Language Technologies Institute

Carnegie Mellon University

Multimodal Affective Computing

Lecture 4: Visual Messages

Louis-Philippe Morency
Jeffrey Girard

Originally developed with help from
Stefan Scherer and Tadas Baltrušaitis

Outline

- Interpersonal Communication
  - Encoder-Decoder Process, Lens Model
  - Elements of interpersonal communication
- Nonverbal visual messages
  - Facial expressions
  - Eye gaze and mutual contact
  - Proxemics and group formations
  - Gestures and body language
  - Practical tools for automatic sensing

Outline

- Nonverbal acoustic messages **[Week 5]**
  - Fundamentals of speech
  - Prosodic manipulation and its meaning
  - Use of varying voice quality
  - Nonverbal vocal expressions
  - Practical tools for automatic sensing

Upcoming Schedule

- Week 4
  - Tuesday: Lecture on Visual Messages
  - Thursday: Discussion (from Week 3 - theories)
- Week 5
  - Tuesday: Lecture on Vocal Messages
  - Thursday: Discussion (visual & vocal messages)
- Week 6:
  - Tuesday: Lecture on Verbal Messages
  - Thursday: Proposal presentations
  - Sunday: Due date for proposal reports
- Week 7:
  - Tuesday: Lecture on Statistical Analysis
  - Thursday Discussion (verbal messages)

TA Office Hours + Tutorials

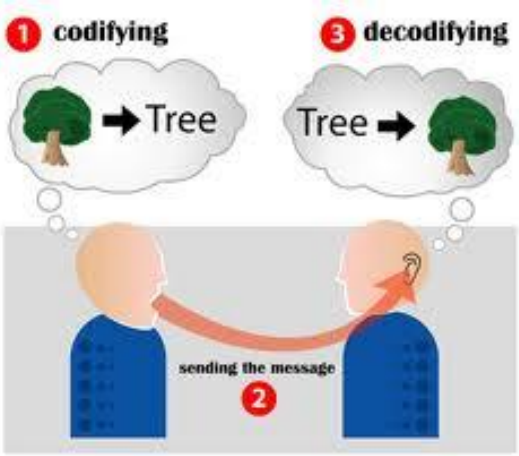
- Monday 2-3pm
- Monday 3-4pm
- Wednesday 2-3pm
- Wednesday 3-4pm

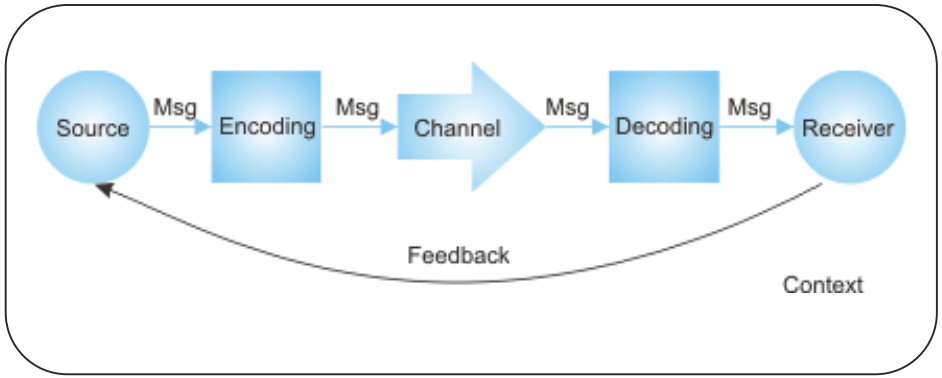
Interpersonal Communication

Interpersonal Communication

Speaker **Listener**

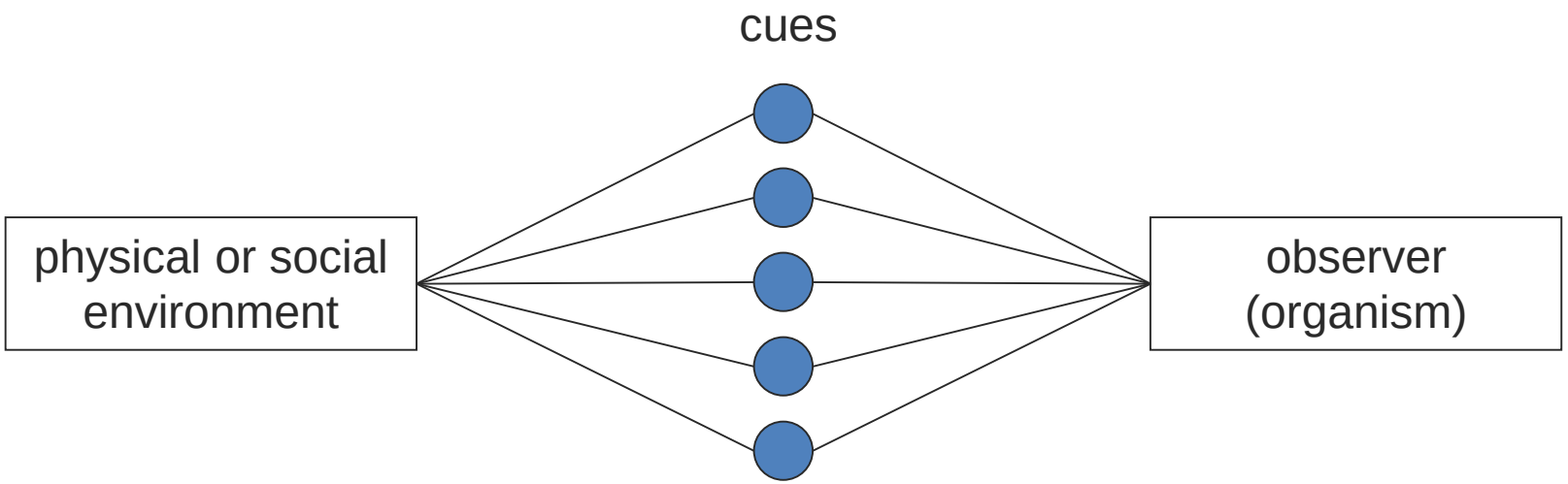
Communication Process: Encoder-decoder





The Lens Model [Brunswick 1956]

- is used in several fields to study how observers correctly and incorrectly use objective cues to perceive physical or social reality

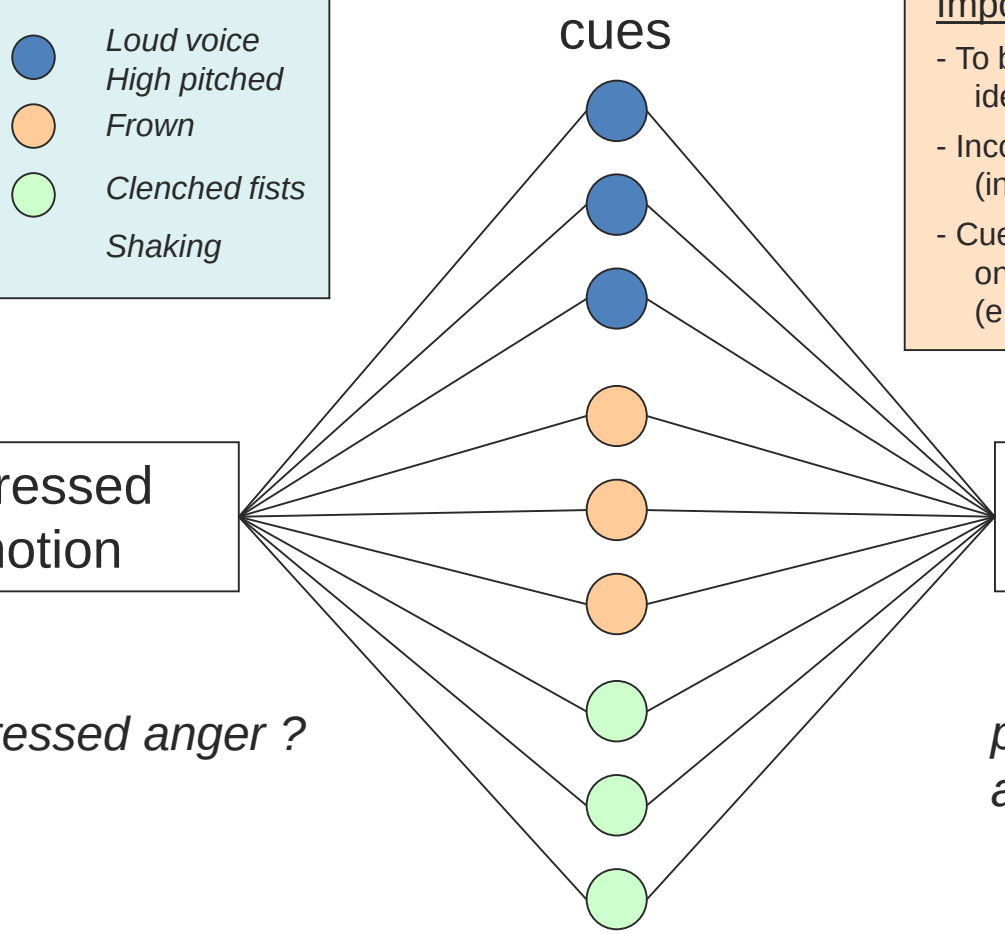


- cues have a probabilistic (uncertain) relation to the actual objects
- a (same) cue can signal several objects in the environment
- cues are (often) redundant

