Machine Processing of Dialogue States; Speculations on ‘Conversational Entropy’

Nick Campbell

nick@tcd.ie

Speech Communication Lab
ADAPT Centre,
Trinity College Dublin
Ireland

Printed in Great Britain at the University Press, Cambridge
(Brooke Crutchley, University Printer)
and published by the Cambridge University Press
(Cambridge, and Bentley House, London)
Agents for Canada and India: Macmillan

WHAT IS LIFE?

The Physical Aspect of the Living Cell

BY

ERWIN SCHRÖDINGER

SENIOR PROFESSOR AT THE DUBLIN INSTITUTE FOR ADVANCED STUDIES

Based on Lectures delivered under the auspices of the Institute at Trinity College, Dublin, in February 1943

CAMBRIDGE
AT THE UNIVERSITY PRESS
1948

life is communication!

second inspiration

norbert wiener

founder of cybernetics

machine communication

THE HUMAN USE OF HUMAN BEINGS

CYBERNETICS
AND SOCIETY

NORBERT WIENER

With a new Introduction by
Steve J. Heims

FAB

'an association in which the free development of each is the condition of the free development of all'

FREE ASSOCIATION BOOKS / LONDON / 1989

“Overview of speech technology results, challenges, trends, promising directions in Social Interactions and Signal Processing”

- this talk presents some thoughts from the Speech Communication Lab in Dublin University as the basis for speculation about such fundamental processes . . .

- The concept of *entropy* was introduced at the beginning of the previous century and has been well-understood by physicists, chemists, and information engineers, among others, but it has failed to take hold in the humanities . . .

a conversation is a living organism
- entropy kills conversation -
laughter reduces entropy by resetting the topic
and so helps keep a conversation alive

- **Erwin Schrodinger** was at TCD when he gave his lectures on “What is life? The Physical Aspect of the Living Cell”

- “Every process, event, happening - call it what you will; in a word, everything that is going on in Nature means an increase of the entropy of the part of the world where it is going on. Thus a living organism continually increases its entropy - or, as you may say, produces positive entropy - and thus tends to approach the dangerous state of maximum entropy, which is of death. It can only keep aloof from it, i.e. alive, by continually drawing from its environment negative entropy - which is something very positive as we shall immediately see. What an organism feeds upon is negative entropy. Or, to put it less paradoxically, the essential thing in metabolism is that the organism succeeds in freeing itself from all the entropy it cannot help producing while alive” **Schrodinger, 1943**

time for laughter

- laughter reduces entropy!
- it is good for you!
- we do it a lot!
- but talking machines don’t know how to laugh . . .

human-machine communication

human-human machine-mediated communication

human-information machine-mediated communication

machine-human communication

speech synthesis

- 1980: Reading Machines . . . *(TTS)*
- 1990: CHATR - concatenative speech synthesis
- 2000: JST/ESP - conversational speech data
- 2010: Herme - conversational devices in public
- 2020: autonomous interactive dialogue systems

laughing robots?

- will a computer/machine/dialogue-system need to laugh?
  - in speech-translation - (at least) - yes . . .
- will a sentient agent need to cry?
  - (I personally doubt it, but . . .)

TTS → ISS:

from ‘speech synthesis’
to ‘*interactive* speech synthesis’

- speech synthesisers currently have no ears/eyes!
- humans couldn’t interact properly without feedback
- talking machines need sensors

- so they can do *social interaction*, not just talking . . .

progression of scientific thought at the SCL

- predominance of nonverbal speech *(from ESP)*
- social interaction vs transfer of propositional content *(d64)*
- role of ‘chat’ in social interaction *(herme)*
- timing & laughter in human-robot conversation *(joker)*
- sensing of participant engagement *(d-ans)*
- monitoring meta-cognition in dialogue *(metalogue)*
- conversational robot-human interaction *(enterface/hmmm)*

‘engagement’ in dialogue

- essential for *‘interactive speech synthesis’*

- a) to know if one has been understood
- b) to sense the listener’s interest
- c) to determine when & what to speak next

- engineering to *“get the message across”*

talking machines

- we have come a long way since 1990!
- siri/cortina/google/++ (they want to own you! *)
- from talking toasters to the kitchen fridge!
    - https://www.youtube.com/watch?v=LRq_SAuQDec
      https://www.youtube.com/watch?v=-B47lut_sVE
    - (* cf recent paper/talks by Steve Young, Cambridge)

chatr (cf LREC'16)

(c) ATR ITL (recovered from 1997)

