# Conversational Informatics: From Conversational Systems to Communication Intelligence

Digital Culture

Andrea Omicini
andrea.omicini@unibo.it

Master in Finanza, Controllo e Auditing + Marketing and Sales / Bologna Business School

Academic Year 2018/2019

# Disclaimer

- a number of different sources were used to prepare these slides
- currently, a possible reference textbook for conversational informatics is [Nishida et al., 2014], even though not all the results it presents are well stabilised
- many terms and concepts in the following slides are taken from [Nishida et al., 2014], even though the reference is not repeated everywhere

# Next in Line...

# Which Sorts of Systems? I

## Artificial and computational systems

- nowadays, most (if not all) artificial systems have a prominent *computational* part
- computational machines
  - have both an *abstract* and a *physical* part
    - where the physical portions are often abstracted away
  - are (mostly) *symbolic*
    - can deal with math, logic, data, information, knowledge
  - are *general-purpose* machines
    - programmable, can be specialised to most purposes

# Which Sorts of Systems? II

## Artificial systems in context

- most artificial systems *participate* to the activities of individuals, groups and societies
- even more, nowadays they are mostly essential to all sorts of human activities

## Socio-technical systems (STS) [Whitworth, 2006]

- artificial systems where both *humans* and *artificial* components play the role of system components
  - from online reservation systems to social networks
- most of nowadays systems are just STS
  - or, at least, cannot be engineering and successfully put to work without a proper socio-technical perspective in the engineering stage

# Which Sorts of Systems? III

## Situated systems

- the *physical* nature of artificial components cannot be always be forgot
- as well as the situatedness in time and space
- along with the influence of the surrounding environment
- most of the interesting systems, nowadays, are *situated systems*, too [Suchman, 1987]

# Which Sorts of Systems? IV

## Spatial systems

Systems where *space* cannot be abstracted away from computation [Shekhar et al., 2016], since space

- is fundamental to the application problem
- is explicitly represented and manipulated
- is essential to express the result of a computation

## Pervasive systems [Zambonelli et al., 2015]

- affecting every aspects of our everyday life
- by spreading through the whole environment where we live and act
- we live surrounded by *pervasive systems* [Grimm et al., 2004]

# Conversational Informatics as a New Paradigm I

## Conversation for intelligent systems

- conversation as a primary means for building *intelligent systems*

! STS could fulfil their potential only by providing full conversational capabilities with humans

# Conversational Informatics as a New Paradigm II

## Conversational informatics

Conversational informatics is an interdisciplinary effort [Nishida et al., 2014] gathering and integrating contributions from

- artificial intelligence
- natural language processing
- speech and image processing
- cognitive science
- conversation analysis

# Conversational Informatics as a New Paradigm III

## Focus

It focusses on

- computer-mediated interaction / communication / coordination
- human-computer interaction (HCI)
- knowledge management and creation

during conversations

# Artificial Intelligence I

## Winter Era

- AI began in 1956 with huge expectations
- basic principles and goals were set from the beginning
- stand-alone computational intelligence as a main target, instead of open-ended interactive systems, mostly to demonstrate the power of computation alone
- lack of raw computational power lead to AI Winter ('80), in spite of some (isolated yet huge) success—Dendral, Macsyma, ALVINN...

# Artificial Intelligence II

## Renaissance

- data-intensive approaches since the Eighties
- machine learning and data mining as the driving forces
- the AI Book by Russell & Norvig summarises the many successes around the agent notion
  - large scale search, knowledge-based systems, intelligent media technology (e.g., language, speech and vision), planning, machine learning and data mining, ...
- practical AI in action
  - new paradigm of computational intelligence leads to huge success—e.g., DeepBlue, IBM Watson, Google Voice Search, Siri
  - progress on traditional AI fields produces autonomous vehicles such as Mars Rovers, Google Self-Driving Car

with more and more data-intensive approaches taking the stage

# From Info- to Techno-plosion I

## Info-plosion

- our society is characterised by the exponential growth of the amount of information on the globe
- when looking at technology, humans are less interested in the computational artefacts than in the information they bring about
- some authors in the conversational informatics area refer to that phenomenon as the info-plosion

# From Info- to Techno-plosion II

## Techno-plosion

- AI accelerates info-plosion and push it towards techno-plosion
- where computational artefacts become pro-active agents within STS
- *autonomous agents* are going to coexist with humans in complex virtual / real / augmented environments
- the need emerges for adequate means of interaction between humans and agents

