UNIVERSITÀ
di VERONA

HCI- Multimodal Systems

Introduction to Multimodal Systems

Cigdem Beyan

A. Y. 2025/2026

Associate Professor

@ Computer Science Dept.

VIPS (Vision, Image Processing and Sound) Research Group

Room: 1.87, Ca’vignal 2

cigdem.beyan@univr.it

Before UNIVR

UNIVERSITÀ di VERONA

Sept. 2004 - June 2008

BEng.
Computer Engineering
Baskent University, Ankara, Turkiye

Sept. 2008 - Aug. 2011

MSc.
School of Informatics, Middle East Technical University, Ankara

Sept. 2011 - July 2015

Ph.D.
School of Informatics, University of Edinburgh, UK

Teaching & Reserach Assistant
Computer Engineering, Baskent University

Sept. 2015 - Mar. 2021

Postdoctoral Researcher
Dept. of Pattern Analysis & Computer Vision (PAVIS), Istituto Italiano di Tecnologia (IIT), Genoa

Mar. 2021 - Aug. 2023

Affiliated Researcher
PAVIS, IIT

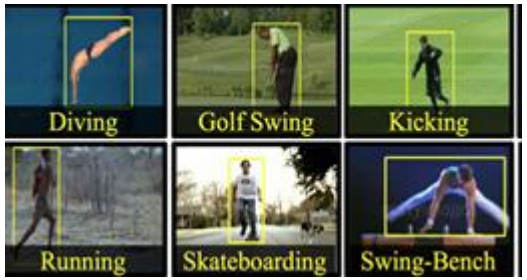
Ricercatore a tempo determinato legge 240/10 art.24-a (RTDA)
Dept. of Information Engineering and Computer Science, University of Trento

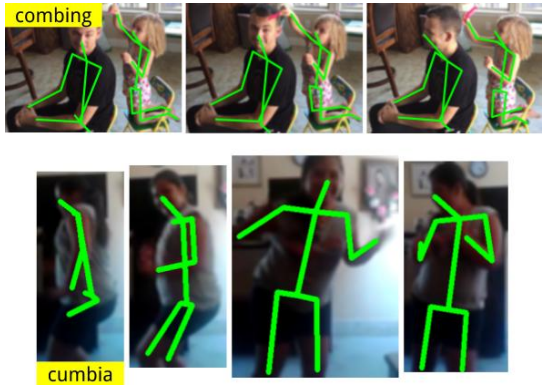
Sept. 2023-

Ricercatore a tempo determinato legge 240/10 art.24-b (RTDB)
Dep. of Management Information and Production Engineering (DIGIP), University of Bergamo

RESEARCH AREA

- **Computer Vision**
  Processing videos / images

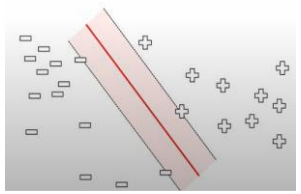


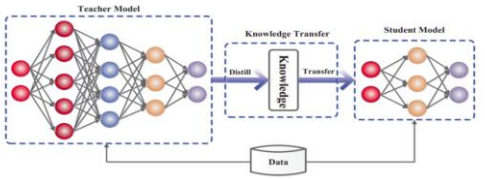


- Audio, text, etc. → **Multimedia**, **Multimodal Learning**

- Human- Computer Interaction

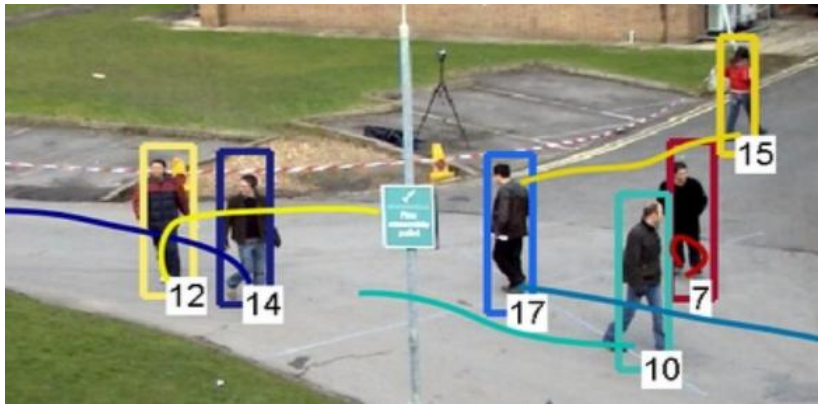
- Machine / Deep Learning Methodologies





My research (related to HCI)

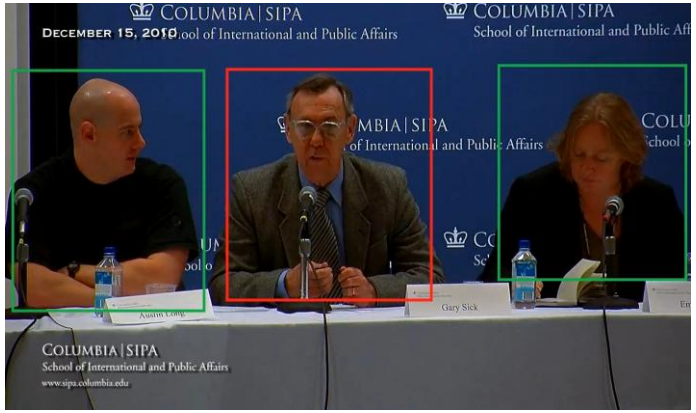
UNIVERSITÀ di VERONA



Multiple people detection and tracking

Pose estimation, pose denoising, predicting missing joints

Human Action Recognition

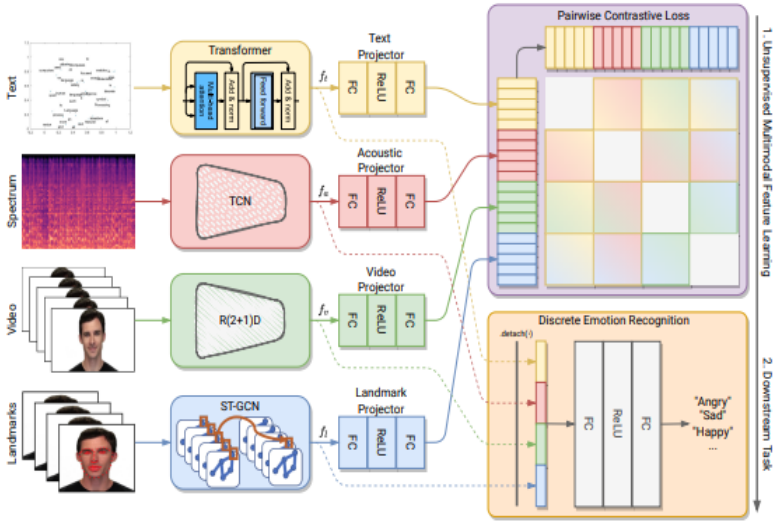


Detecting Nonverbal Signals (e.g., gaze, speaking activity, gestures, prosody)

My research (Related to HCI)

UNIVERSITÀ di VERONA

Human-Object Interaction Analysis



Multimodal Emotion Recognition

Social Traits Detection

Content

UNIVERSITÀ di VERONA

- 12H (6 weeks) of theory with me
  - This part of the course is called: Multimodal Systems
- Topics
  - Nonverbal behaviour in communication
  - Automated analysis of body: movement, gestures, facial expressions, and speech.
  - Data capturing techniques, extracting features, and automatic analysis
  - Social artificial intelligence: example applications, social psychology, organizational psychology, and social robotics.
  - Affective computing: theories of emotion, emotion recognition systems, applications of emotion recognition in HCI.
  - Integration of multimodal nonverbal cues: fusion techniques, e.g., late and early fusion.

Exam– Multimodal Systems

UNIVERSITÀ
di VERONA

- Multimodal systems (12h) will be evaluated through an ORAL exam.
- There are no Lab. works or assignments corresponding to these 12h.
- The ORAL exam will be about the course materials (slides) as well as the readings supplied by the lecturer.
  - The materials will be distributed in **Moodle**.

Schedule and readings

- Tentative Schedule
- Readings – No book but there are suggested readings linked to the slides and distributed from Moodle.

