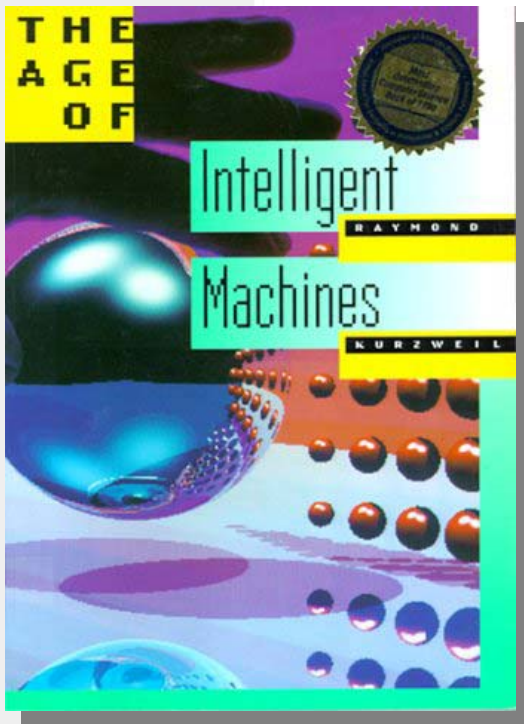




# Informed Speculation



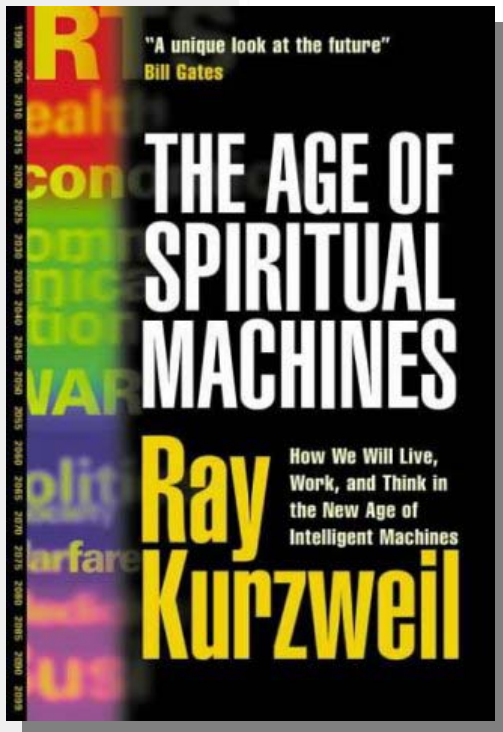
“A PC will have the computational power of the human brain by 2019, and will be equivalent to 1000 human brains by 2029.”



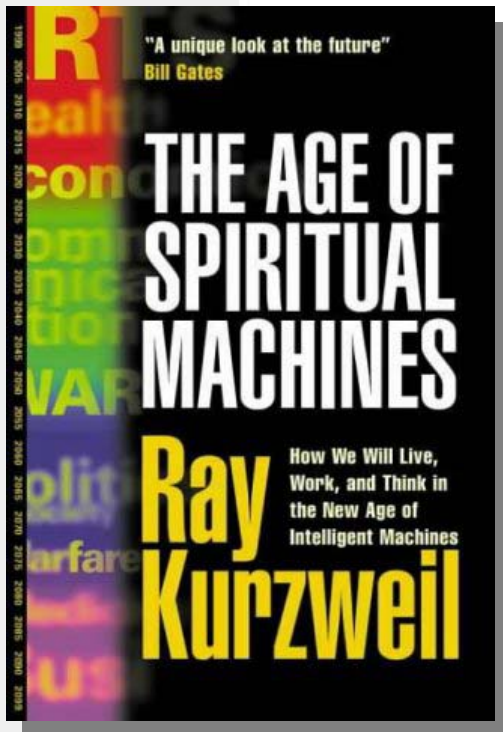
“translating telephones allow two people across the globe to speak to each other even if they do not speak the same language”

“speech-to-text machines translate speech into a visual display for the deaf”

“telephones are answered by an intelligent answering machine that converses with the calling party to determine the nature and priority of the call”



- “the majority of text is created using continuous speech recognition”
- “ubiquitous language user interfaces”
- “most routine business transactions take place between a human and an animated visual presence that looks like a human face”
- “pocket-sized reading machines for the visually impaired”
- “listening machines for the deaf”
- “translating telephones commonly used for many language pairs”

