HCI- Multimodal Systems

Modelling Nonverbal Behaviors

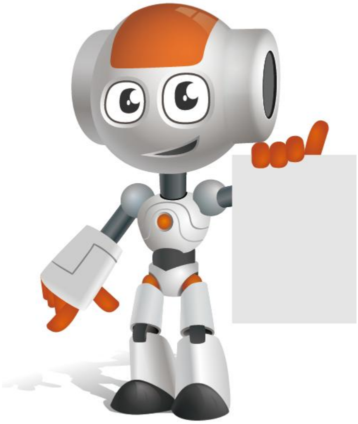
Cigdem Beyan

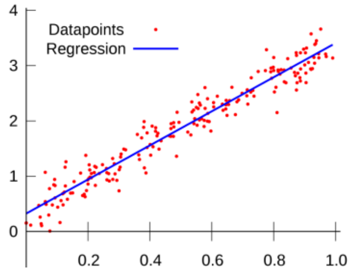
A. Y. 2025/2026

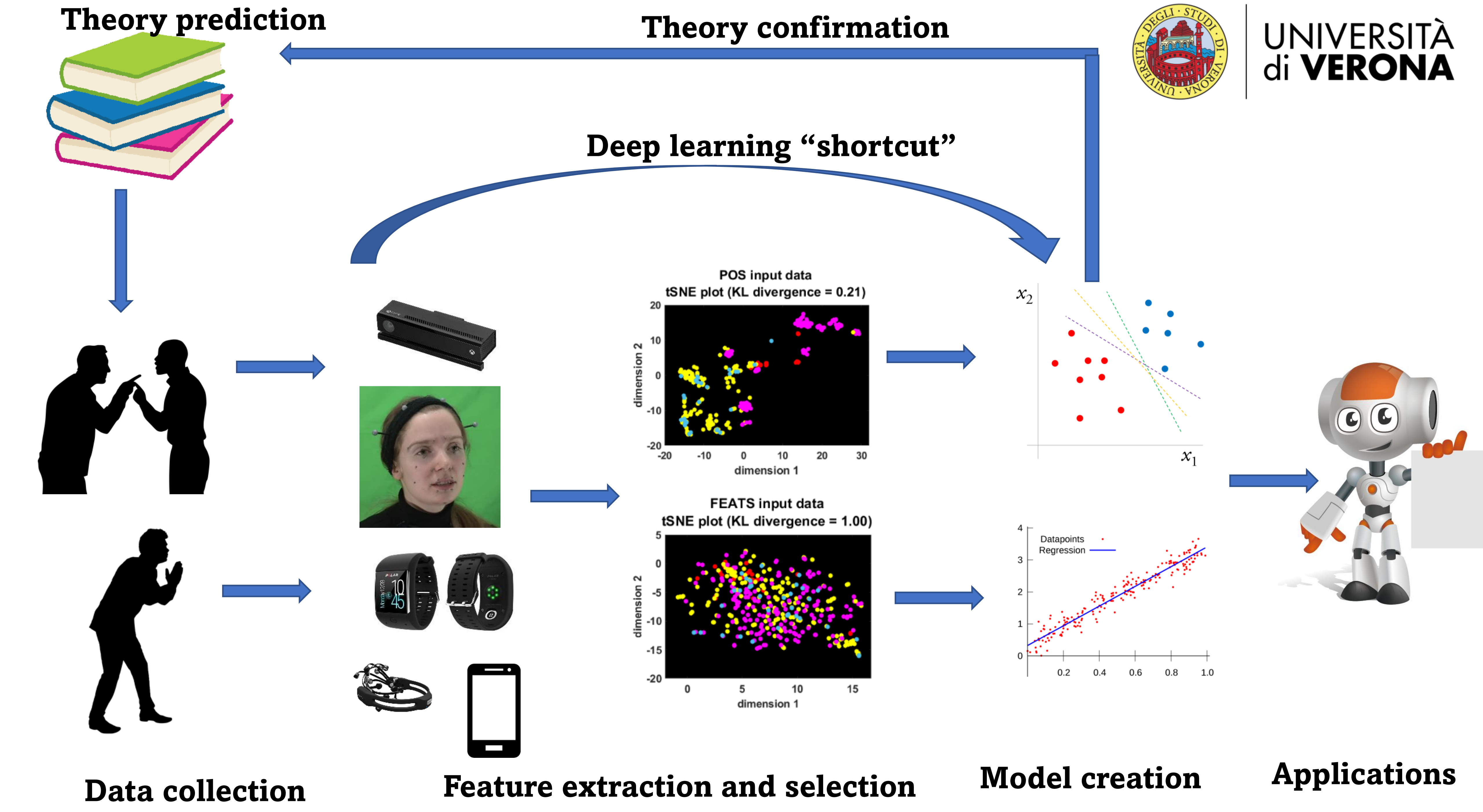
Today

- The pipeline of nonverbal behaviour analysis
  - From data acquisition and annotation to automatic recognition/synthesis
- More about nonverbal features/cues
  - Frequently used features (and sensors)
    - Popular coding schemes
    - How to relate them to emotions
- A brief introduction to (some of) emotion theories

Modeling nonverbal behaviors







Theory prediction
Theory confirmation
UNIVERSITÀ di VERONA
Deep learning "shortcut"
POS input data
tSNE plot (KL divergence = 0.21)
dimension 2
dimension 1
FEATS input data
tSNE plot (KL divergence = 1.00)
dimension 2
dimension 1
x_2
x_1
Datapoints
Regression
Data collection
Feature extraction and selection
Model creation
Applications

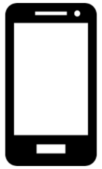
Theory prediction (literature review)

UNIVERSITÀ di VERONA

- Are there predictions (assumptions, theories, etc.) regarding a particular social and affective phenomenon?
  - Can we use these theories to design an experiment and collect data?
- Are there any observable nonverbal behavioural (NB) cues described in the literature?
  - What kind of data should we collect? Which modalities should we focus on?
  - Are there specific NBs described that can be detected?
  - Are there any available standardized datasets?

Data Collection

- **Setting:** In the wild (ecological setting) and/or in lab conditions
- **Sensors:** Number and type of sensors, privacy, intrusiveness, sensor synchronization
- **Ethical & Privacy Issues:**
    - Is the procedure ethically acceptable? (Consult Local Ethical Committee)
    - Are “fragile” groups involved? (e.g., children)
- **Consent Form:**
    - Which data will be collected and for what purpose
    - Is the data anonymous or not?
    - Who has access to the data?
    - Each participant needs to agree and sign it
- **How much data is enough?**

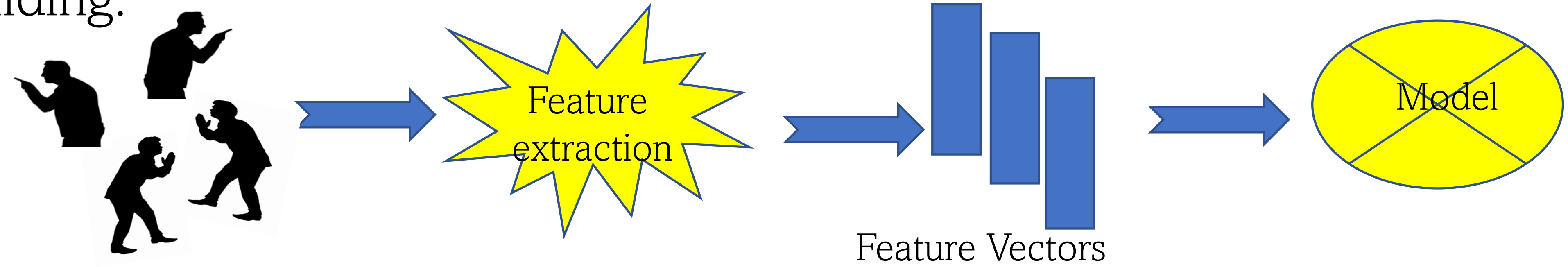


Data Collection - Annotation

- No annotations (**labels**), **no learning**!
- You can perform **clustering** if you do not have labels.
- **Self-Reports:** before session, after session, during session
- **External Observers:** Annotation by external observers
- **Psychological Questionnaires:** Inline with the theories used to build the experimental setting
    - E.g., for personality traits detection: using **Big Five Inventory Questionnaires:** Measures the five major dimensions of personality (openness, conscientiousness, extraversion, agreeableness, and neuroticism).

Feature Extraction & Selection

- Low-level features = **Nonverbal Cues**
- **Several libraries** to compute low-level features
  - We will be seeing some of the libraries
  - This stage can be skipped if "deep learning" is applied
    - Aka data-driven modelling
- Often the extracted features (so-called candidates-features) are too many and some need to be eliminated
  - Feature selection (you can use any strategy of machine learning)
- Output: the data represented by feature vectors, which is to be used for model building.

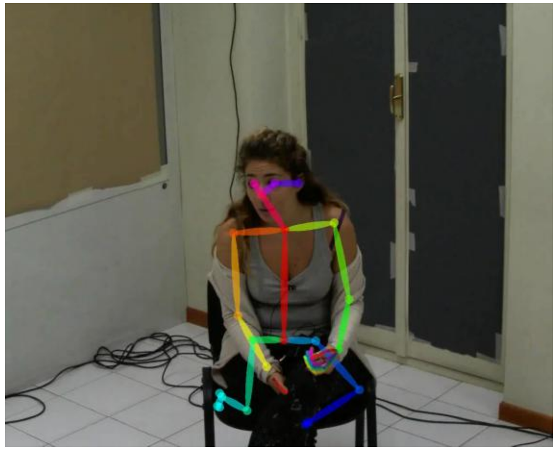


Example Nonverbal Features

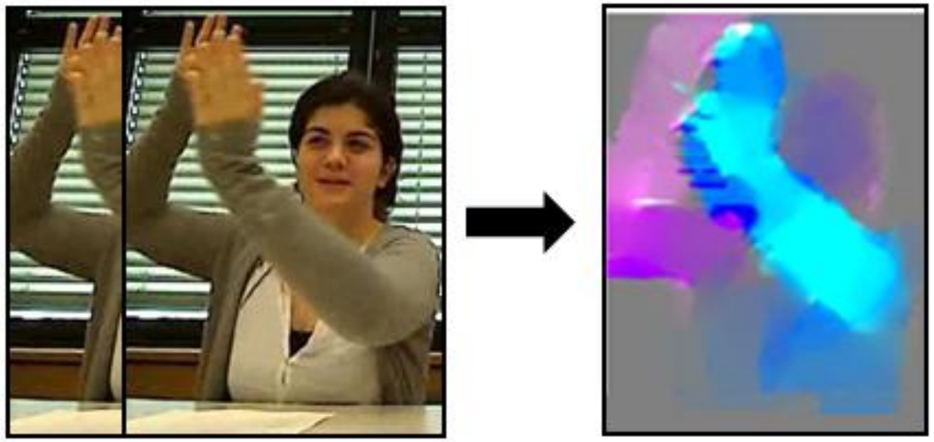
Body Activity	Eye Gaze and Visual Focus of Attention	Facial Expressions	Vocal Behavior	Physical Appearance	Proxemics
-Body Orientation -Self-touching -Hand behind head	-Mutual Gaze -Fast blink -Gaze dynamic	-Smiling -Lip-pout -Tense-mouth -Facial action units	-Pitch - Loudness - Speaking rate - Turn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

Body Activity

- Refers to postures as well as movements of upper body, arms and hands, includes gesture detection.
- Requires detecting human and/or their body parts (e.g., arms, upper body).



Body Posture



Optical Flow Images

Face-Touch Detection

Eye Gaze & Visual Focus of Attention

- Is a point in space where a person is looking at (possibly corresponding to an object).