$\rightarrow$ communication intelligence

# Next in Line...

# Conversation for Humans I

## Conversation is easy for humans

- conversation (in any form) is apparently the most important social activity for humans—more general, for social animals
- humans engage themselves in conversations using all their skill sets without apparent effort
- however, this has not produced a single, comprehensive framework for investigating conversation

# Conversation for Humans II

## Viewpoints

- verbal communication
- nonverbal communication
- social discourse
- narratives and content flow
- cognitive processes

# Stories and Narratives

## Storytelling

- conversations as exchanges of small information items that build large stories
- pieces of conversations are brought in and out, back and forth from conversations to individual lives
- to understand facts, to revision beliefs, to interpret and predict events, to build new stories
- storytelling has been pointed out as the source of human intelligence: not just to communicate, but to recognise and understand the world
- storytelling plays an essential role in the cognition process

# Conversation in a Social Discourse I

## Social mechanisms for conversation

- conversation might be either *focussed* or *unfocussed*
- typical conversations have *roles*—e.g., speaker, addressee, side participants, bystanders, eavesdroppers
- conversations have *spatial arrangements* for participants
- the *engagement* in a conversation may vary—for different participant at a given time, for the same participant over time

# Conversation in a Social Discourse II

## Speech act theory

- speech act theory analyses utterances of conversation with respect to social implications
  - **locutionary acts** — as the actual action of making an utterance
  - **illocutionary acts** as the social acts brought about by given utterances
  - **perlocutionary acts** — as the social effect caused by the given utterances
- since it focuses more on the individual communication acts in isolation, speech act theory is precious yet limited in scope and results

# Conversation in a Social Discourse III

## Joint activity theory

- focus on *interactive aspects*, such as collaborations, coordination among participants, which should be of primary importance in conversation
- basic hypothesis
  1. language is fundamentally used for social purposes
  2. language use is a species of *joint action*
  3. language use always involves speaker's *meaning* and addressee's *understanding*
  4. the basic setting for language use is face-to-face conversation
  5. language use often has *more than one layer* of activity
  6. the study of language use is both a *cognitive* and a *social* science

# Conversation in a Social Discourse IV

## Action ladder

- according to joint activity theory, participants of conversation build a *layered space* of new information on the common ground using shared coordination devices
- actions of participants consist of multiple levels of abstraction, or action ladders, in which the actions on the upper levels are realised by those on the lower levels

# Conversation in a Social Discourse V

A+B

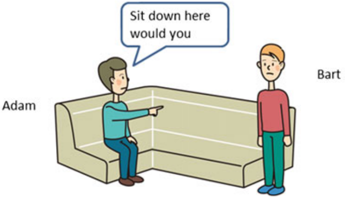
| Level | A's actions in progress | B's actions in progress |
|---|---|---|
| 4 | Proposal and consideration | |
| 3 | Signaling and recognition / meaning and understanding | |
| 2 | Presentation and identification | |
| 1 | Execution and attention | |

A

| Level | A's actions in progress |
|---|---|
| 4 | A is proposing to B |
| 3 | A is asking B |
| 2 | A is presenting to B a signal |
| 1 | A is executing for B's perception |

B

| Level | B's actions in progress |
|---|---|
| 4 | B is considering A's proposal |
| 3 | B is recognizing A's request |
| 2 | B is identifying A's signal |
| 1 | B is attending to A's articulation |




[Nishida et al., 2014]

# Conversation in a Social Discourse VI

*A* -> *B*: "please sit down"

| | |
|---|---|
| Perlocutionary act | *A* is trying to get *B* sit down. |
| Illocutionary act | *A* is asking *B* to sit down. |
| Locutionary act | *A* is saying to *B* "Please sit down." |
| Rhetic act | *A* is using the words *please*, *sit*, and *down* with a certain sense and reference. |
| Phatic act | *A* is uttering the words *please*, *sit*, and *down*. |
| Phonetic act | *A* is producing the noises that constitute "Please sit down." |

downward evidence

upward completion

[Nishida et al., 2014]

# Making Sense of a Conversation with Multiple Modalities

## Using body to talk

- in a conversation, the speaker usually *coordinates* multiple parts of his/her *body*, including his or her eyes, face, hands, head, and torso
- the speaker does this to *express* what he/she wants to communicate
  - e.g., pointing to refer to a portion of the surrounding environment
- complementing or even substituting utterance
- integrating all the diverse forms of communication allow humans to make sense out of a conversation