Master's Thesis Opportunities

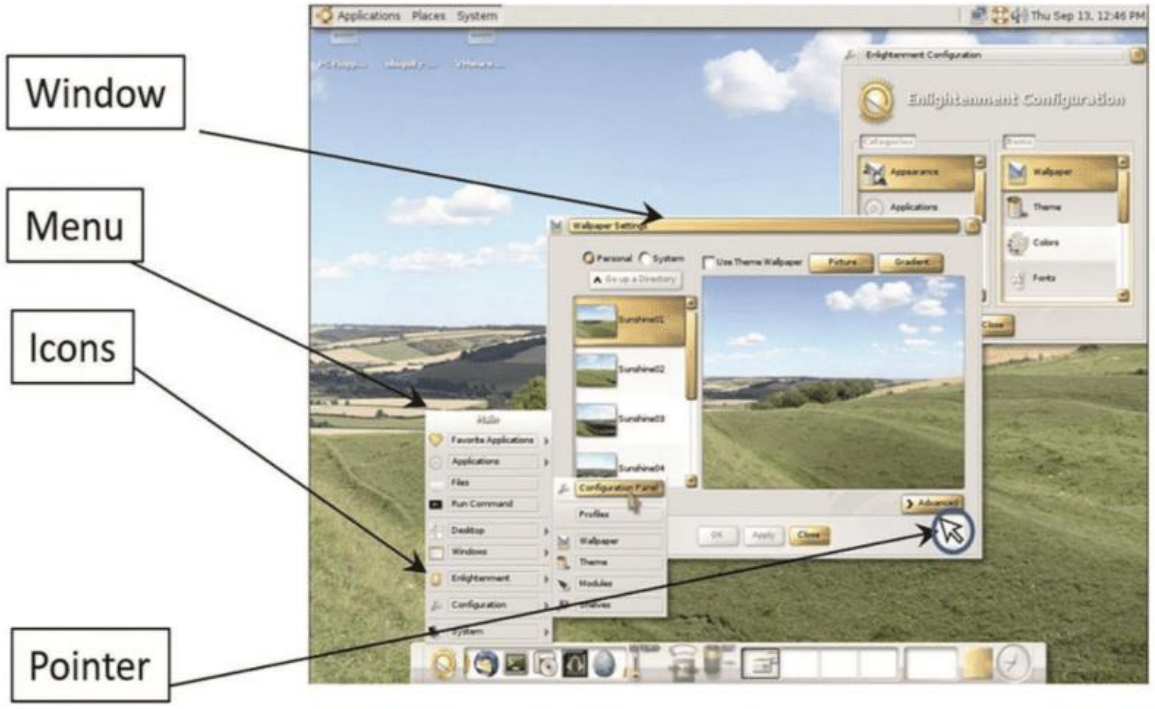
- If you are interested in the content of these 12 hours, I am offering Master's thesis opportunities on similar topics,
  - where we also focus on the **machine learning/deep learning aspect** as well.
- The goal is to develop a fully automated affective/social computing framework.

Today

- Multimodal interfaces
  - A bit of history
- What is multimodal systems?
  - What is multimodal
  - What is modality
  - Multimedia vs. multimodal systems
- Motivations for multimodality
- Verbal & Nonverbal communications
- Social signal processing & Affective computing
  - Applications, examples, in practice

GUIs & WIMP

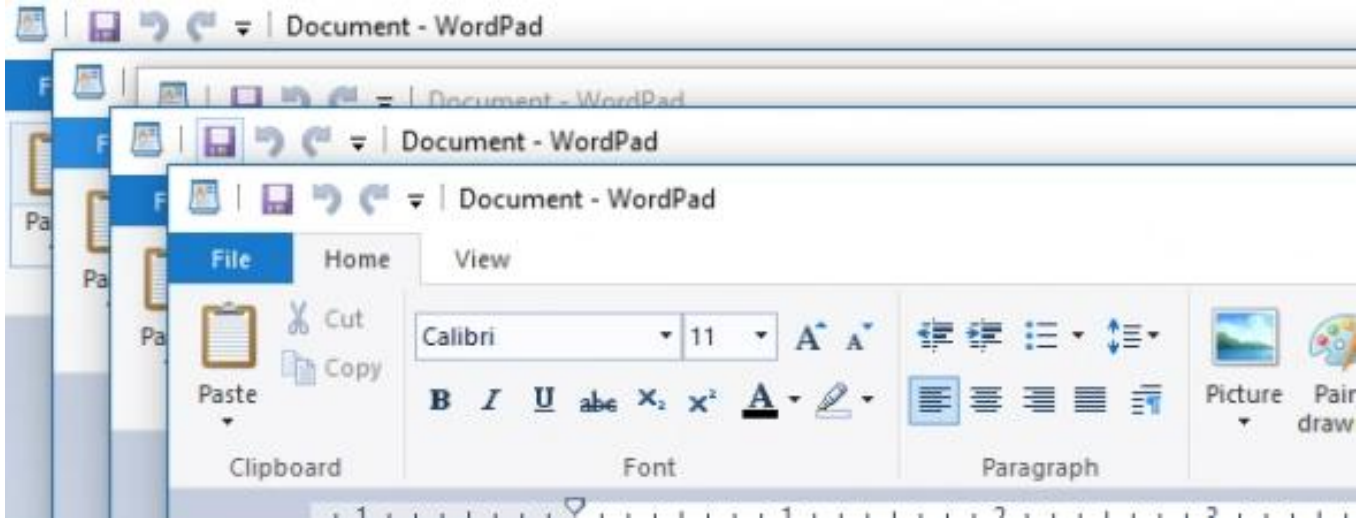
- The most traditional, consolidated, commonly used, and widespread interfaces are **graphical user interfaces** (GUIs).
- These adopt the so-called **WIMP** paradigm, i.e., **W**indows, **I**cons, **M**enus, and **P**ointing devices.



WIMP

UNIVERSITÀ di VERONA

- WIMP interfaces have become so prevalent since
  - They are very good at abstracting workspaces and documents,
  - They are analogous to documents as paper sheets or folders,
  - Their basic representations as rectangular regions on a 2D flat screen make them a good fit for system programmers,
  - Suitable for multitasking work environments.



Post-WIMP

- WIMP interfaces are not optimal for complex tasks such as computer-aided design or for applications needing more **natural interaction paradigms**, e.g., **interactive games**.
- Post-WIMP interfaces (Van Dam, 1997) aim at overcoming such problems and consist of **widgetless user interfaces**, including virtual reality systems, and user interfaces based on gestures, speech, and physical controls.
- Post-WIMP interfaces integrate input from **several sensory channels** and produce **multimedia output**.

POST-WIMP

- Wall-size displays
- Multi-touch displays
- Vehicle applications
- Public kiosks

POST-WIMP

- Virtual reality, augmented reality, mixed reality

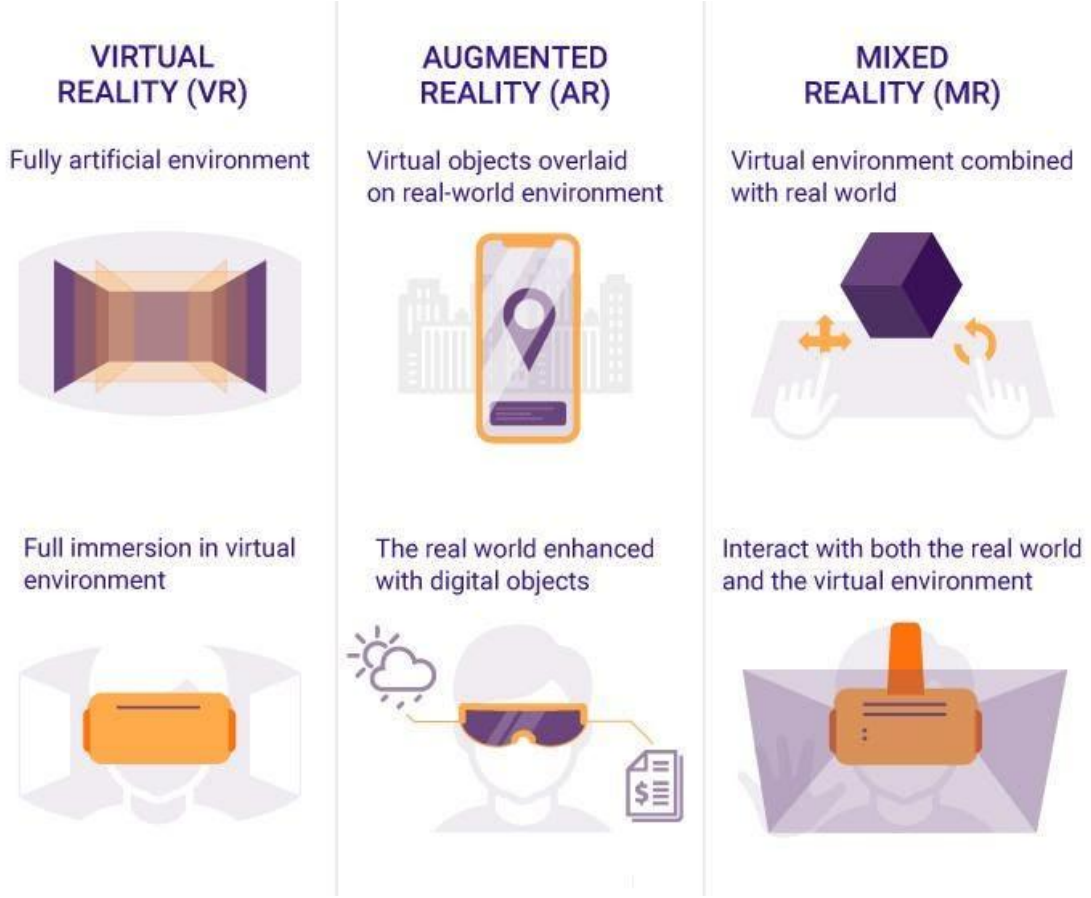


Image credit: https://www.forbes.com/sites/quora/2018/02/02/the-difference-between-virtual-reality-augmented-reality-and-mixed-reality/