Example: Emotional communication

Example:

Vocal cues	*Loud voice, High pitched*
Facial cues	*Frown*
Gestures	*Clenched fists*
Other cues …	*Shaking*



Important issues:

- To be measured, cues must be identified a priori
- Inconsistencies on both sides (indiv. diff., broad categories)
- Cue utilization could be different on the left and the right side (e.g. important cues not used)

Expressed emotion

Emotional attribution

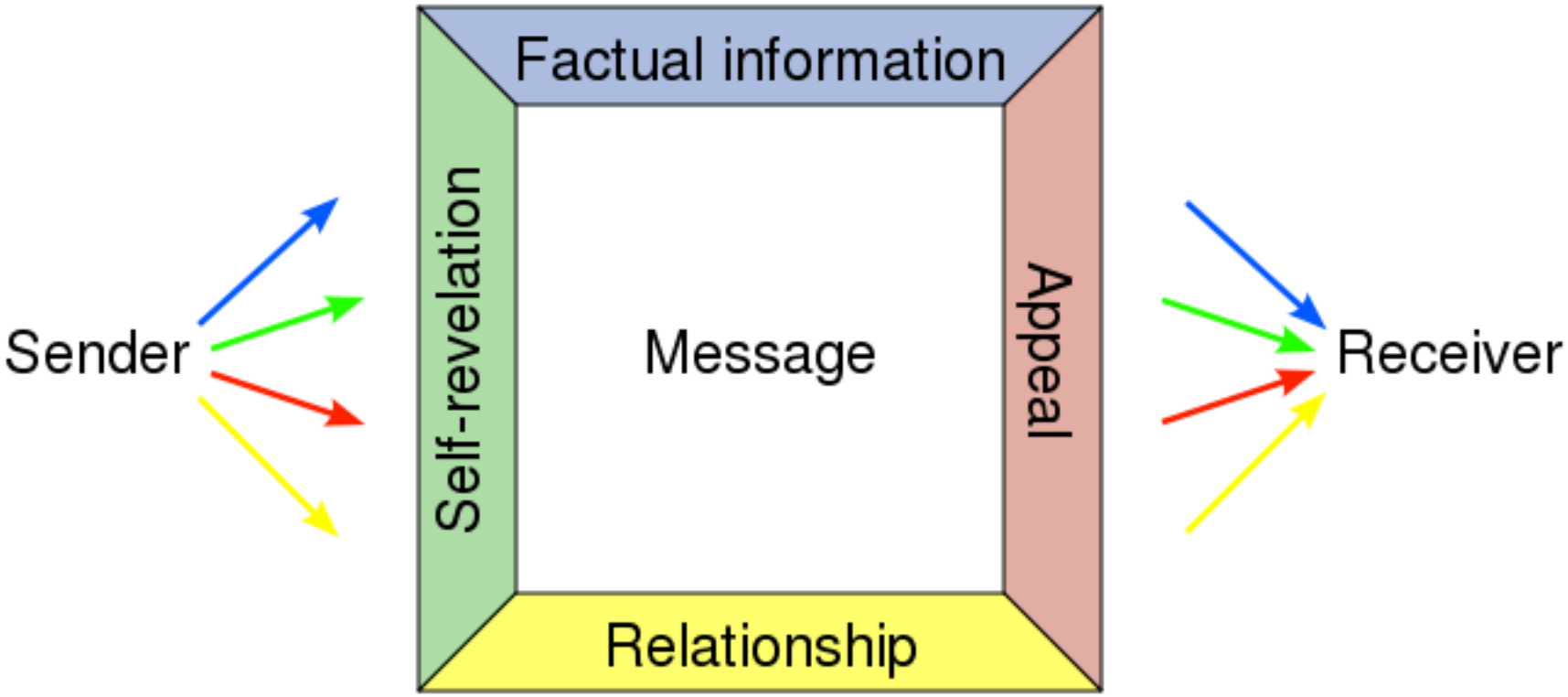
expressed anger ?

perception of anger?

encoder

decoder

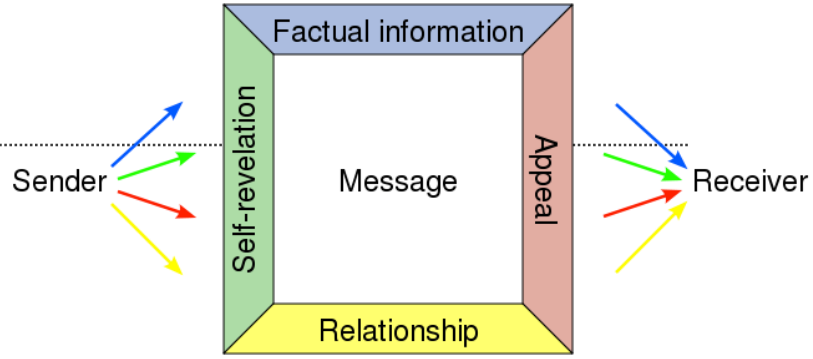
Four-Sides Model



Friedemann Schulz von Thun

Source: Wikipedia

Four-Sides Model



"There is something green in my soup."

Sender:

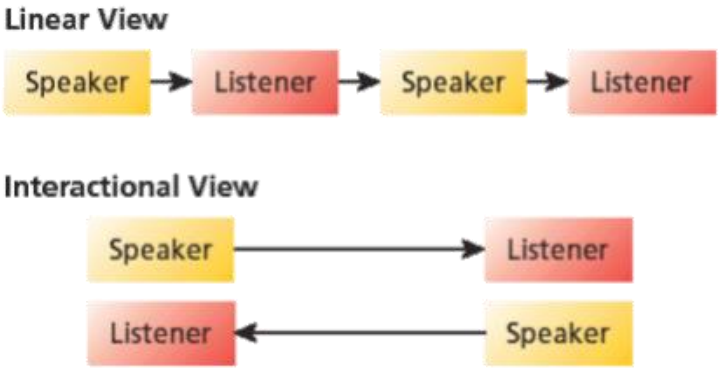
Factual:	There is something green.
Self-revelation:	I don't know what it is.
Relationship:	You cooked, you should know what it is.
Appeal:	Tell me what it is!

Receiver:

Factual:	There is something green.
Self-revelation:	You do not know what the green item is, and that makes you feel uncomfortable.
Relationship:	You think my cooking is questionable.
Appeal:	I should only cook what you know in the future.

"If you don't like what I cook, you should cook it yourself!"

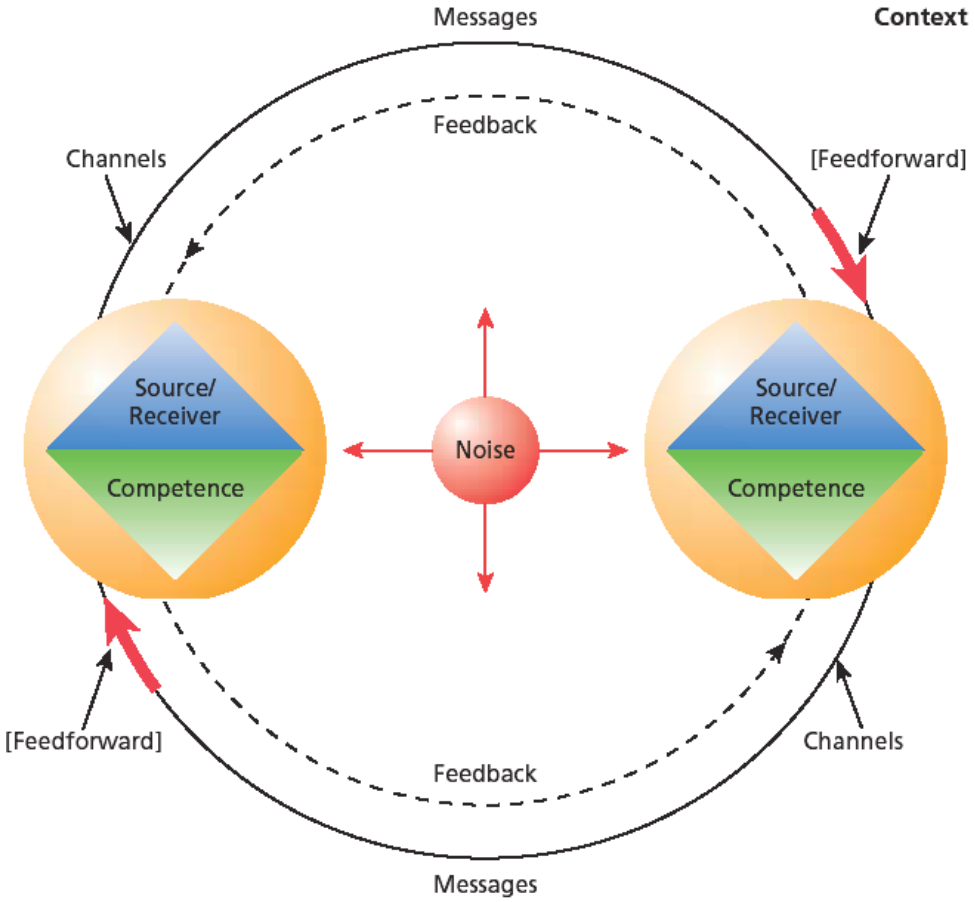
Different Views of Interpersonal Communication