CHATR Speech Synthesis

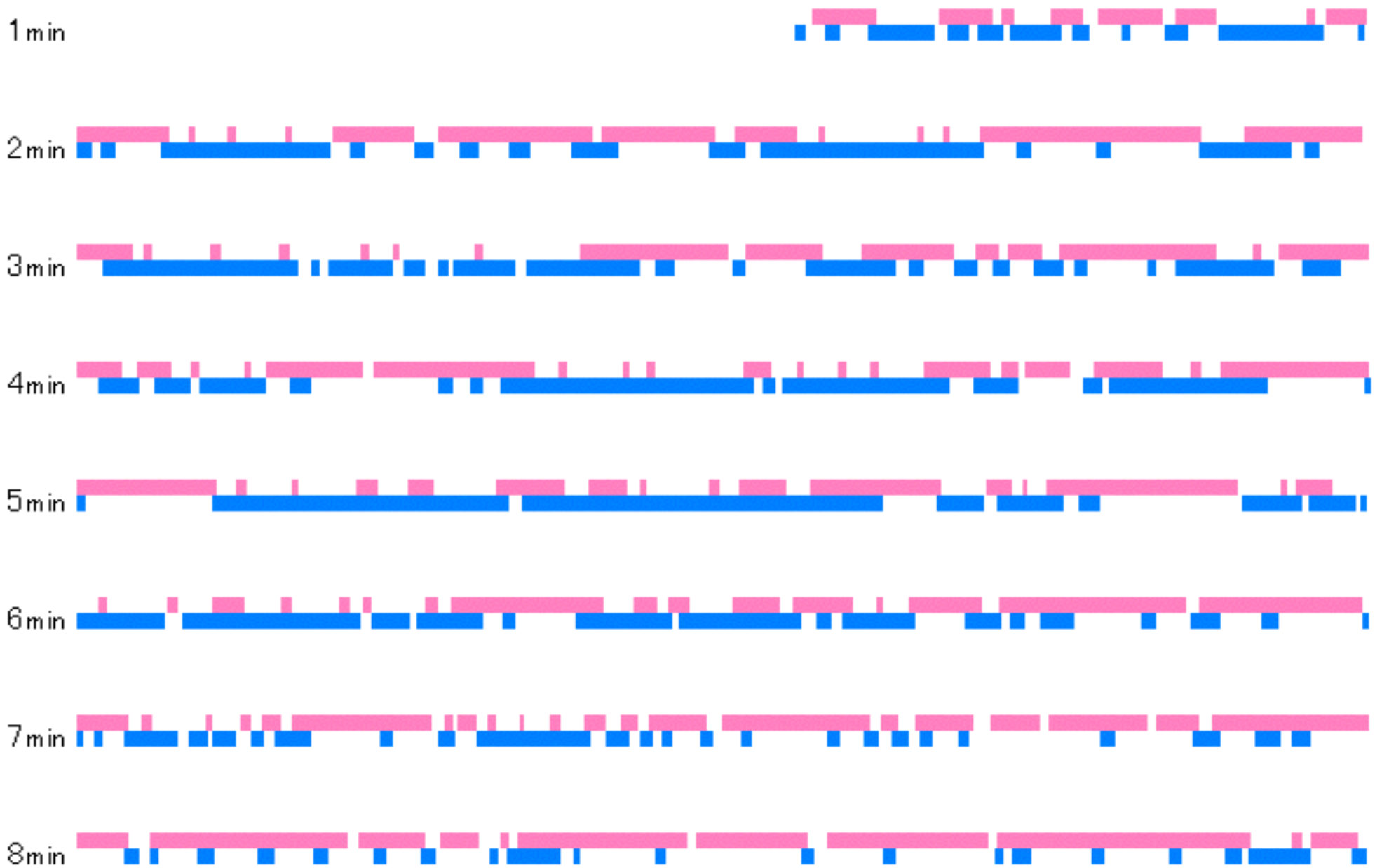
音声合成自動作成

take the original tour

20TH ANNIVERSARY EDITION

www/speech-data.jp/chatr/

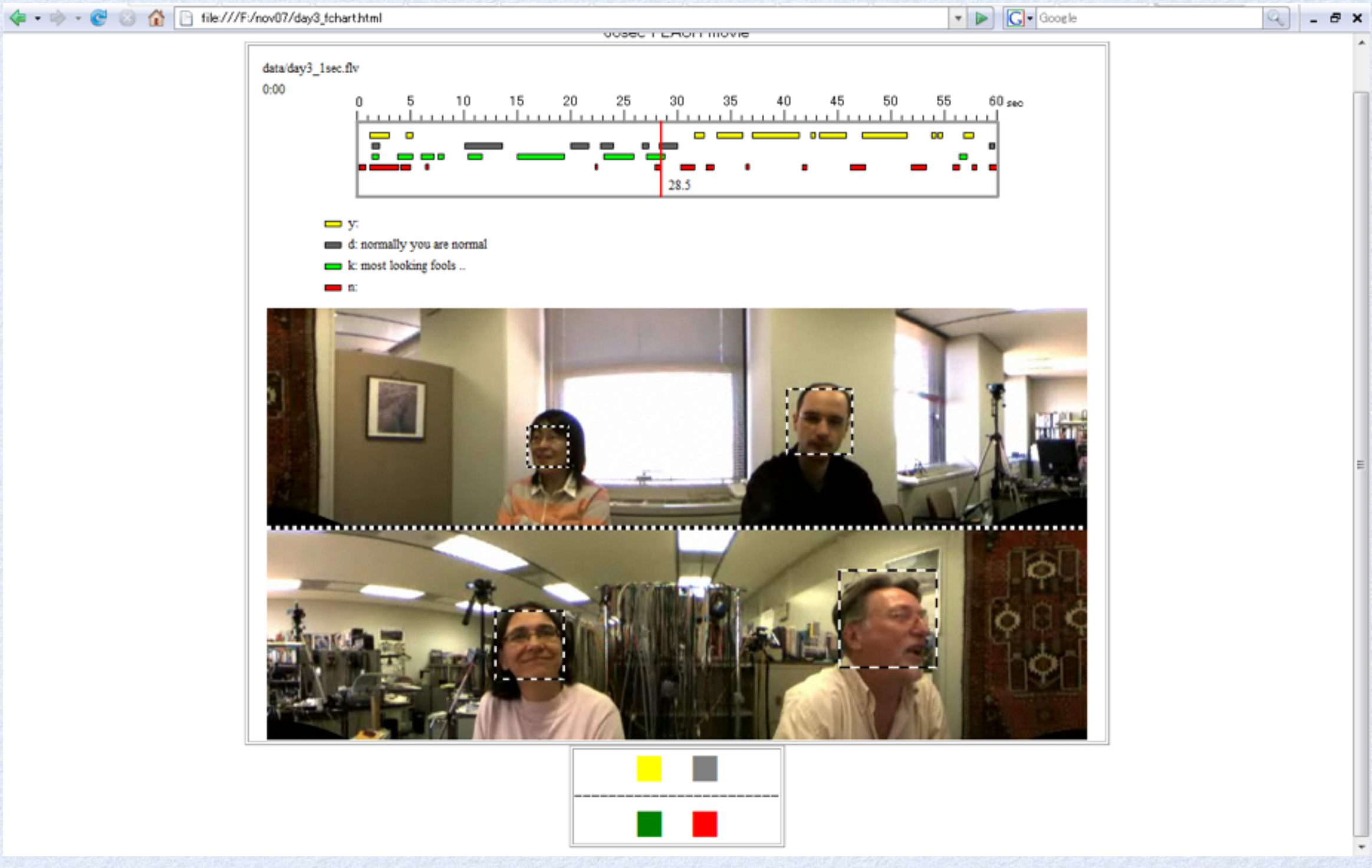
jst/esp short vs long utterances



(these data from telephone speech)

OUR MOST IMPORTANT DEVICES
(FOR SPEECH PROCESSING !)

PATTERNS OF SPEECH ACTIVITY



the d64 : mega-multi-modal

d-ans

- The D-ANS Corpus: the Dublin-Autonomous Nervous System Corpus of Biosignal and Multimodal Recordings of Conversational Speech

d-ans

- The D-ANS Corpus: the Dublin-Autonomous Nervous System Corpus of Biosignal and Multimodal Recordings of Conversational Speech

Herme - a talking machine

- Herme ‘chatted people up’ when they visited the Science Gallery in Dublin

- She collected natural human-machine conversation data over a period of three months

- DO NOT LOOK TOO CLOSELY AT THIS PHOTOGRAPH!

photo May 2011
with permission

listening machines

and so

on to entropy

conversational entropy

- Francesca Bonin's PhD (TCD 2015) :
  - the relation between social signals and discourse phenomena such as topic changes
  - immediately after a topic change there is a significant drop in social activity,
    - tentative hypothesis: "The interactional entropy of a segment x is defined as the number of occurrences of social signals in x"