Post-WIMP

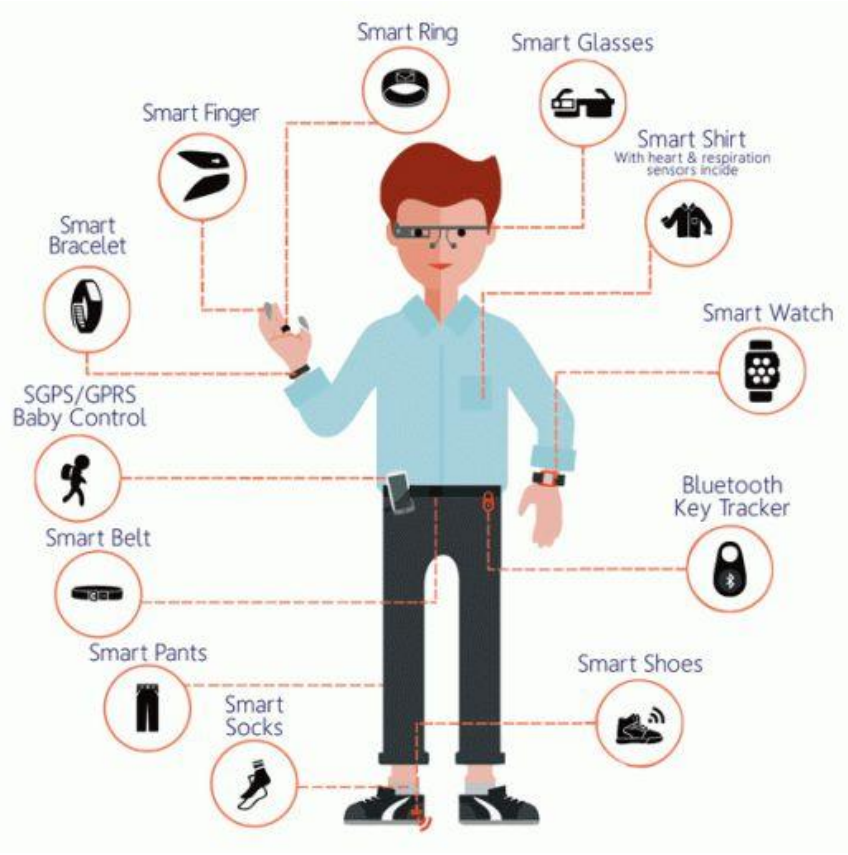
UNIVERSITÀ di VERONA

- **Pervasive Computing:** Integrating computation into everyday objects and activities, making computing accessible anytime and anywhere.
  - Computers exiting everywhere: embedded into fridges, washing machines, door locks, cars, furniture, people
  - Mobile portable computing devices
  - Wireless communication
  - E.g., Wearable health trackers
- **Ubiquitous Computing:** A vision where computing devices are seamlessly integrated into the environment and used naturally without conscious thought
  - is more user and application-driven compared to Pervasive computing, still both keywords are used interchangeably.
  - E.g., smart home devices, like lights and thermostats, automatically adjust based on your preferences without needing direct control.

Post-WIMP

- **Disappearing Computing:** The idea that technology becomes so embedded and integrated into the environment that it effectively "disappears" from the user's perception.
  - Smart fabrics and wearables
  - Voice-activated assistants (e.g., Alexa)
  - Smart buildings
  - Augmented reality contact lenses….

Image credit: www.kio.tech/en-us/blog/augmented-reality-smart-contact-lenses



Post-WIMP

- **Ambient Intelligence:** Environments that are sensitive and responsive to the presence of people, utilizing embedded systems and artificial intelligence to enhance user experience.
    - E.g., Automated lighting, climate control, traffic lights that adapt in real time to traffic conditions, cars that communicate with each other

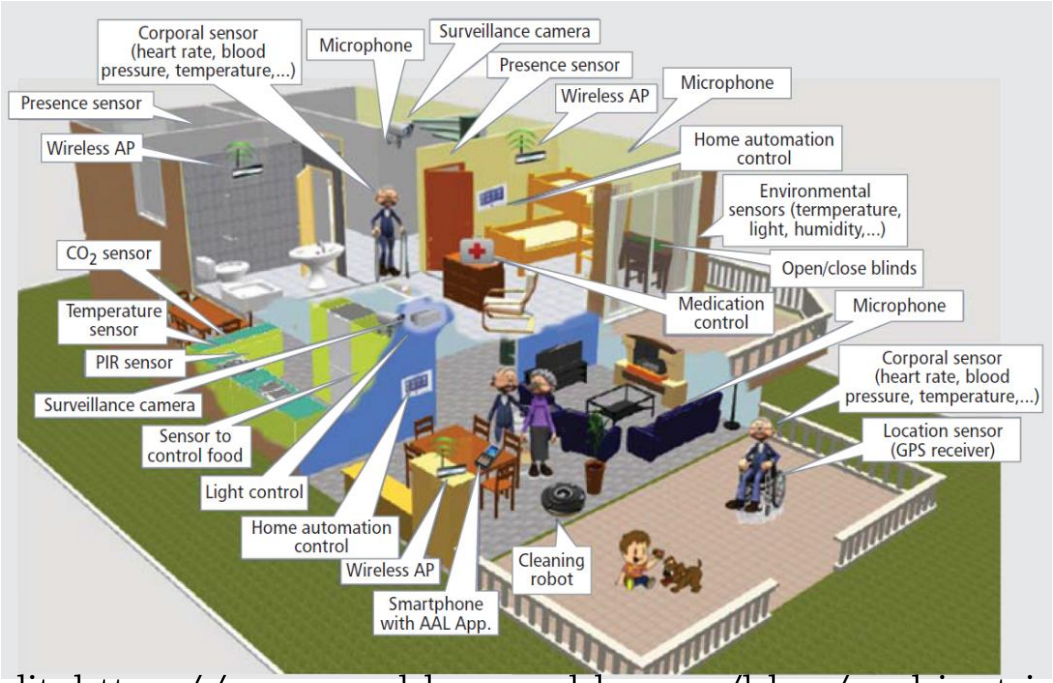


Image credit: https://ammanublog.weebly.com/blog/ambient-intelligence

Post-WIMP Interfaces

UNIVERSITÀ di VERONA

- Are grounded on models and theories from psychology, physiology, biomechanics, neurosciences, cognitive and social sciences …
- Exploit the human **sensory** system;
- Exploit **multiple sensory** channels;
- Exploit the deep **interconnections** between the **sensory** channels;
- Exploit **non-verbal communication**;
- Exploit **affect** and **emotion**;
- Exploit **social signals**;
- Exploit theories from **art** and **human sciences** (music, choreography, theatre, cinema).

Multimodal Interfaces

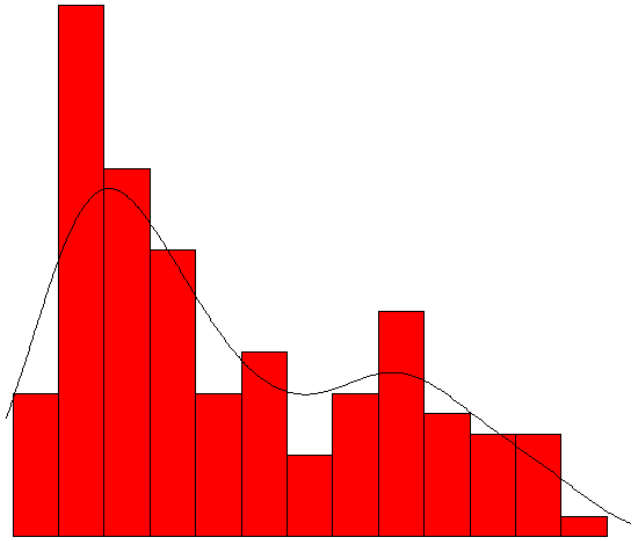
UNIVERSITÀ di VERONA

- Post-WIMP interfaces are often multimodal interfaces, that is they exploit multiple sensory modalities.
  - Sensory modality: the sensory channel through which information is perceived; it refers to the **type of communication channel** used for transferring or acquiring information.

  - Multimodal: In our context refers to the integration of information from several different **sensory channels**.

Multimodal Systems

- What is multimodal?



Multimodal distribution: Multiple modes, i.e., peaks in a probability density function

MULTIMODAL SYSTEMS

UNIVERSITÀ di VERONA

- What is modality?
  - Refers to the way in which something is expressed or perceived.

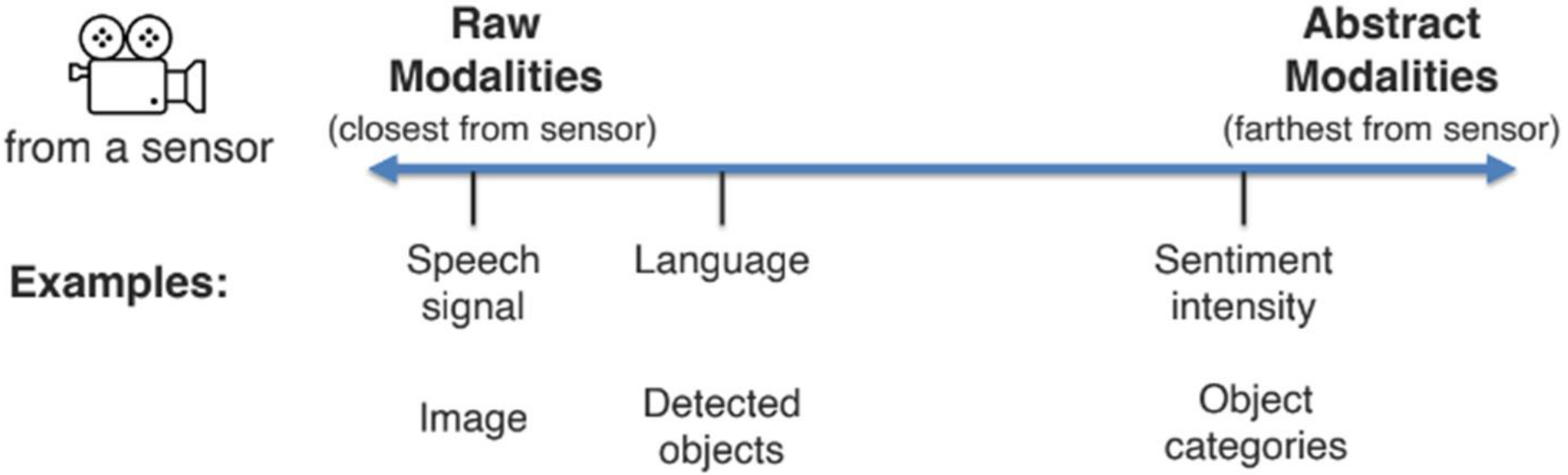


Image Credit: L.P. Morency, Multimodal machine learning lecture notes, CMU.

