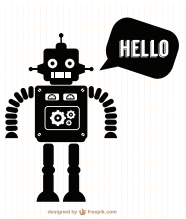
# Why, When and How to Converse with a Robot




**Prof. Roger K. Moore**

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

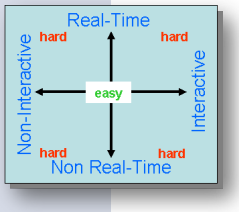
SHEFFIELD ROBOTICS

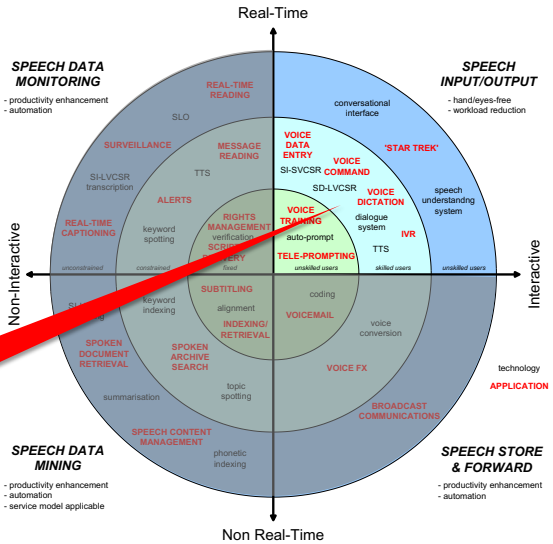


1



# Spoken Language Technology




Real-time interactive systems



2

# Talking with Machines

Command and Control Systems

Dictation Systems

Interactive Voice Response (*IVR*) Systems

Voice-Enabled Personal Assistants

Embodied Conversational Agents (*ECAs*)

Autonomous Agents (*robots*)



3



# Talking with Machines

Command and Control Systems

Dictation Systems

Interactive Voice Response (*IVR*) Systems

Voice-Enabled Personal Assistants

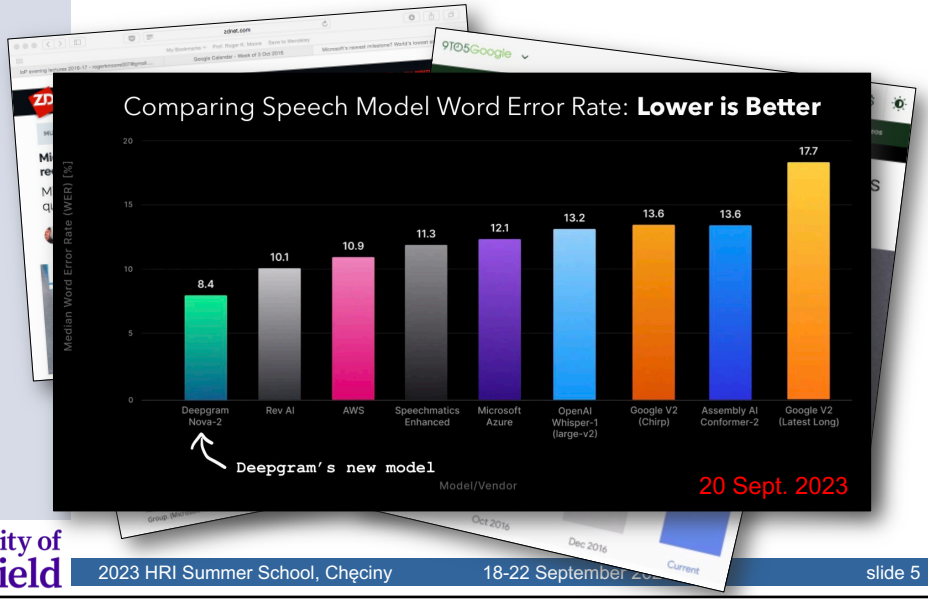
Embodied Conversational Agents (*ECAs*)

**Autonomous Agents (*robots*)**



4

# Amazing Progress






5



# Amazing Progress






- Tens of millions of Alexa-enabled devices were sold worldwide over the 2017 Christmas holiday season
- Google Assistant is available on over 225 home-control brands and more than 1,500 devices
- Apple's Siri has 41.4 million monthly active users in the U.S. as of July 2017

https://medium.com/swlh/the-past-present-and-future-of-speech-recognition-technology-cf13c179aaf



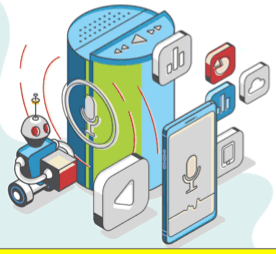
6

# Where Are We Going?

## GLOBAL VOICE ASSISTANT MARKET


**North America**
Largest Market
By Region (2019E)

**APAC**
Fastest-Growing Market
By Region (2020-2030)



**2019 Market Size $1,723.6 million**

**2030 Market Size $26,872.6 million**

**Market Growth Rate (2020-2030) 29.7 %**

https://www.psmarketresearch.com/market-analysis/voice-assistant-market



7

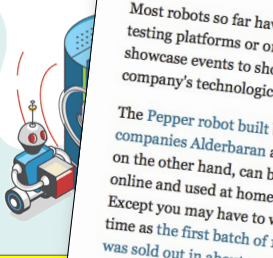


# Where Are We Going?


GLOBA

North
Largest
By Regi

ng Market
20-2030)



**Market Growth Rate (2020-2030) 29.7 %**

BUSINESS INSIDER UK Tech Finance Politics Strategy All
TECH
More: Innovation robot pepper

**The next generation of Siri-like assistants will be robots living in your home**

ANTONIO VILLAS-BOAS
JUN. 22, 2015, 6:40 PM 19

FACEBOOK LINKEDIN TWITTER

Most robots so far have remained on testing platforms or on stages at showcase events to show off a company's technological ability.

The Pepper robot built by Japanese companies Alderbaran and Softbank, on the other hand, can be bought online and used at home right now. Except you may have to wait a long time as the first batch of 1000 Peppers was sold out in about a minute, and it's only available in Japan.


Pepper, the personal digital assistant robot.

https://www.psmarket...
-analysis/voice-assistant-market



8

# Are We Nearly There Yet?

https://youtu.be/pIEFKWYW8bg

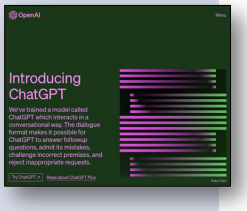

https://youtu.be/JepKVUym9Fg

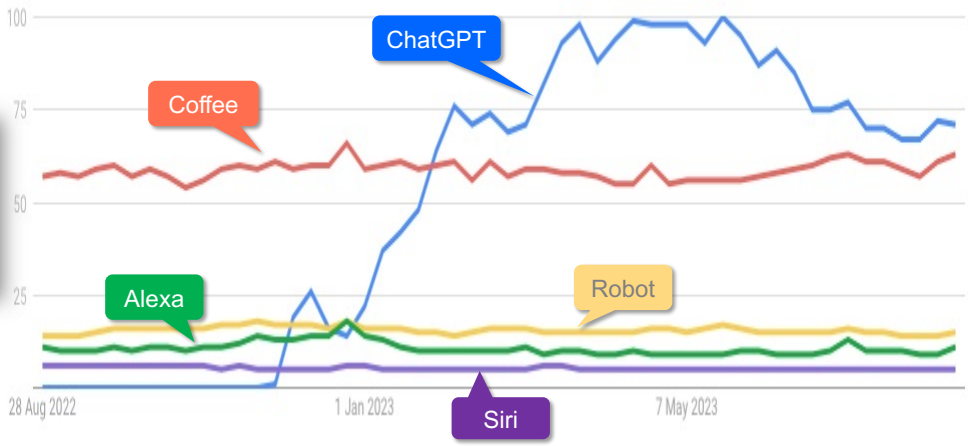


9



# Recent Trends



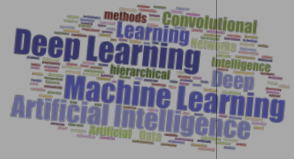






10

# The 'State-of-the-Art'

- There is steady year-on-year progress
- Improvements come from:
  - increase in available computer power
  - latest machine-learning paradigms
  - huge real-world training corpora
  - open tools / benchmark testing / challenges
- 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*)
  - inefficient (*trained on more data than a person hears in a lifetime*)
  - biased (*trained on un-curated data*)
  - untrustworthy (*e.g. hallucinates misinformation*)
  - ecologically damaging (*model training has a significant carbon footprint*)
- It is easy to underestimate the richness and complexity of spoken language interaction
- It is not 'natural' to talk to a machine!
- The availability of open tools and data is de-skilling



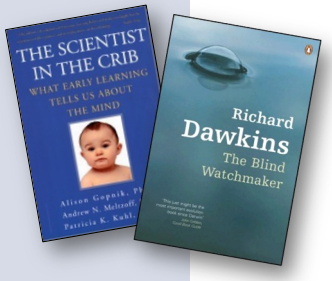



11



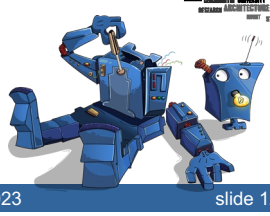
# Speech is not just another Modality

*"Spoken language is the most sophisticated behaviour of the most complex organism in the known universe!"*





Simply interfacing state-of-the-art speech technology with a state-of-the-art robot does not lead to effective human-robot interaction (*talking and listening is not enough*)