laughing later in the topic

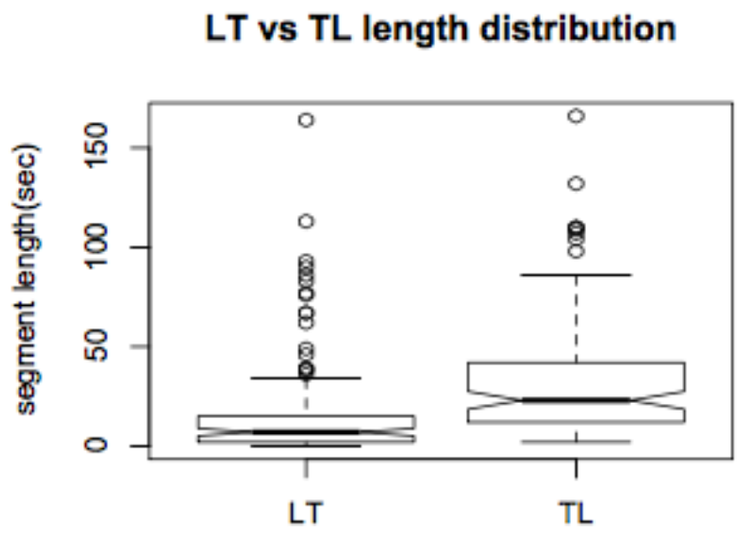


Figure 1: Topic boundary neighbourhood (left) and LT-TL comparison (right)

Figure 2: Inter/intra topic segmentation

** from Francesca Bonin, Nick Campbell, Carl Vogel, "Temporal distribution of laughter in conversation" in Proceedings of the Third Interdisciplinary Workshop on Laughter and other Non-Verbal Vocalisation in Speech*

topic changes & laughter

- She clearly showed that after a topic change a decrease of interactional entropy occurs, and concludes that this information might be used to better understand the discourse structure via non-linguistic information such as laughter, overlaps, backchannels, and silence, and thereby shed new light upon the discourse functionality of social signals.

- simply put: *maybe machines don't need to listen!*

non-verbal interaction processing

- the world is a very noisy place - ASR is not perfect
- syntactic/semantic processing is not yet mature - machine understanding is not perfect
- people don't often say exactly what they mean - the language itself is not perfect
  - and yet people cope!
    - non-verbal processing helps . . .

practical application

- by observing the amount of non-verbal behaviour in speech (particularly laughter) we can estimate the likelihood of a forthcoming topic change

- the system can be aware of its environment through sensing movement and dynamics of vocal activity.

- *It doesn't need to listen to the speech content per se.*

- we can thereby enable our delivery device to interrupt a conversation at a timely point without being aware of the linguistic content of any conversations

where does entropy come in?

- machines must know when to speak
- when to deliver their content
- how to parse the response
- but they shouldn’t be “always listening”!

back to the (iot) fridge

- it’s located in the social centre of the house
- it’s always connected and never turned off
- it has a fixed environment and can learn:
  - who is where
  - what is happening
  - what needs to be said
  - when to say it - timing is essential!

delivering content

- situation update:
  - Granny's in the bath & the water has gone cold . .
- system response:
  - tell someone! (who/when?)
  - but first sense the context (is it timely to speak?)
    - hence: *Interactive Speech Synthesis*

latest update - earlier this month:

eNTERFACE'16

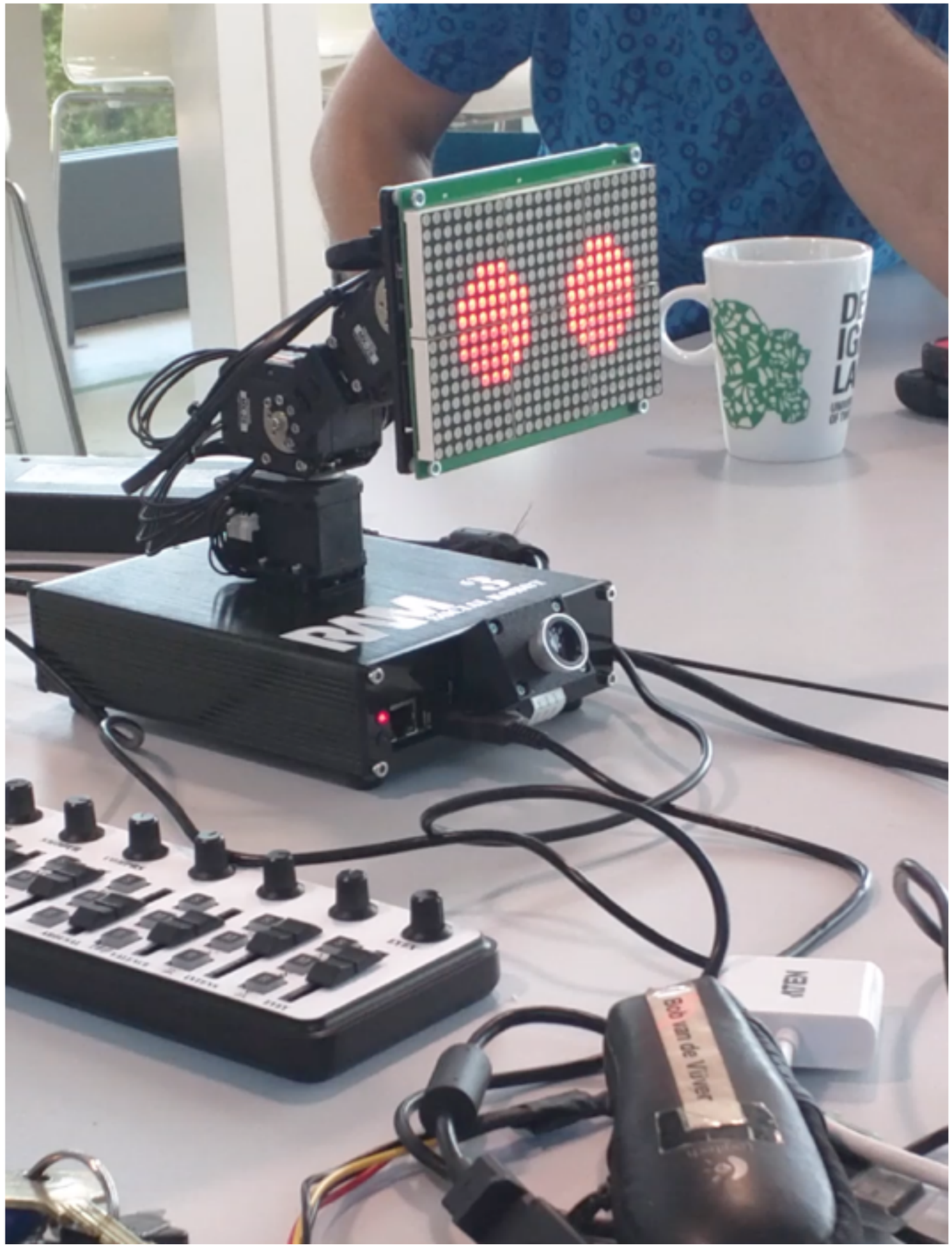
18 Jul – 12 Aug 2016 (DesignLab), hosted by Human Media Interaction, University of Twente