Multimodal Systems

- What is multimodality?

Related to sensory modalities; touch, speech, what you see (attention)….

Image Credit: L.P. Morency, Multimodal machine learning lecture notes, CMU.

Multimodal Systems

- Multimodality vs. Multimedia
  - **Multimodality**
    - *Modality* refers to a certain type of information and/or the representation of the information
    - *Sensory modality*

  - **Multimedia**
    - *Medium* is the instrument for storing or communicating information

A TV show is a *medium* that uses *auditory* and *visual modalities*.

Multimodal Systems

UNIVERSITÀ di VERONA

Multimodal systems are "systems that support a user communicating with an application by using different modalities such as voice (in a human language), gesture, handwriting, typing, audio-visual speech, etc."

W3C Multimodal Interaction Working Group, Multimodal Interaction Requirements, W3C NOTE 8 January 2003 http://www.w3.org/TR/mmi-reqs/W3C is The World Wide Web Consortium

Multimodal Systems

A multimodal HCI system is simply one that responds to inputs in more than ***one modality*** *or* ***communication channel*** *(e.g., speech, gesture, writing, and others).*

Jaimes, A., Sebe, N., 2007. Multimodal human–computer interaction: A survey. Computer Vision and Image Understanding, 108 : 116–134.

Multimodal Systems

UNIVERSITÀ
di VERONA

- The user typically provides input in one or more modalities, and receives output in one or more modalities.
- **Input** may be classified as
  - **Sequential:** input is received on a single modality, though that modality can change over time.
  - **Simultaneous**: input is received on multiple modalities, and treated separately.
  - **Composite:** input is received on multiple modalities at the same time and treated as a single, integrated "composite" input.
- The **output** generated by a multimodal system can take various forms, e.g., audio, visual, haptic feedback, lighting, …

Example Multimodal Systems

- Observing the user(s) and gathering, analyzing, and integrating information from several modalities;
- Building internal representations of the user(s), e.g., in terms of cognitive and/or emotional states (e.g., goals, beliefs, mood, emotion, …).
    - A virtual therapist system tracks a user's facial expressions, tone of voice, and body language to assess their emotional state. It builds an internal model to understand their stress levels and mental health, tailoring responses and suggestions accordingly.

Example Multimodal Systems

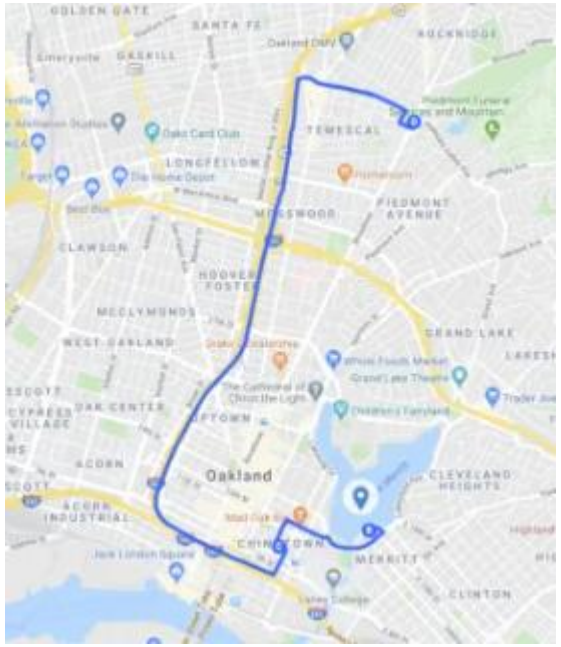
UNIVERSITÀ di VERONA

- Generating real-time multimedia feedback for the user(s), based on the analysis of the input, the internal models, and the tasks at hand;
  - A smart fitness coach app analyses a user's exercise performance through video and sensor data. It provides real-time feedback through audio prompts, visual cues on the screen, and haptic vibrations to correct form and enhance performance.
- Providing the user(s) with a multimodal interface to the machine for the execution of complex tasks needing natural interfaces.
  - A smart home assistant allows users to control their home environment using voice commands, touchscreens, and gestures. Users can adjust lighting, set thermostats, and control entertainment systems through a natural, integrated interface.

Motivations for Multimodality

1) Human-human communication is multimodal: Unimodal communication is an artifact of communication technology

2) Input and output by the most effective means



- Certain kinds of content are most easily expressed in specific non-verbal modalities.
- But some information is better suited to verbal communication, e.g., “the left bank of the river.”
  - Ex: drawing the borders of a region.

Motivations for Multimodality

3) Adapting to the environment

- Multimodal interfaces enable rapid adaptation to changing environments, by switching to the most suitable modality or by complementing different modalities.
- Adaptation to changes in the physical environment: e.g., adaptation to ambient noise, to darkness/brightness, …
- Adaptation to changes in the social environment: single user vs. multiple users, social interaction, interfaces for collaborative applications, …

Motivations for Multimodality

UNIVERSITÀ di VERONA

4) Task performance and user preference

- Many empirical studies showed that multimodal interfaces improve task performance and are preferred by users over unimodal interfaces.
- Ex: Clear advantages over unimodal speech for map-based tasks (Oviatt, 1996);
  - Faster than GUI for map-based tasks (Cohen et al., 1998);
  - Faster than GUI for drawing applications (Nishimoto et al., 1995);
  - User preference for speech and gesture in object manipulation tasks (Hauptmann, 1989).

Motivations for Multimodality

UNIVERSITÀ di VERONA

5) Error handling

- Mode switching: use of an alternate modality to escape error spirals in a modality (e.g., speech and pointing).
- Cross-modal compensation: use of information from a modality to compensate errors in another modality (e.g., compensation of audio information with visual information in user's localization in Tangible Acoustic Interfaces).
- Multimodal confirmation: use of information from a modality for confirming the results obtained by analysis of data from another modality.

Ten Myths of Multimodal Interaction
(Oviatt, 1999)

UNIVERSITÀ di VERONA

1. If you build a multimodal system, users will interact multimodally.
    - Users tend to intermix unimodal and multimodal interactions.
2. Speech and pointing are the dominant multimodal integration patterns.
    - Modalities that transmit **written input**, manual **gesturing**, and **facial expressions** can generate symbolic information that is more richly expressive than simple object selection.
3. Multimodal input involves simultaneous signals.
    - Multimodal signals often do not co-occur temporally; much of multimodal interaction involves **sequential use of modalities**.

Ten Myths of Multimodal Interaction
(Oviatt, 1999)

4. Speech is the primary input mode in any multimodal system that includes it.
    - Speech is neither the exclusive carrier of important content nor does it have temporal precedence over other modalities. These can convey information that is not present in the speech signal, e.g., spatial information and manner of action.
5. Multimodal language does not differ linguistically from unimodal language.
    - Multimodal language is briefer, syntactically simpler, and less disfluent than users' unimodal speech.

Ten Myths of Multimodal Interaction
(Oviatt, 1999)

UNIVERSITÀ di VERONA

6. Multimodal integration involves redundancy of content between modes.
    - **Complementarity** of content may be more significant in multimodal systems than **redundancy**.

7. Individual error-prone recognition technologies combine multimodally to produce even greater unreliability.
    - In an appropriately flexible multimodal interface, people determine how to use the available input modalities most effectively; mutual disambiguation of signals may contribute to a higher level of robustness.

Ten Myths of Multimodal Interaction
(Oviatt, 1999)

UNIVERSITÀ di VERONA

8. All users' multimodal commands are integrated in a uniform way.
    - When users interact multimodally, there actually can be individual differences in integration patterns. Systems should **adapt** to a user's **dominant integration pattern**.
9. Different input modes are capable of transmitting comparable content.
    - Modalities differ in the type of information they transmit, their functionality during communication, the way they are integrated with other modes, and in their basic suitability to be incorporated into different interface styles.

Ten Myths of Multimodal Interaction
(Oviatt, 1999)

UNIVERSITÀ di VERONA

10. Enhanced efficiency is the main advantage of multimodal systems.
    - Their main advantages may be found in other aspects, such as **decreased errors**, **increased flexibility**, or **increased user satisfaction**.

Guidelines for multimodal systems
(Reeves et al., 2004)

UNIVERSITÀ di VERONA

- Be **consistent** – in system output, presentation, and prompts, enabling shortcuts, state switching, …
- Provide good **error prevention** and **error handling**, make functionality clear and easily discoverable.
- Multimodal systems should be designed for the **broadest range of users** and contexts of use.
    - Support the best modality or combination of modalities anticipated in changing environments (e.g., office vs. car).
- Designers should take care to address **privacy** and **security** issues in multimodal systems. E.g., provide non-speech alternatives in a public context.