12

# Speech is not just Audible Text


communicative

ostensive

situated

embodied

pragmatic

grounded

Speech is …

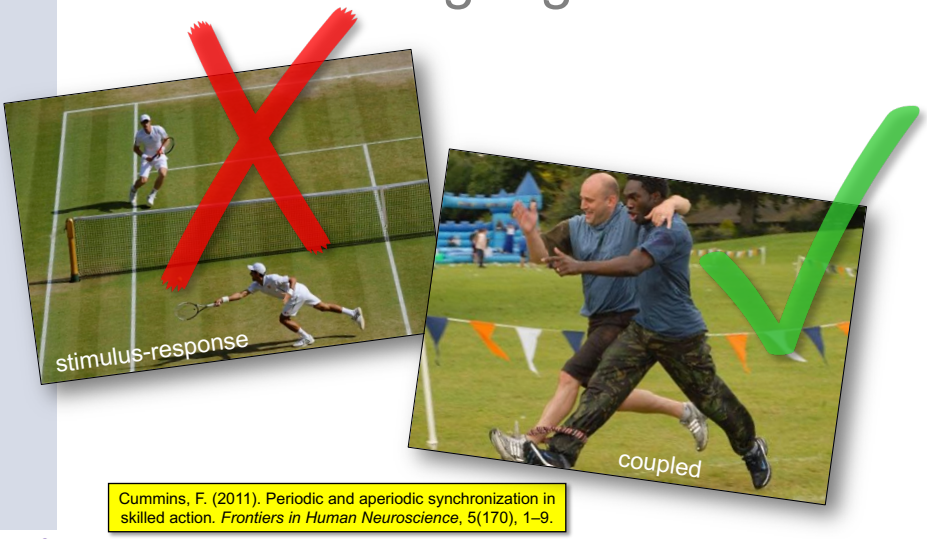
- variable
- ambiguous
- effortful
- contrastive
- prosodic
- adaptive
- context-dependent
- meaningful
- referential
- indexical
- rhetorical
- paralinguistic
- personalised
- affective
- multimodal
- contaminated



13



# What is Language Like?




Cummins, F. (2011). Periodic and aperiodic synchronization in skilled action. *Frontiers in Human Neuroscience*, 5(170), 1–9.

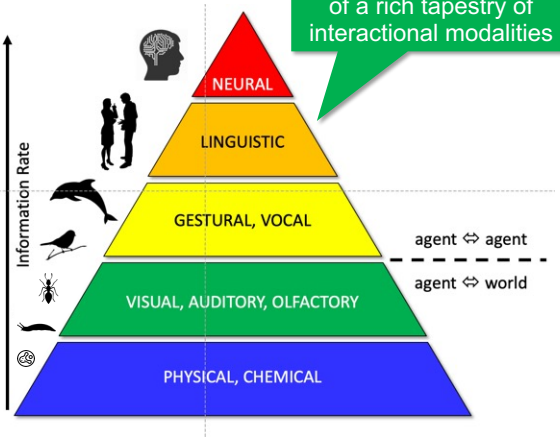


14

# A Unifying Framework

Spoken language is part of a rich tapestry of interactional modalities

Information theory provides a means for unification



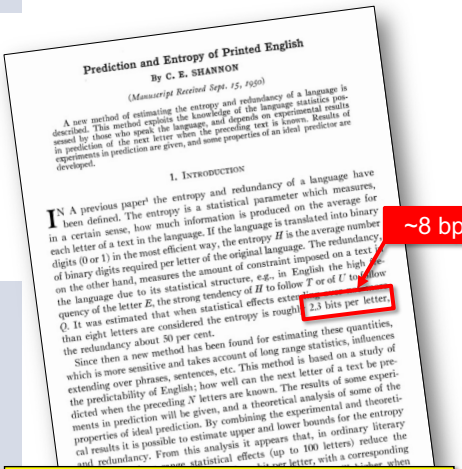

Moore, R. K. (2023). Pragmatics, Synchronics and Energetics in Spoken Language – an Information Theoretic Perspective. In *Workshop on Limits and Benefits of Information-Theoretic Perspectives in Spoken Communication*, Dublin.



15



# The Information Rate of Spoken Language



Prediction and Entropy of Printed English

By C. E. SHANNON

(Manuscript Received Sept. 15, 1950)

A new method of estimating the entropy and redundancy of a language is described. This method exploits the knowledge of the language statistics possessed by those who speak the language, and depends on experimental results in prediction of the next letter when the preceding text is known. Results of experiments in prediction are given, and some properties of an ideal predictor are developed.

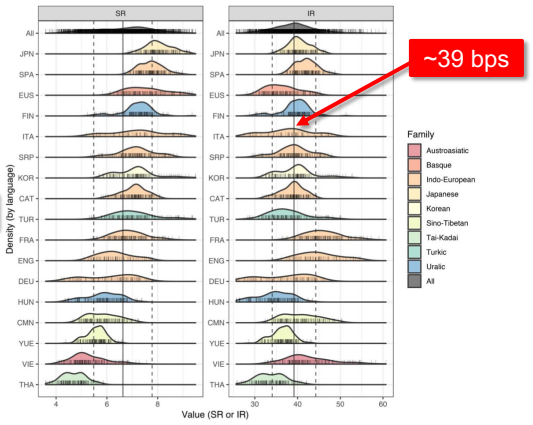
1. Introduction

In a previous paper the entropy and redundancy of a language have been defined. The entropy is a statistical parameter which measures, in a certain sense, how much information is produced on the average for each letter of a text in the language. If the language is translated into binary digits (0 or 1) in the most efficient way, the entropy $H$ is the average number of binary digits required per letter of the original language. The redundancy, on the other hand, measures the amount of constraint imposed on a text in the language due to its statistical structure, e.g., in English the high frequency of the letter $E$, the strong tendency of $H$ to follow $T$ or of $U$ to follow $Q$. It was estimated that when statistical effects extend over not more than eight letters the entropy is roughly 2.3 bits per letter, the redundancy about 50 per cent.

Since then a new method has been found for estimating these quantities, which is more sensitive and takes account of long range statistics, influences extending over phrases, sentences, etc. This method is based on a study of the predictability of English; how well can the next letter of a text be predicted when the preceding $N$ letters are known. The results of some experiments in prediction will be given, and a theoretical analysis of some of the properties of ideal prediction. By combining the experimental and theoretical results it is possible to estimate upper and lower bounds for the entropy and redundancy. From this analysis it appears that, in ordinary literary [illegible] statistical effects (up to 100 letters) reduce the

~8 bps

Shannon, C. E. (1951). Prediction and entropy of printed English. *The Bell System Technical Journal*, 50–64.




~39 bps

Coupé, C., Oh, Y., Dediu, D., & Pellegrino, F. (2019). Different languages, similar encoding efficiency: Comparable information rates across the human communicative niche. *Science Advances*, 5(9).



16

# The Information Rate of Spoken Language

- Of course, spoken language is *not* a fixed code with a constant information rate
- What people say is conditioned on critical *causal* factors such as ...
  - their situated and embodied circumstances: '*pragmatics*'
  - the temporal evolution of events: '*synchronics*'
  - the level of effort that they are prepared to devote to their behaviour: '*energetics*'
- In other words, the information rate in spoken language varies as a function of the pragmatic, synchronic and energetic context

Moore, R. K.. (2023). Pragmatics, synchronics and energetics in spoken language – an information theoretic perspective. In *Workshop on Limits and Benefits of Information-Theoretic Perspectives in Spoken Communication* (LBIT), Dublin.

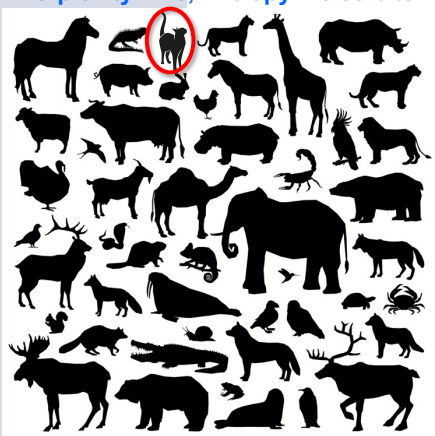


17



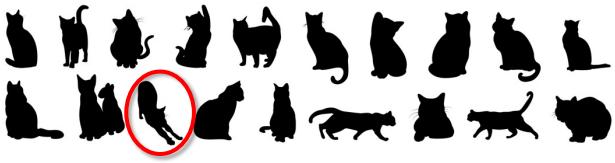
# Pragmatics: *situated/embodied context*

**Perplexity = 48, Entropy = 5.58 bits**



"*Ooh, look at that cat*"

**Perplexity = 20, Entropy = 4.32 bits**



"*Ooh, look at that cat … the one stretching*"

**Perplexity = 2**
**Entropy = 1 bit**

"*Ooh, look at that cat … the one on the left*"

**Perplexity = 1**
**Entropy = 0 bits**

"*Ooh, look at that*"

Silhouette images designed by brgfx / macrovector / Freepik



18

# Pragmatics → Energetics

**Entropy = 1 bit**

Ambiguity
*(Shannon's 'noisy channel' metaphor)*

**Entropy = 1 bit**

"*Ooh, look at that cat … the one in front and slightly to the left*"

**Transmitted Information = 83 bits**

"*Ooh, look at that … I think it's a cat, or maybe two cats with one sitting in front of the other as I can see what looks like three ears with one head lower than the other and slightly to the left?*"

**Transmitted Information = 450 bits**

Zipf, G. K. (1949). *Human Behavior and the Principle of Least-Effort.* Cambridge: Addison-Wesley Press.

Silhouette images from Freepik



19



# Energetics: *motivational context*

**Transmitted Information ∝ Effort**

"*Why do I have to shout?*"

"*I … DO … NOT … KNOW!*"
" *I do not know*" **47 bits**
*"I don't know"*
*"I dunno"*
*"dunno"*
[ə̃ə̃ə̃] **3 bits**

Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics*, 31, 373-405.