Chong et al., Detecting Attended Visual Targets in Video, CVPR 2020
Code: https://github.com/ejcgt/attention-target-detection

Eye Gaze & Visual Focus of Attention

- Can be detected by estimating a) gaze direction……

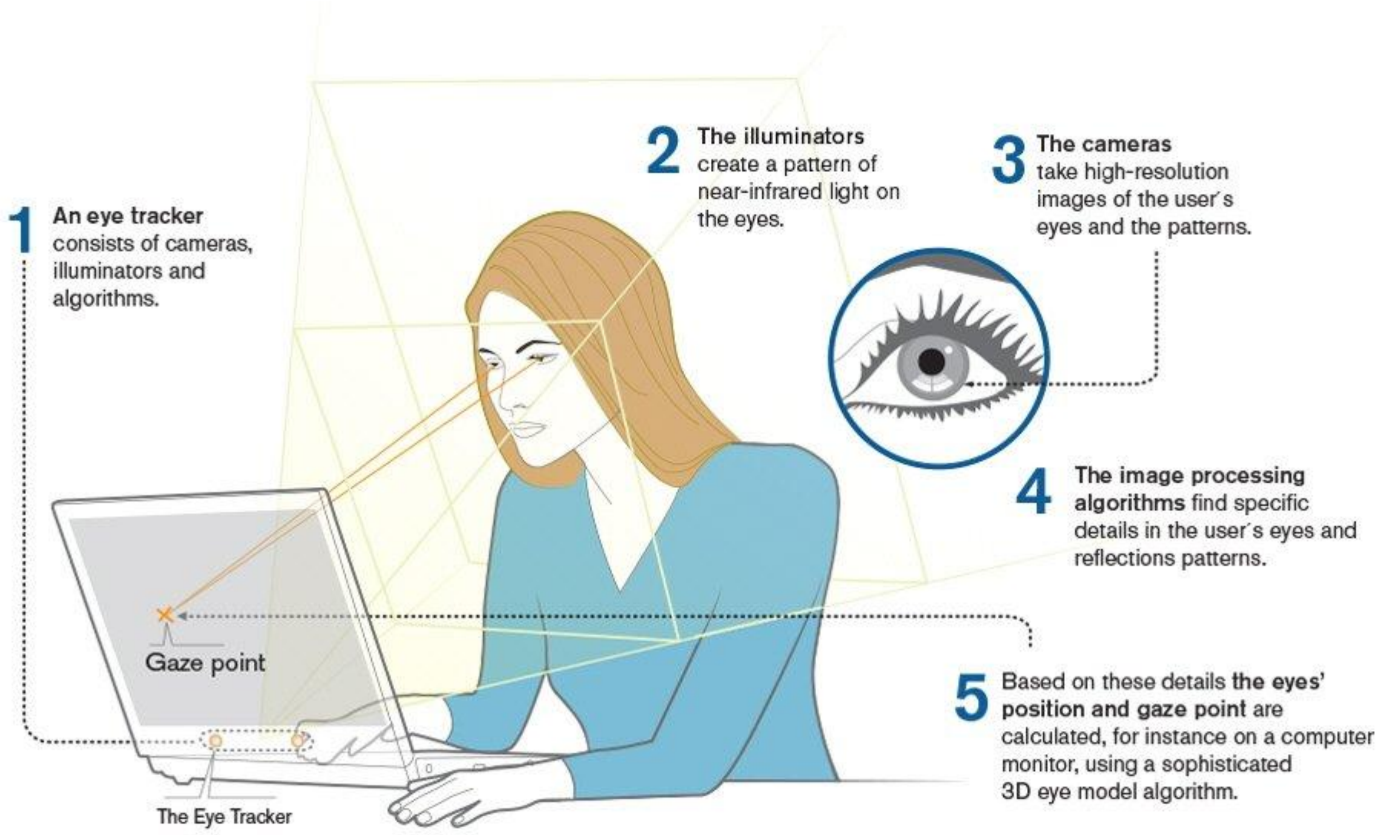
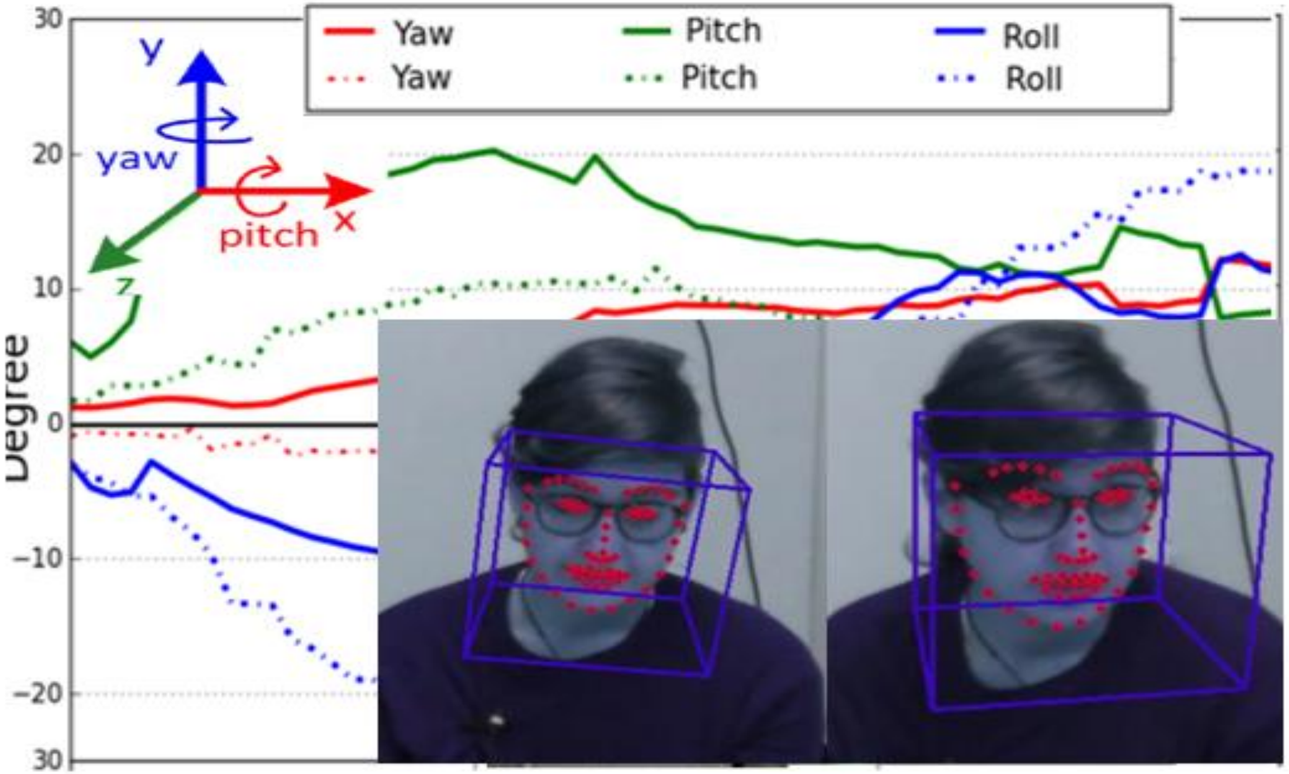


Image credit: TOBII eye tracker

Eye Gaze & Visual Focus of Attention

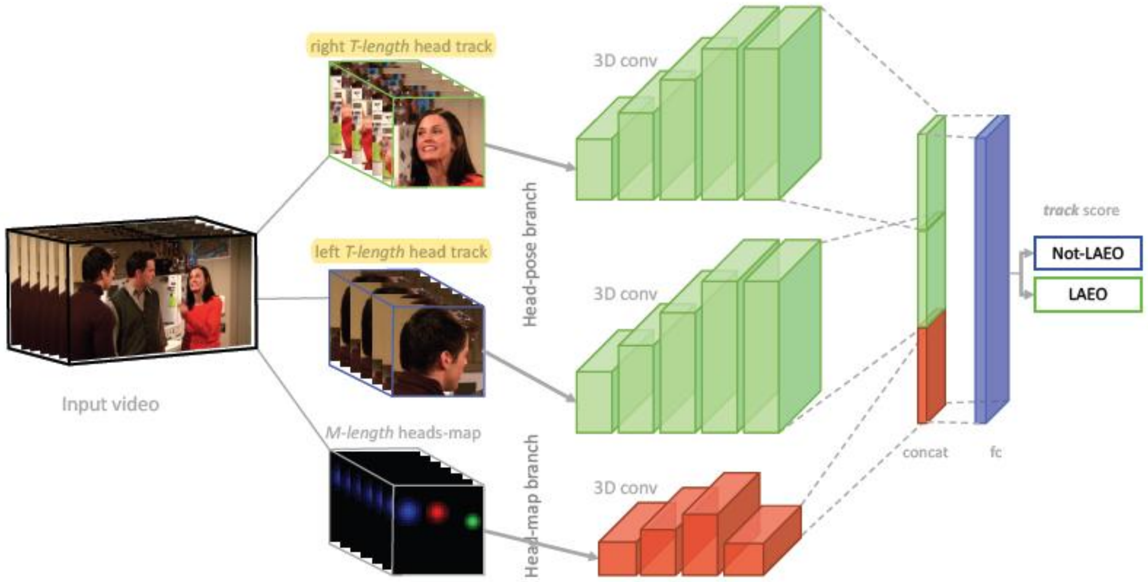
- Can be detected by estimating a) gaze direction, b) head pose……..



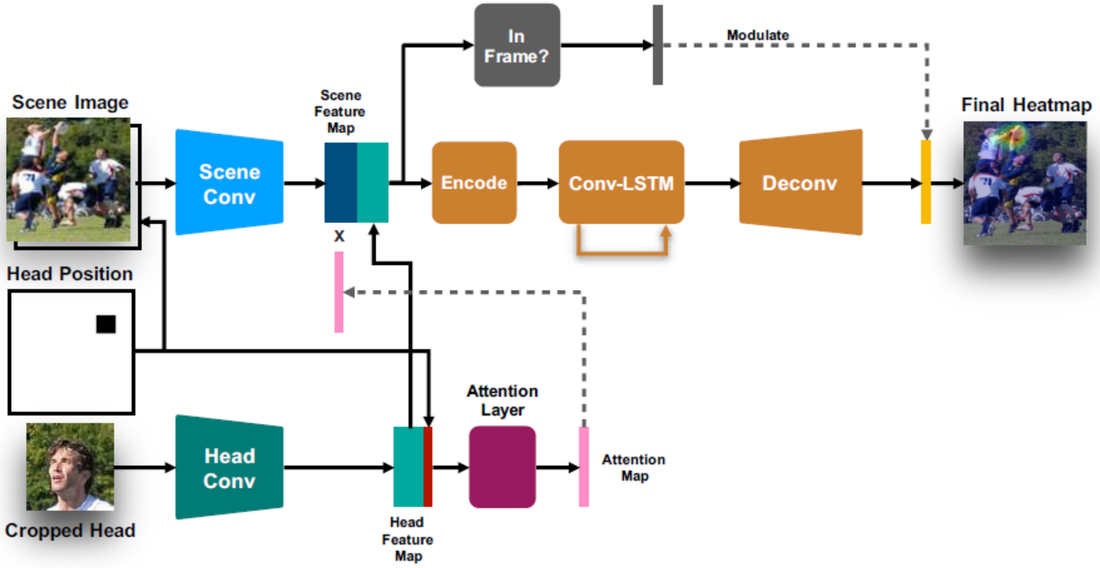
Baltrusaitis et al. OpenFace: An open source facial behavior analysis toolkit.
https://github.com/TadasBaltrusaitis/OpenFace

Eye Gaze & Visual Focus of Attention

- Can be detected by estimating a) gaze direction, b) head pose, and more recently c) directly from images.



Manuel J. Mar´ın-Jimenex et al. LAEO-Net++, 2020.



Chong et al., Detecting Attended Visual Targets in Video, CVPR 2020.