# Cognitive Process I

## Modelling the cognitive process underlying conversations

- early attempts in cognitive linguistics, focussing mostly on the cognitive apparatus required to interpret verbal items in a conversation
- *mind reading* is pointed out as an essential in conversations, and more generally in social life
- as the ability to understand motivations, beliefs, goals of the other participants while they speak—*theory of mind*

! some other primates have exhibited some theory of mind

# Cognitive Process II

## Emotions

- emotion plays an important role in human intelligence and hence in communication
- different frameworks for emotion classification and composition were developed and proposed, and their effects on conversation discussed and exemplified
- in particular, investigating how emotions relates to goal-oriented activities—in particular, how they may affect goal-oriented conversation

# Next in Line. . .

# Focus on. . .

# Lines of Research for Conversational Systems I

## Early history

- in the 1960s, work on text-based natural language dialogue systems begins
  - the first computer programs capable of answering questions typed in English appear
  - simple yet ground-breaking
- in the 1980s, speech recognition systems were first developed, towards a more natural user interface were developed
- in the 1990s, multimodal dialogue systems allowed users to combine speech with nonverbal input – e.g., gestures, gazes – to interact with a computational system
- in 1987, Apple's Knowledge Navigator previewed the role of AI within an embodied conversational agent

# Lines of Research for Conversational Systems II

## Understanding and generating stories

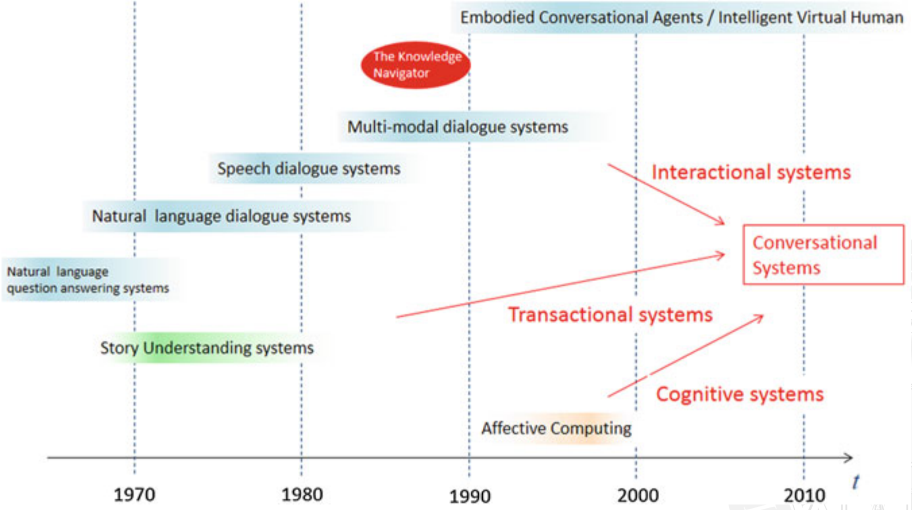
- other AI approaches focussed on how narratives and stories could be understood and generated
- towards *transactional systems*
- such approaches are critical to address content management of conversational systems
- not just telling stories
- but also learning from conversations, and retaining (pieces of) accumulated conversations in memory, for future (re)use

# Lines of Research for Conversational Systems III

## Cognitive systems

- since the 1990s, focus on cognitive system exhibiting more human-like intellectual behaviours resulting from autobiographical dynamic memory, emotion and theory of mind
- among others, *affective computing* became a visible line of research

# Lines of Research for Conversational Systems IV




[Nishida et al., 2014]

# Lines of Research for Conversational Systems V

## Integration—as usual

It can be easily observed that

- without proficient interaction, stories may be told poorly
- without stories, conversation is boring
- without a cognitive model, we cannot build an attractive conversational agent

As a result, integration is considered necessary for conversational systems to be successful

# Early Examples of Natural Language Dialogue Systems I

## Baseball (1961)

- natural language question answering system built to answer questions about baseball games
- based on a small dictionary and programmed grammar
- produces database retrieval commands to answer information requests in natural language (English)