"*Huh? Speak more clearly!*"

Silhouette images from Freepik



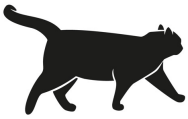
20

# Energetics → Synchronics

Lindblom, B. (1990). Explaining phonetic variation: a sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modelling* (pp. 403–439). Kluwer Academic Publishers.

"*Ooh, look at that cap*"

"*Ooh, look at that cart*"



"*Ooh, look at that rat*"

Localised Contrastive Effort *(in time)*

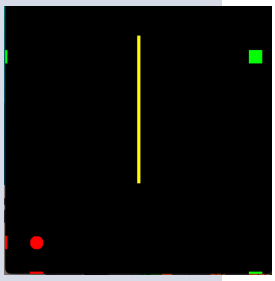
Silhouette images from Freepik



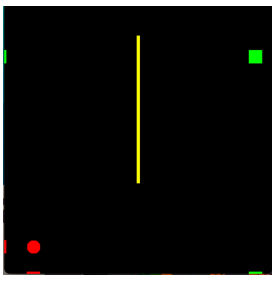
21



# Pragmatics → Synchronics

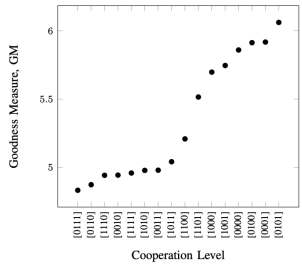


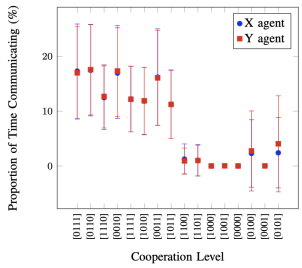
trapped by a local minimum



overcoming trap by communicating

Moore, R. K. (2023). Local minima drive communications in cooperative interaction. *Proceeding of the AISB* Convention, Swansea.



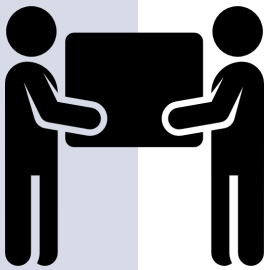






22

# Synchronics: *temporal context*

"*To me*" … "*To you*"



"*Look at that …*"

"*Three, two, one … GO!*"

Mutual Information
*(multi-modal)*

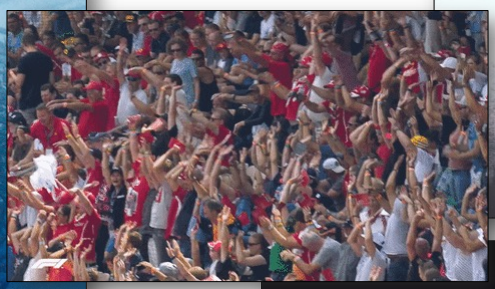
Silhouette images designed by kjpargeter / Freepik

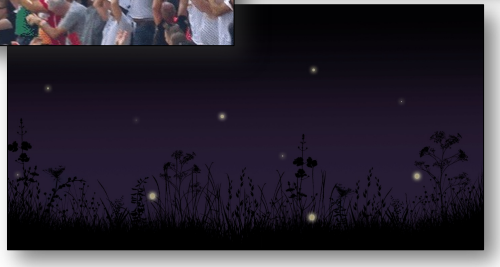


23



# Synchronics: *temporal context*



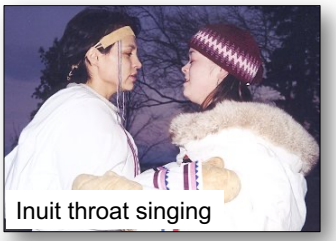


Strogatz, S. H. (2012). *Sync: How Order Emerges from Chaos In the Universe, Nature, and Daily Life*. Hatchette Book Group.



24

# Synchronics: *temporal context*



alignment → accommodation → entrainment

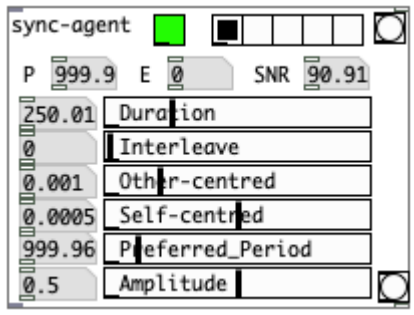
Cummins, F. (2014). Voice, (inter-)subjectivity, and real time recurrent interaction. *Frontiers in Psychology*, 5, 760.



25

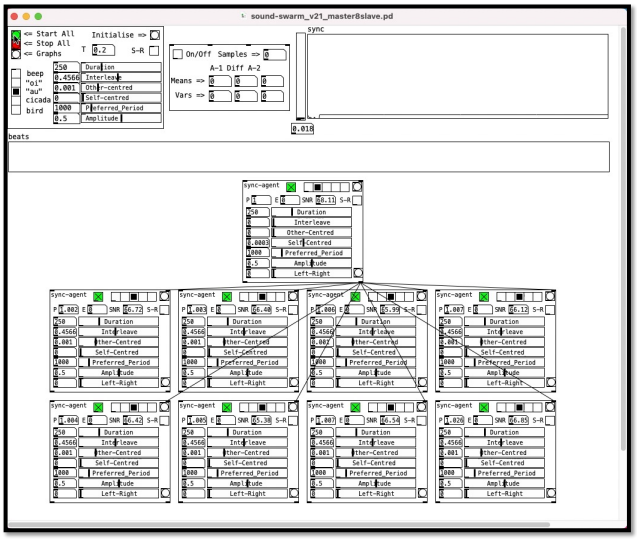


# Synchronics: *temporal context*




Vocalising/Listening 'Agent'

Moore, R. K. (2023). A control systems perspective on entrainment. *RITMO Entrainment Workshop*, Oslo.

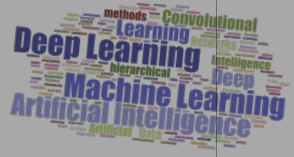




26

# The 'State-of-the-Art'

- There is steady year-on-year progress
- Improvements come from:
  - increase in available computer power
  - latest machine-learning paradigms
  - huge real-world training corpora
  - open tools / benchmark testing / challenges
- 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*)
  - inefficient (*trained on more data than a person hears in a lifetime*)
  - biased (*trained on un-curated data*)
  - untrustworthy (*e.g. hallucinates misinformation*)
  - ecologically damaging (*model training has a significant carbon footprint*)
- It is easy to underestimate the richness and complexity of spoken language interaction
- It is not 'natural' to talk to a machine!
- The availability of open tools and data is de-skilling

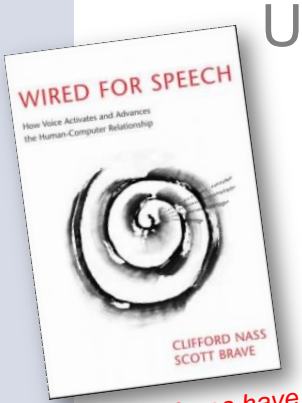





27



# Usability Issues




*"Voice interfaces have become notorious for fostering frustration and failure."*

noise
accents
understanding
privacy
embarrassment
alternative GUIs
task familiarity
limited functionality

*How many 'real' users?*


Moore, R. K., Li, H., & Liao, S.-H. (2016). Progress and prospects for spoken language technology: what ordinary people think. *INTERSPEECH* (pp. 3007–3011). San Francisco, CA.



28

# Alexa is not a Robot!

(*and a robot may not be a personal assistant*)


*"A robot is an actuated mechanism programmable in two or more axes with a degree of autonomy, moving within its environment, to perform intended tasks."*

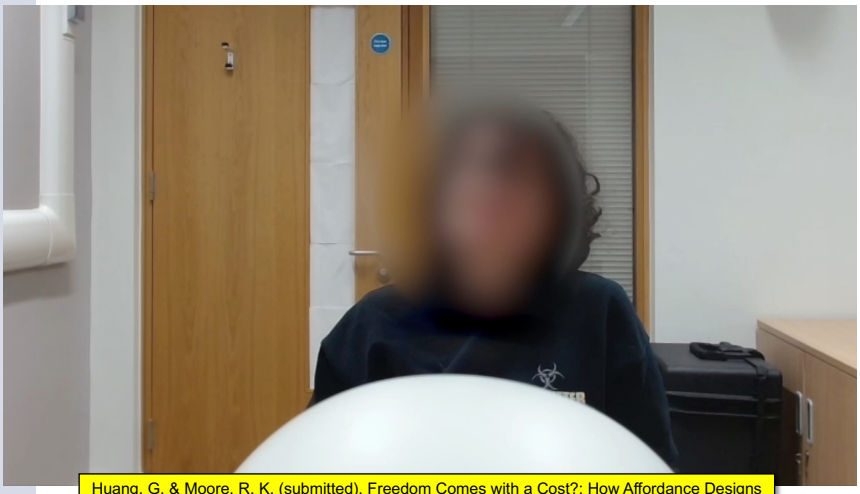
http://www.leorobotics.nl/definition-robots-and-robotics



29



# Talking to a Robot

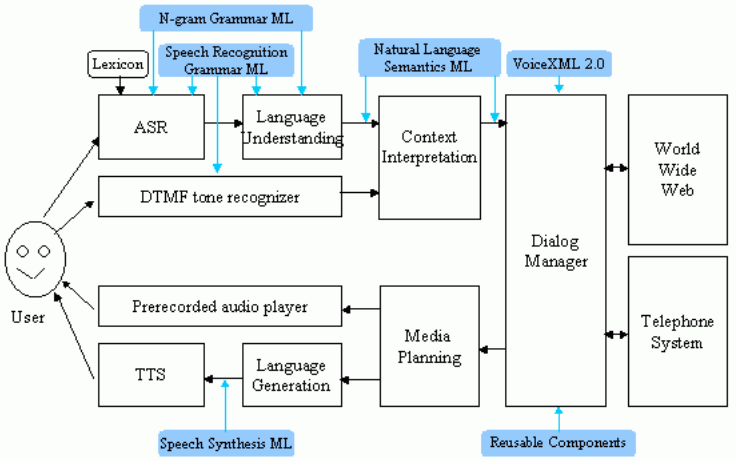


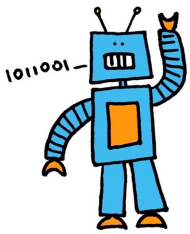
Huang, G. & Moore, R. K. (submitted). Freedom Comes with a Cost?: How Affordance Designs Affect Users' Experience with a Conversational Social Robot. *Frontiers in Robotics and AI*.