Facial Expressions

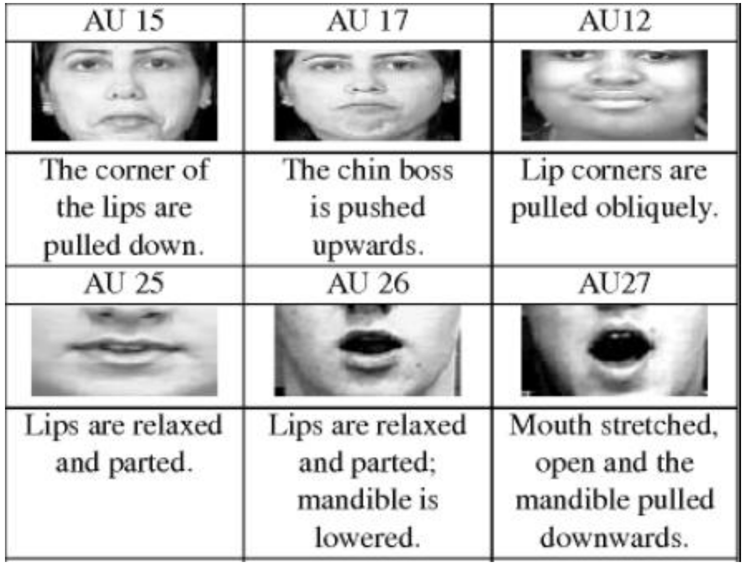
UNIVERSITÀ di VERONA

- Faces are important source of information, and frequently used in affective computing.
- Few SSP works relied on facial expressions, such as to detect roles.
- Detecting facial landmarks (facial key points) and recognizing facial action units…

Facial landmarks from 300W Dataset

Facial Expressions

- The Facial Action Coding System by Paul Ekman Group, is a comprehensive, anatomically based system for describing all visually discernible facial movement.
- It breaks down facial expressions into individual components of muscle movement, called Action Units (AUs).



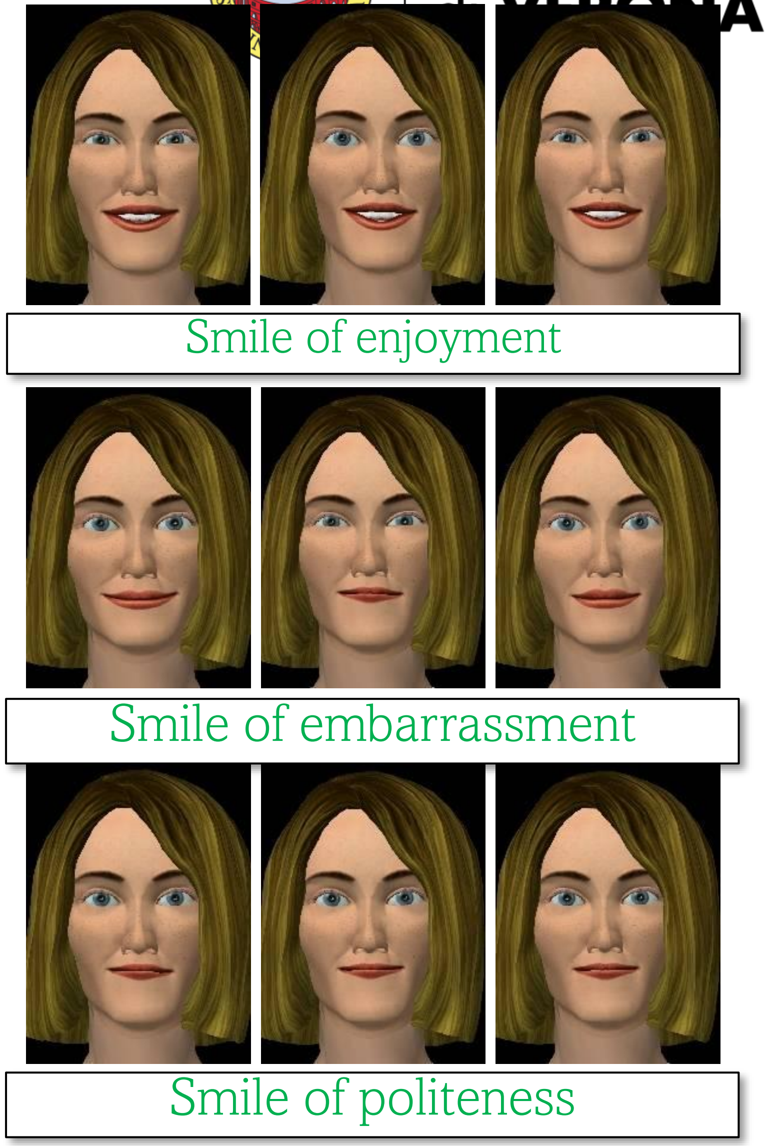


Image credit: Ochs et al., Cognitive processing 2012

Vocal Behavior

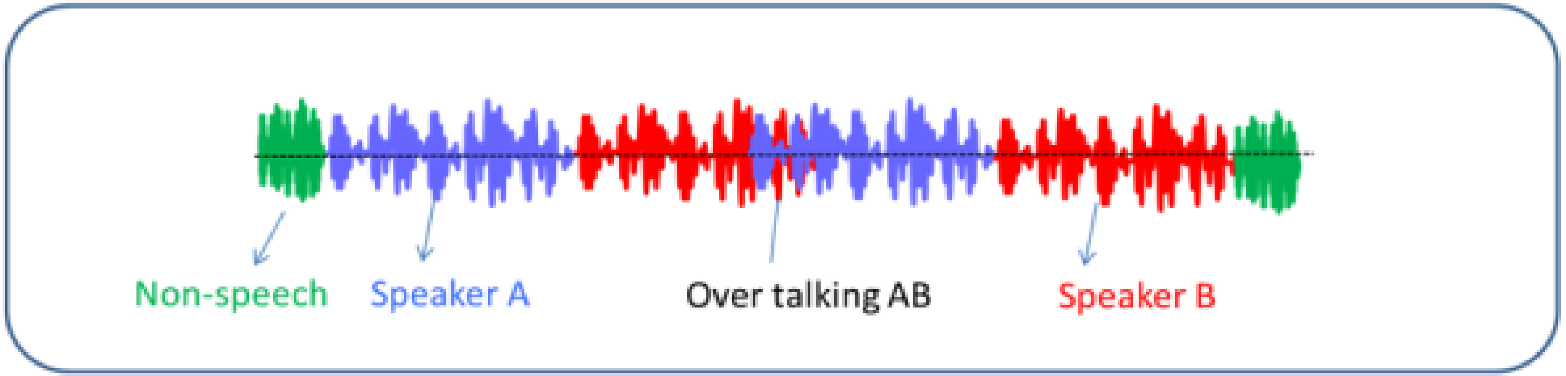
Speaking Activity

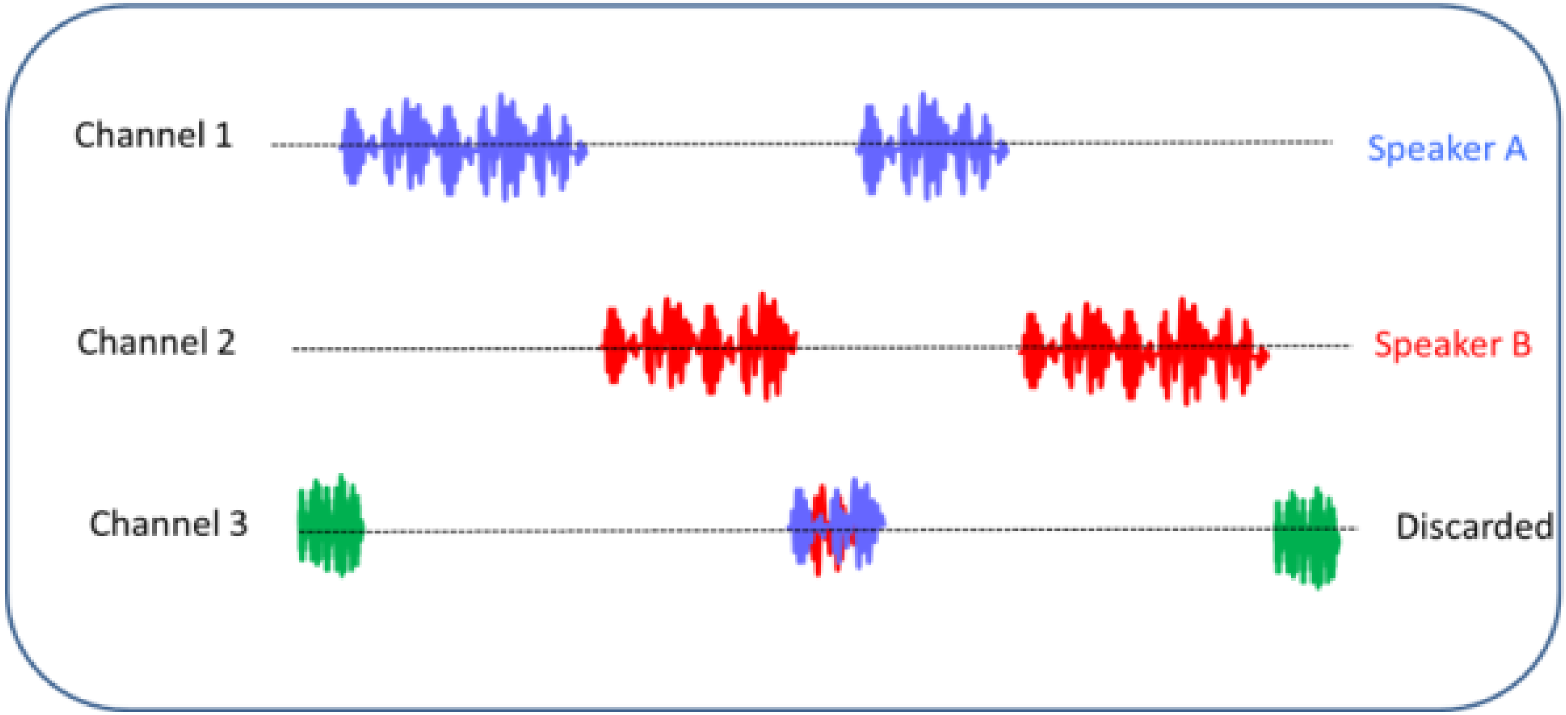
Prosody

Vocal Behavior

Speaking Activity

Speaker Diarization

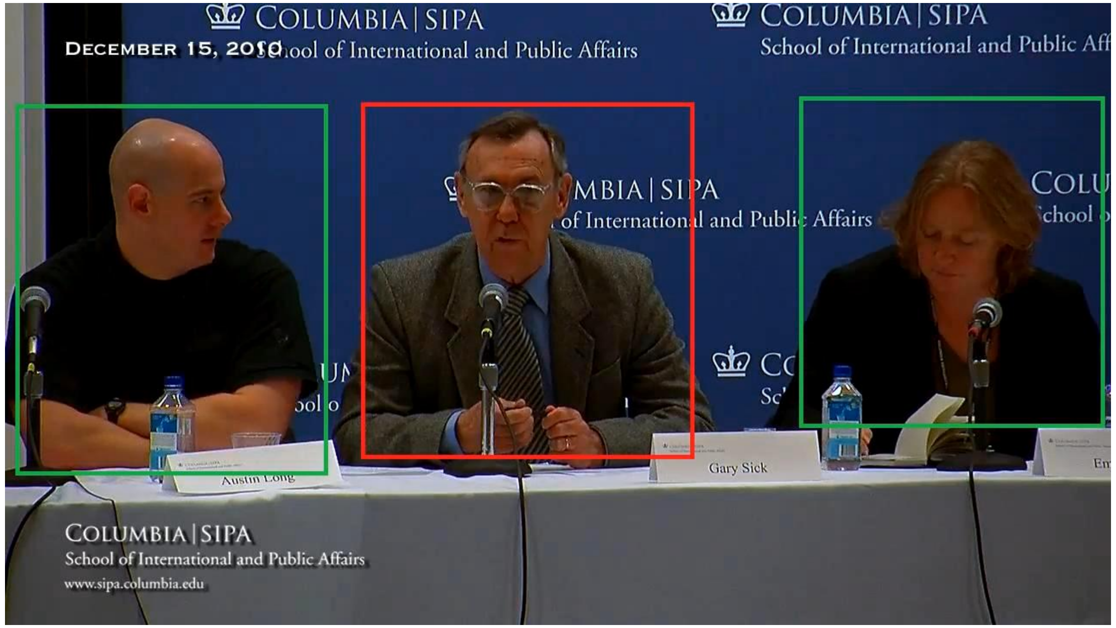




Vocal Behavior

Speaking Activity

- statistical properties of turns (e.g., total length, mean, maximum, fraction….),
- speaking time length,
- overlapping speech time,
- turn-taking order,
- number of successful / unsuccessful interruptions,
- speaker floor grabs,
- fraction of time non-overlapping speech accounts for
- ……



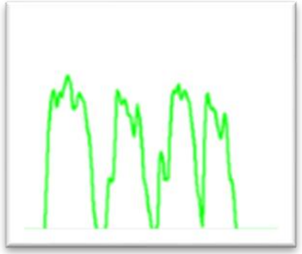
Red box indicates speaking
Green boxes indicate non-speaking

Muhammad et al., ICCV 2020 & WACV 2021,
Beyan et al., Trans. Multimedia 2021.