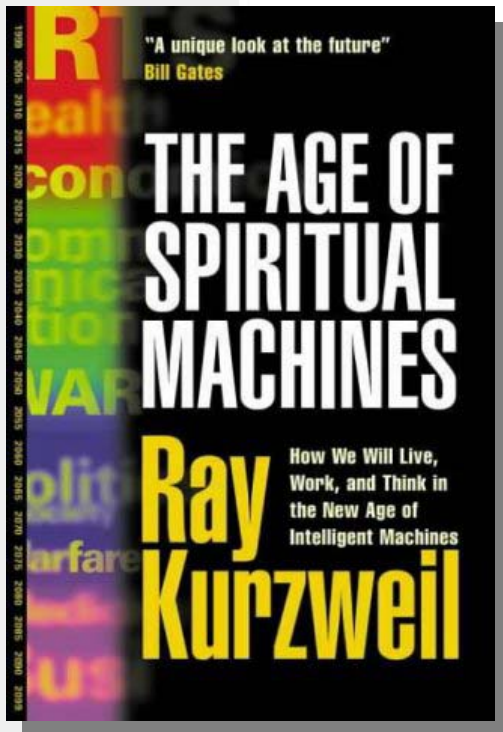


“most interaction with computing is through gestures and two-way natural-language spoken communication”

“deaf persons read what other people are saying through their lens displays”

“the vast majority of transactions include a simulated person”

*“people are beginning to have relationships with automated personalities”*



“implants are used to provide input and output between the human user and the world-wide computer network”

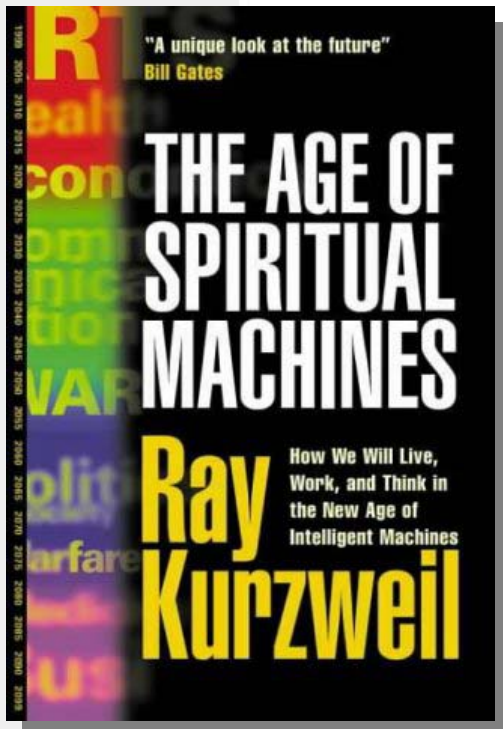
“automated agents are learning on their own”

“the majority of communications involving a human is between a human and a machine”

“growing discussion on what constitutes being human”



# aurix 2049 - 2099



*“no longer any clear distinction between humans and computers”*

# How Likely is Any of This ?

“The industry has yet to bridge the gap between what people want and what it can deliver.”

“Reducing the ASR error rate remains the greatest challenge.”



Xuedong "X.D." Huang

**Microsoft**

**‘Making Speech Mainstream’, X. D. Huang, Microsoft Speech Technologies Group, 2002.**



**Talk**knowledge  
to**kn**o**le**d**z**i

“After sixty years of concentrated research and development in speech synthesis and text-to-speech (TTS), our gadgets, gizmos, executive toys and appliances still do not speak to us intelligently.”