30

# ‘Standard’ Speech Interface



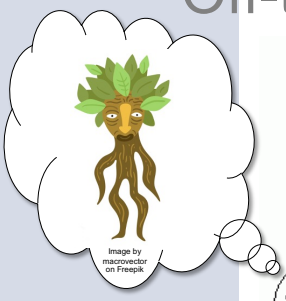

Introduction and Overview of W3C Speech Interface Framework, http://www.w3.org/TR/voice-intro/

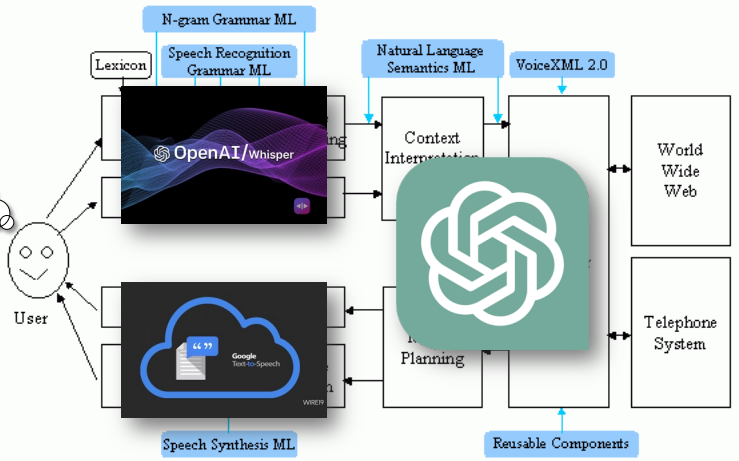


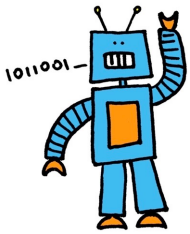
31



# ‘Off-the-Shelf’ Speech Components



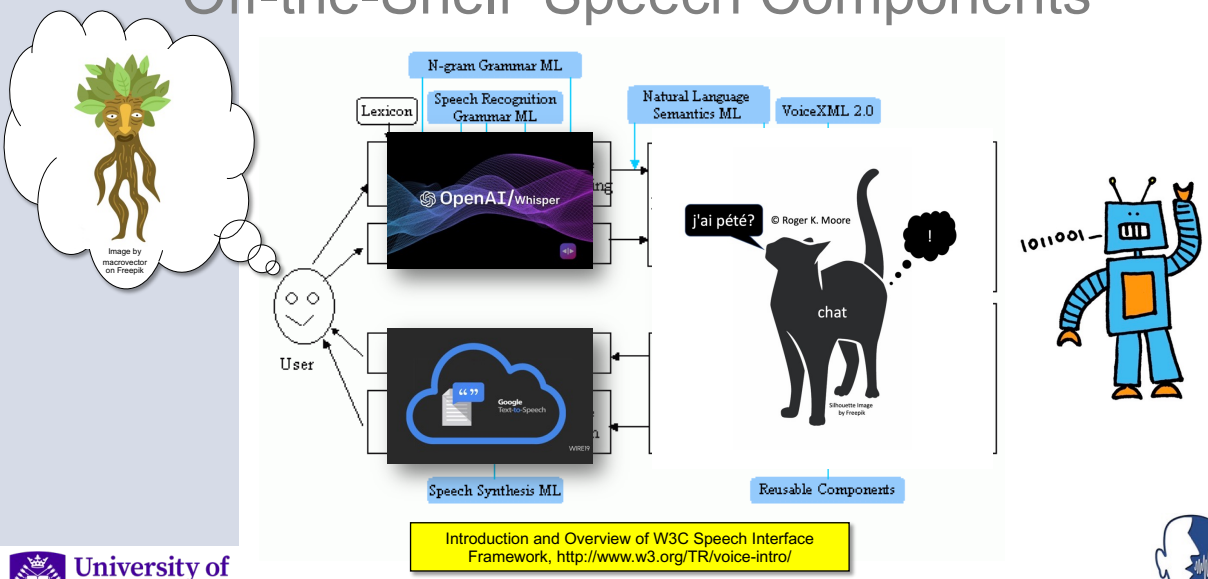




Introduction and Overview of W3C Speech Interface Framework, http://www.w3.org/TR/voice-intro/



32

# 'Off-the-Shelf' Speech Components



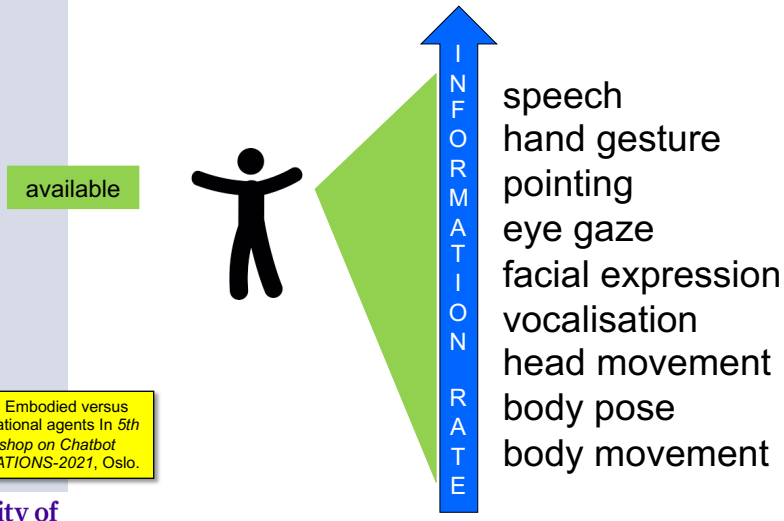

Introduction and Overview of W3C Speech Interface Framework, http://www.w3.org/TR/voice-intro/



33



# Multi-Modal Communication

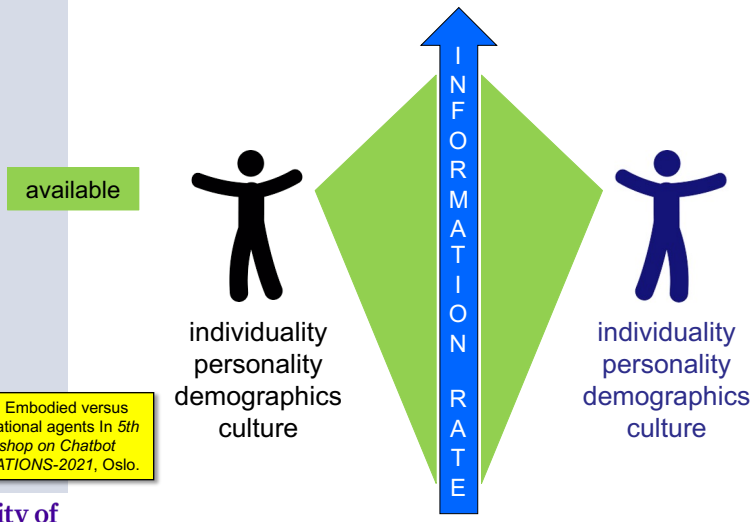



Moore, R. K. (2021). Embodied versus disembodied conversational agents In *5th International Workshop on Chatbot research: CONVERSATIONS-2021*, Oslo.



34

# Multi-Modal Human-Human Interaction



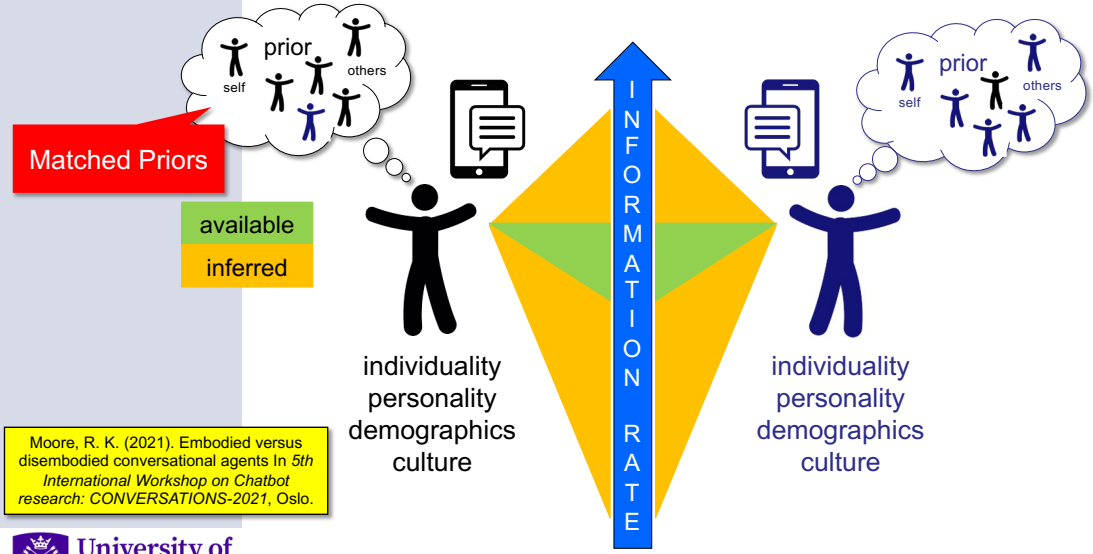

Moore, R. K. (2021). Embodied versus disembodied conversational agents In *5th International Workshop on Chatbot research: CONVERSATIONS-2021*, Oslo.