Search

Home Calls Projects & Teams Important dates Author's kit Workshop News

HMI (uTwente) social robot

sensitive
reactive
autonomous
aware
responsive
conversational
cute

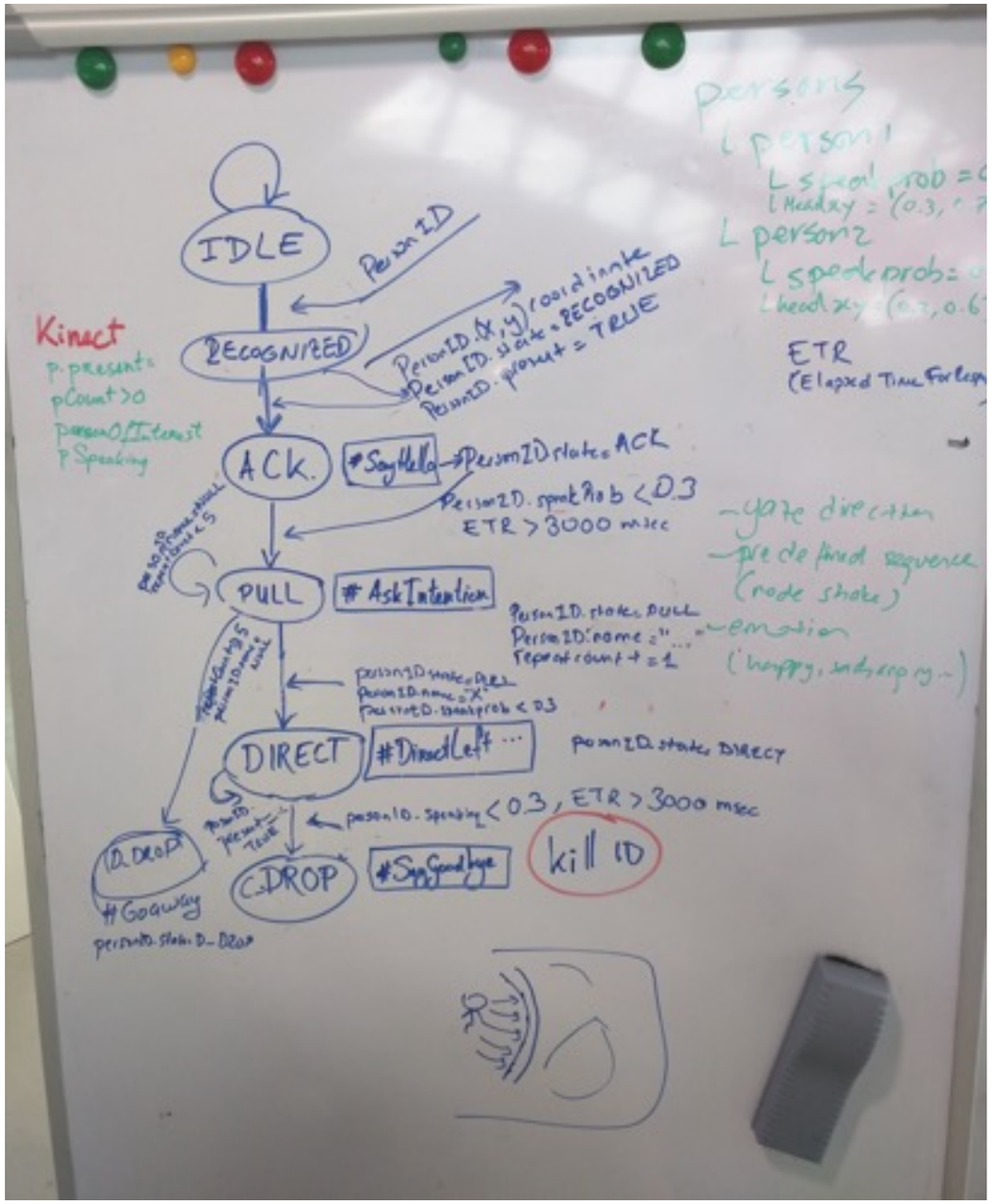


HMMM @ eNTERFACE ’16

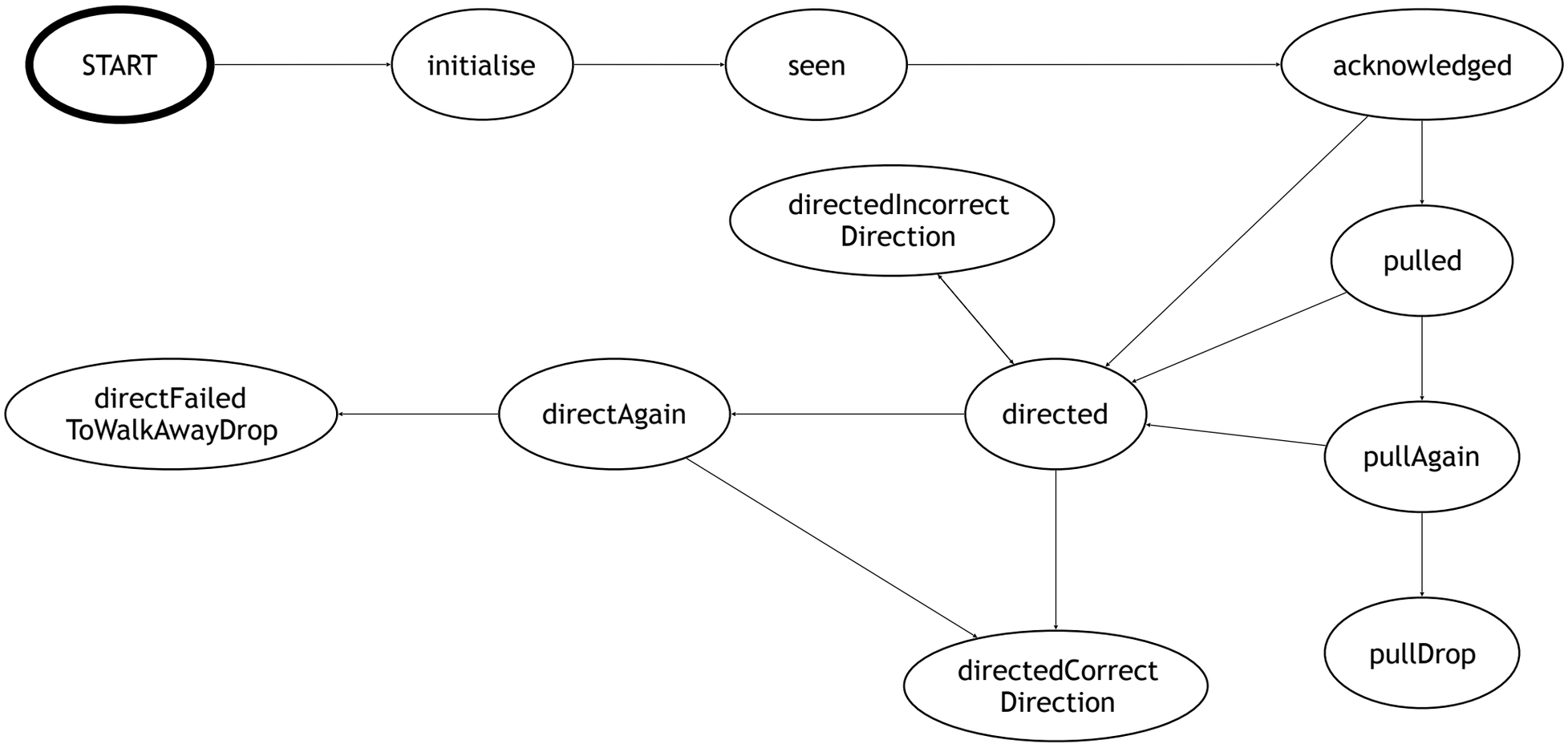
- Heterogeneous Multi-Modal Mixing :