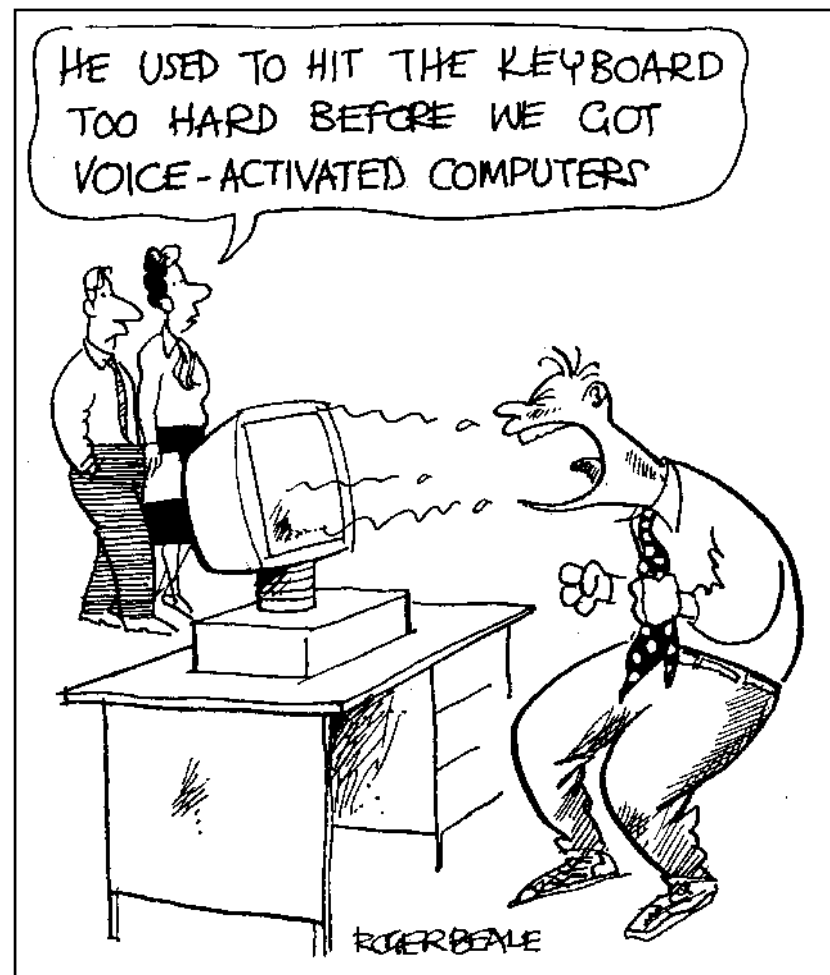
**‘Fiction and Reality of TTS’, Speech Technology Magazine, vol.7, no.1, Jan/Feb 2002.**

# aurix R&D Challenges

The Corpus Facility Portal

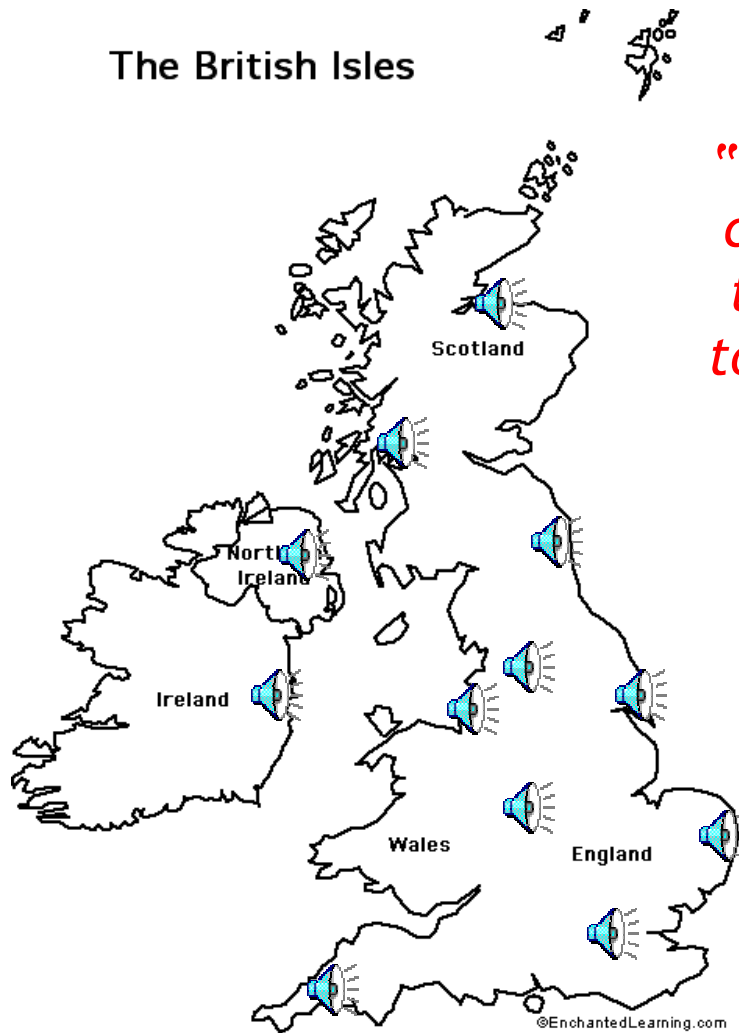


School of Linguistics  
& Applied Language Studies



# R&D Challenges

The British Isles

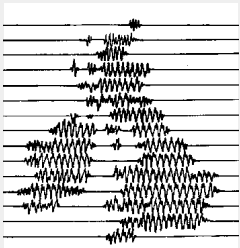


*“When a sailor in a small craft faces the might of the vast Atlantic Ocean today, he takes the same risks that generations took before him”*