35



# Mono-Modal Human-Human Interaction

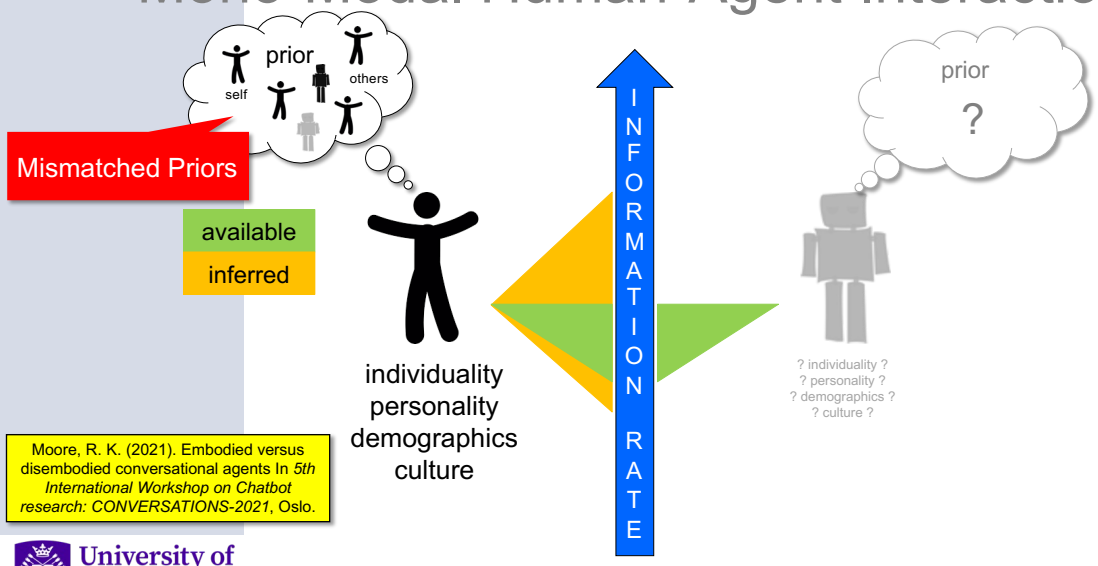



Moore, R. K. (2021). Embodied versus disembodied conversational agents In *5th International Workshop on Chatbot research: CONVERSATIONS-2021*, Oslo.



36

# Mono-Modal Human-Agent Interaction



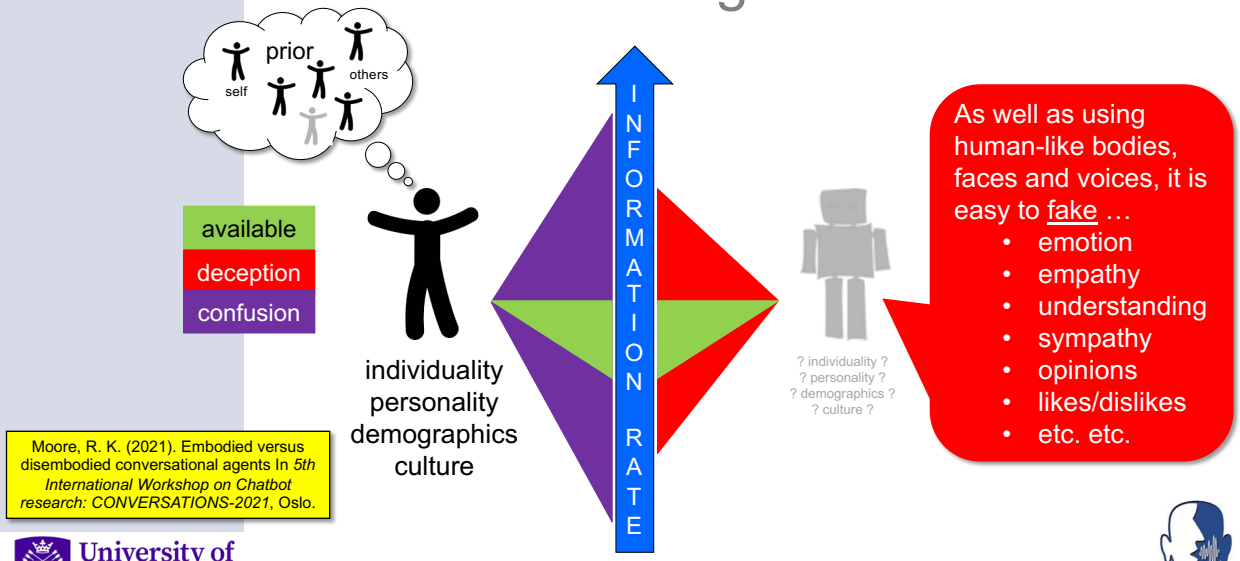

Moore, R. K. (2021). Embodied versus disembodied conversational agents In *5th International Workshop on Chatbot research: CONVERSATIONS-2021*, Oslo.



37



# Multi-Modal Human-Agent Interaction

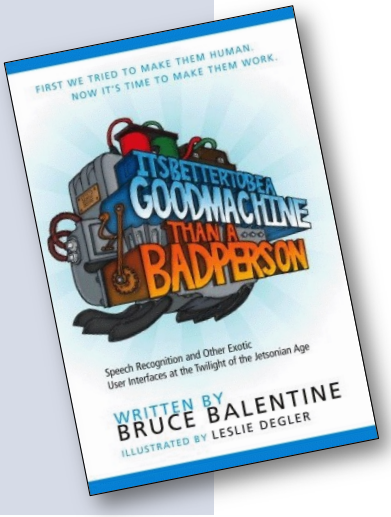



Moore, R. K. (2021). Embodied versus disembodied conversational agents In *5th International Workshop on Chatbot research: CONVERSATIONS-2021*, Oslo.



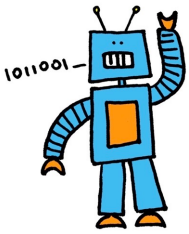
38

# Mismatched Interlocutors




- Spoken language interaction between human beings is founded on shared experiences, representations and priors (*i.e. mutual situatedness/embodiment*)
- So, is there a fundamental limit to the language-based interaction that can take place between mismatched interlocutors?

Moore, R. K. (2016). Is spoken language all-or-nothing? Implications for future speech-based human-machine interaction. In K. Jokinen & G. Wilcock (Eds.), *Dialogues with Social Robots – Enablements, Analyses, and Evaluation*. Springer Lecture Notes in Electrical Engineering (LNEE).

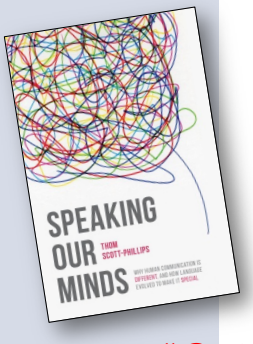


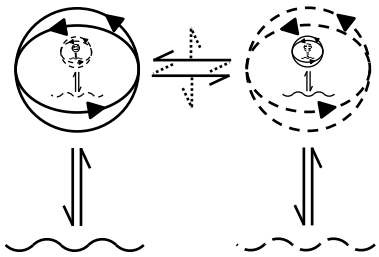



39



# Matched Interlocutors



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).

Human-Human Languaging

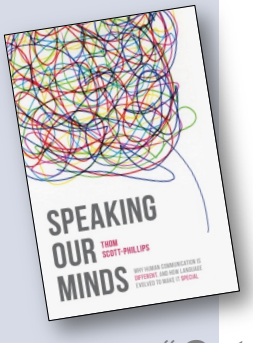
=

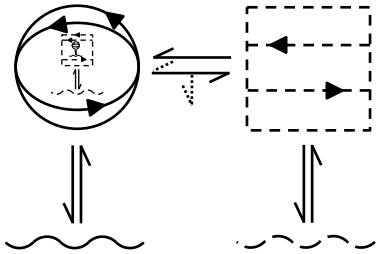
"*Ostensive Inferential Recursive Mind-Reading*"



40

# Mismatched Interlocutors



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).

Human-Robot Languaging

=

"~~Ostensive~~ *Inferential* ~~Recursive~~ ~~Mind-Reading~~"

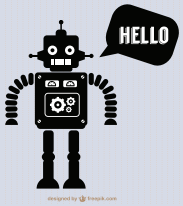
T. Scott-Phillips: *"The assumption of continuity between a fully coded communication system at one end, and language at the other, is simply not justified."*



41



# Talking to a Machine

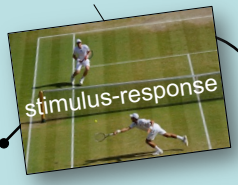



Philips, M. (2006). Applications of spoken language technology and systems. In M. Gilbert & H. Ney (Eds.), *IEEE/ACL Workshop on Spoken Language Technology* (SLT)