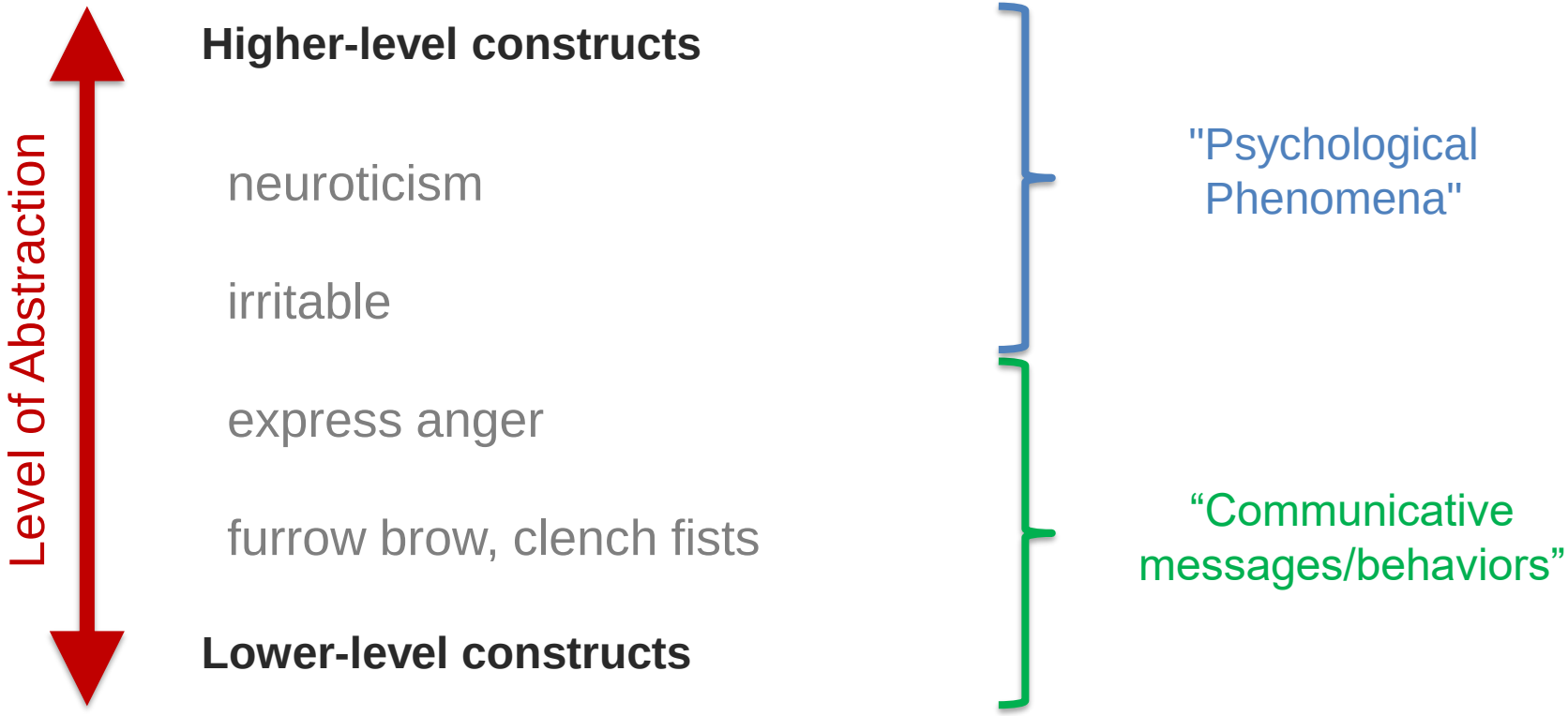
What is missing?

Multimodal Communicative Behaviors

Elements of Interpersonal Communication



Multimodal Communicative Messages



Nonverbal Communication

Nonverbal Communication

Nonverbal Communication

Visual

- **Gestures**
  - Head gestures
  - Eye gestures
  - Arm gestures
- **Facial expressions**
  - 32 action units
  - Smile, frowning
- **Eye contact**
  - Head gaze
  - Eye gaze
- **Proxemics**
  - Dyads
  - Group formations
- **Appearance**

Nonverbal Communication

Other nonverbal cues:

- ❑ Vocal behaviors (next week)
  - Vocal expressions
  - Prosody
  - Voice quality
- ❑ Haptics
  - Touch gestures
- ❑ Olfactics
  - Scents
  - Odors

Visual

- **Gestures**
  - Head gestures
  - Eye gestures
  - Arm gestures
- **Facial expressions**
  - 32 action units
  - Smile, frowning
- **Eye contact**
  - Head gaze
  - Eye gaze
- **Proxemics**
  - Dyads
  - Group formations
- Physical appearance

Roles of Nonverbal Messages [Burgoon, 1985]

- Adults generally rely more on nonverbal than verbal cues in determining social meaning
- The more nonverbal cues are at odds with verbal ones, the more adults rely on nonverbal
- Children who already learned language rely more on verbal than nonverbal cues
  - Children begin life relying on nonverbal signals
  - Sometime prior to puberty, they return to greater belief in nonverbal signals
- Individual have consistent biases in channel reliance (verbal vs nonverbal)

Body Posture and Gestures

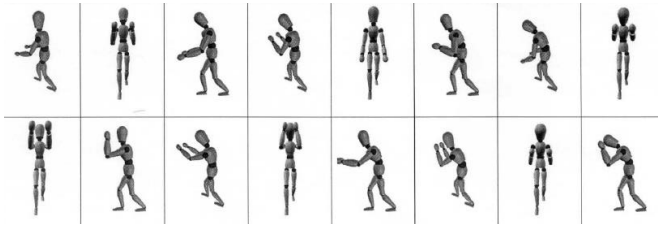
Body Posture and Emotions

Positive or negative?

[Body Cues, Not Facial Expressions, Discriminate Between Intense Positive and Negative Emotions, H. Aviezer et al., 2012]

Body Posture and Emotions

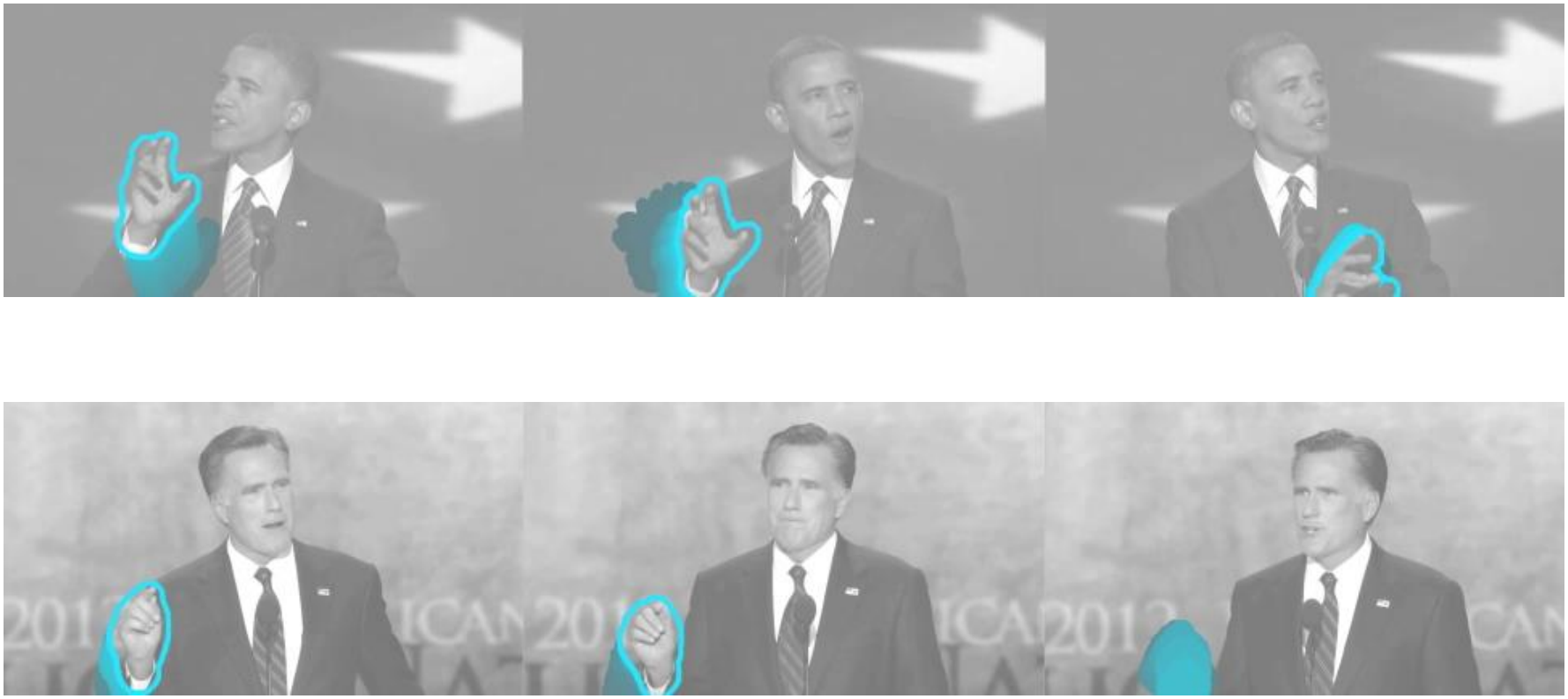
Body Posture and Emotions



Emotion	Frequent posture features
Anger	Head backward, no chest backward, no abdominal twist, arms raised forwards and upwards, shoulders lifted
Joy	Head backward, no chest forward, arms raised above shoulder and straight at the elbow, shoulders lifted
Sadness	Head forward, chest forward, no abdominal twist, arms at the side of the trunk, collapsed posture
Surprise	Head backward, chest backward, abdominal twist, arms raised with straight forearms
Pride	Head backward or lightly tilt, expanded posture, hands on the hips or raised above the head
Fear	Head backward, no abdominal twist, arms are raised forwards, shoulders forwards
Disgust	Shoulders forwards, head downwards
Boredom	Collapsed posture, head backwards not facing the interlocutor

Coulson, Mark. 2004. “Attributing Emotion to Static Body Postures: Recognition Accuracy, Confusions, and Viewpoint Dependence.” Journal of Nonverbal Behavior 28 (2): 117–139.