- *“Realising fluent, multi-party, human-robot interaction with a mix of deliberate conversational behaviour and bottom-up (semi)autonomous behaviour”*

- for a ‘receptionist’ robot



how to handle humans

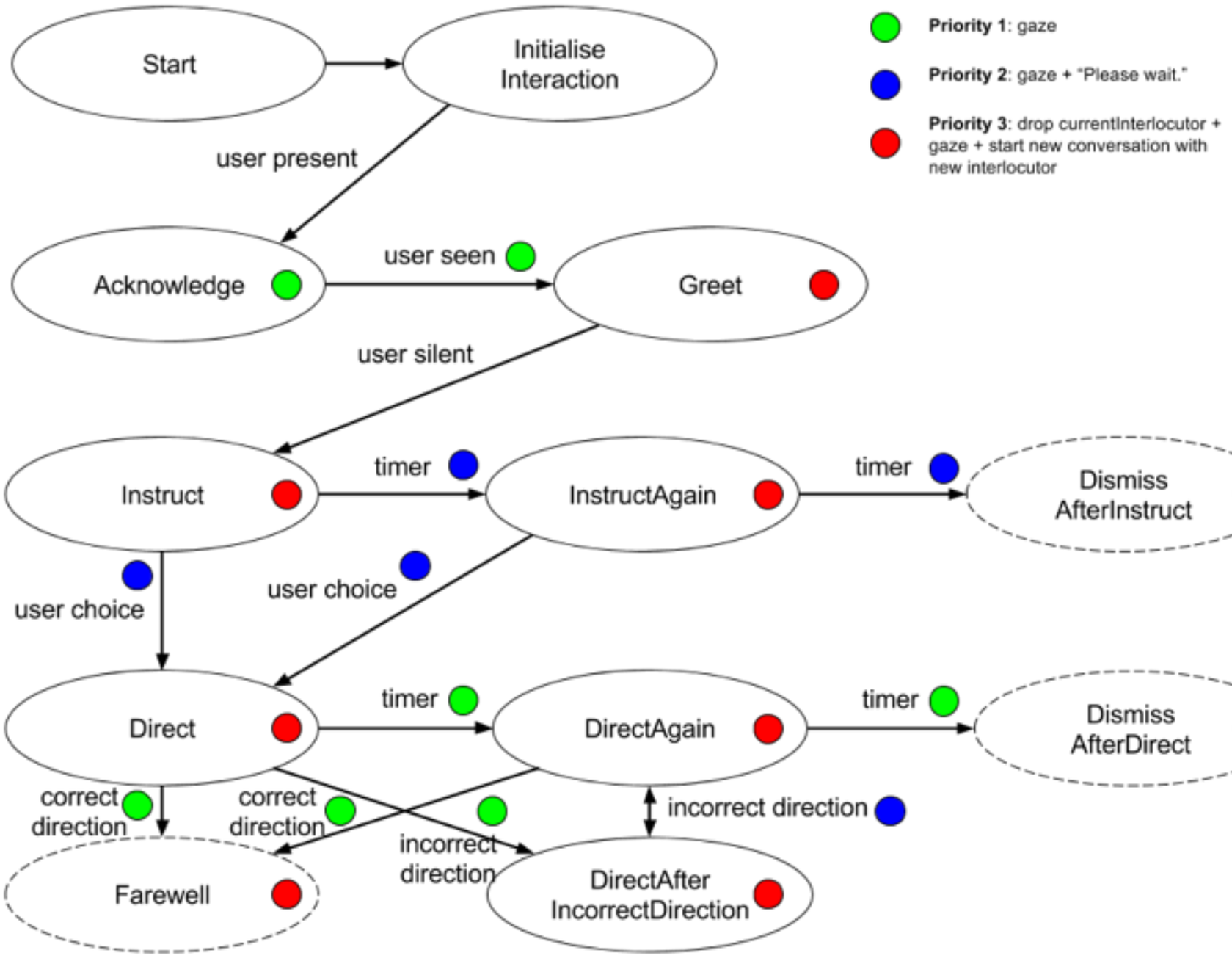


if you are a ‘receptionist’ and there’s only one of them!

situation awareness

- the receptionist robot (r3d3) usually idles . . .
- if it senses the presence of a human it acts:
  - acknowledge/examine/resolve/remove/idle
  - multiple humans can cause problems
    - so queue them & do small-talk (active idling)

while not idling:



* image copyright eNTERFACE-HMMM

how interactive?