Standard British English  
(what dictionaries capture)

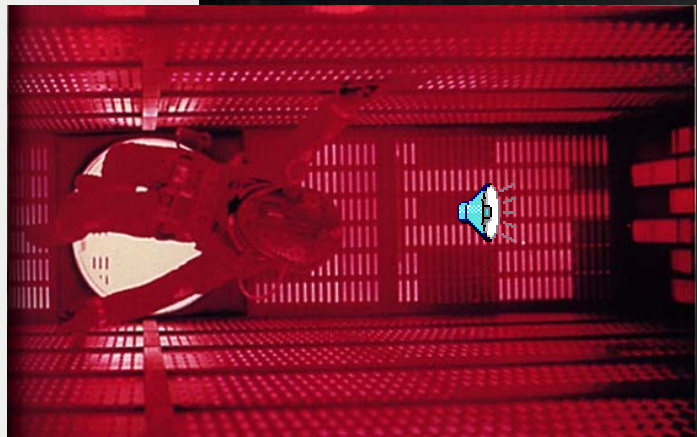
## ABI



*Accents of the  
British Isles*



# aurix R&D Challenges



A composite image featuring the movie title "2001: A SPACE ODYSSEY" at the bottom, a glowing green sphere in the center, and a vertical stack of logos on the right: rhetorical, AT&amp;T Natural Voices, NUANCE, ScanSoft, and Elanspeech. Each logo is accompanied by a small speaker icon.



aurix

# R&D Challenges



## ASR

Accept input that is:

- spontaneous
- emotional
- whispered
- accented
- disfluent
- interrupted
- contaminated
- from the elderly
- from the young
- OOV rich

## TTS

Deliver output that is:

- communicative
- intelligible
- understandable
- acceptable
- appropriate
- expressive
- personalised
- sympathetic
- reactive

## Dialogue

Provide interaction that is:

- usable
- effective
- efficient
- engaging
- cost-effective
- enjoyable
- multi-modal
- portable

# How Good Does it Need to be ?

“It has to work perfectly”

... as good as a human

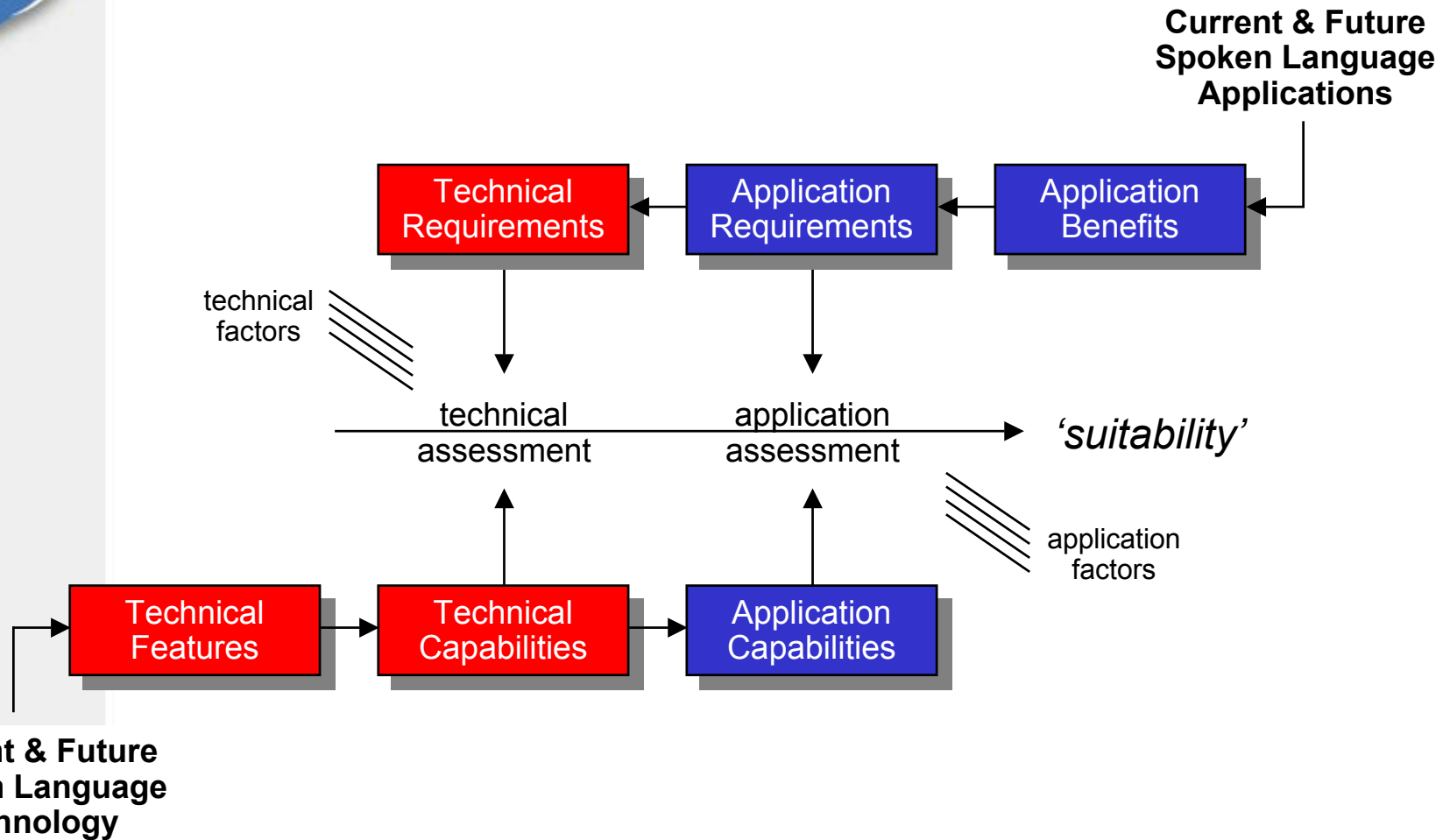
... better than a human (!)