Presidential Candidate Gestures



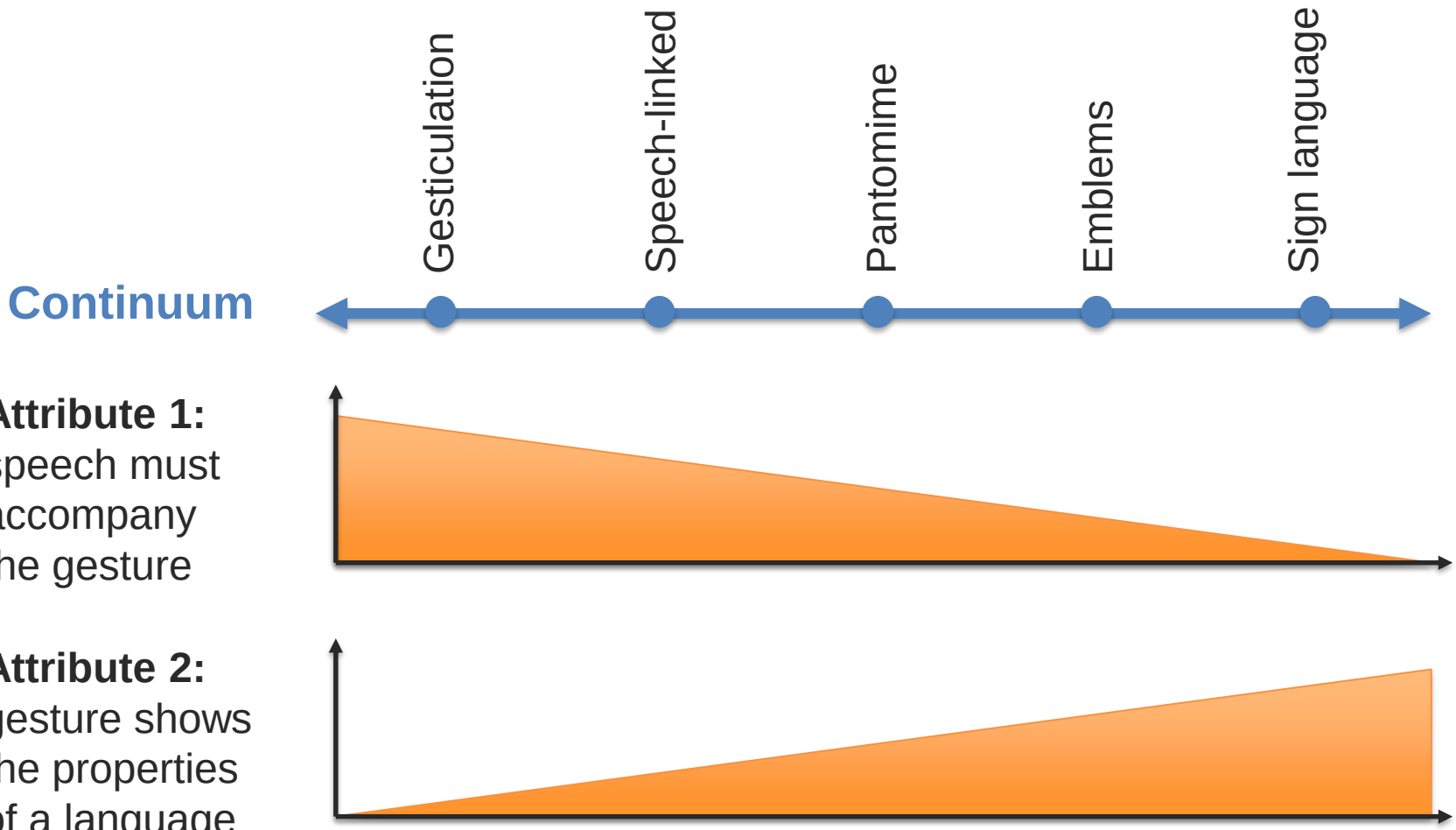
https://archive.nytimes.com/www.nytimes.com/interactive/2012/10/02/us/politics/what-romney-and-obamas-body-language-says-to-voters.html?_r=1

Five types of Body Gestures [Ekman and Friesen, 1969]

Primarily expressed using hands and arms

	Name and Function	Examples
	**Emblems** directly translate words or phrases.	"OK" sign, "Come here" wave, hitchhiker's sign
	**Illustrators** accompany and literally "illustrate" verbal messages.	Circular hand movements when talking of a circle, hands far apart when talking of something large
	**Affect displays** communicate emotional meaning.	Expressions of happiness, surprise, fear, anger, sadness, disgust
	**Regulators** monitor, maintain, or control the speaking of another.	Facial expressions and hand gestures indicating "Keep going," "Slow down," or "What else happened?"
	**Adaptors** satisfy some need.	Scratching head

Kendon's Continuum [McNeill, 1992]



Attribute 1: speech must accompany the gesture

Attribute 2: gesture shows the properties of a language

Kendon's Continuum [McNeill, 1992]

Gesticulations: Accompanied by speech. Usually synchronous with the co-expressive speech. But properties unlike language.

➡ **4 dimensions:**

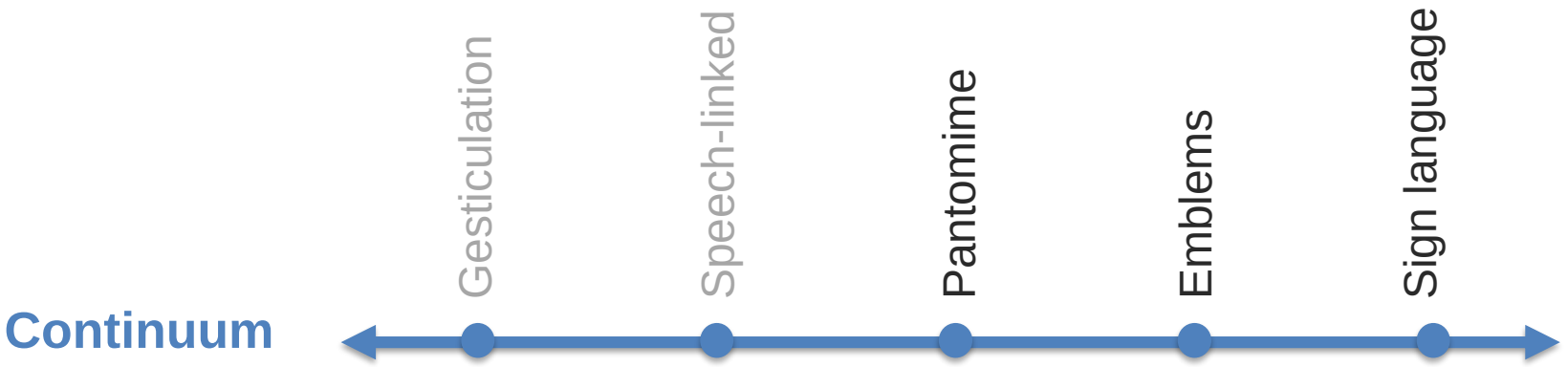
- ❑ **Iconic:** gestures that present images of concrete entities and/or actions (e.g., grasping gesture)
- ❑ **Metaphoric:** Gestures are not limited to depictions of concrete events (e.g., holding a memory or idea)
- ❑ **Deictic:** entails locating entities and actions in space vis-à-vis a reference point (including abstract pointing)
- ❑ **Beats:** signaling the temporal locus of something the speaker feels to be important (e.g., flicks of the hand)

Kendon's Continuum [McNeill, 1992]

Speech-linked: Accompanied by speech. Gesture is part of the sentence itself. They occupy a gap that fills a grammatical slot

"Sylvester went [gesture of an object flying out laterally]"

Kendon's Continuum [McNeill, 1992]



Pantomime: Gesture or sequence of gestures conveying a narrative line, with a story to tell, produced without speech.

Emblems: Conventionalized signs, which potentially refer to a complete verbal translation. Meaningful without speech, but can occur with speech.