Usability

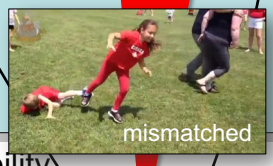
Like a Human


Add NL/Dialog




Structured Dialog

*'Habitability Gap'*

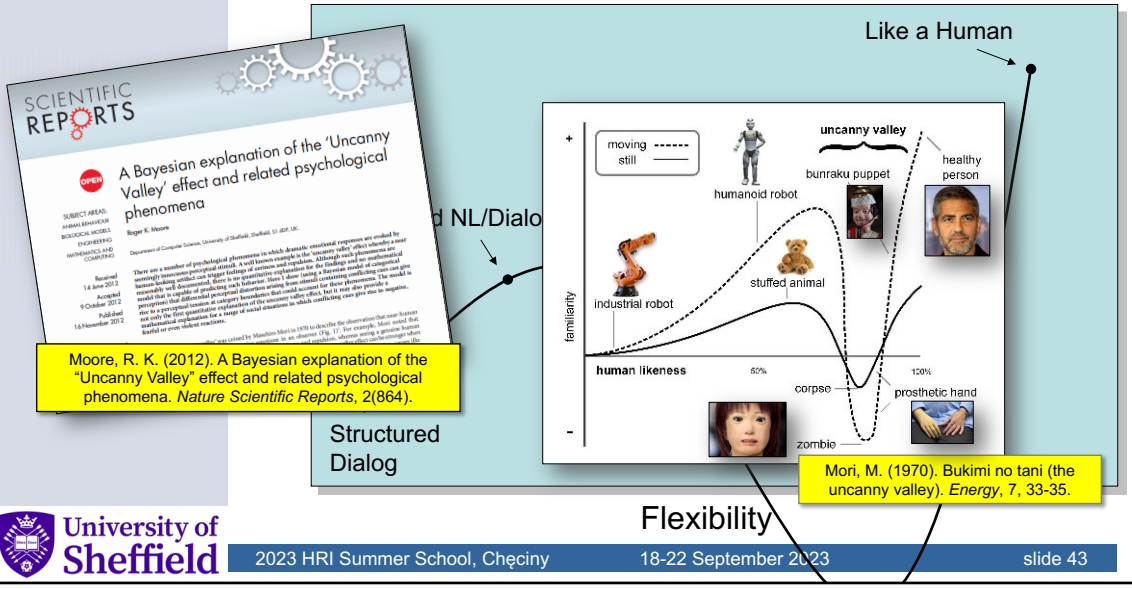



Flexibility



42

# Talking to a Machine




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

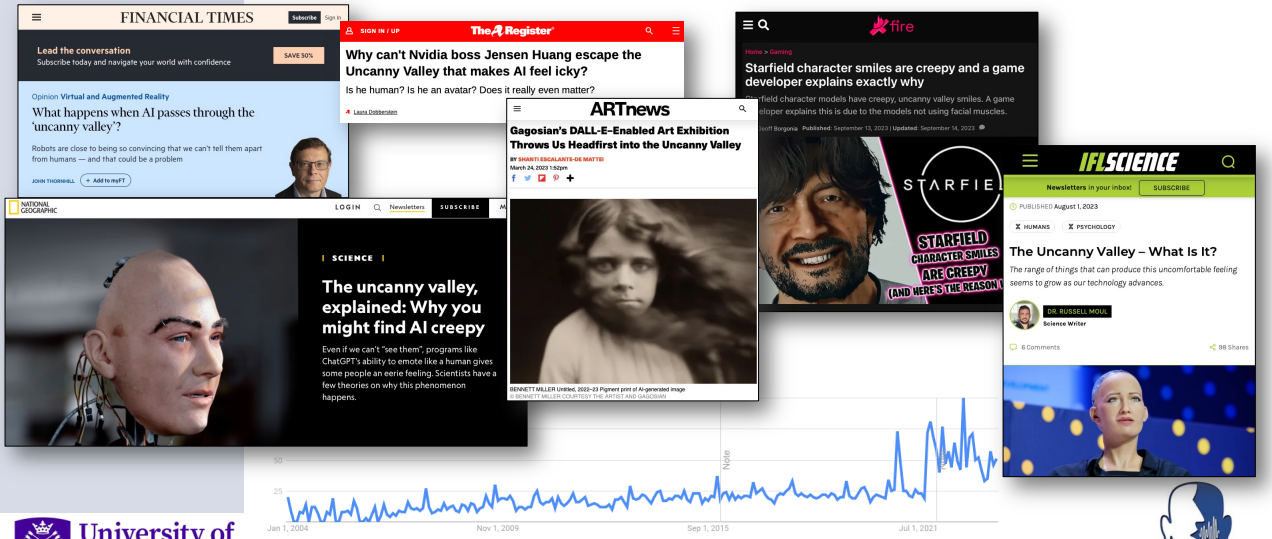
Mori, M. (1970). Bukimi no tani (the uncanny valley). *Energy*, 7, 33-35.



43



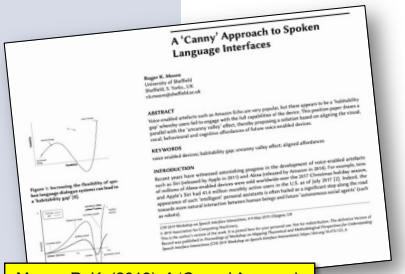
# The Uncanny Valley






44

# A 'Canny' Approach




Moore, R. K. (2019). A 'Canny' Approach to Spoken Language Interfaces. *CHI-19 Workshop on Mapping Theoretical and Methodological Perspectives for Understanding Speech Interface Interactions*, Glasgow.

- Minimise conflicting cues by aligning 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*)
  - communicative (*i.e. how it interacts*)
- This means that an agent's capabilities will be capped (*but in a meaningful way*)
- It also implies that a voice-enabled agent should have an 'appropriate' voice (*and a chatbot should use an appropriate font*!)

uncanniness ↓ ➤ affinity ↑ ➤ rapport ↑



45



# Deploying an 'Appropriate' Voice

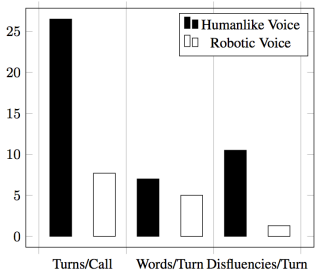
"Welcome to the route planning service - how can I help you?"

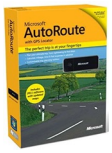
WoZ

Voice Changer

normal

*"Welcome to the route planning service - how can I help you?"*



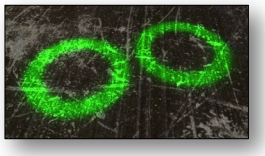




Moore, R. K., & Morris, A. (1992). Experiences collecting genuine spoken enquiries using WOZ techniques. *5th DARPA Workshop on Speech and Natural Language*, 61–63.



46

# ‘Appropriate’ Voices


Wilson, S., & Moore, R. K. (2017). Robot, alien and cartoon voices: implications for speech-enabled systems. *1st Int. Workshop on Vocal Interactivity in-and-between Humans, Animals and Robots* (VIHAR-2017), 40–44.



47

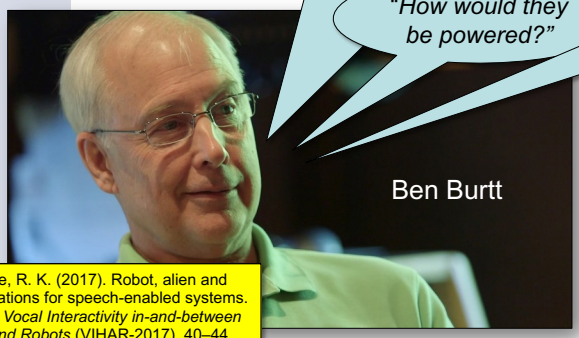


# ‘Appropriate’ Voices

*“I usually first think, if these objects, places, robots or machines really existed what would they sound like?”*

*“How would they be powered?”*

*“What would be the actual physics of how they work?”*



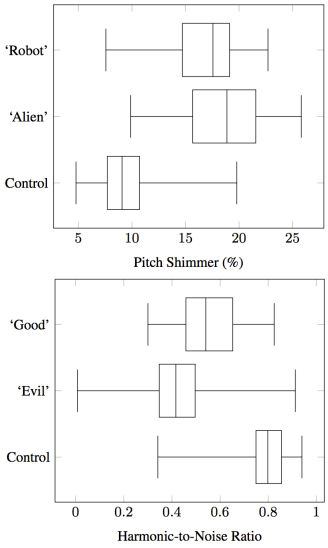

Wilson, S., & Moore, R. K. (2017). Robot, alien and cartoon voices: implications for speech-enabled systems. *1st Int. Workshop on Vocal Interactivity in-and-between Humans, Animals and Robots* (VIHAR-2017), 40–44.

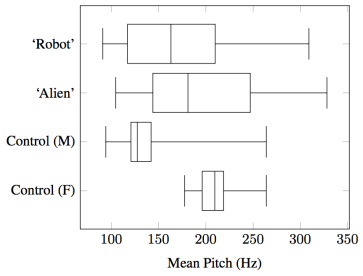


48

# Designer Voices

| Technique | Method |
|---|---|
| Time reversal | *delay line* |
| Speed change | *delay line* |
| Tremolo | *modulated amplitude* |
| Vibrato | *modulated pitch* |
| Ring modulation | *multiplication of two signals* |
| Comb filter | *short delayed version added to the original* |
| Echo | *long delayed version added to the original* |
| Flanger | *delay-modulated version added to the original* |
| Chorus | *multiple flangers with different delays* |
| Phaser | *phase-modulated version added to the original* |
| Reverberation | *convolution with room acoustic* |
| Pitch shift | *homomorphic filtering* |
| Harmony | *pitch-shifted version added to the original* |
| Filtering | *frequency shaping* |
| Formant shift | *altered vocal tract length* |
| Vocoding | *linear prediction analysis-synthesis* |



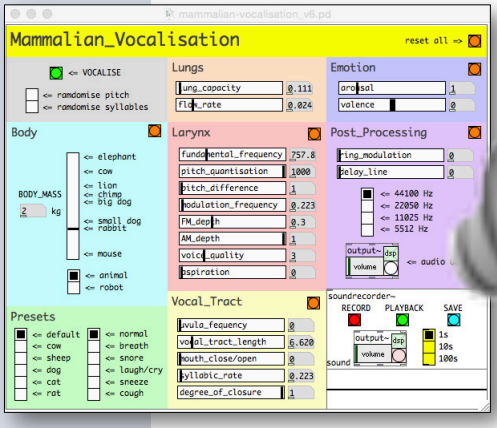

Wilson, S., & Moore, R. K. (2017). Robot, alien and cartoon voices: implications for speech-enabled systems. *1st Int. Workshop on Vocal Interactivity in-and-between Humans, Animals and Robots* (VIHAR-2017), 40–44.