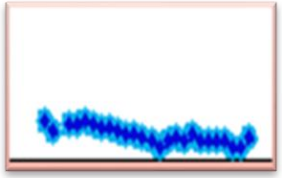
Vocal Behavior

Prosody

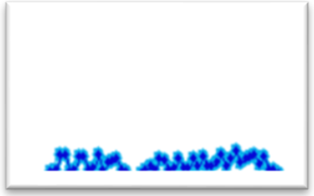
Loudness (the energy)

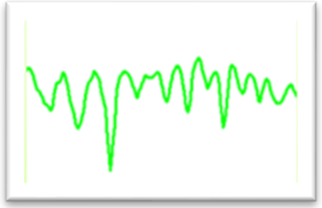


Intonation (the fundamental frequency)

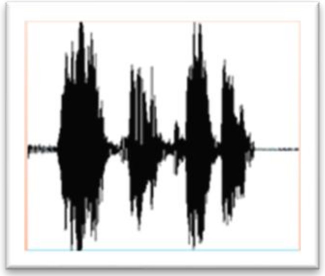


Vocal variability (jitter, shimmer)





Speaking rhythm (the utterance timing)



Vocal Behavior

Prosody

- statistical of signal energy
- fundamental frequency (variation, maximum, mean)
- spectral features (formants, bandwidths, spectrum intensity)
- Speaking rate (e.g., number of syllables per second)
- **Mel-Frequency** Cepstral Coefficients
- **Spectrograms** (w/ deep learning)
- ....

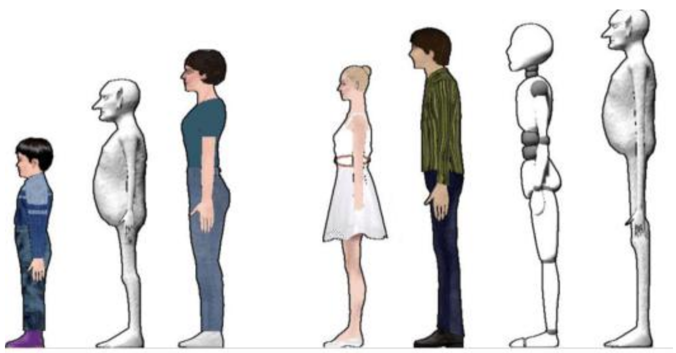
openSMILE
audio feature extraction

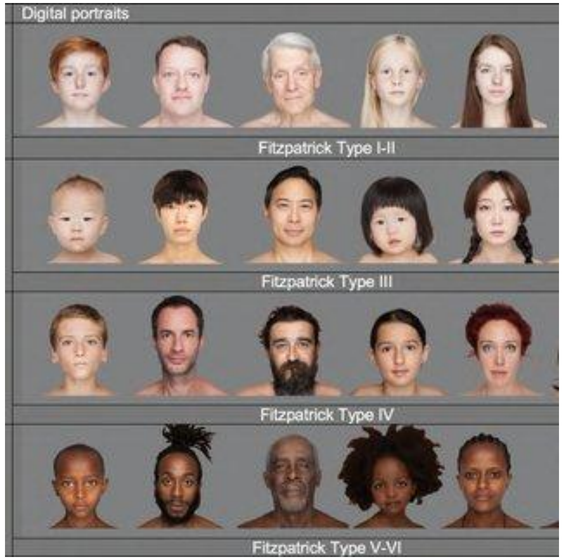
release v3.0.0 | release date october 2020 | downloads 4.9k | docs 3.0

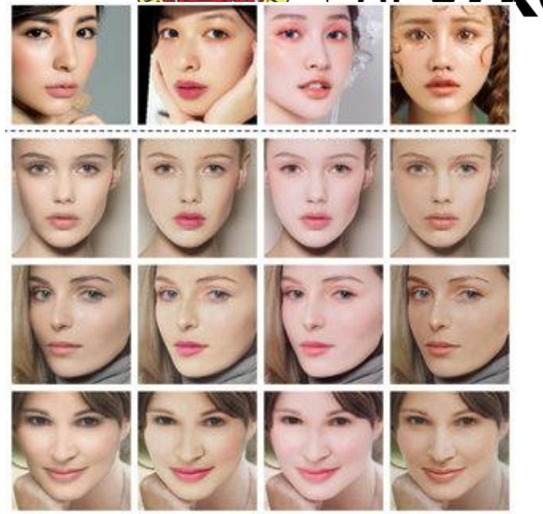
...gy/amplitude related LLDs	Group
...m of auditory spectrum (loudness)	Prosodic
...g. HNR, shimmer (local)	Voice quality
14 spectral LLDs	Group
Alpha ratio (50–1000 Hz/1–5 kHz)	Spectral
Hammarberg index	Spectral
MFCCs 1–4	Cepstral
Formants 1, 2, 3 (rel. energy)	Voice quality
Harmonic difference H1-H2, H1-A3	Voice quality
Spectral flux	Spectral
Spectral slope (0–500 Hz, 0–1 kHz)	Spectral
...frequency related LLDs	Group
F_0 (linear and semi tone)	Prosodic
Jitter (local), Formant 1 (bandwidth)	Voice quality
...rmants 1, 2, 3 (frequency)	Voice quality
...mants 2, 3 (bandwidth)	Voice qu...

Eyben et al., The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for Voice Research and Affective Computing, 2016

Physical Appearance







height, body shape, skin / hair color, clothes, make up…

- Physical attractiveness, sympathy, and appreciation.
- Social relation detection, hirability (job interviews).
- Feature learning from RGB images!!!

Molla et al., 2017, Egocentric Mapping of Body Surface Constraints
Yan & Suk, 2020, Skin Color Perception in Portrait Image and AR-based Humanoid Emoji
Huang et al., 2020, Real-World Automatic Makeup via Identity Preservation Makeup Net
Ahed J Abugabah et al., 2020, Learning Context-Aware Outfit Recommendation

Proxemics

UNIVERSITÀ di VERONA

- People's use of physical space.
- It studies the social meaning of interpersonal distances, especially when the movement of people is not constrained and, therefore, social, cultural, and psychological phenomena are the only factors underlying the physical distance between two individuals.

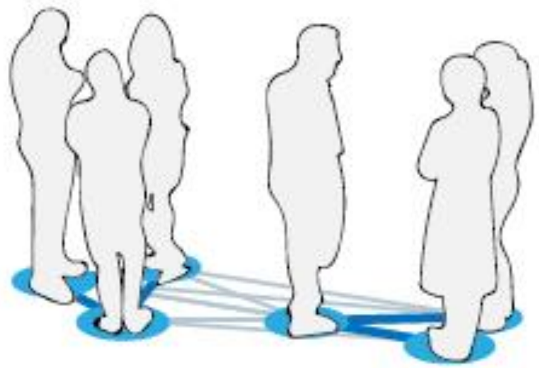
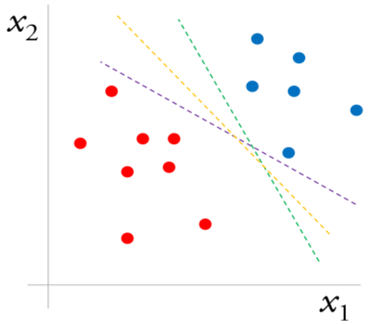


Image Credit: Swofford et al., DANTE, ACM HCI, 2020.

Model Creation

- **Recognition/Classification/Detection**
  - simple heuristics or rule-based system
  - Machine/deep learning



- **Main types of models**
  - **Classification:** a mapping from the input variables to discrete/categorical output variable
    - to find the decision boundary that divides the set into classes
    - e.g., emotion labels
  - **Regression:** a mapping from the input variables to continuous output variable
    - to find the best-fit line
    - e.g., intensity
  - **Clustering** (rarely)

Model Creation

- Examples of supervised machine/deep learning methods:
  - Support Vector Machine (SVM)
  - Random Forests (RF)
  - Convolutional Neural Network (CNN)
  - Long short-term memory (LSTM-RNN)

Model Creation

- **Synthesis**
  - Direct data retargeting: capture the values and map them to a virtual model
  - Procedural approach: Theory driven, or rule-based
    - E.g., if the virtual agent is happy, it should move the corner of the mouth up (as in smile)
  - Data-driven (machine learning)
    - Model tries to learn (generalizes) from many examples of the same expression.
    - In the past HMM was very popular
    - Now, generative deep models (e.g., variational autoencoders, diffusion models, GANs)

Model Creation: multimodality and modality fusion

- Early Fusion & Late Fusion
  - Multimodal features are combined in the feature extraction level and the model is being learned through the combined features.
    - Summation, concatenation…..
  - Individual models per modality and at the decision level the predictions are merged.
    - Majority voting, weighted decisions

Model Creation: multimodality and modality fusion

- **Challenges of multimodal data:**
  - Some sensors are temporally unavailable
    - Missing data
  - Noisy data
  - Redundant data
  - One modality dominates the other
  - Need for synchronization/calibration
    - E.g., lips movement synch with audio, gestures synch with speech

Model evaluation

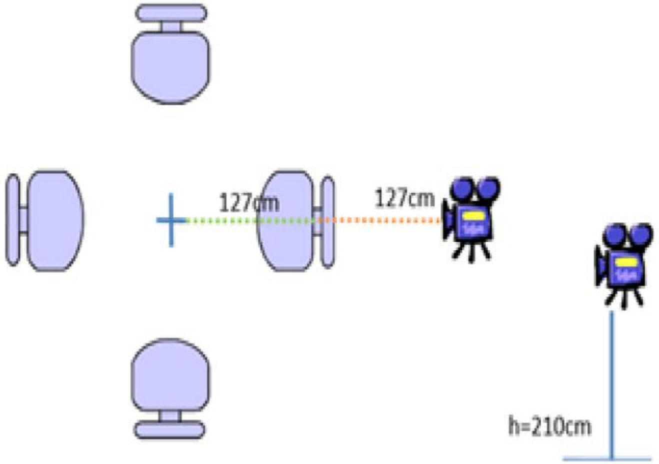
- If your model is based on classification or regression, it means you have annotations that can be used for model evaluation.
  - You can employ traditional evaluation metrics such as **accuracy**, **mean squared error**, **F1-score**, and **confusion matrix**, among others.
- For synthesis:
  - (Online) studies with a high number of naïve participants.
  - Questionnaires, e.g., Likert scale
  - Several specific questionnaires, e.g., RoSas (facial expressions, gestures, body posture, and vocal inflections)
  - Using questionnaires focusing on perception of the interaction, e.g., human-likeness, naturalness, believability