- social intelligence
- context awareness
- content knowledge ***
- response sensing ability
- being aware!

Herme’s tricks

- Herme speaks in triads:
- she keeps control of the conversation - without understanding - but by knowledge of the context
- visitors respond instinctively to her utterances - little understanding on her part is required!

herme's (triadic) chat-lines

- hello? hi
hello
hi

- my name is hur-mi. h e r m e
hur-mi
what's your name?

how old are you?
 - really?
I'm nearly seven weeks old

- do you have an i d number
i need an i d number to talk to you
i d numbers are on your right
thank = you

- are you from dublin?
 - really
I'm from the Speech Communication Lab
                    here in Tee See Dee

- tell me about you
 - really?
owe?

- tell me something else
owe
really

- why are you here today?
really
why

- do you like the exhibition
really
why?

=> i like your hair

- do you know any good jokes?
tell me a funny joke
ha ha haha ha
tell me a knock knock joke
who's there
who?
who
ha ha haha ha
- I know a joke
what's yellow and goes through walls
a ghost banana
ha ha hehe he.
ho hoho ho ho

- thanks for your help
goodbye, see you later
goodbye

encouraging laughter

- we laugh when something is funny - ha ha!
- but we laugh more often when we are embarrassed or when we have understood a point in a conversation or have achieved a completion
- laughter punctuates normal conversation

laughter & entropy

- Bonin's work (and others) showed us that laughter functions in conversation as an entropy-killer

- it signals the natural thematic structure and provides break-in points for topics to be reset

- there's probably a use for that . . .

social constraints

- social constraints that apply between humans can be used in machine-generated dialogues to facilitate conversational interaction

- Herme did no processing of ‘meaning’ - very little was necessary - on either side!

- so when our delivery device, the interactive speech synthesiser, does its stuff, *when we have control of the context,* then we can use social expectations to minimise the processing load

minimum entropy?

SCL@SCSS

- **Speech Communication Lab**
- **School of Computer Science & Statistics**
- **Trinity College Dublin, Ireland**

- with thanks to Science Foundation Ireland

thank you

he also said:

- The disintegration of a single radioactive atom is observable (it emits a projectile which causes a visible scintillation on a fluorescent screen). But if you are given a single radioactive atom, its probable lifetime is much less certain than that of a healthy sparrow. Indeed, nothing more can be said about it than this: as long as it lives (and that may be for thousands of years) the chance of its blowing up within the next second, whether large or small, remains the same.

- **Erwin Schrodinger** “What is life? The Physical Aspect of the Living Cell”, TCD, 1943

third inspiration:

Scientific Papers of Josiah Willard Gibbs, Volume 1/Chapter II

< Scientific Papers of Josiah Willard Gibbs, Volume 1

←Graphical Methods in the Thermodynamics of Fluids

Scientific Papers of Josiah Willard Gibbs, Volume 1 *by Josiah Willard Gibbs*

A Method of Geometrical Representation of the Thermodynamic Properties of Substances by means of Surfaces

On the Equilibrium of Heterogeneous Substances, Part 1→

[33]

II.

A METHOD OF GEOMETRICAL REPRESENTATION OF THE THERMODYNAMIC PROPERTIES OF SUBSTANCES BY MEANS OF SURFACES.

[*Transactions of the Connecticut Academy*, II. pp. 382–404, Dec. 1873.]

THE leading thermodynamic properties of a fluid are determined by the relations which exist between the volume, pressure, temperature, energy, and entropy of a given mass of the fluid in a state of thermodynamic equilibrium. The same is true of a solid in regard to those properties which it exhibits in processes in which the pressure is the same in every direction about any point of the solid. But all the relations existing between these five quantities for any substance (three independent relations) may be deduced from the single relation existing for that substance between the volume, energy, and entropy. This may be done by means of the general equation,

$$d\epsilon = td\eta - pdv, \tag{1}$$[1]

$$\text{that is, } p = -\left(\frac{d\epsilon}{dv}\right)_\eta, \tag{2}$$

$$\text{that is, } t = \left(\frac{d\epsilon}{d\eta}\right)_v, \tag{3}$$

where v, p, t, ϵ, and η denote severally the volume, pressure, absolute temperature, energy, and entropy of the body considered. The subscript letter after the differential indicates the quantity which is supposed constant in the differentiation.

Addendum: graphical representation of the free energy of a body - can this be applied to conversation states?

- The figure shows a plane of constant volume, passing through the point A that represents the body's initial state.

- The curve MN is the section of the "surface of dissipated energy". AD and AE are, respectively, the energy () and entropy () of the initial state. AB is the "available energy" (now called the Helmholtz free energy) and AC the "capacity for entropy" (i.e., the amount by which the entropy can be increased without changing the energy or volume).