... better than the alternatives



**“...If you’d like to hear all of your options again, press 49. If you’ve forgotten why you called in the first place, press 50.”**

# 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.

# How Good Does it Need to be ?

## Data Entry

- for ASR with voice correction to be better than 'Soft Typing', the WER needs to go from 18% to 5%
- which is about twice the human WER

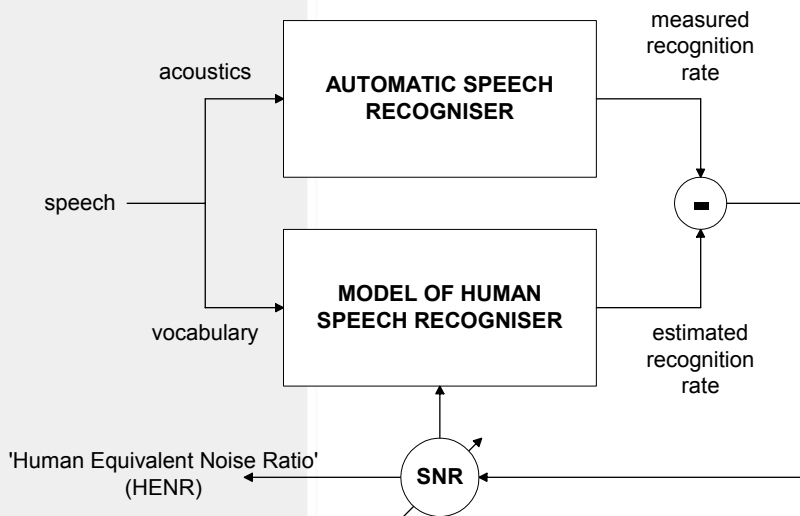
## Command Input

- for ASR to beat a dedicated 256-button keypad, the WER needs to go from 11% to 2%
- which is about twice the human WER

**'How good does ASR have to be ... and when will it be that good?', R. K. Moore, 2003.**



# When Will it be Good Enough ?



**'Evaluating Speech Recognisers', R. K. Moore, IEEE Trans. Acoustics, Speech and Signal Processing 25, pp 178-183, 1977.**

# When Will it be Good Enough ?

... in about 15-20 years

... assuming that the present rate of incremental progress can be sustained

... and that's another story!

## Conclusions

How good is the technology now ?

What's in store for the future ?

What are the R&D challenges ?

How good does it need to be ?

When will it be good enough ?

## 20/20 Speech Ltd.

Science Park, Malvern, Worcs., WR14 3SZ, UK

Tel: +44 1 684 585101 Fax: +44 1 684 585151

<http://www.2020speech.com>

<http://www.aurix.com>



# Questions