Example: Emergent Leader Detection

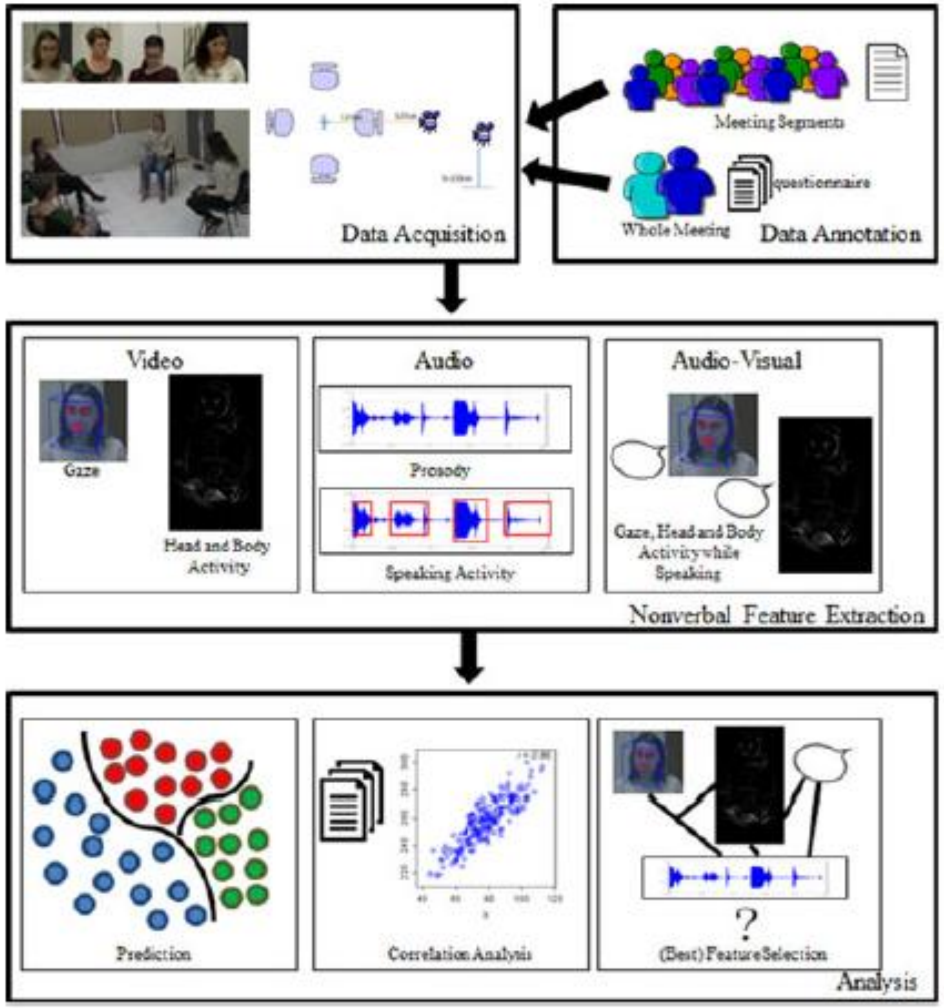
- How to detect (emergent) leaders by analysing nonverbal behaviours?
- **Theory prediction:** what is an emergent leader, which social traits it is related to (e.g., extroversion), and which nonverbal behaviors can be relevant?
- **Dataset:**
  - We need a conversational environment, in which we can test decision-making
    - a meeting scenario, a group of 4, 4 unacquainted persons
  - The group solves a "survival task": participants are presented with a dramatic survival situation (e.g., a plane crash in the desert), a single decision has to be taken by the group
  - 16 meeting sessions, up to 30 minutes each, in the wild
  - 4 cameras located behind each person, 1 camera to take overall scene, personal lapel microphones

Example: Emergent Leader Detection

- **Dataset:** Labels
  - Questionnaires self-assessment (pre-experimental)
  - Questionnaires participants evaluate others (post-experimental)
  - External observers' annotation without questionnaires
  - Classification: the most leader, the least leader, none of the two (3 classes)



Example: Emergent Leader Detection



- Feature extraction
  - **Visual:** visual focus of attention, head/body activity
  - **Audio:** speaking activity, prosody
- **Model creation:** Support Vector Machines
- **Model evaluation**
- **Classification metrics:** class accuracies
- **Correlation analysis** (Pearson's Correlation Coefficient): SYMLOG-friendliness scores of all participants and leaders only versus the corresponding NF, one per each time.

Example: Emergent

- Features

VFOA Based NFs
The total time that p_i is being watched by the other participants.
The total time that p_i is mutually looking at any other participants (mutual engagement (ME)).
The total time that p_i is being watched by any other participants while there is no ME.
The total time that p_i looks at other participants.
The total time that p_i initiates the MEs with any other participants.
For p_i the total time intercurrent between the initiation of ME with any other participants.
The total time that p_i is looking at any other participants while there is no ME.
Ratio between the TW_i and TL_i.
Head/body activity based NFs
The total time that the head/body of p_i is moving.
The total number of head/body activity turns for p_i where each turn represents a continuous head/body activity.
Average head/body activity turn duration for p_i.
Standard deviation ($std.$) of head activity in x, y dimensions and the $std.$ of body activity for p_i.

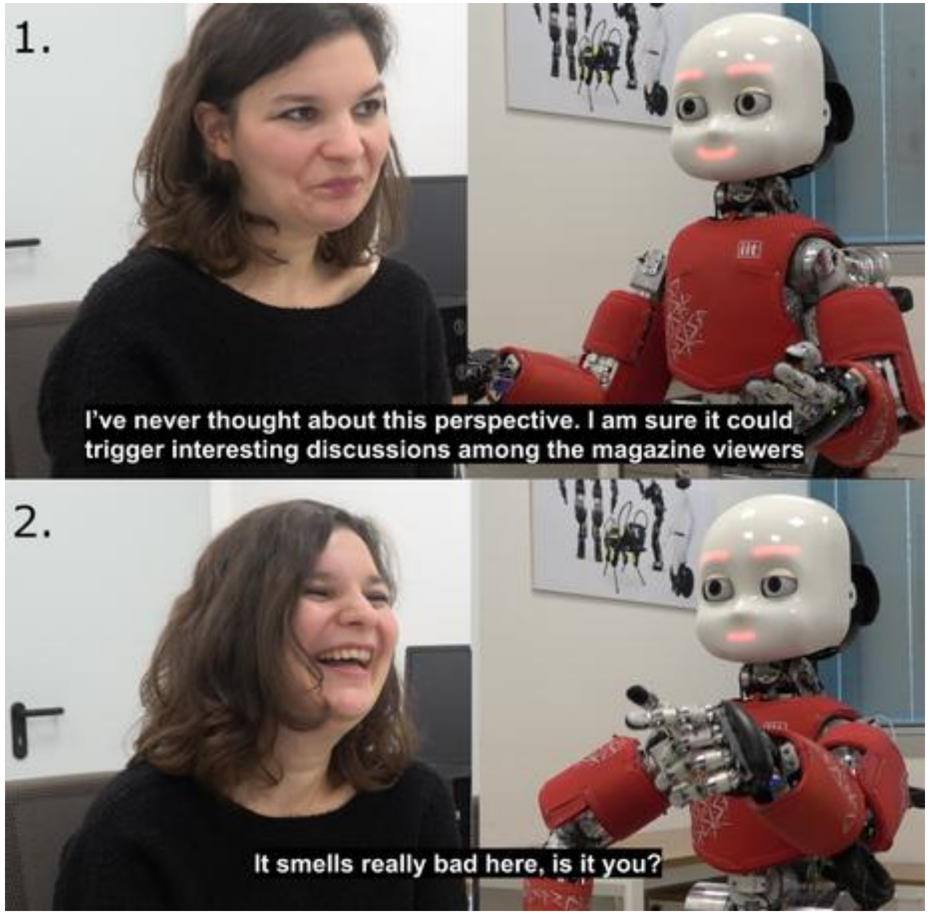
Speaking-Act based NFs
The total speaking length of p_i when at least one other participant is also speaking.
The total speaking length of p_i when nobody is speaking.
Ratio between $TMSL_i$ and $TSSL_i$.
The total number of speaking turns of p_i without utterances (a turn lasts minimum 2 seconds).
The average speaking turn durations of p_i.
The total number of successful interruptions of p_i: p_i starts talking when p_j is speaking and p_j finishes his/her turn before p_i does.
The total number of unsuccessful interruption of p_i: p_i starts talking when p_j is speaking when p_i finishes his/her turn p_j is still speaking.
Being successfully interrupted: p_j starts speaking when p_i is speaking, p_i finishes his turn while p_j is still speaking.
Being unsuccessfully interrupted: p_j starts speaking when p_i is speaking, p_j finishes his turn while p_i is still speaking.
The total time that p_i speaks first after another speaker.
Ratio between the TSI_i and BSI_i.
p_i floor grab: similar to TSI_i but if everybody in the group stops speaking.
Ratio between the total speaking length of p_i (no matter p_i speaks alone or at the same time with the others) to silence.
Ratio between TSI_i and the total number of turns p_i has.
Ratio between TUI_i and the total number of turns p_i has.

Prosodic based NFs
[Total, minimum, maximum, median, mean, standard deviation] speaking energy of p_i when there is no overlapping speech segment.
[Total, minimum, maximum, median, mean, standard deviation] speaking energy of p_i.
[Minimum, maximum, median, mean, standard deviation] pitch for p_i when there is no overlapping speech segment.
[Minimum, maximum, median, mean, standard deviation] pitch for p_i.

Beyan, Cigdem, et al. "Prediction of the leadership style of an emergent leader using audio and visual nonverbal features." *IEEE Trans. on Multimedia* 2017.

Example 2: Comfortability Analysis in HRI

- What are nonverbal cues of (low) comfortability? Is it possible to detect it automatically?
- **Theory prediction:** what is the definition of comfortability? In which scenarios do people feel uncomfortable?
- **Dataset:** Data collection in an ecological setting. humanoid robot used to elicit low comfortability:
  - The robot asks "difficult" questions and performs a set of actions to induce low/high comfortability
  - The participants' reactions are recorded
  - The participants are unaware of the purpose of the study



Lechuga et al., "Comfortability Analysis Under a Human-Robot Interaction Perspective", JCSR 2020.

Example: Comfortability Analysis in HRI

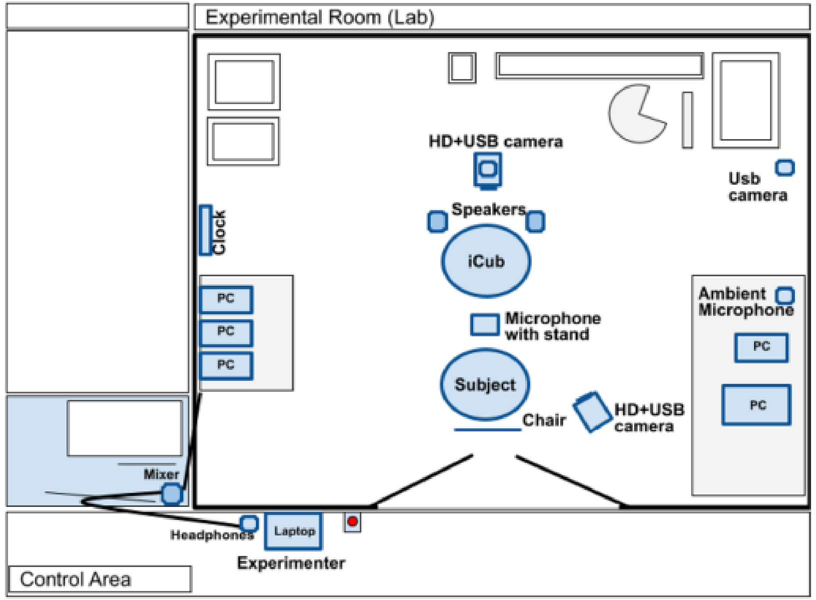
Lechuga et al., “Comfortability Analysis Under a Human-Robot Interaction Perspective”, JCSR 2020.

Example: Comfortability Analysis in HRI

UNIVERSITÀ di VERONA

- The robot's speech is entirely scripted, the timing is controlled by another person behind the scenes.
- Some scripts and behavior of the robot is to make a person feel uncomfortable, such as:
    - Interrupt when talking to ask: And what about...?
    - Ohh ... sure [WRONG Name], Whatever you say
    - iCub looks at the wall's clock several times while listening
    - [Name], you mean that you do not have the courage to pursue your dreams?
    - 20 seconds silence (solely looking at the participant's face and floor). Then say: "Buffalo buffalo Buffalo buffalo buffalo buffalo Buffalo buffalo" without any meaning behind.
    - Now that I know you more, allow me to ask: Aren't you ashamed of having those naive goals and poor results?

Example: Comfortability Analysis in HRI



- Each session is about 17-54 minutes.
- 29 participants, all engaged with a social robot before.
- Different nationalities, almost gender balance, similar age.
- **Sensors:** camera, microphone, physiological data (galvanic skin response, temperature, heart rate, and hand motion (accelerometer, gyroscope, magnescope))

Example: Comfortability Analysis in HRI

UNIVERSITÀ di VERONA

- Annotations:
  - **Personality traits and attitude towards robots:** to study whether there is a significant correlation between Comfortability and people's personality and/or attitude towards robots, the participants were asked to complete the Ten Item Personality Measure (TIPI) and the Robotic Social Attributes Scale (RoSAS) questionnaires some days in advance.
  - **Comfortability self-report:** After the interview, the participants were asked to fill up a questionnaire to report their Comfortability regarding their experience. They were asked how they had felt when the robot performed a specific action. These questions appeared in a randomized manner and were scored following a 7-point Likert scale (i.e., with 1 being Extremely Uncomfortable and 7 being Extremely Comfortable).