# A Voice for the '*MiRo*' Robot



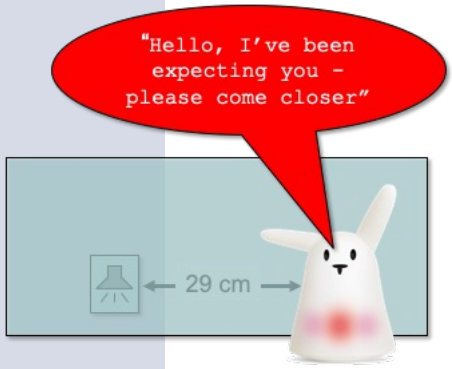

Body mass = 2 kg
Breathing rate = 0.7 Hz
Fundamental frequency = 760 Hz
Vocal tract length = 6.6 cm

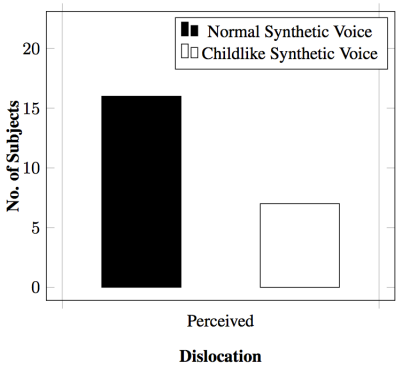
Moore, R. K. (2016). A real-time parametric general-purpose mammalian vocal synthesiser. In *INTERSPEECH* (pp. 2636–2640). San Francisco, CA.

Moore, R. K., & Mitchinson, B. (2017). A biomimetic vocalisation system for MiRo. In *Living Machines 2017*. Stanford, CA.

# Assessing Appropriateness

*(using the 'ventriloquist effect')*



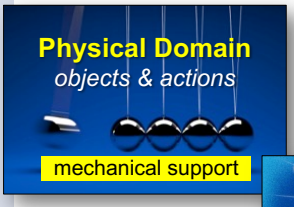

Moore, R. K., & Maier, V. (2012). Visual, vocal and behavioural affordances: some effects of consistency. *5th International Conference on Cognitive Systems - CogSys 2012*, 76.



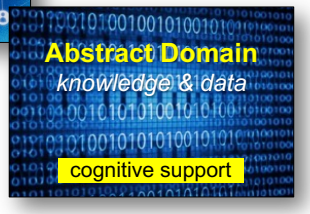
51



# Why Use Spoken Language?








*hands-free manipulation*

*eyes-free observation*

*collaborative coordination*

*the management of joint attention*

*multi-modal interaction*

*the expression of empathy*

*the expression of companionship*

*verbal + non-verbal signalling*

*the expression of emotion*

*sense making*

*access to information*

*problem solving*

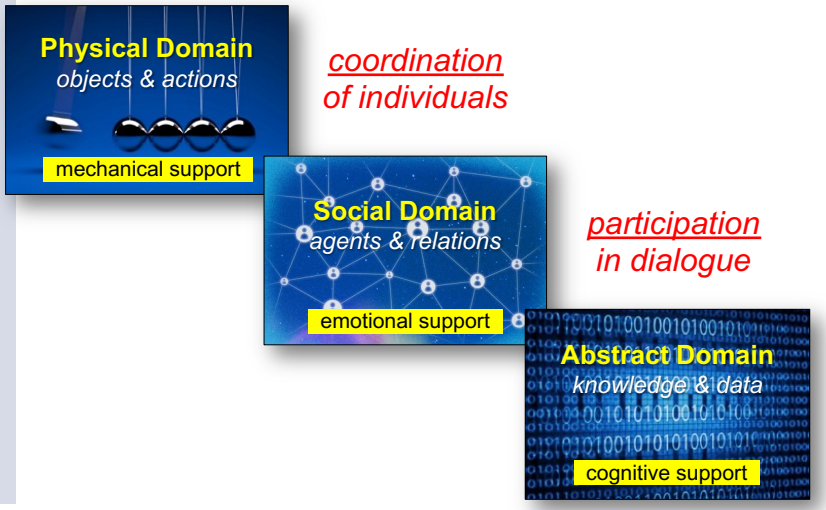
*high-rate communications*

Moore, R. K. (2019). Talking with Robots: Opportunities and Challenges. *International Conference Language Technologies for All (LT4All)*.



52

# When to Use Spoken Language?



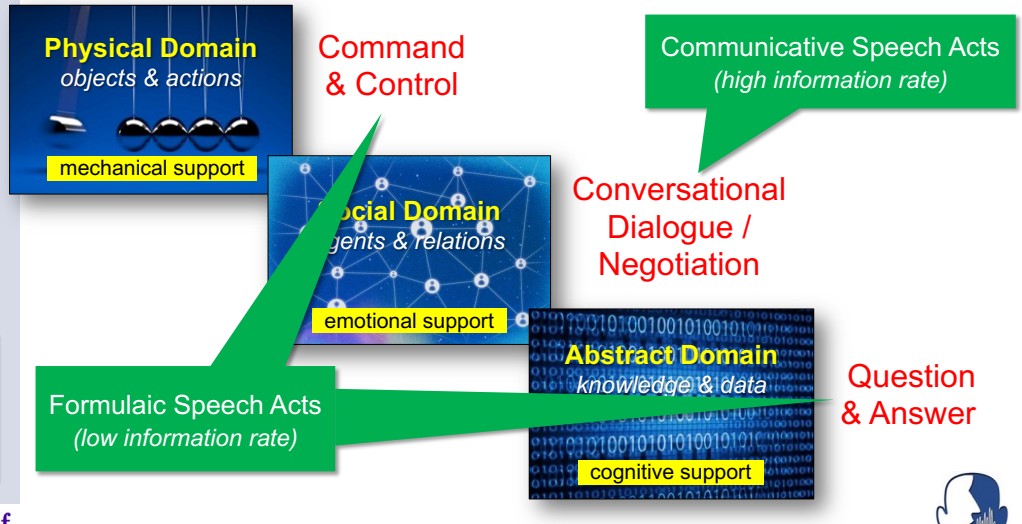

Moore, R. K. (2019). Talking with Robots: Opportunities and Challenges. *International Conference Language Technologies for All (LT4All).*



53



# How to Use Spoken Language?




Moore, R. K. (2019). Talking with Robots: Opportunities and Challenges. *International Conference Language Technologies for All (LT4All).*



54

# The Way Forward?


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.

To progress from talking and listening devices to intelligent communicative machines, we need to move beyond …

- speech
- words
- meaning
- communication
- dialogue
- one-off interactions



55



# The Way Forward?


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.

To progress from talking and listening devices to intelligent communicative machines, we need to move beyond …

- speech
- words
- meaning
- communication
- dialogue
- one-off interactions

Spoken language is multimodal, distributed, and optimised to the physical and temporal context.



56

# The Way Forward?


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.

To progress from talking and listening devices to intelligent communicative machines, we need to move beyond …

- speech
- words
- meaning
- communicati
- dialogue
- one-off interactions

Agents need to know who is talking, where they are, what they have said, why they have said it, what to say back, when to say it and how to say it.



57



# The Way Forward?


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.

To progress from talking and listening devices to intelligent communicative machines, we need to move beyond …

- speech
- words
- meaning
- communication
- dialogue
- one-off interactio

The behaviour of a living system is determined by personality, drives, needs, beliefs, desires, intentions and emotions. Hence an 'intelligent' machine needs to be able to interpret and express the consequent paralinguistic phenomena.



58

# The Way Forward?


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.

To progress from talking and listening devices to intelligent communicative machines, we need to move beyond …

- speech
- words
- meaning
- communication
- dialogue
- one-off interactio

Interaction between social agents is more than exchanging messages using some shared code. Behaviours are crafted to support continuous, coordinated joint action: i.e. continuous adaptive coupling between talking-listeners and listening-talkers.





# The Way Forward?


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.

To progress from talking and listening devices to intelligent communicative machines, we need to move beyond …

- speech
- words
- meaning
- communicatio
- dialogue
- one-off interactions

Interaction is grounded in social relationships, e.g. status and trust. These act as priors and influence the way people interact. An autonomous social agent is not a neutral partner.

# The Way Forward?


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.

To progress from talking and listening devices to intelligent communicative machines, we need to move beyond …

- speech
- words
- meaning
- communication
- dialogue
- one-off interactions

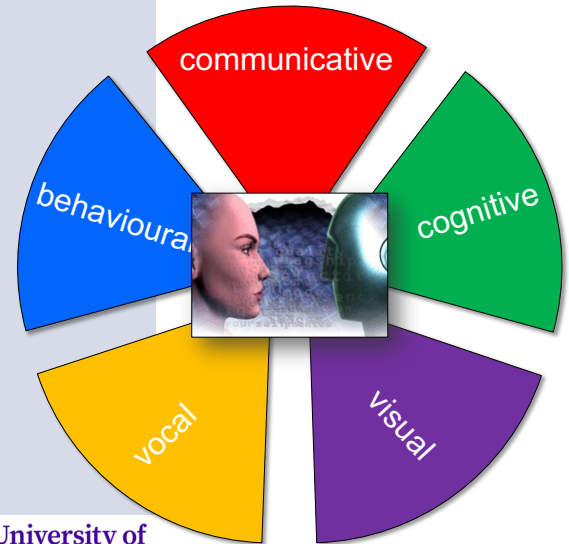
Users retain person/context-specific memories of previous interactions, and are able to draw on this in order to converse efficiently.