Guidelines for multimodal systems
(Reeves et al., 2004)

UNIVERSITÀ di VERONA

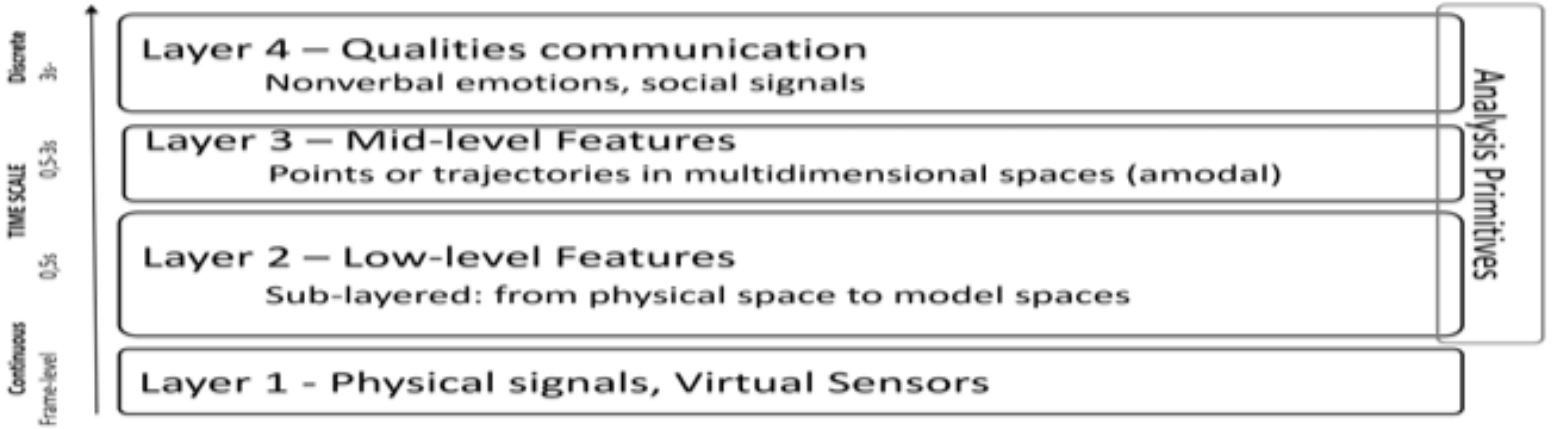
- Maximize human cognitive and physical abilities, based on an understanding of users' **human** information processing abilities and **limitations**.
- Modalities should be integrated in a manner compatible with user preferences, context, and system functionality.
    - E.g., match the output to acceptable user input styles, such as constrained grammar or unconstrained natural language.
- Multimodal interfaces should **adapt to different users**, as well as different contexts of use. Capture individual differences (e.g., age, preferences, skill, impairment) in a user profile and use it for interface settings.

Frameworks

UNIVERSITÀ di VERONA

- Frameworks and conceptual models have been proposed for multimodal systems.
  - These are not architectures, rather they are a level of **abstraction** above an architecture.
  - Indeed they do not indicate how components are allocated to hardware devices and the communication among devices.
- Frameworks have been developed for both multimodal systems focusing on verbal communication and for multimodal systems focusing on non-verbal communication.
- **We will be focusing on** nonverbal communication.

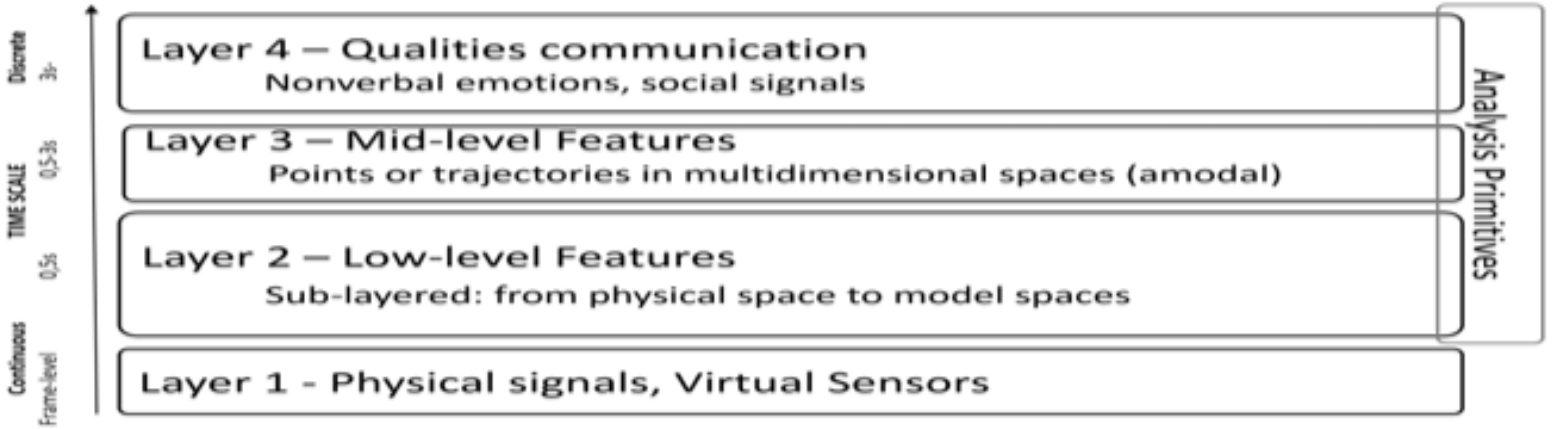
Example Framework



Layer 1: RGB-D sensor such as Kinect, providing 3D trajectories of specific body parts, the silhouette of the tracked bodies, and captured depth image.

Camurri, A., Volpe, G., Piana, S., Mancini, M., Niewiadomski, R., Ferrari, N., Canepa, C., 2016. The Dancer in the Eye: Towards a Multi-Layered Computational Framework of Qualities in Movement. In Proceedings of MOCO.

Example Framework

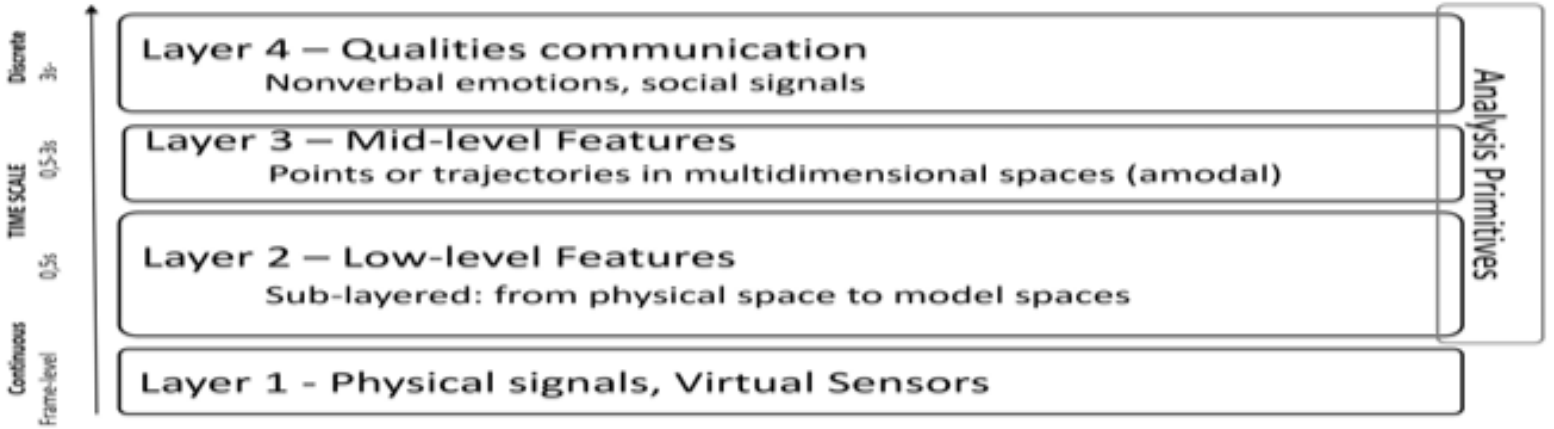


Layer 2: Receives raw data and extracts a collection of features e.g., characterizing movement locally such as velocity, acceleration, and jerk.

Layer 3: When layer 2 analysis is local in time, layer 3 deals with structure aspects of the data such as it computes features describing one single movement unit. E.g., calculating suddenness, which is defined as a rapid change in the velocity.

Camurri, A., Volpe, G., Piana, S., Mancini, M., Niewiadomski, R., Ferrari, N., Canepa, C., 2016. The Dancer in the Eye: Towards a Multi-Layered Computational Framework of Qualities in Movement. In Proceedings of MOCO.

Example Framework



Layer 4: Using the features extracted in layer 3 with e.g., a machine learning method for decision making such as detecting the expressive qualities, emotions, internal stages, etc.

Instead of machine learning, there are also methods relying on rule-based decision-making and/or correlation analysis.

Camurri, A., Volpe, G., Piana, S., Mancini, M., Niewiadomski, R., Ferrari, N., Canepa, C., 2016. The Dancer in the Eye: Towards a Multi-Layered Computational Framework of Qualities in Movement. In Proceedings of MOCO.

MULTIMODAL SYSTEMS- MODALITIES

UNIVERSITÀ di VERONA

- Natural language – spoken, written
- Visual – images, videos, RGB, infrared (thermal), depth…
- Audio – voice, sound, music
- Biological signals – Electroencephalogram (EEG), Electrocardiogram (ECG) and Galvanic Skin Response (GSR)
- Haptics – touch
- Motion capture system's data
- …..

MULTIMODAL SYSTEMS- MODALITIES

UNIVERSITÀ di VERONA

The whole body communicates by using:

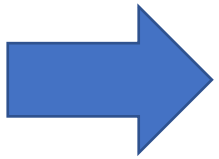
- Prosody, intonation (nonverbal vocal signals)
- Non-speech sounds (e.g., laughter, cry)
- Gesture (hand and arm movements, expressivity of the movements)
- Facial action (e.g., smile, frown, skin color)
- Gaze (e.g., eyes and head movements)
- Body orientation and posture (trunk and leg movements),
- Proxemics, i.e., distance
- Actions (fight, flight, withdraw)
- Touch (pressure, movement, position)

Other cues (specific sensors may be needed):

- Pupillary dilation
- Respiration (sound, rhythm, air volume)
- Heart rate, pulse
- Temperature
- Electrodermal response, perspiration
- Muscle action activation

Multimodal Communication & Behaviors

- Verbal Communication



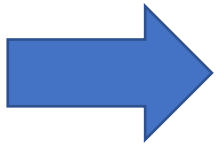
What you say?