# Early Examples of Natural Language Dialogue Systems II

## LUNAR (1973)

- Lunar Natural Sciences Natural Language Information System (LUNAR)
- a natural language question answering system that can answer queries about the rock samples brought from the moon, featuring
  - *syntactic analysis* using heuristic and semantic information to select the most likely syntactic analysis from candidates
  - (domain specific) *semantic representation* used as an intermediate representation to produce database queries
  - Augmented Transition Network Grammar (ATNG) to analyse inputs by combining syntactic and semantic analysis
- able to answer complex queries such as

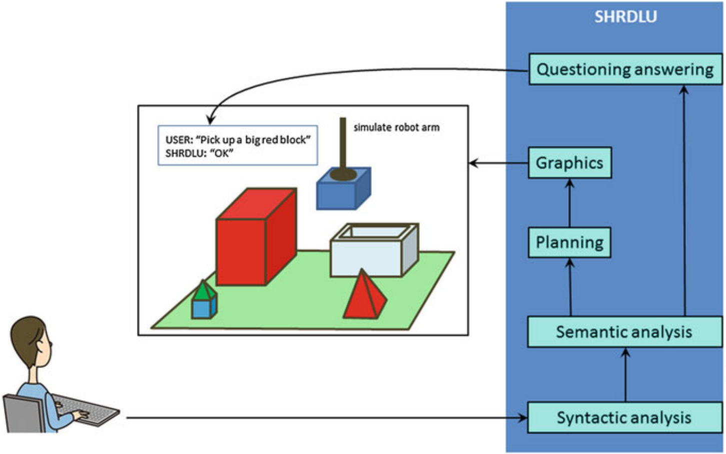
```
Give me all lunar samples with Magnetite
In which samples has Apatite been identified
How many samples contain Titanium
Which rocks do not contain Chromite and Ulvospinel.
```

# Early Examples of Natural Language Dialogue Systems III

## SHRDLU (1972)

- interacting program able to understand the meaning of natural language expressions with respect to the *block world*
- while limited in scope, first approach prototyping the *entire process of natural language understanding* in computational terms
- combining all required components (natural language processing, planner, . . . ) in a virtual agent behaving as if it understands the meaning of natural language
- e.g., given the utterance "pick up a big red block", SHRDLU can determine which physical object is referred to and what goal the user wants it to achieve

# Early Examples of Natural Language Dialogue Systems IV




[Nishida et al., 2014]

# Early Examples of Natural Language Dialogue Systems V

## ELIZA (1966)

- no extensive grammar, no semantic knowledge—just
  - identification of keywords
  - detection of a minimal context
  - selection of proper transformations
  - response generation when no keywords are found in the input
- basically, cunning pattern matching—but funny, and brought some fans to AI at the time

# Speech Dialogue Systems and Multimodal Interfaces I

## HEARSAY-II (1980)

- HEARSAY-II allowed the user to actually speak to the system
- based on a *blackboard architecture*
  - a shared database (blackboard)
  - a collection of cooperating processes
- information representation at seven levels of abstraction, from the parameter level to the phrase level
- knowledge sources processing incoming signal concurrently, generating hypothesis for interpretation at the different levels of abstractions, that should converge in a single coherent interpretation
- changes handled by the *blackboard monitor*
- later replaced by the goal blackboard and goal processor, dedicated to *goal- oriented processing*

# Speech Dialogue Systems and Multimodal Interfaces II

## Put-That-There (1980)

- speech dialogue systems extended to multimodal interfaces
- so as to use more than one modality of communication in human–computer interaction
- Put-That-There allowed the user to arrange simple shapes on a large graphics display surface by *voice* and a simultaneous *pointing gesture*

# Speech Dialogue Systems and Multimodal Interfaces III

## Multimodal interaction systems

- fusion as the key issue
- robust and efficient algorithms to interpret multimodal input from heterogeneous sensors by handling errors and missing data
- to produce multimodal output, output from different knowledge sources are tailored and synchronised
- this involves content selection and organisation, coordinated distribution of information on available modalities, modality-specific content realisation, and laying out generation results
- *presentation* may be characterised as a knowledge-based, goal-directed activity under constraints
- planning algorithm to assemble the components into a coherent structure of presentation

# Embodied Conversational Agents & Intelligent Virtual Humans I

## The Knowledge Navigator (1987) & Peedy (1997)

- The Knowledge Navigator video (TKNV) pioneered the idea of a visual human-like embodied conversational agent capable of multimodal conversational interaction
- anthropomorphism is debatable, but surely can allow the users to apply proficient inter-human communication skills
- Peedy, the Conversational Personal Assistant, is a conversational system in which a parrot-agent named Peedy helps the user select songs from a collection of audio CDs—and embodied the key concepts introduced in TKNV