Example: Comfortability Analysis in HRI

UNIVERSITÀ di VERONA

Nonverbal Features: OpenFace and OpenPose packages were used.

- **Facial action units:** the mean and standard deviation of each intensity of the sixteen AUs were included for each three-second video clip
- The person's **upper body positions** of some key-points in the arm. The mean and standard deviation of these key-points coordinates were computed per each three-second clip.
- The **eye gaze direction** vector in world coordinates for each eye (i.e., the x, y and z coordinates for the left eye and the x, y, and z coordinates for the right eye), the eye gaze angle direction averaged for both eyes. The mean and standard deviation of all these features were considered per each three-second video clip
- The **location of the head** with respect to the camera in millimeters (i.e., the x, y and z coordinates; where a positive Z is being further away from the camera) and the rotation of the head in world coordinates with the camera being the origin (i.e., the x, y and z coordinates representing the pitch, yaw, and roll respectively). The mean, standard deviation, velocity, and acceleration of the head location and rotation were for each three-second video clip.

Let’s delve into some of the (popular) nonverbal cues and discuss them for affective computing

Facial expressions

- How to measure and compare two (facial) expressions?

beautiful smile

bright smile

smart smile

- How to show that two expressions by 2 different persons are “the same”?
- How can the machine classify facial expressions?
- Can we describe a facial expression in a more objective manner?

Facial Action Coding System (FACS)

- By Paul Ekman and Wallace Friesen, 1978:
- Allows expert coders to manually measure facial expressions by breaking them down into component movements of individual facial muscles (i.e., action units)
- Inspired by human anatomy
- Encodes any visible change on the face (muscle deformation, apparition of wrinkles/bulges, folds), and secondary movements (propagation of muscle contraction)
- Taxonomizes human facial movements by their appearance on the face

Facial Action Coding System (FACS)

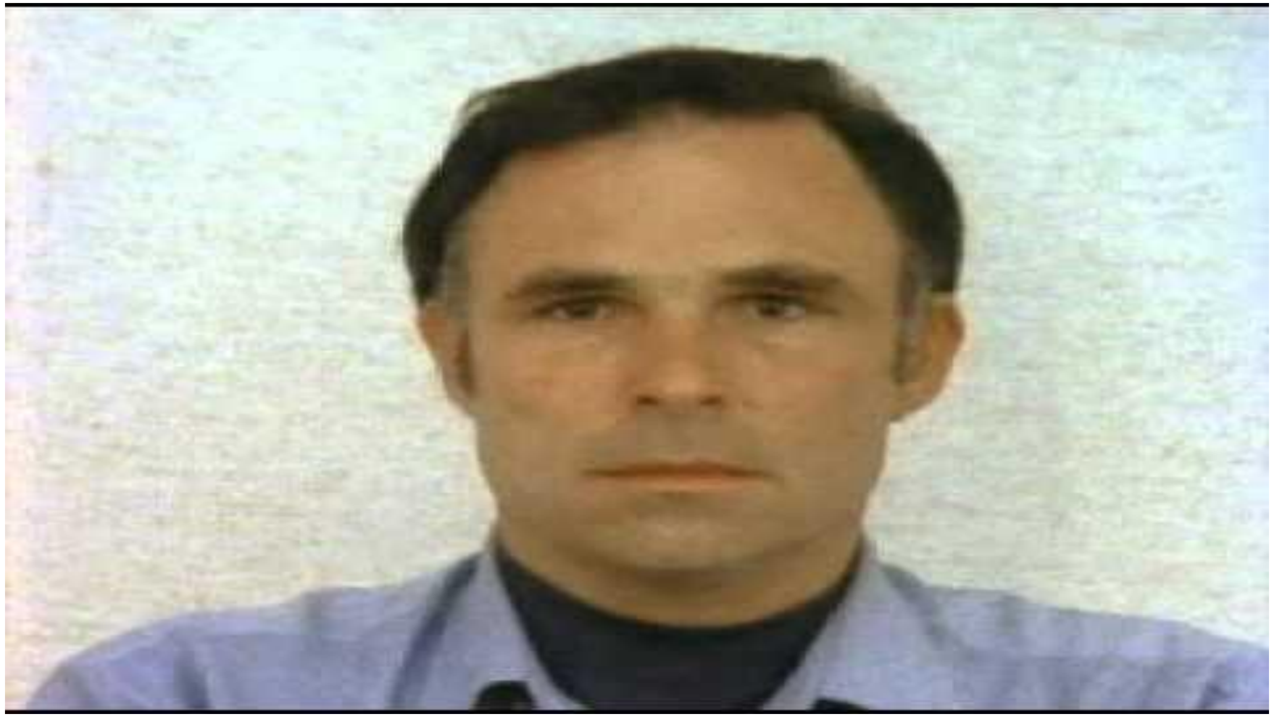
UNIVERSITÀ di VERONA

- FACS can be applied by humans (certified annotators).
- Can be used to describe any facial expressions (not only emotions)
- Using FACS, nearly any anatomically possible facial expression is coded by deconstructing it into specific Action Units (AU) and their temporal segments producing the expression.
- AUs are independent of any interpretation and can be used for any higher-order decision-making process, e.g., recognition of emotions.

Facial Action Coding System (FACS)

- FACS include:
- Action Units (AUs): the fundamental actions of individual muscles or groups of muscles
  - Face: 46 AUs
  - Head: 14 AUs
  - Gaze: 11 AUs
- Intensities of AUs are annotated by appending a letter (A to E) to the Action Unit number, meaning that intensity can be expressed in 5 levels scale:
  - A: Trace
  - B: Slight
  - C: Marked or Pronounced
  - D: Severe or Extreme
  - E: Maximum
- certain AUs can be Left (L) or Right (R)

Facial Action Coding System



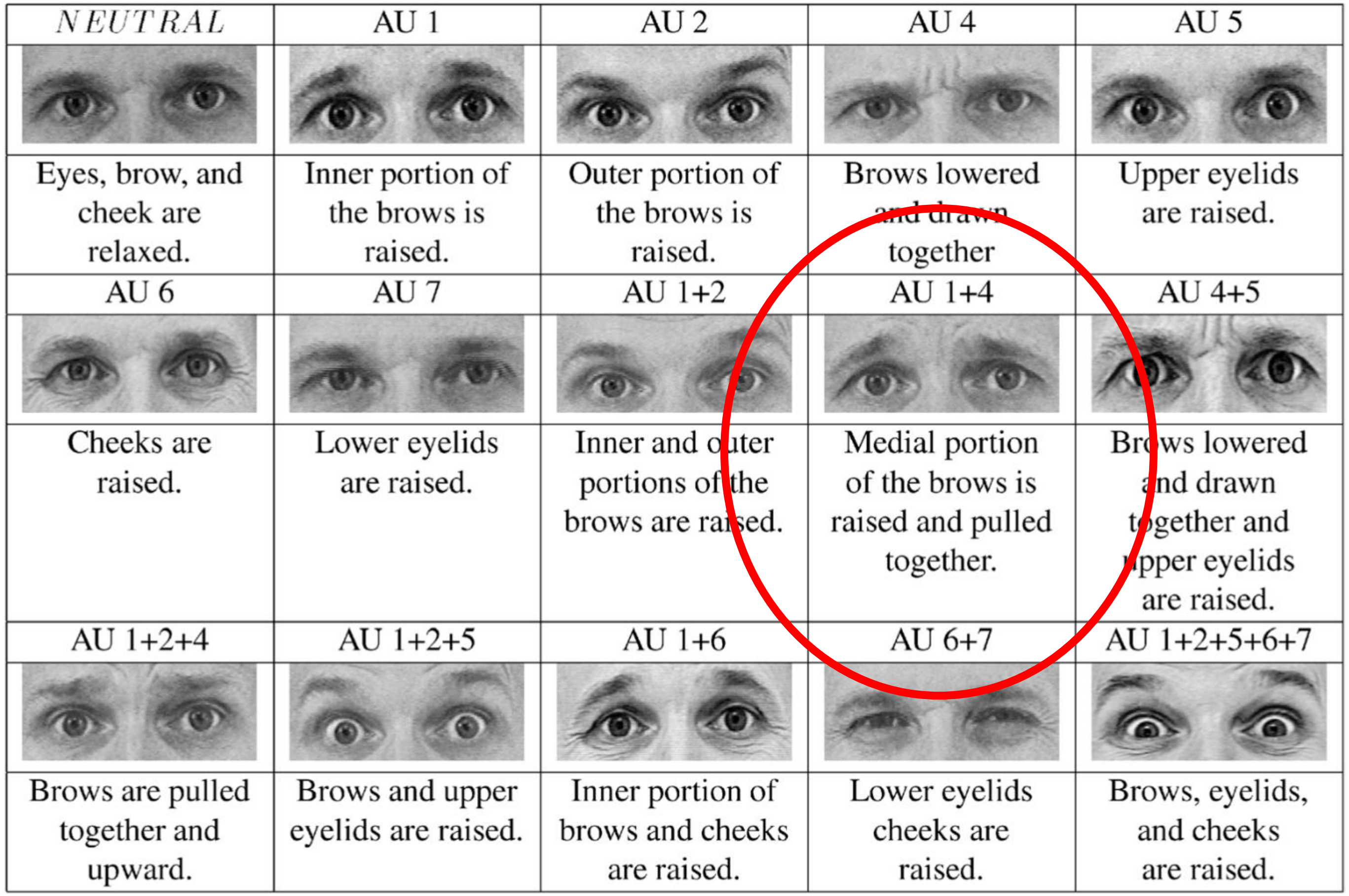
https://www.youtube.com/watch?v=-G7IRRydpVA

Upper Face Action Units					
AU 1	AU 2	AU 4	AU 5	AU 6	AU 7
Inner Brow Raiser	Outer Brow Raiser	Brow Lowerer	Upper Lid Raiser	Cheek Raiser	Lid Tightener
*AU 41	*AU 42	*AU 43	AU 44	AU 45	AU 46
Lid Droop	Slit	Eyes Closed	Squint	Blink	Wink

Lower Face Action Units					
AU 9	AU 10	AU 11	AU 12	AU 13	AU 14
Nose Wrinkler	Upper Lip Raiser	Nasolabial Deepener	Lip Corner Puller	Cheek Puffer	Dimpler
AU 15	AU 16	AU 17	AU 18	AU 20	AU 22
Lip Corner Depressor	Lower Lip Depressor	Chin Raiser	Lip Puckerer	Lip Stretcher	Lip Funneler
AU 23	AU 24	*AU 25	*AU 26	*AU 27	AU 28
Lip Tightener	Lip Pressor	Lips Part	Jaw Drop	Mouth Stretch	Lip Suck