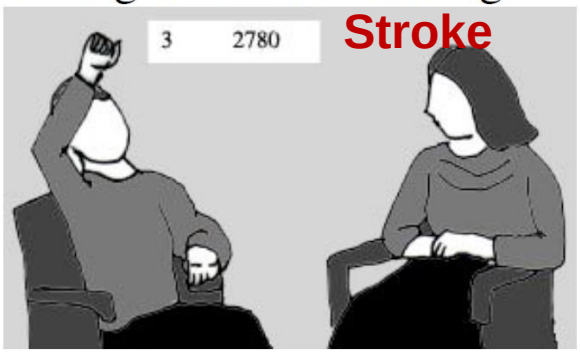
Sign language: Not accompanied by speech. Simultaneously speaking and signing can mutually interfere. Has its own linguistic structure.

Gesture Phases - Example

"so he gets a / hold of a big [oak tree / and he **bends it way back**]"

so he gets a / hold of a big

[oak tree / and he



bends it way ba

ck]

5th phase: retraction

Co-expressiveness and Synchrony

- Stroke phase of the gesticulation is synchronous with the co-expressive speech about 90% of the time.
  [Shuichi Nobe's studies]

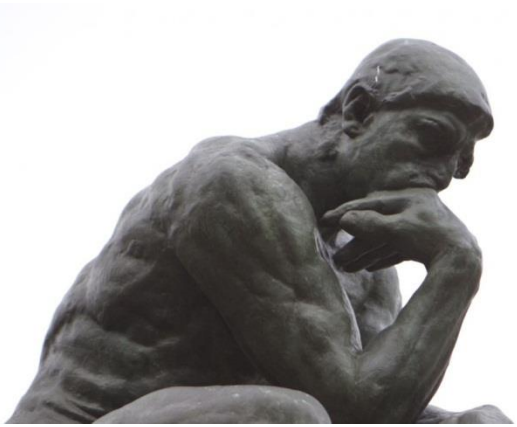
*"…and he tries going **up through** it this time…"*

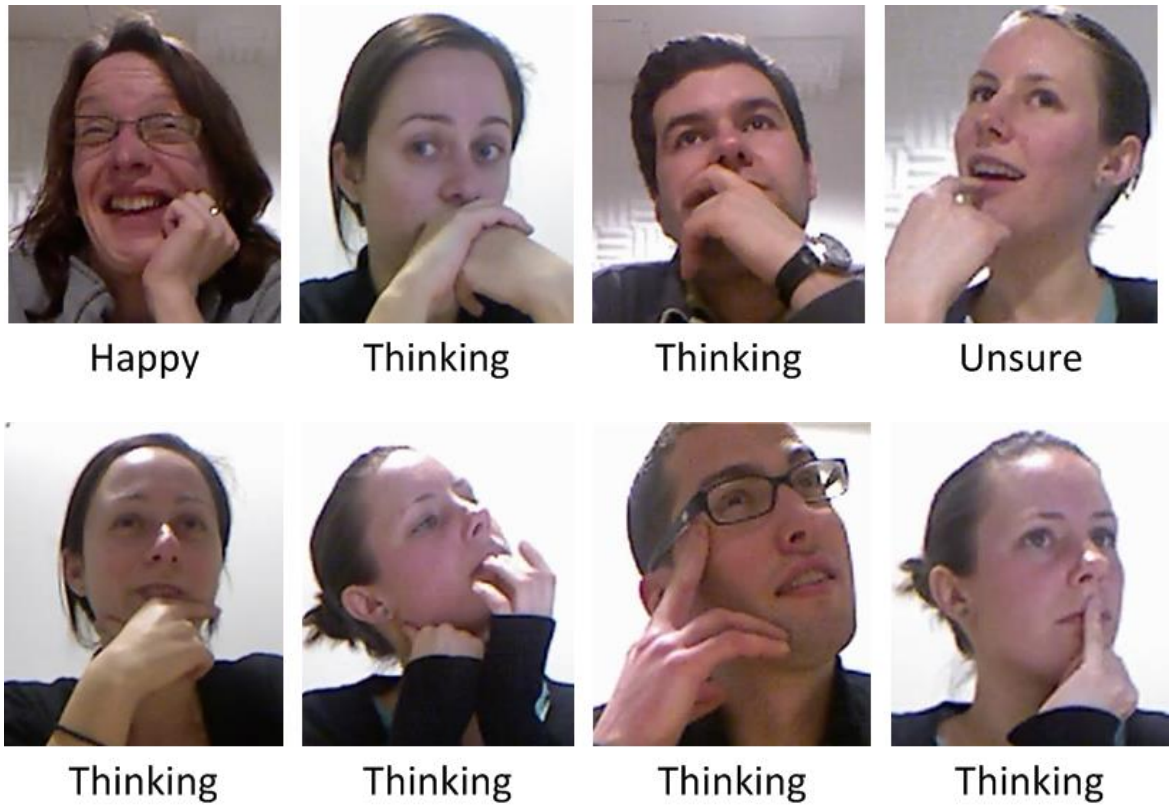
The gesture integrates both linguistic components:

1) Hand raises upward
2) Fingers spread outwards to create an interior space

- Gesticulations rarely if ever follow their co-expressive speech (Kendon 1972).

Hand over face gestures

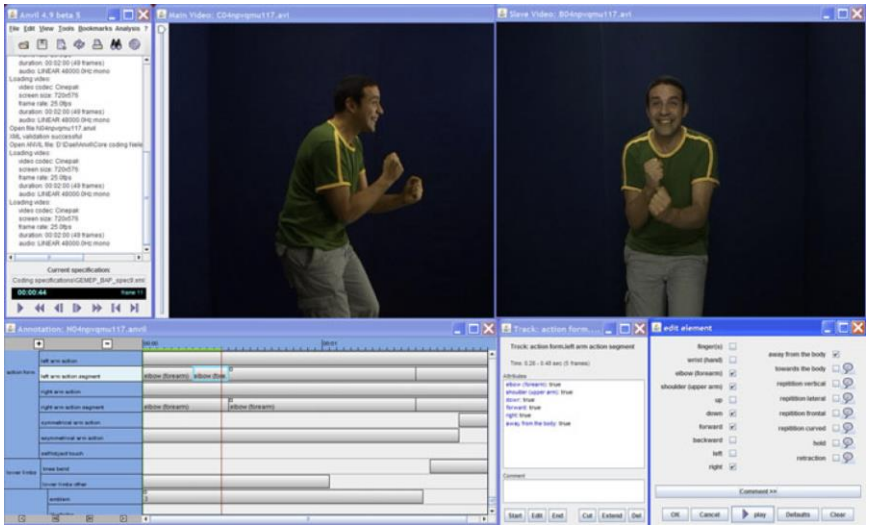




[**Automatic Detection of Naturalistic Hand-over-Face Gesture Descriptors**, M. Mahnoud et al., 2014]

Body Action and Posture Coding System (BAP)

Latest attempt at standardizing the coding of body postures and gestures



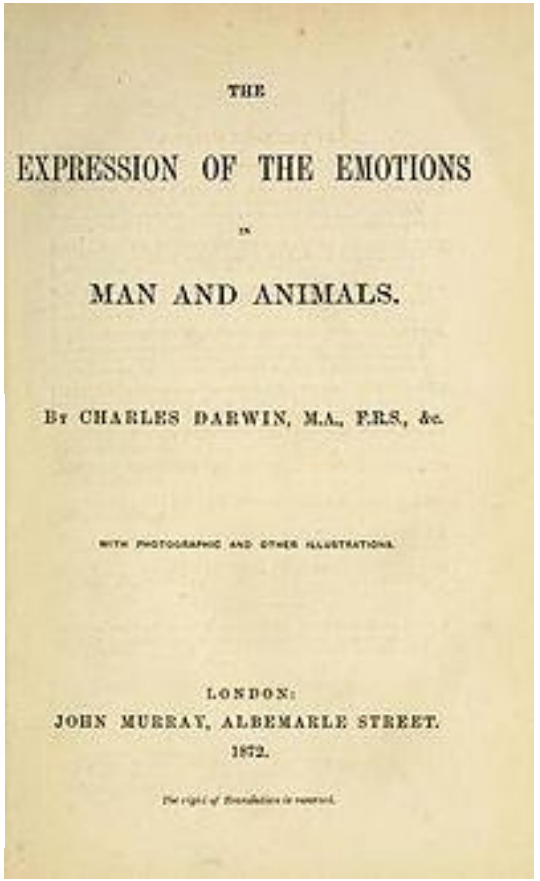
- 141 behavioral codes
- Multi-level coding system:
  - Anatomical (body parts)
  - Form (Movement)
  - Functional (emblems, …)
- Coded the GEMEP dataset
  - 10 professional actors
  - 18 affective states

Dael, N., Mortillaro, M., & Scherer, K. R. (2012). The Body Action and Posture coding system (BAP): Development and reliability. *Journal of Nonverbal Behavior, 36*, 97-121.