# Embodied Conversational Agents & Intelligent Virtual Humans II

## Rea (1999)

- Rea is an embodied conversational agent that can effectively use a human-like body for verbal and nonverbal communication
- the system actually implements conversational interactions, in contrast to employing conversation as a metaphor of interface
- focused on effective use of nonverbal communication media as social signals, such as gaze or hand gestures, to achieve communication functions, such as acknowledgment of the communication partner or taking turns
- implemented the ideas in the real-estate domain, where Rea was characterised as a real-estate agent who shows users the features of various houses
- full symmetry between input and output modalities were intended to allow Rea to participate in human-computer conversation with equal footing

! KQML (standard markup / agent communication language)was employed at the system level so that heterogeneous modules could interact with each other without constraints by exchanging messages in a common representation language

# Embodied Conversational Agents & Intelligent Virtual Humans III

## Methodological progress after Rea

- specific script/markup languages for specifying the behaviour of embodied conversational agents—e.g., AIML, Microsoft Agent, MPML, STEP, BML
- corpus-based generation of behaviours in which the agent's behaviour is generated in two phases
  - first one or more corpus is built by collecting and annotating a sufficient amount of observational data for the target communicative behaviours
  - then, parameter values are determined based on the analysis of the data accumulated in the corpora

# Story Understanding/Generation Systems

## Knowledge to understand stories

- after early experiences, it was soon clear that some background knowledge (structure) was required in order to understand stories
- e.g., *scripts* to represent knowledge of stereotypical scenes
- or, *plans* to understand goal-oriented stories
- theory of *dynamic memory* was used to learn from stories

# Cognitive Computing I

## Believability and lifelikeness of interacting agents

- interacting agents should be believable in order to gain sufficient sense of presence and reality
- a believable agent can provide the illusion of life and permit the audience's suspension of disbelief
- e.g., in the Oz project (1994), virtual worlds are inhabited by interactive characters within which the user experiences a story from a first person perspective
- e.g., The Artificial Life Interactive Video Environment (ALIVE) system allows the user to engage in entertainment interaction with an animal-like agent

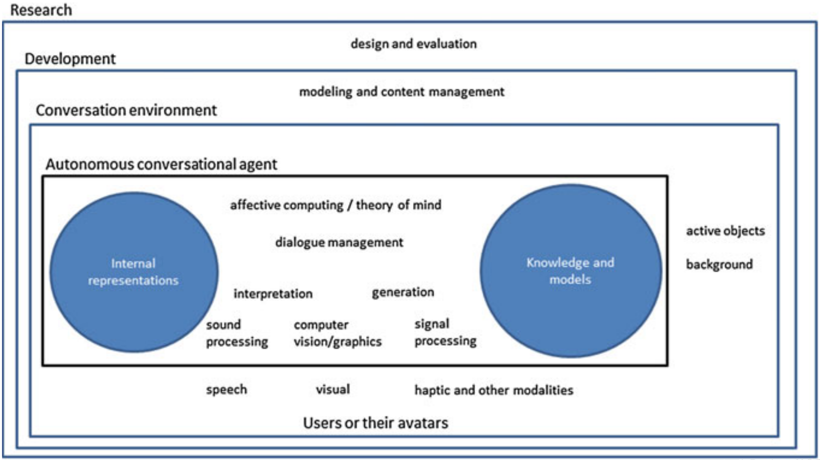
# Cognitive Computing II

## Affective computing

- *emotion* plays a critical role in communication and decision making
- e.g., an emotionless machine intelligence might be vulnerable; it may not recognise a danger that is not logically deducible, which even a young child could easily identify
- in communication, the ability to express and interpret emotion is mandatory for an intelligent agent to engage *empathically*
- according to Picard (1997), *emotional intelligence* consists in the core of human intelligence, argued that computational intelligence needs the ability to recognise and express emotions in order to be a genuinely intelligent partner
- Affective Reasoner (1992) used the OCC model to realise emotional natural language dialogues