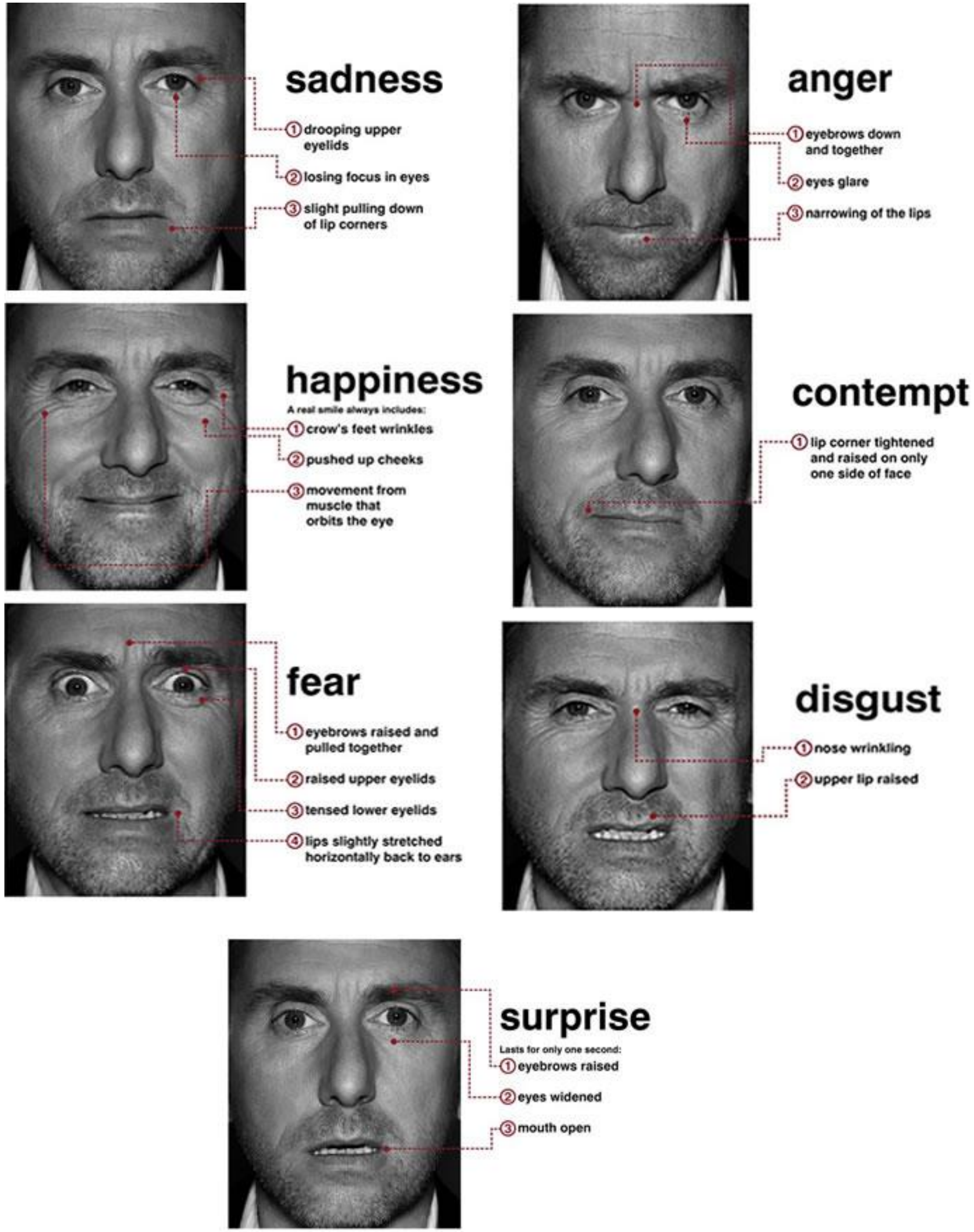
UNIVERSITÀ di VERONA

It is possible to have combinations of Aus. But, some combinations are not "intuitive"…



UNIVERSITÀ di VERONA

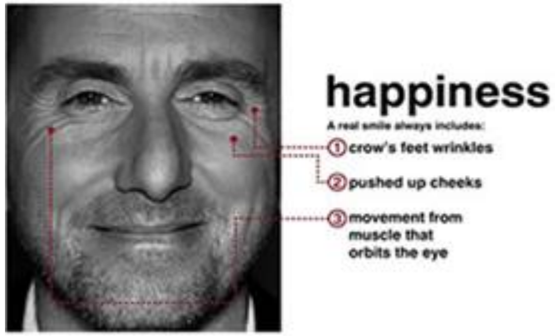
Ekman, 2007



Examples(!) of expressions in FACS notation

Emotion ⬍	Action units ⬍
Happiness	6+12
Sadness	1+4+15
Surprise	1+2+5B+26
Fear	1+2+4+5+7+20+26
Anger	4+5+7+23
Disgust	9+15+17
Contempt	R12A+R14A

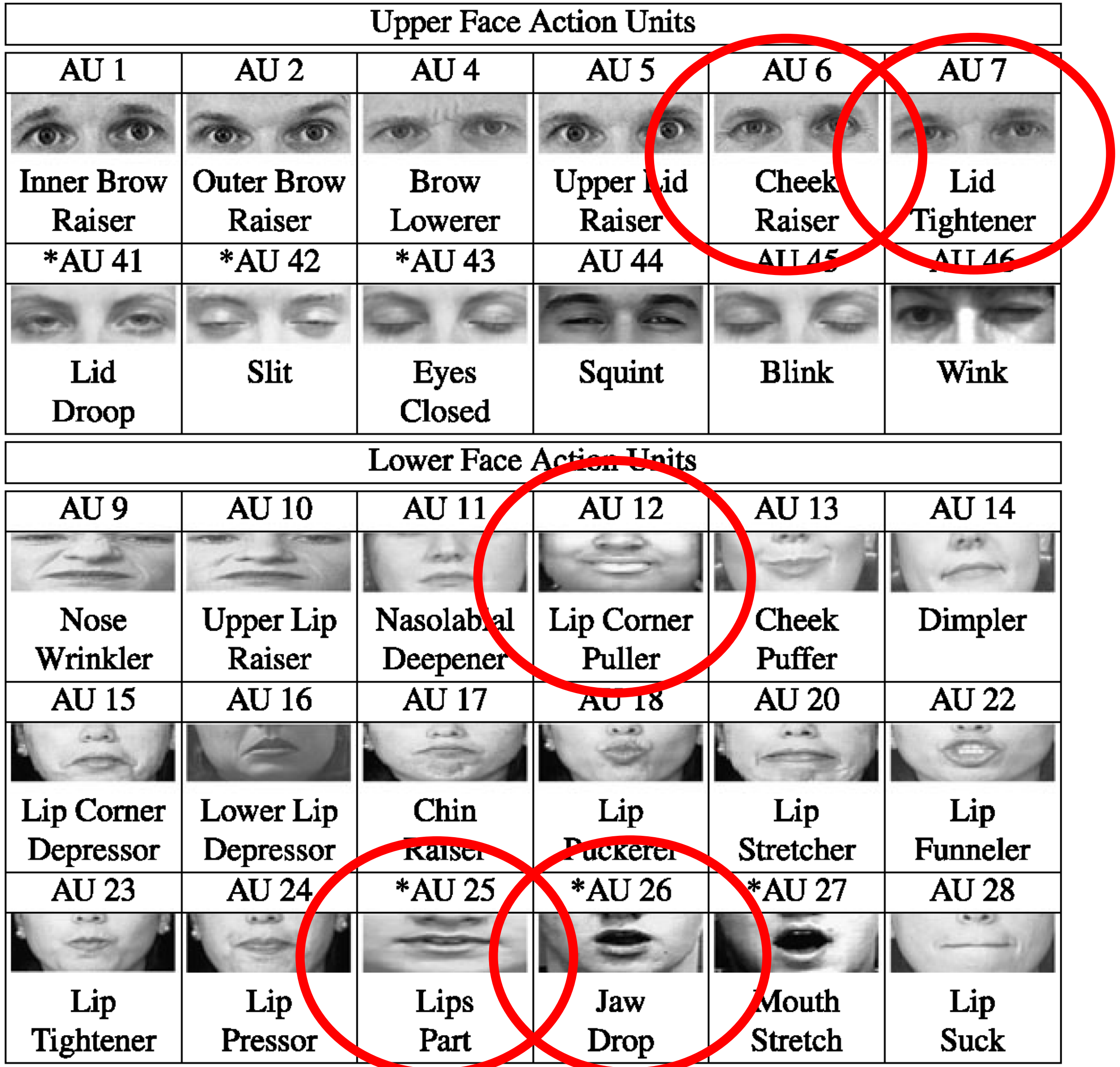
A prototypical "happy face" is a combination of…



AU 6 + 7 which are muscle activities around the eyes

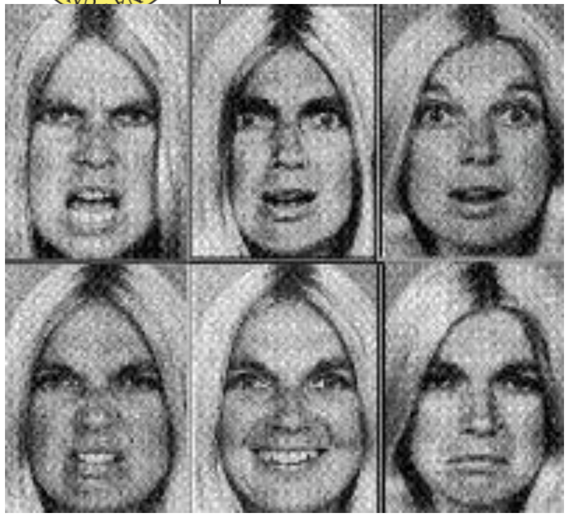
AU 12 – which is a zygomatic major activity – the most important element of the smile

They can be accompanied by AU25 and AU26



EMOTIONS

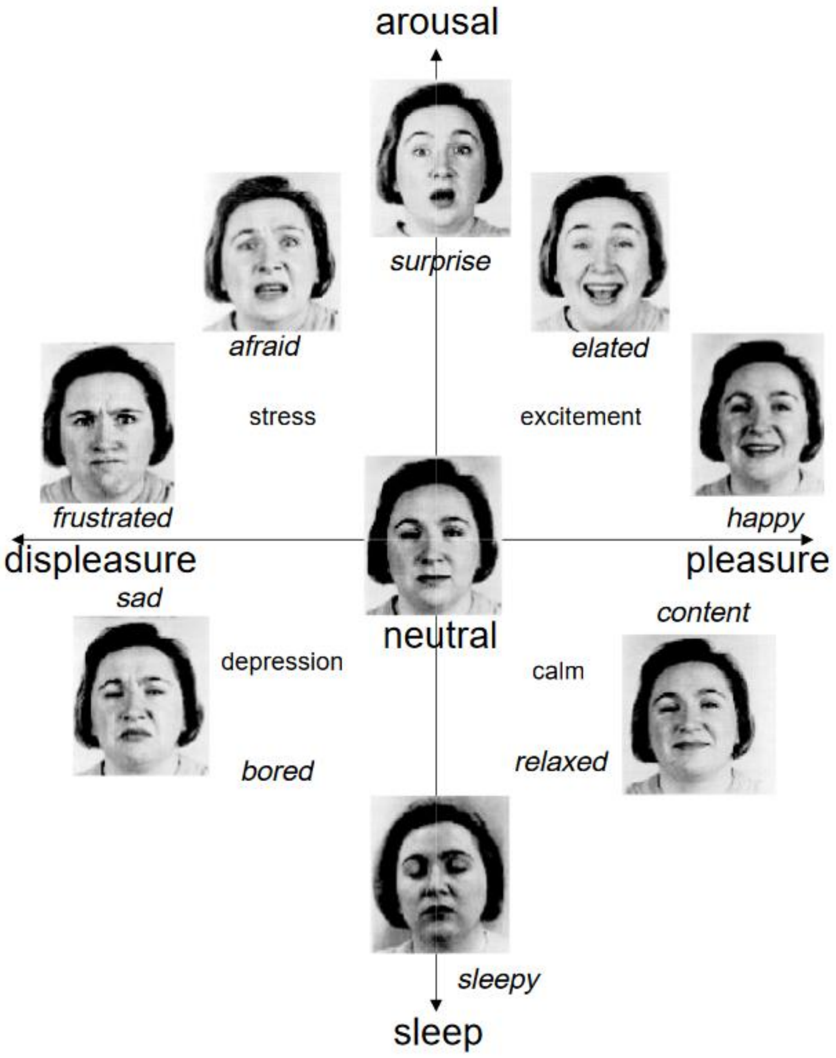
- **Discrete Approach (Ekman)**
  - Anger, fear, sadness, happiness, disgust, surprise
  - Universally recognized (Ekman)
  - basic emotion = family of related states (Ekman 75)
- **Real-life emotions**
  - Are often complex and involve several emotions
  - Theories (Ekman 75, Plutchik 80, Ekman 92, Scherer 84)
  - Lost luggage study (Scherer 98)
  - TV interviews : Belfast & EMoTV corpora (Douglas-Cowie et al. 2005)



(Ekman 75)

EMOTIONS

- If there are an infinite number of emotions, then there can be an infinite number of expressions
- **Russell 2D model** is based on this idea such that
  - some easily distinguishable expressions are close to each other in this model.
  - Emotions are defined based on two fundamental dimensions: **valence** and **arousal.**
- **Valence:** measures how positive or negative an emotion is.
  - E.g., happiness has a positive valence, while sadness has a negative valence.
- **Arousal:** intensity associated with an emotion.
  - E.g., excitement represents high arousal, while calmness represents low arousal.