- The words
- How you phrase them? Syntax

Multimodal Communication & Behaviors

- Verbal Communication

- Vocal Nonverbal Communication



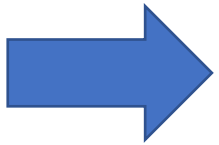
How you say?

- Prosody
  - Intonations
  - Voice quality
- Vocal expressions
  - Laughter
  - Pause fillers “hmm”

Multimodal Communication & Behaviors

- Verbal Communication

- Vocal Nonverbal Communication



How you say?

- Prosody
  - Intonations
  - Voice quality
- Vocal expressions
  - Laughter
  - Pause fillers “hmm”

VERBAL / NONVERBAL COMMUNICATION

1. Fighting
2. Discussing
3. Blaming each other
4. ?……

Image inspired from: Prof. Alessandro Vinciarelli, The University of Glasgow

VERBAL / NONVERBAL COMMUNICATION

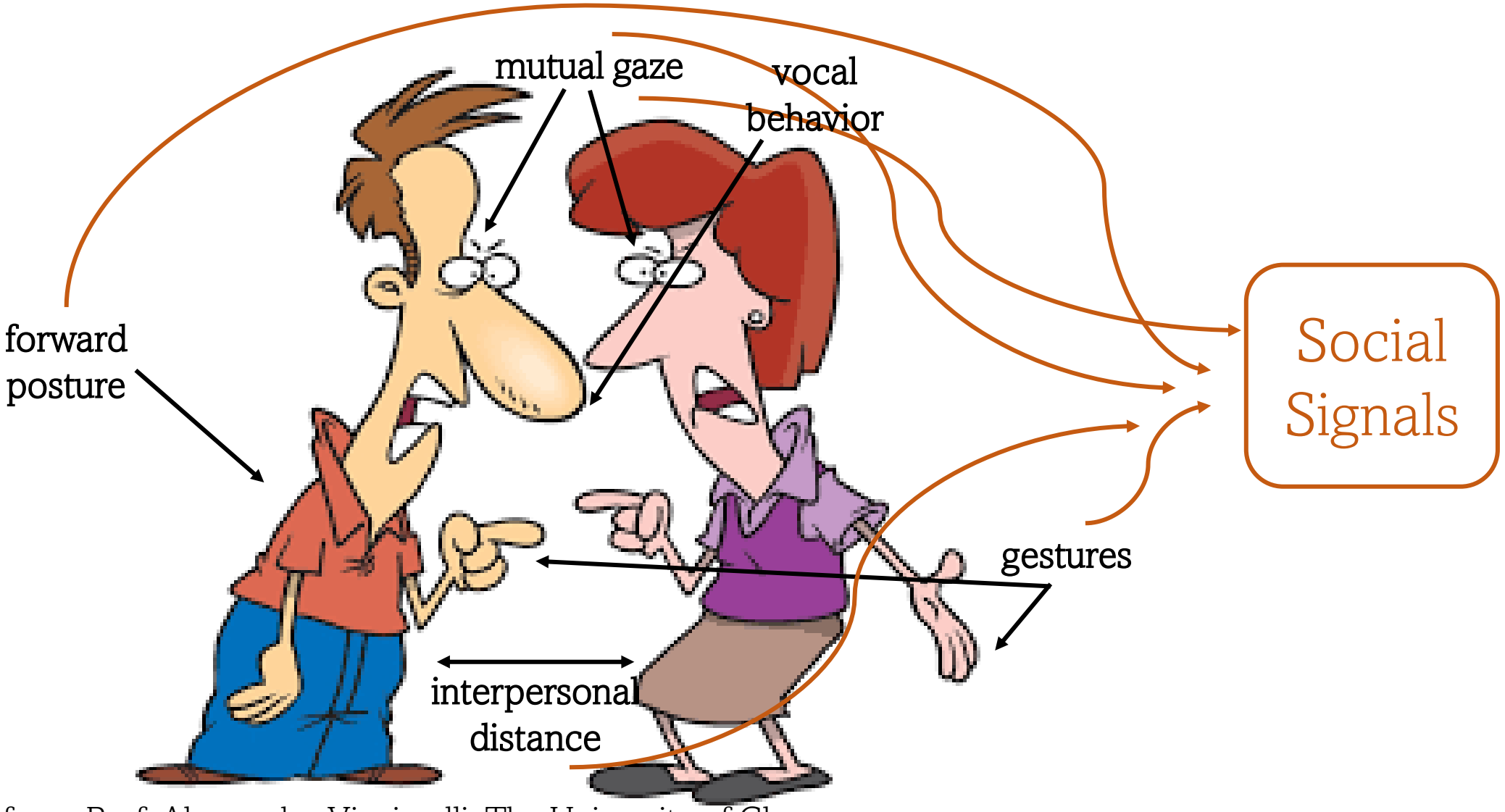
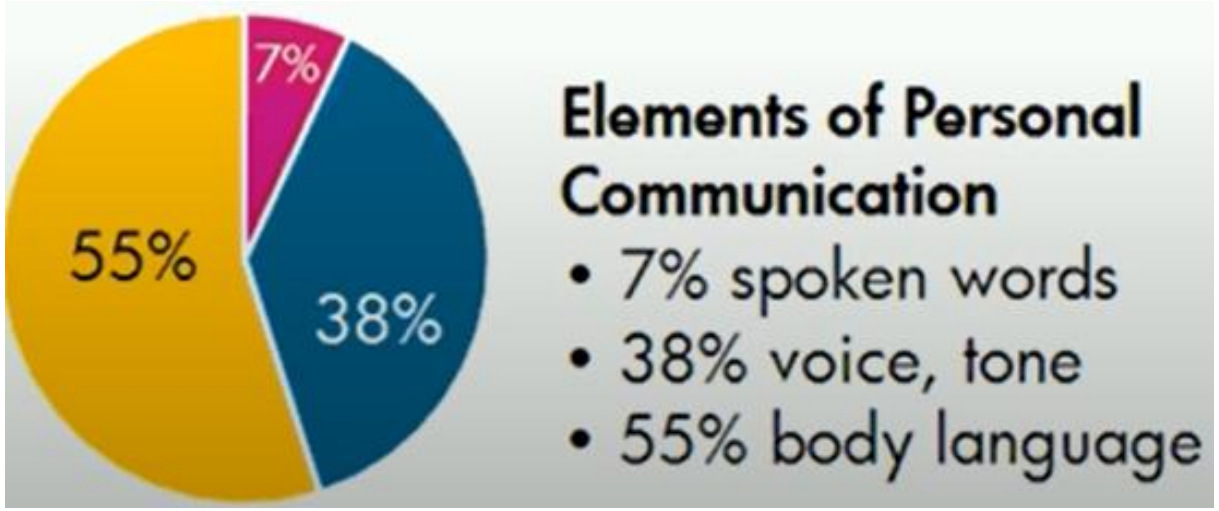


Image inspired from: Prof. Alessandro Vinciarelli, The University of Glasgow

Verbal/ Nonverbal Communication

- *Albert Mehrabian* identified 3 elements of communication:
  - Spoken words, voice/tone, and body language.
  - All 3 parts must support each other.
  - 93% of communication is NOT carried by words.



Why nonverbal communication?

- To study human (human behavior understanding) ➔ Psychology
  - so that we can build human-friendly multimedia interaction systems ➔ **Computer Science**
- To build "natural" human-computer interaction, e.g., with robots ➔ **Computer Science**

NONVERBAL COMMUNICATION

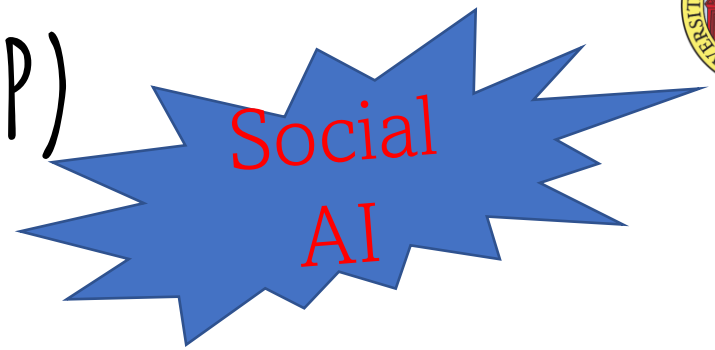
- Nonverbal behavioural cues are the **physical**, **machine-detectable** traces of **social and psychological phenomena**.
- Nonverbal communication occurs with every encounter, mostly **subconsciously**.
- What nonverbal behaviour may communicate?
  - Emotions
  - Personality
  - Social attitudes
  - Interpersonal relations
  - Attention, engagement, flow
  - Other internal cognitive/physiological states, e.g., tiredness, stress….
  - Psychological diseases and well-being, e.g., depression
  - Cultural/social background
  - ….

Nonverbal Behavioral Cues

They are the physical, machine detectable traces of *social and psychological phenomena*.

Body Activity	Eye Gaze and Visual Focus of Attention	Facial Expressions	Vocal Behavior	Physical Appearance	Proxemics
-Body Orientation -Self-touching -Hand behind head	-Mutual Gaze -Fast blink -Gaze dynamic	-Smiling -Lip-pout -Tense-mouth -Facial action units	-Pitch - Loudness - Speaking rate - Turn taking	-Clothes -Make up -Gender -Age	-Distance -Velocity -Direction of Flow -Seating

Social Signal Processing (SSP)



An interdisciplinary domain at the crossroad between technology and human sciences (psychology, sociology, etc.).