Facial expressions

Facial Expressions – some history

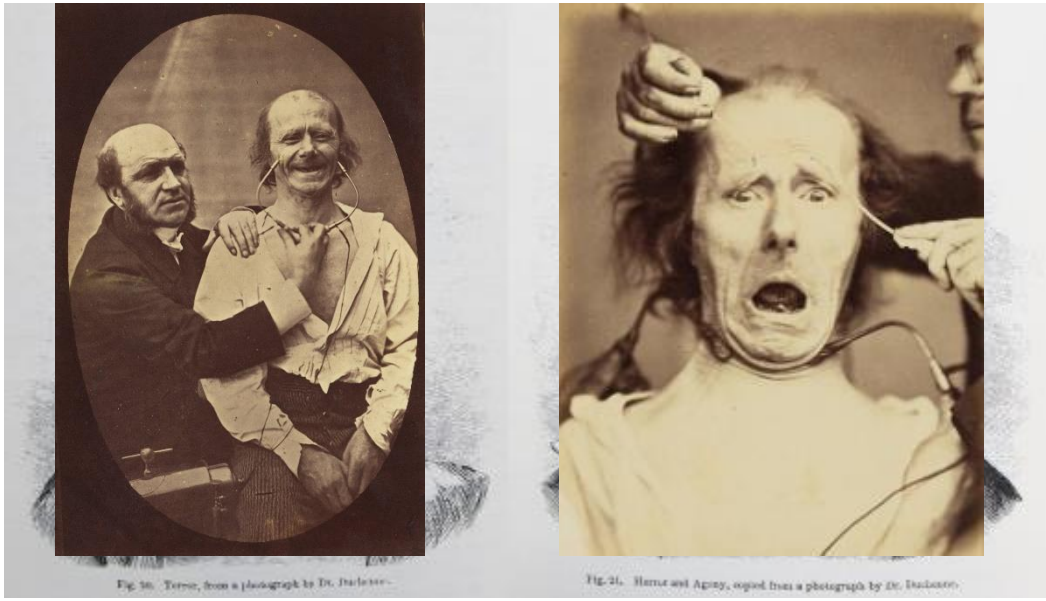
- Charles Darwin



[**Expressions of the emotions in man and animals**, C. Darwin, 1872], republished in 1999 with Paul Ekman as editor

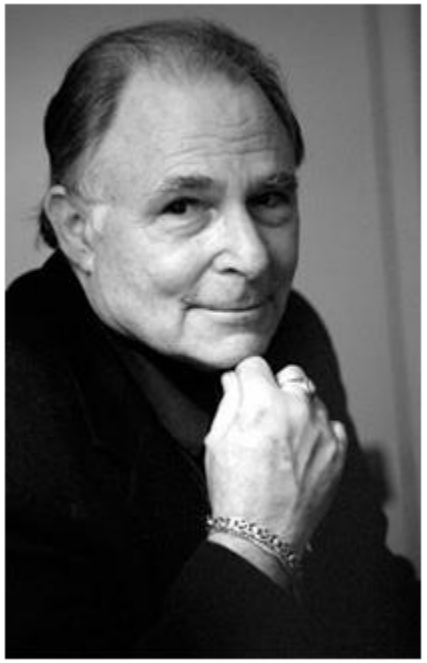
Facial Expressions – some history

- Duchenne photographs
- First use of photography in psychology experiments



Facial Expression – some history

- Paul Ekman and six basic emotions
  - Happiness, Sadness, Fear, Surprise, Disgust, Anger
- Universal
  - Recognition
  - Expression
- Main character in “Lie to Me” based on his research



[**An Argument for Basic Emotions**, P. Ekman, 1992]

Facial Expressions of emotion

Which emotion does each face convey?

These are the six basic/universal/prototypical emotions

Facial Expressions of emotion

- Many more expressions of emotions exist, but they are less universal and more culture specific:
  - Frustration
  - Interest
  - Confusion
  - Boredom
  - Embarrassment
  - Etc.
- They also often require a video rather just an image to be recognised

Facial Action Coding System

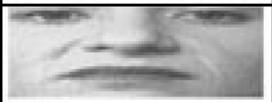
- Comprehensive taxonomy of facial expressions
- Codes action presence and intensity (at 5 levels)
- Correspond to individual or groups of muscles
- 32 Action Units in total
  - Some more rarely occurring than others
- Action descriptors for head and eye actions
- Provides us with signals rather than messages
  - Objective way to quantify expression

[**Facial Action Coding System: A Technique for the Measurement of Facial Movement**, P. Ekman and W. Friesen, 1978]

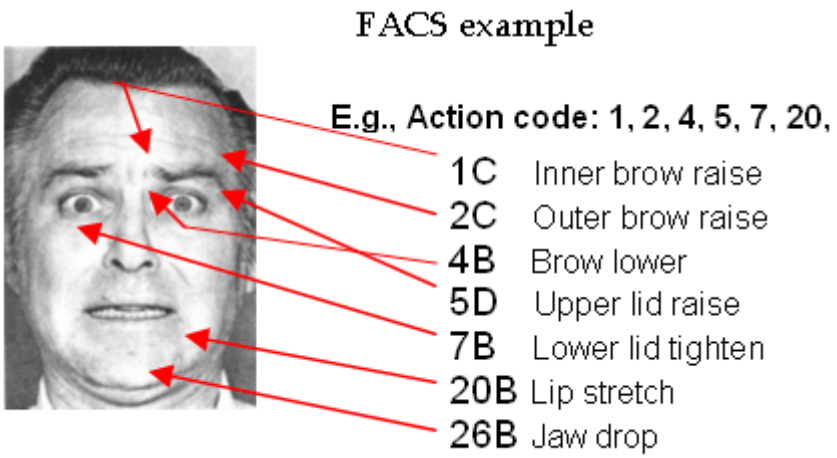
Facial Action Coding System

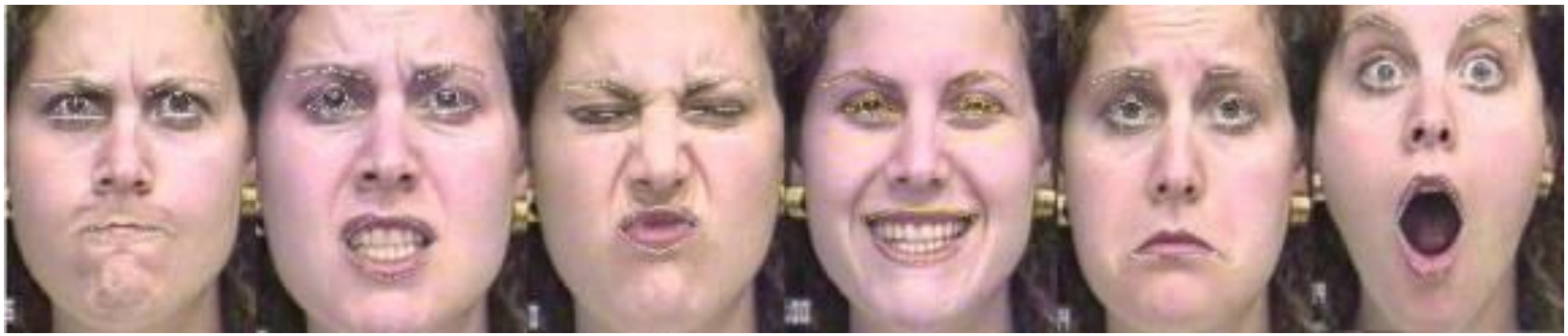
Upper Face Action Units					
AU 1	AU 2	AU 4	AU 5	AU 6	AU 7
					
Inner Brow Raiser	Outer Brow Raiser	Brow Lowerer	Upper Lid Raiser	Cheek Raiser	Lid Tightener

Facial Action Coding System

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

Facial Action Coding System





Anger Fear Disgust Happy Sad Surprise

Facial Action Coding System

Table 1

Action Units Portrayed in Each Emotion Expression Posed, Based on the Emotion-Facial Action Coding Scheme (Ekman & Rosenberg, 1997)

Emotion expression	Action units	
	Female	Male
Anger	4 + 7 + 17 + 24	4 + 7 + 17 + 23
Disgust	4 + 7 + 9 + 16 + 19 + 25 + 26	4 + 7 + 9 + 16 + 19 + 25 + 26
Fear	1 + 2 + 4 + 5 + 15 + 16 + 21 + 25 + 27 + 58	1 + 2 + 4 + 5 + 25 + 26 + 58
Happiness	6 + 7 + 12 + 25	6 + 7 + 12 + 25
Sadness	1 + 4 + 7 + 15 + 16 + 17 + 21	1 + 4 + 15
Surprise	1 + 2 + 5 + 25 + 27 + 38	1 + 2 + 5 + 25 + 26 + 38
Contempt	7 + R10	7 + R10 + 25
Pride	12 + 53	12 + 53
Shame	15 + 43 + 54	43 + 54
Embarrassment	6 + 14 + 24 + 43 + 51 + 54	14 + 43 + 52 + 54

Expression dynamics

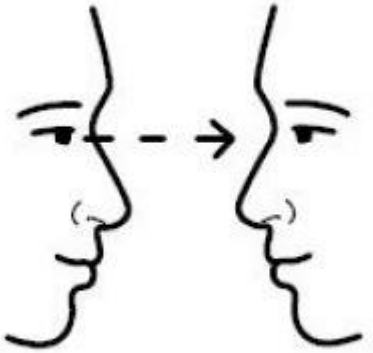
- Not only the configuration of AUs matter
- Temporal characteristics carry information as well
  - Onset, peak, offset
- Can use to distinguish posed and spontaneous expressions

[**How to Distinguish Posed from Spontaneous Smiles using Geometric Features**, M. Valstar et al., 2007]