Breazeal, 2003

EMOTIONS

- Can we infer all emotions in all circumstances from human facial movements?
- Are facial expressions a comprehensive solution for understanding emotions?

Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements

ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Psychological Science in the Public Interest
2019, Vol. 20(1) 1–68
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1529100619832930
www.psychologicalscience.org/PSPI
SAGE

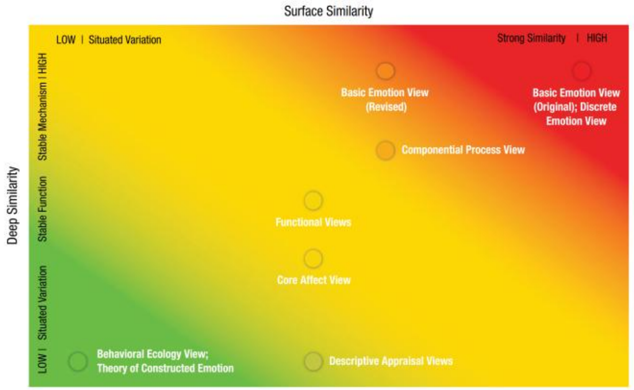


Fig. 1. Explanatory frameworks guiding the science of emotion: the nature of emotion categories and their concepts. The information in the figure is plotted along two dimensions. The horizontal dimension represents hypotheses about the similarities in surface features shared by instances of the same emotion category (e.g., the facial movements that express instances of the same emotion category). The vertical dimension represents hypotheses about the similarities in the mechanisms that cause instances of the same emotion category (e.g., the neural circuits or assemblies that cause instances of the same emotion category). The colors represent the type of emotion categories proposed in each theoretical framework. Approaches in the green area describe ad hoc, abstract categories; those in the yellow area describe prototype or theory-based categories, and those in the red area describe natural-kind categories.

https://journals.sagepub.com/doi/pdf/10.1177/1529100619832930

Is it soo…. easy?

In real life faces are rarely encountered in isolation

Aviezer et al., 2009

Context

a

Aviezer et al., 2009

Context

b

Context

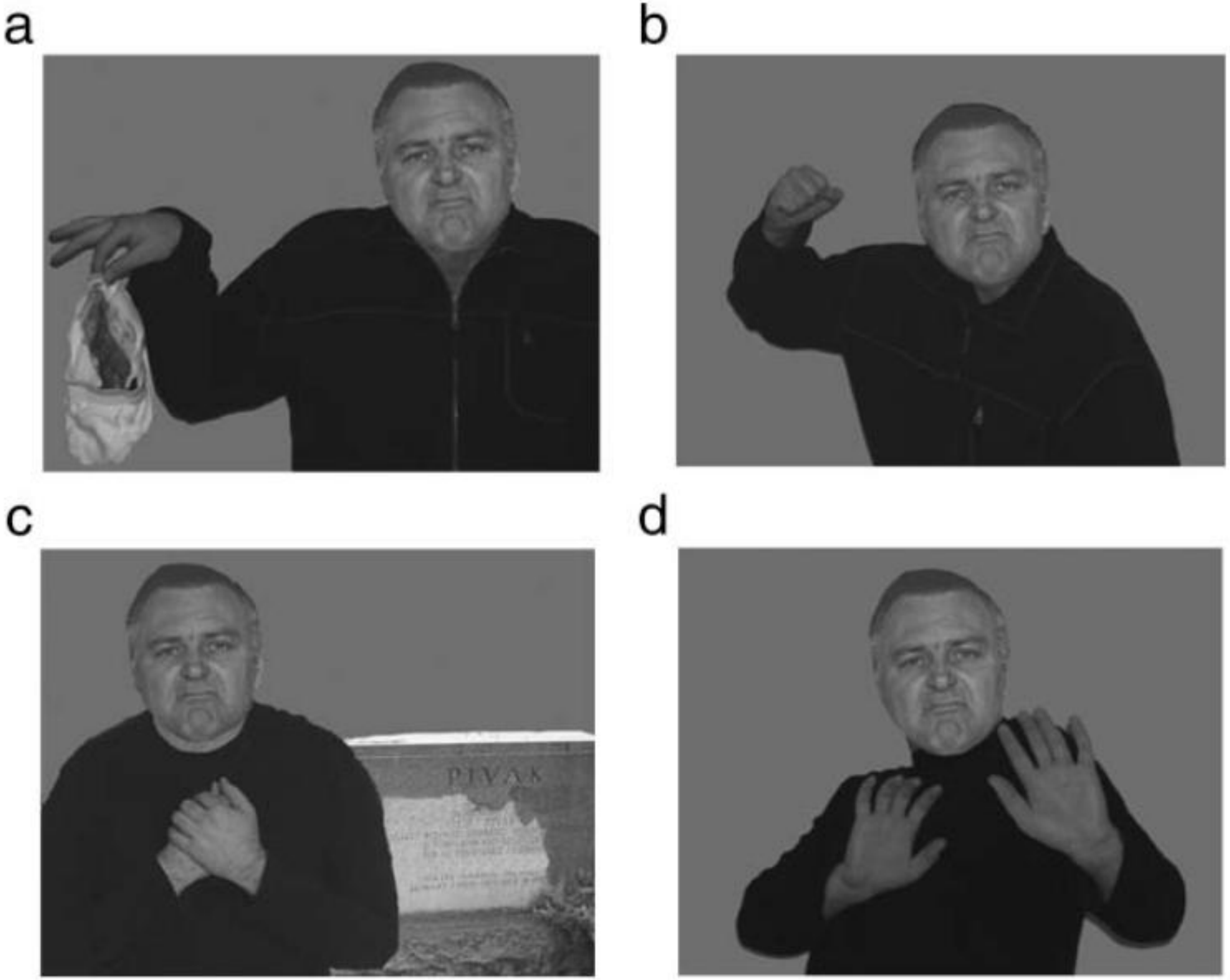
C

Context

d

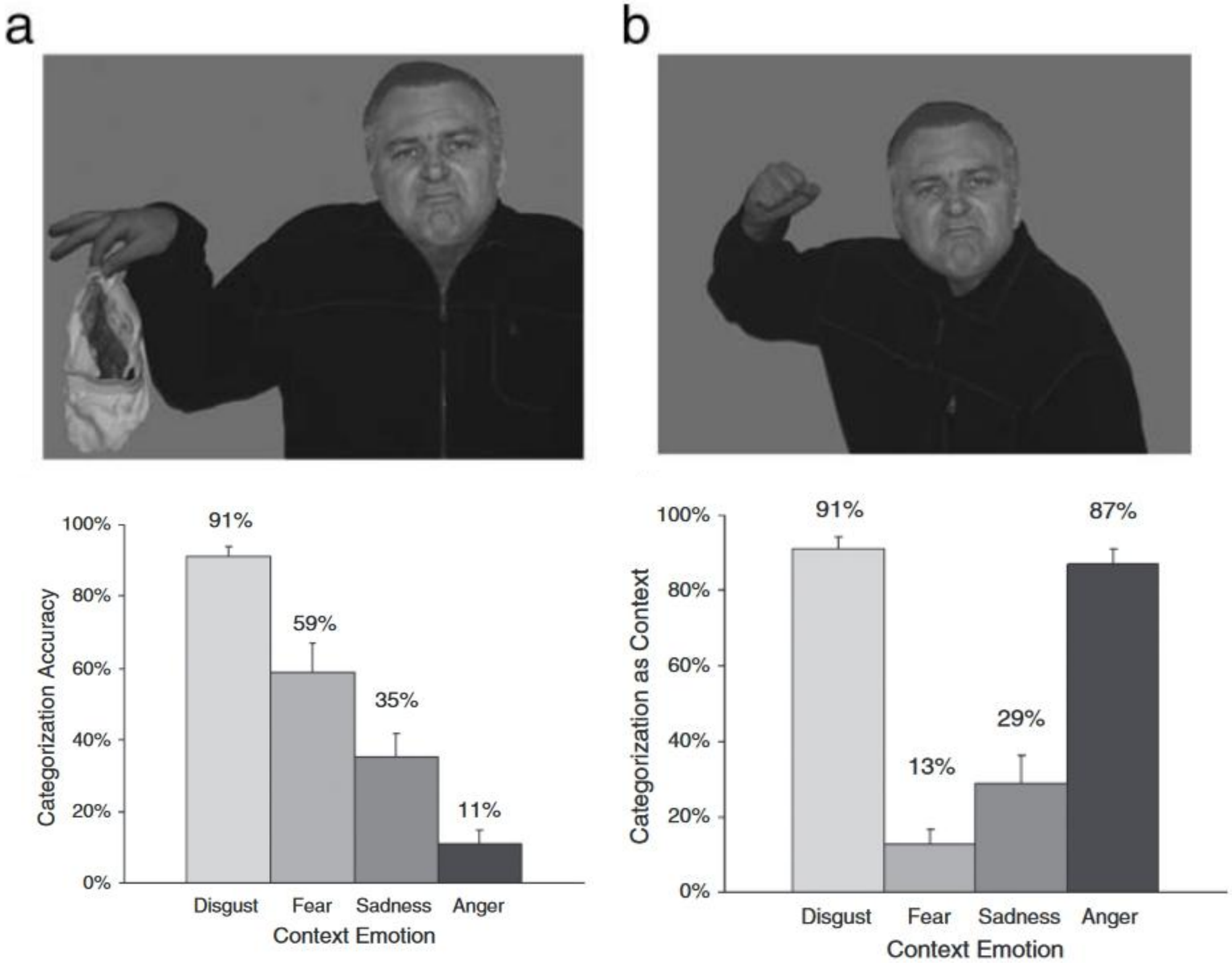
Context

UNIVERSITÀ di VERONA



Context: categorical approach

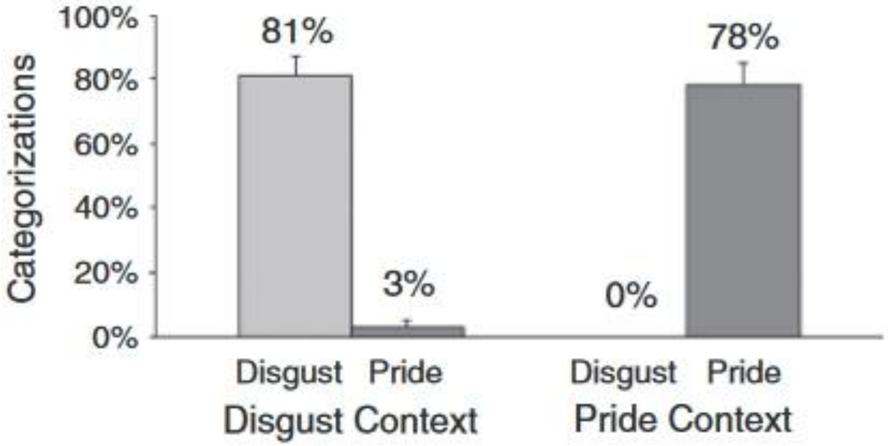
the percentage of times the face was categorized as "disgust"



the percentage of times the face was categorized as expressing the context emotion

Context: dimensional approach

“Context” is understood broadly:
- gestures and postures
- object manipulations



Switching between positive and negative valence context

EMOTIONS: TAKE HOME MESSAGES

UNIVERSITÀ di VERONA

- Widely used theories of emotion claim that basic facial expressions signal
  - discrete emotions, or affective dimensions
- In both cases, the information is thought to be directly "read out" from the face
- In contrast, studies show that:
  - identical facial configurations **<u>convey</u>** different emotions (and dimensional values) depending on the (affective) context
  - the perception of basic facial expressions is not context-invariant and can be altered by context
- Similarity between the expressions (target label, perceived label) matters. Labels of similar expressions are more easily "exchanged"
- In real life, faces are rarely encountered in isolation, and the context in which they appear is very informative
- WHICH "EMOTION RECOGNITION" SOFTWARE USES THE CONTEXT INFO?

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