SSP research integrates social psychology concepts into Artificial Intelligence and focuses on automatic detection, interpretation and synthesis of social signals represented in terms of nonverbal behavioral cues (gestures, eye gaze, body posture, postures, vocal characteristics) capable to convey socially-relevant information.

Social Signals

- Produced during interactions that:
  - play a part in the information and adjustment of relations and interactions between agents (human and artificial)
  - provide information about the agents

Pantic et al. 2011

- Communicative or informative signal that conveys information about social actions, social interactions, emotions, attitudes, and relationships.

Poggi & D'Errico, 2010

- Are events in which actually or virtually present agents exchange an array of communicative and informative signals performed by one agent in relation to one or more other agents.

Pantic & Vinciarelli

Why Social signal processing?

The ability to recognize human social signals and social behaviors are essential when building social robots, human-robot interaction, or interactive systems.

3 main problems:
Modeling: identification of the principles and laws
Analysis: automatic detection and interpretation
Synthesis: automatic generation of artificial social signals

What is NOT social signal processing?

Is this emotion recognition?

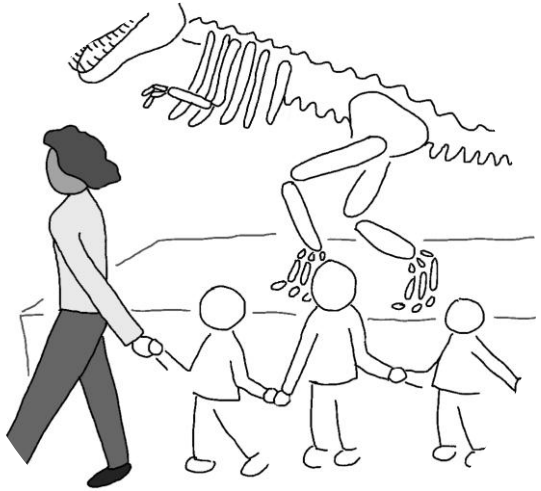
- The capacity of inferring affective/emotional states is a basic attribute for socially aware systems (intelligent systems)
  - An alternative/parallel inference channel focused on nonverbal behavior.
- Shown that there is a subconscious communication channel that conveys information about the attitude, regulates roles, and is used to take decisions.
- The channel is active even when the person in the interaction is NOT experiencing any EMOTIONAL STATE!!!

What is social signal processing?

- Consists of non-linguistic, largely unconscious signals about the social situation.
- It communicates social relation rather than the emotion.
- It happens over longer time frames than typical linguistic phenomena or emotional displays.

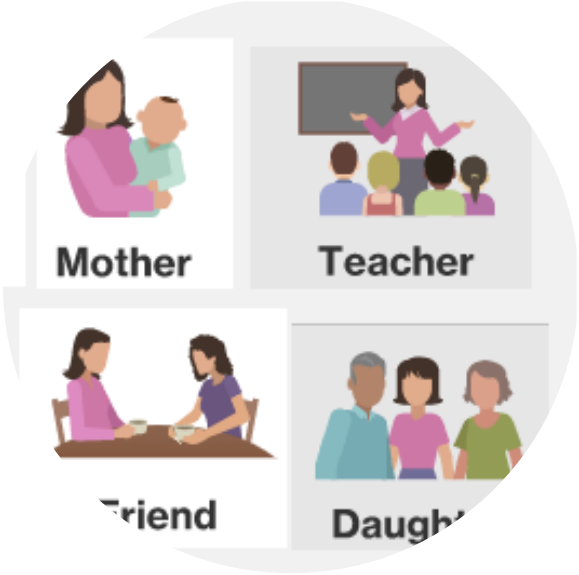
SSP Research Topics

- Measuring group efficacy during teamwork, brainstorming
- Identifying roles (e.g., leadership) during teamwork
- Analyzing visitor behavior during museum visit
- Discover behavior patters in spectator crowds
- Monitor social interaction during parties
- Measuring the effectiveness of a politicians



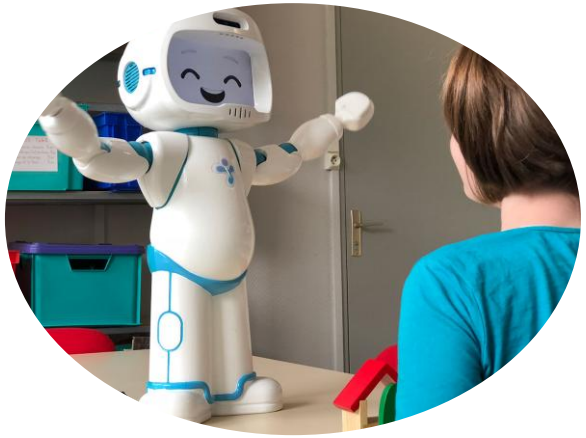
SSP Research Topics

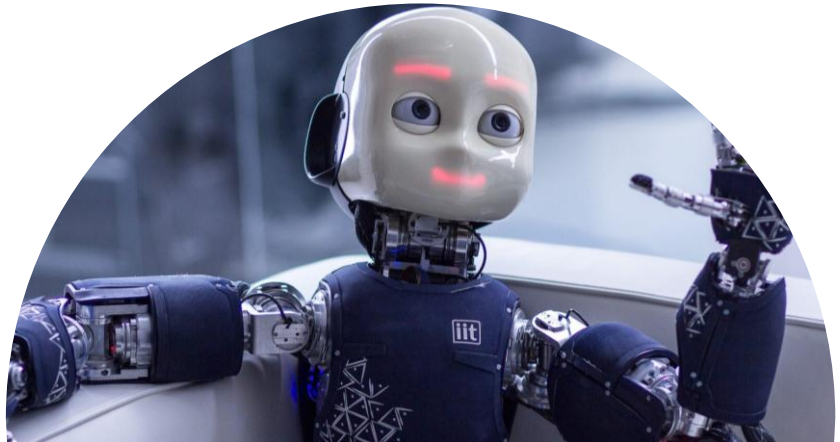
- Detection of personality traits
- Identification of emergent leaders
- Dominance, competence, liking and hirability detection
- Social roles detection (protagonist, follower, supporter…)
- Social relations detection (birthday party: birthday person, parents, friends, guests)
- Engagement detection



SSP Research Topics

- Social Robots
- Virtual reality
- Health – therapy, e.g., autism, training of social skills (e.g., mediator)
- Education and e-learning, e.g., online virtual classes
- Surveillance and prevention
- Social robotics and other embodied interfaces
- Entertainment





Affective Computing

- .. is the study and development of systems and devices that can *recognize*, *interpret*, *process*, and *simulate* human affects.

  Rosalind Picard

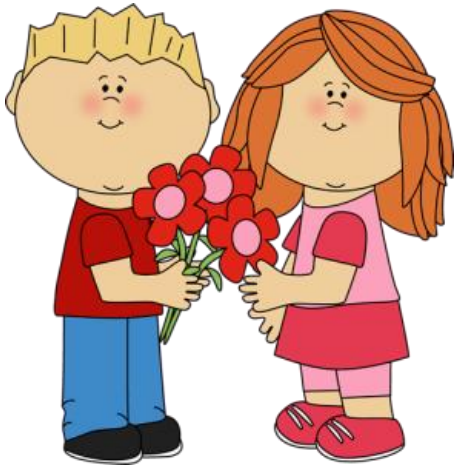
- **Interdisciplinary field of research**

- " ..computing that relates to arises from or deliberately influences emotion or other affective phenomena…"

- How can emotions be generated in computers, be recognized by computers, and be expressed by computers?"