Eye gaze

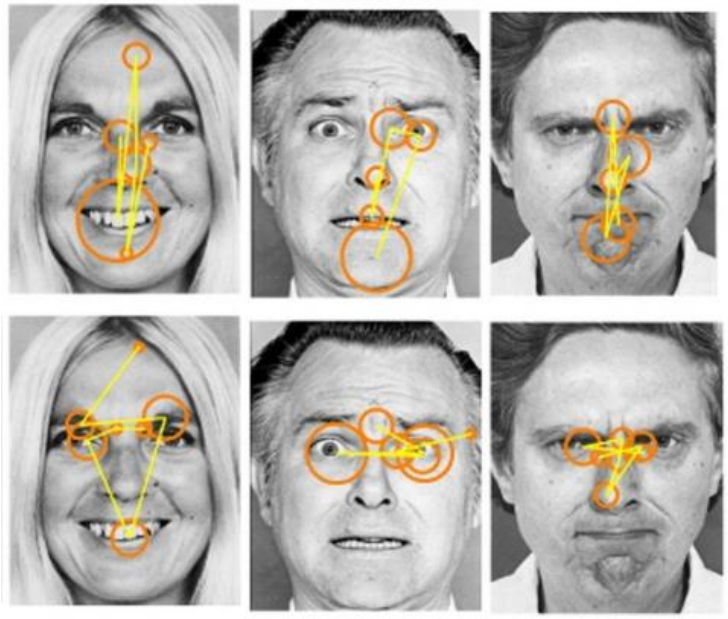
Eye Contact

- Monitor Feedback
- Secure Attention and Interest
- Regulate or Control Conversation
- Signal Nature of Relationship
- Signal Status and Aggression
- Compensate for Distance
- Joint attention



Eye gaze and healthcare

- Eye behavior is indicative of certain psychological conditions especially depression and psychosis
- Where children look in expressive faces reveals a lot about autism



[**Neural basis of eye gaze processing deficits in autism**, K. Pelmphrey et al., 2005]

[**Relationship between abstract thinking and eye gaze pattern in patients with schizophrenia**, J. Oh et al., 2014]

Proxemics and Group Formations

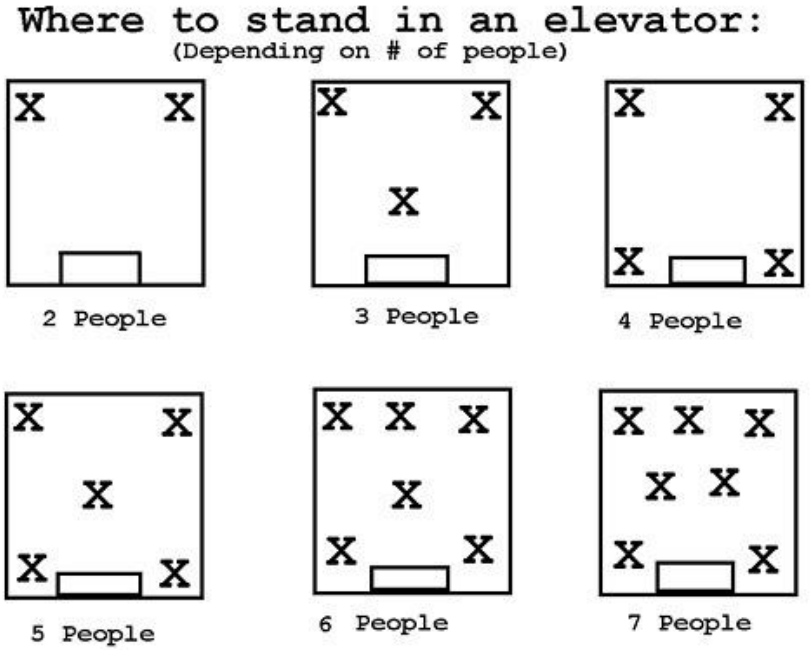
Proxemics

Proxemic Distances

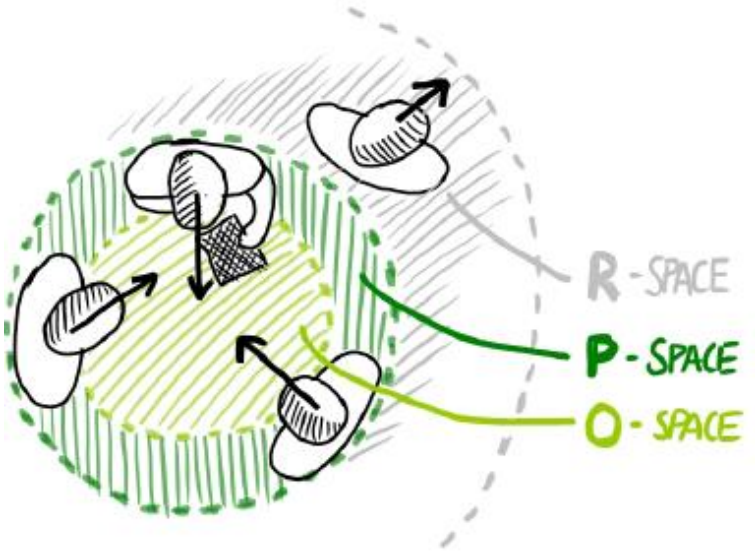
Relationship		Distance
Intimate relationship		Intimate distance 0 — 18 inches close phase — far phase
Personal relationship		Personal distance 1½ — 4 feet close phase — far phase
Social relationship		Social distance 4 — 12 feet close phase — far phase
Public relationship		Public distance 12 — 25+ feet close phase — far phase

Elevators

- Proxemics are very clearly observed in elevators (lifts)



Group Formation: F-formation [Kendon, 1992]



F formation (Facing formation):

o-space is a convex empty space surrounded by the people involved in a social interaction;

p-space is a narrow stripe that surrounds the o-space, and that contains the bodies of the participants;

r-space is the area beyond the p-space

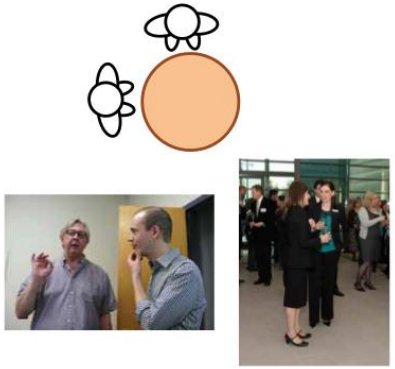
Illustration from Marquardt, Nicolai & Hinckley, Ken & Greenberg, Saul. Cross-device interaction via micro-mobility and F-formations. UIST'12

Group Formation: F-formation [Kendon, 1992]

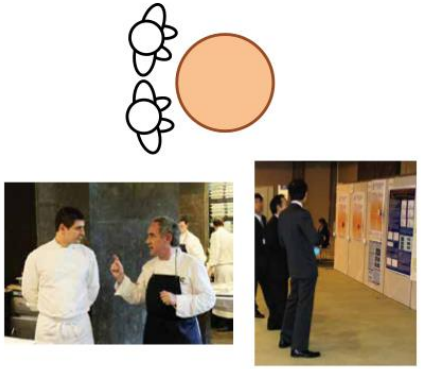
2 people formations:

vis-à-vis arrangement

L arrangement



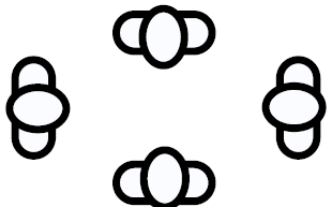
Side-by-side arrangement



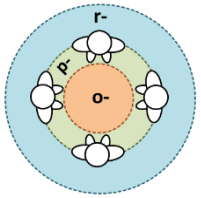
3+ people formations:

Semi-circular

Rectangular



Circular



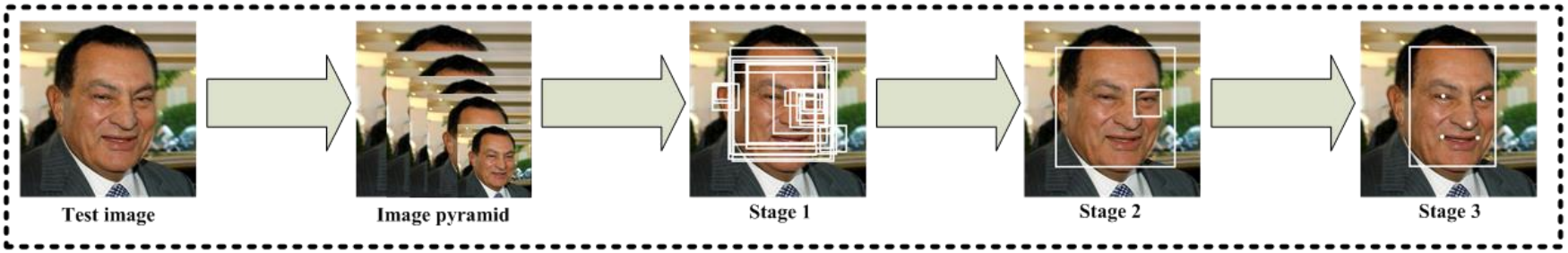
Setti et al. F-formation detection: individuating free-standing conversational groups in images. *PloS one*, *10*(5), 2015