# Focus on. . .

# Issues

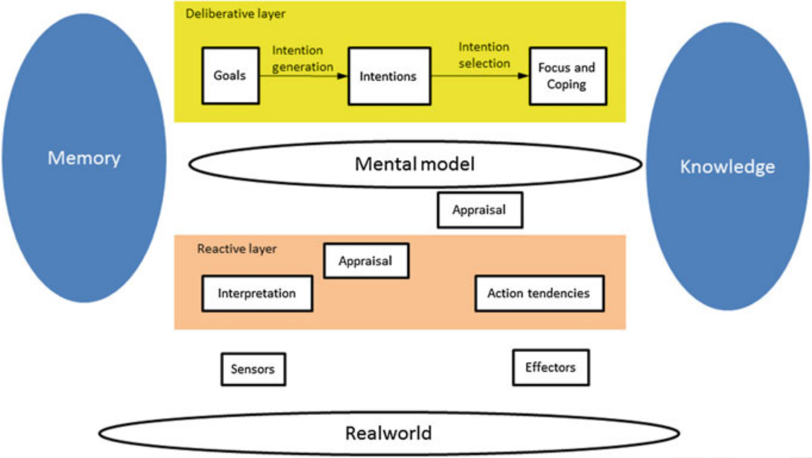



[Nishida et al., 2014]

# Architecture I

- key issue: identifying major components of conversational systems and their organization
- components from insights from existing systems
- to support a complex structure of invocation and coordination among components, a control structure is required—e.g., blackboard

# Architecture II




[Nishida et al., 2014]

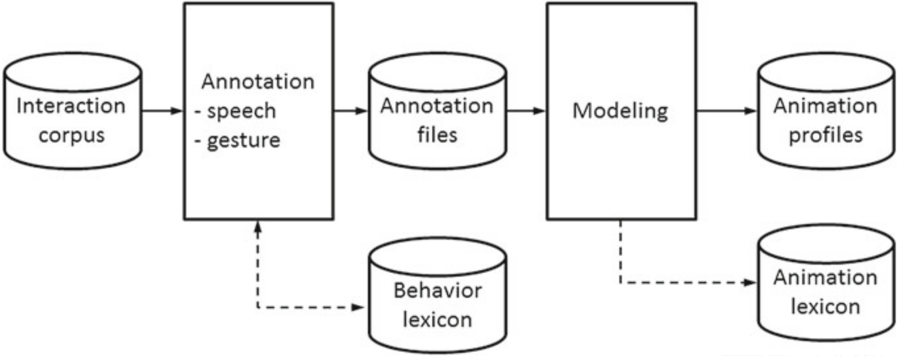
# Scripts & Markup Languages

- general requirements include easy specification of the behaviour of conversational systems
- scripts are procedural, markup languages are declarative
- markup languages allow for partial specification of behaviours, which may be easy for system builders while increasing computational cost
- BML and FML result from a standardisation forum to allow researchers to share data and codes

# Conversation Corpus for Behaviours I

- a corpus-based approach implements a *data-driven approach* and bases target behaviours of conversational agents on varieties observed in existing conversations
- in corpus-based generation of conversation agent behaviours, one first creates an interaction corpus or a database containing instances of actual conversation behaviors, from which specification of interactions is generated in a markup language
- interaction corpora in systems like AMI, CHIL, VACE

# Conversation Corpus for Behaviours II



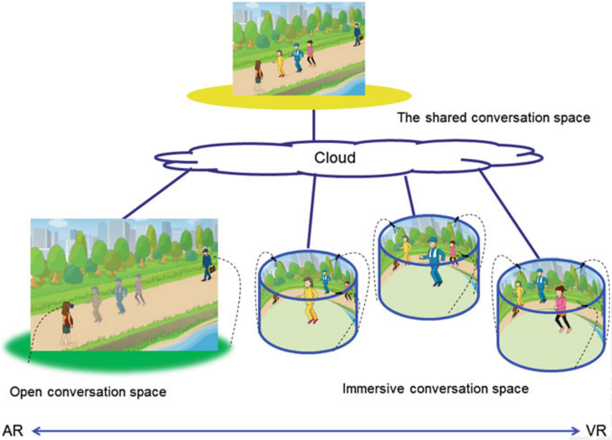

[Nishida et al., 2014]

# Conversation Corpus for Behaviours III

- the impact of a corpus-based approach depends on the quality of a corpus and resulting model
- machine learning techniques may be applied to corpora to generate interactional behaviours of conversational agents

# Focus on. . .

# Smart Conversation Space I




[Nishida et al., 2014]