Affective Computing- motivations

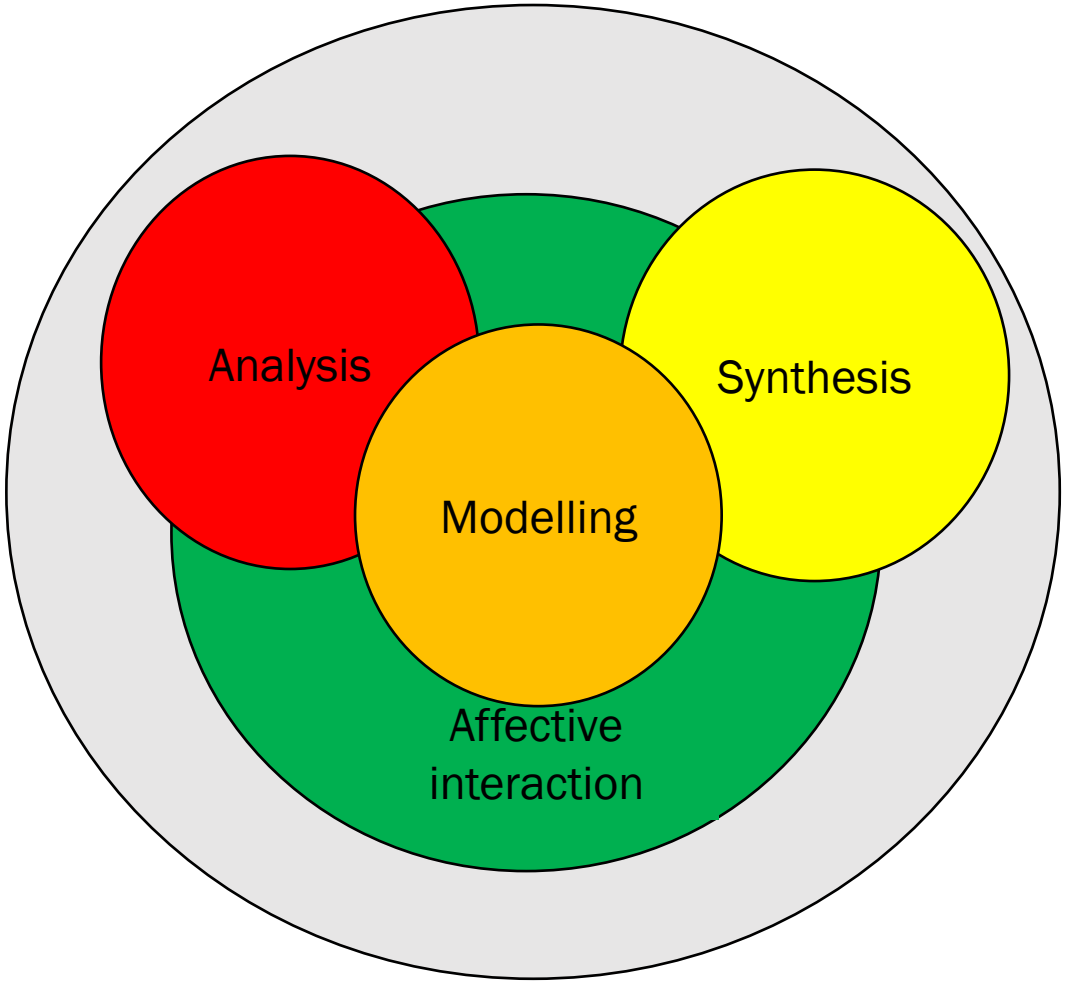
People “see” emotions everywhere…



Affective Computing (Ac)

Four main directions of the research:

- to analyze, detect/recognize affective states from observable symptoms,
- to model affect, emotion, sentiment,
- to communicate affective states by means of (multi)-modal signals,
- to build (interactive) affective applications.



AC Research topics

Emotion recognition

- Voice e.g., depression detection
- Text e.g., hate speech detection
- Sentiment analysis
- Facial expressions
- Physiology e.g., detecting stress from skin conductance
- Touch

Synthesis

- Emotional speech
- Emotional facial expressions
- Affect-induced behavioral variation in robots and agents
- Multisensory mapping e.g., emotions expressed by colors or sounds

AC Research topics

Modeling

- modeling emotional impact on decision making
- modeling emotion elicitation, e.g., appraisals
- emotion intensity

The ethics of “giving” machines emotional capabilities

(Interactive) applications

- health – therapy, e.g., positive computing
- games/entertainment, e.g., frustration detection
- Education and e-learning, e.g., boredom detection

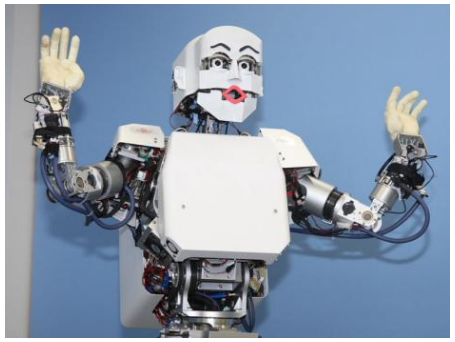
AC APPLICATIONS

Affective Agents and robots
- emotionally intelligent agents and robots

More effective & realistic interaction
- user modeling
- affect-adaptive interfaces and affective HCI

Serious games

Therapy
- applications for autistic children
- VR Treatment Environments (e.g., self-help anxiety management)



Kobian, 2009

Arts & Entertainment
- interactive (affective) games
- affect-based music retrieval

E-learning

Assistance and Caregiving
- empathic agents

Economics and Marketing
- user profiling and content choice

Well-being and Positive Psychology:
- emotion, moods, and personality traits monitoring

SSP & AC in Practice (1)

- Measuring **social distancing** during COVID19. Implemented and tested in COOP grocery shops in Genoa.
- The automatic estimation of the inter-personal distance from an image, and the characterization of related people aggregations,
- Measuring social distance (SD) is not only a geometrical problem, but it also implies a deeper understanding of the social behavior in the scene.

- Scene Geometry Understanding
- People detection/ pose estimation
  - 2D people detection
  - 2D and 3D pose estimation
- Interaction analysis
- SD estimated in a single frame as the interpersonal distance between people

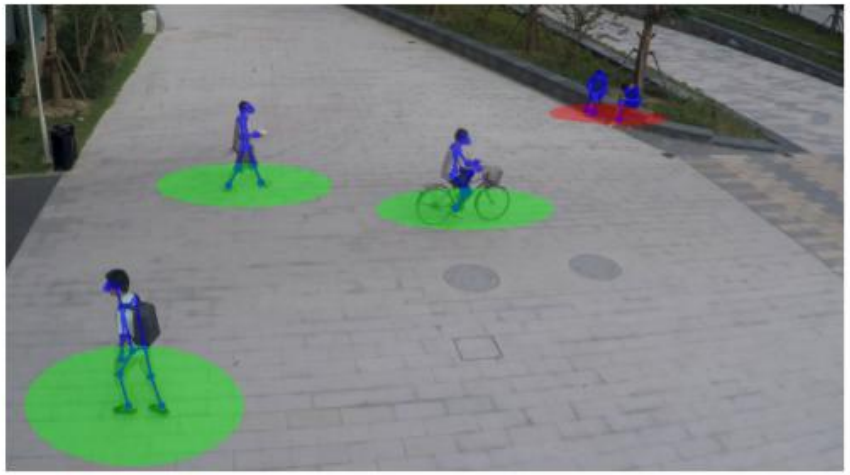
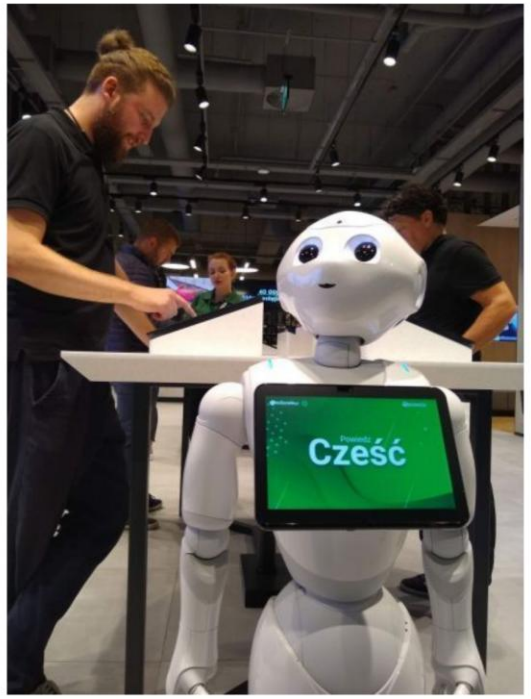


Image and text:
M. Cristani et al., (2020), The Visual Social Distancing Problem, IEEE Access

SSP & AC in Practice (2)

- **Social Robots:** to study social and affective interaction with robots
  - recognition/synthesis/induction

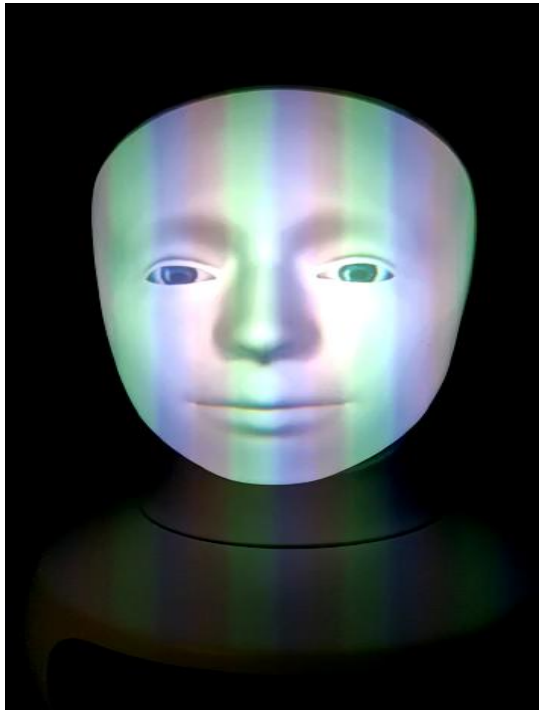
- commercial products e.g., Pepper (SoftBank/Aldebaran)
  - able to recognize emotions from human facial expressions and voice prosody
  - can react by showing e.g., empathy
  - no facial expressions, but expressive movements, posture, voice modulation
  - used in supermarkets…



POLAND Humanoid robot Pepper has joined the Eobuwie.pl team in its Stary Browar store in Poznań. Similar robots are to be introduced to the company's stores in Wrocław and Warsaw.

The **robot sales assistant** debuted in June in the Eobuwie.pl store in Forum Gdańsk. "Our company is technology focused, which inspires us and allows us to take on even the most ambitious projects. That is why we decided to check out how a humanoid robot would cope as a sales assistant and whether it would meet the customer service standards of our stores," explains Marcin Grzymkowski, the CEO of Eobuwie.pl. "You can collect your order from it and it explains what to do step by step, which makes the shopping procedure clearer for the customer. Additionally, he will talk about our store's concept and our original Esize.me service. Pepper's presence guarantees easy shopping. Children engage and play with him and parents can shop undisturbed and take advantage of the full services available in the store," adds Konrad Jezierski, the director of

SSP & AC in Practice (3)



FurHat Robot
video by D. Pasquali

Different embodiments…

Shujie Zhou et al., 2020

Cigdem Beyan
cigdem.beyan@univr.it
https://cbeyan.github.io/