61



# Bringing It All Together




Need to take a holistic approach to …

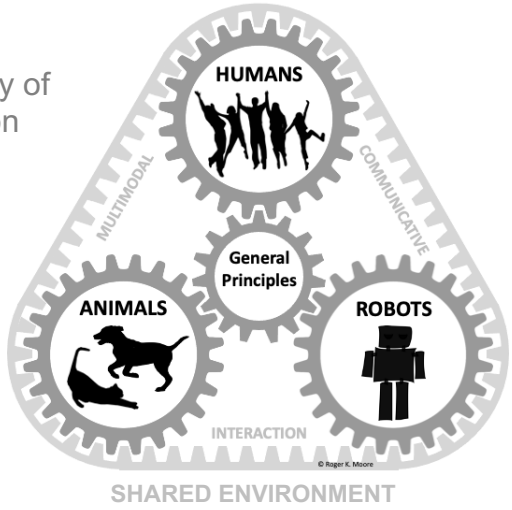
- designing and implementing interactional affordances
- accommodating user priors



62

# Bringing It All Together

Need an overarching theory of communicative interaction



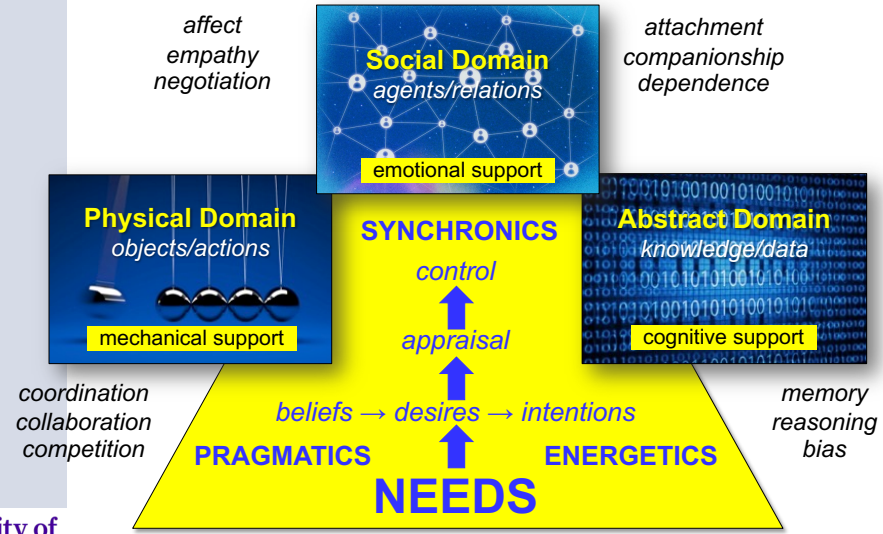

Moore, R. K. (2021). Où les êtres Sociaux: vers une théorie de l'interaction communicative. *3e Colloque International - Objets Animés, Humains, Animaux : Partenaires de Soins Tendres.*



63



# Bringing It All Together





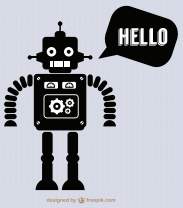

64

# Bringing It All Together






# Summary



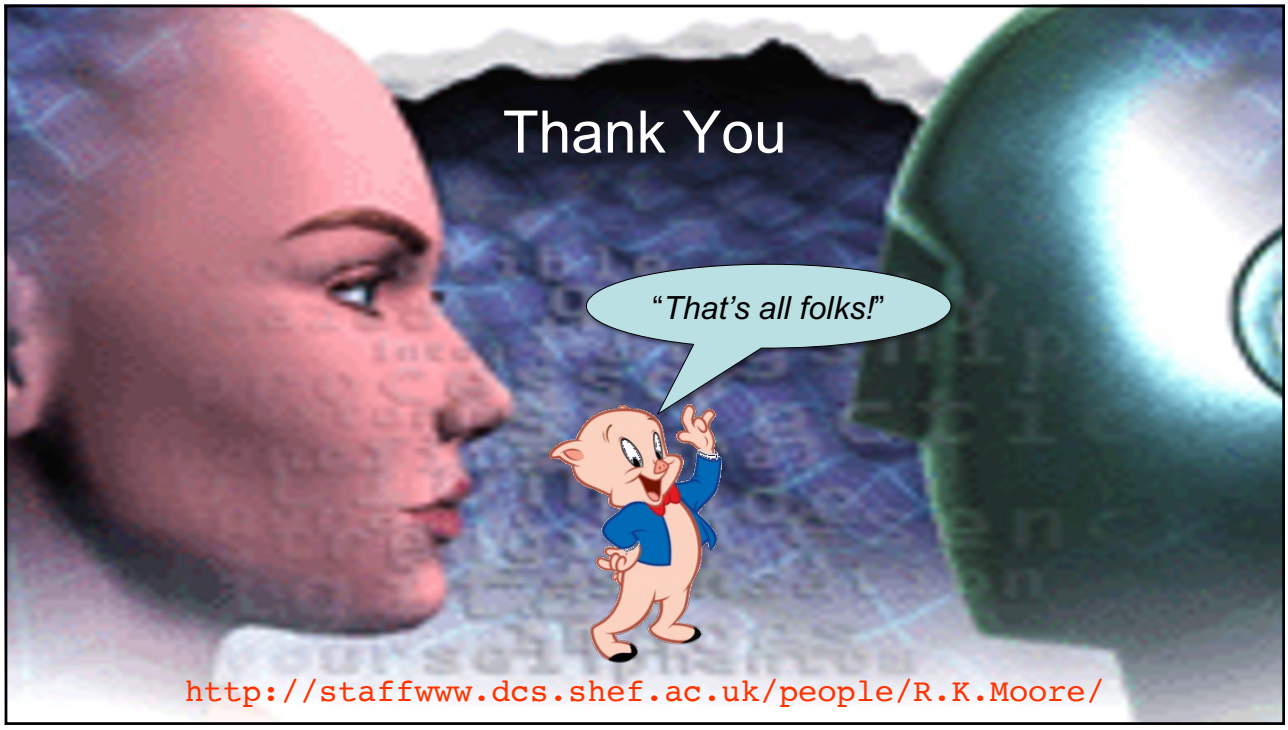

- Next-generation language-based interactive systems need to have a much deeper understanding of user behaviour
- Embodied systems need to facilitate continuous adaptive multimodal incremental closed-loop coupling between users and the physical, social and abstract worlds
- The visual, vocal and behavioural affordances of speech-enabled devices need to reflect their actual abilities
- We need to understand how language might function between mismatched agents
- We need to be able to measure (*and thus optimise and predict*) 'usability' for language-enabled systems
- We must guard against inappropriate use-cases (*e.g. infantalising users*)

# Where to Find Out More



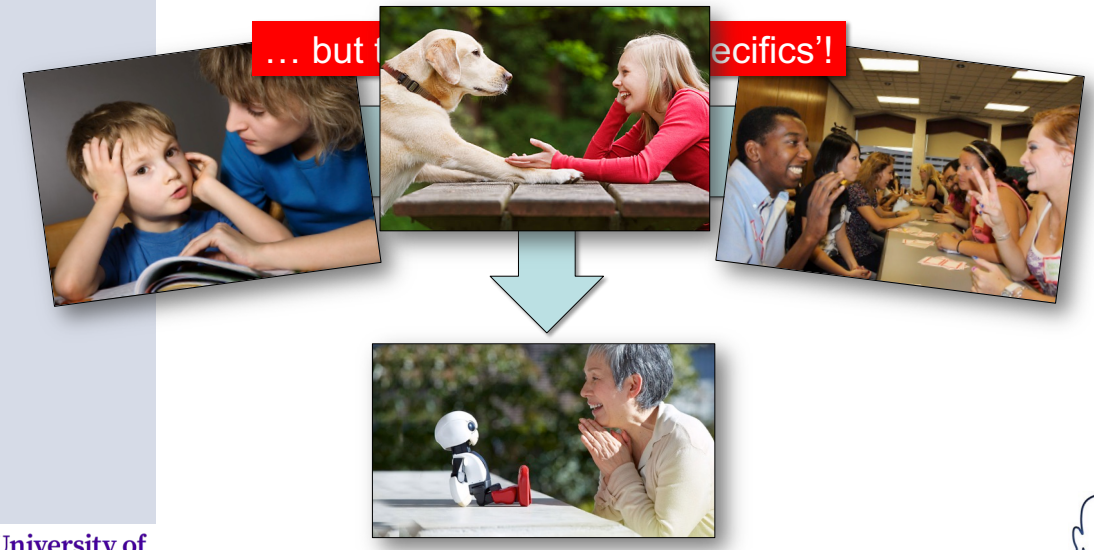

67

# Thank You




68

# Mismatched Interlocutors






69



# Abstract


- It is often taken for granted that speech provides a 'natural' means by which users can interact with an autonomous agent such as robot - especially if it is humanoid in design.
- Indeed, the ready availability of off-the-shelf spoken language tools such as Google's speech-to-text/text-to-speech or OpenAI's Whisper/ChatGPT makes it relatively easy to implement a ready-made solution.
- However, not only does such an approach risk creating a chimera - a potentially confusing and inappropriate compilation of misaligned visual, vocal and behavioural affordances - but it also facilitates a casual misrepresentation/faking of agency (e.g. by having the robot express apparently personal likes, dislikes and opinions using the pronoun "I").
- This talk will address these issues, and provide an insight into why, when and how users might wish to converse with a robot.
- It will also give an appreciation of the necessary design considerations appropriate to any given use-case.
- The talk will conclude with a discussion of the potential limits for language-based interaction between 'mismatched' agents such as humans and robots.

- 90 mins = ~50-60 slides




70