# Smart Conversation Space II

- open conversation space enhanced by augmented reality (AR) where people can move around talking to each other with reference to objects and events in the space
- an immersive conversation space where each participant is completely encircled by large screens and the behaviour of the participant is sensed and reflected on his/her image in the shared conversation space

# Immersive Collaborative Interaction Environment I

- an immersive conversation space uses a series of large screens encircling the user to provide rich ambient audio-visual information so that s/he may be involved and feel closely attached to the illustrated scene
- the user in return responds with diverse social signals such as voice, body motions, altered gaze, facial expressions, and so on

# Next in Line...

# Dangers I

## The dangers of AI?

- it looks quite cool these days to chit-chat about the dangers of AI and autonomous machines: robots will take the power (steal our jobs, win our women / men, kill everybody), *yeah*, sooooon.
- however, the perils of autonomous systems are not the same posed by conversational informatics at its peak

# Dangers II

## Sociality vs. reasoning

- driven by our troublesome experience with social networks, everybody knows about *cognitive biases* such as confirmation bias and Dunning–Kruger effect
  - basically, we do not easily let facts stand in our way
- we are hyper-social animals: “We never think alone”
- reasoning evolved *after* our ability to interact socially
- we privilege social interaction over reasoning
- and, *looking right in front of our group* is more important than *being really right*—whatever this means
  - the very structure of our judiciary process should tell us a lot about this

# Dangers III

## Dangers of conversational informatics

- when social interaction becomes human interaction within STS the same issues may occur
- in particular when, for the info-plosion, we are overwhelmed by data, information, knowledge—*facts* of some sort, difficult to be understood and interpreted to compose the overall picture
- what happens after the techno-plosion, when we do no longer interact with some database or web page to look for some fact, but instead we engage in a conversation with a conversational agent?
- we are probably going to *trust* the conversational agent by default—knowledge acquired conversationally is trusted at a deeper level than any other knowledge
- e.g., fabricating false evidence is nowadays so easy, making it accepted as real could be much easier via conversational systems

# Dangers IV

## What could we do?

- the point is that info- and techno-plosion changed / are changing our surrounding environment into the knowledge-intensive, hyper-social environment we nowadays live in *too fast* for us to evolve and be ready for it
- of course, the first measure regard the way in which we *build* those systems that could endanger us—at least, for the aspects we really can control
- so, for instance, one simple advice would be *not* to build conversational agents like us, and make them instead ready – their conversational intelligence, first of all – for *rational interaction*
- e.g., empathy is ok in order to better understand humans along with their motivations, but should not overcome rationality

# References I


Grimm, R., Davis, J., Lemar, E., Macbeth, A., Swanson, S., Anderson, T., Bershad, B., Borriello, G., Gribble, S., and Wetherall, D. (2004).
System support for pervasive applications.
*ACM Transactions on Computer Systems*, 22(4):421–486.

Nishida, T., Nakazawa, A., Ohmoto, Y., and Mohammad, Y. (2014).
*Conversational Informatics. A Data-Intensive Approach with Emphasis on Nonverbal Communication*.
Springer Japan, Tokyo.

Shekhar, S., Feiner, S. K., and Aref, W. G. (2016).
Spatial computing.
*Communications of the ACM*, 59(1):72–81.

Suchman, L. A. (1987).
*Plans and Situated Actions: The Problem of Human-Machine Communication*.
Cambridge University Press, New York, NYU, USA.

Whitworth, B. (2006).
Socio-technical systems.
In Ghaou, C., editor, *Encyclopedia of Human Computer Interaction*, pages 533–541. IGI Global.

# References II

Zambonelli, F., Omicini, A., Anzengruber, B., Castelli, G., DeAngelis, F. L., Di Marzo Serugendo, G., Dobson, S., Fernandez-Marquez, J. L., Ferscha, A., Mamei, M., Mariani, S., Molesini, A., Montagna, S., Nieminen, J., Pianini, D., Risoldi, M., Rosi, A., Stevenson, G., Viroli, M., and Ye, J. (2015).
Developing pervasive multi-agent systems with nature-inspired coordination.
*Pervasive and Mobile Computing*, 17:236–252.
Special Issue "10 years of Pervasive Computing" In Honor of Chatschik Bisdikian.

# Conversational Informatics: From Conversational Systems to Communication Intelligence

Digital Culture

Andrea Omicini
andrea.omicini@unibo.it

Master in Finanza, Controllo e Auditing + Marketing and Sales / Bologna Business School

Academic Year 2018/2019