Automatic visual behavior analysis

Automatic analysis of visual behavior

- Face detection
- Face tracking
  - Facial landmark detecion
- Head pose
- Eye gaze tracking
- Facial expression analysis
- Body pose tracking

Face Detection – Multi-Task CNN [SPL 2016]



Stage 1: candidate windows are produced through a fast Proposal Network

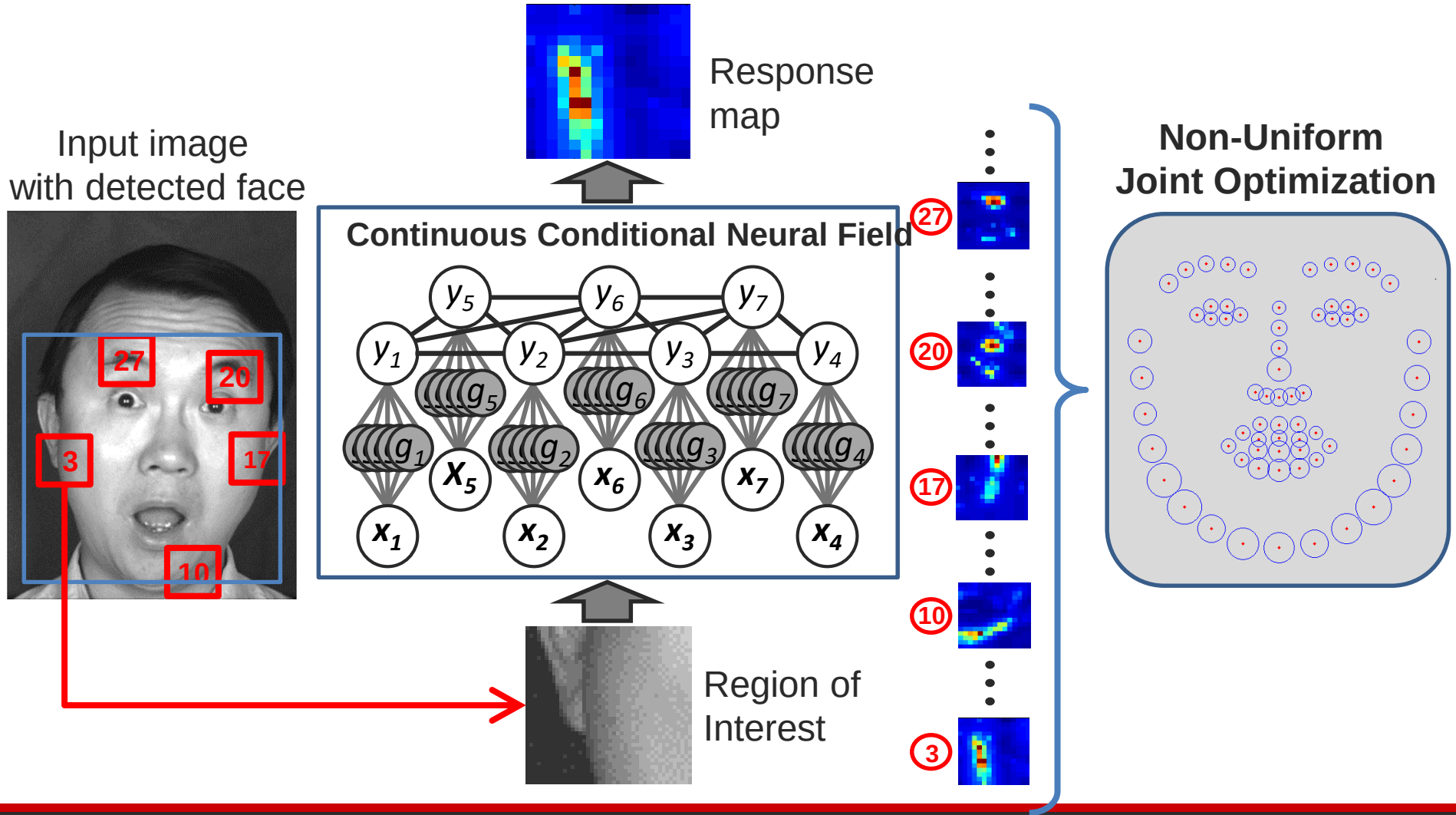
Stage 2: refine these candidates through a Refinement Network

Stage 3: produces final bounding box and facial landmarks position

Existing software (face detection)

- Multi-Task CNN face detector
  - https://kpzhang93.github.io/MTCNN_face_detection_alignment/index.html
- OpenCV (Viola-Jones detector)
- dlib (HOG + SVM)
  - http://dlib.net/
- Tree based model (accurate but very slow)
  - http://www.ics.uci.edu/~xzhu/face/
- HeadHunter (accurate but slow)
  - http://markusmathias.bitbucket.org/2014_eccv_face_detection/
- NPD
  - http://www.cbsr.ia.ac.cn/users/scliao/projects/npdface/

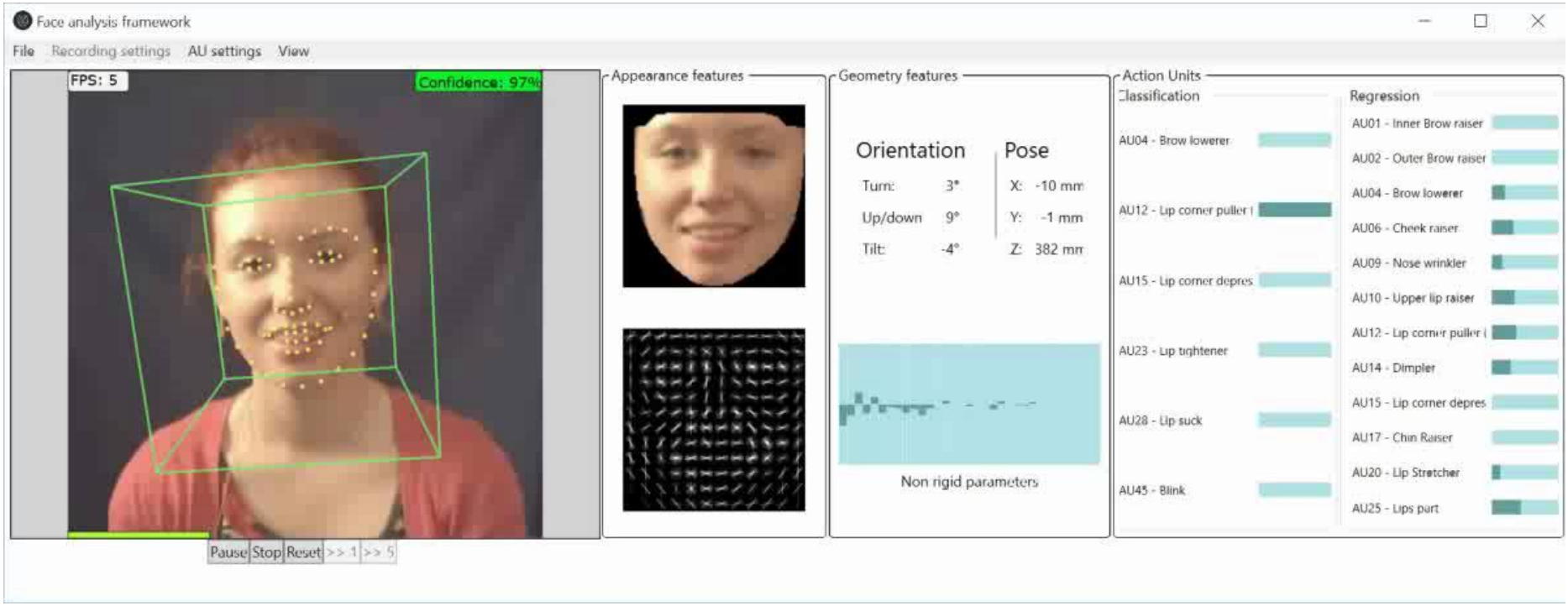
Facial Landmarks: Constrained Local Neural Field



Existing software (facial landmarks)

- OpenFace: facial features
  - https://github.com/TadasBaltrusaitis/OpenFace
- Chehra face tracking
  - https://sites.google.com/site/chehrahome/
- Menpo project (good AAM, CLM learning tool)
  - http://www.menpo.org/
- IntraFace: Facial attributes, facial expression analysis
  - http://www.humansensing.cs.cmu.edu/intraface/
- OKAO Vision: Gaze estimation, facial expression
  - http://www.omron.com/ecb/products/mobile/okao03.html (Commercial software)
- VisageSDK
  - http://www.visagetechnologies.com/products/visagesdk/
  - (Commercial software)

Facial expression analysis



[**OpenFace: an open source facial behavior analysis toolkit**, T. Baltrušaitis et al., 2016]

Existing Software (expression analysis)

- OpenFace: Action Units
  - https://github.com/TadasBaltrusaitis/OpenFace
- Shore: facial tracking, smile detection, age and gender detection
  - http://www.iis.fraunhofer.de/en/bf/bsy/fue/isyst/detektion/
- FACET/CERT (Emotient API): Facial expression recognition
  - http://imotionsglobal.com/software/add-on-modules/attention-tool-facet-module-facial-action-coding-system-facs/ (Commercial software)
- Affdex
  - http://www.affectiva.com/solutions/apis-sdks/
  - (commercial software)

Gaze Estimation – Eye, Head and Body

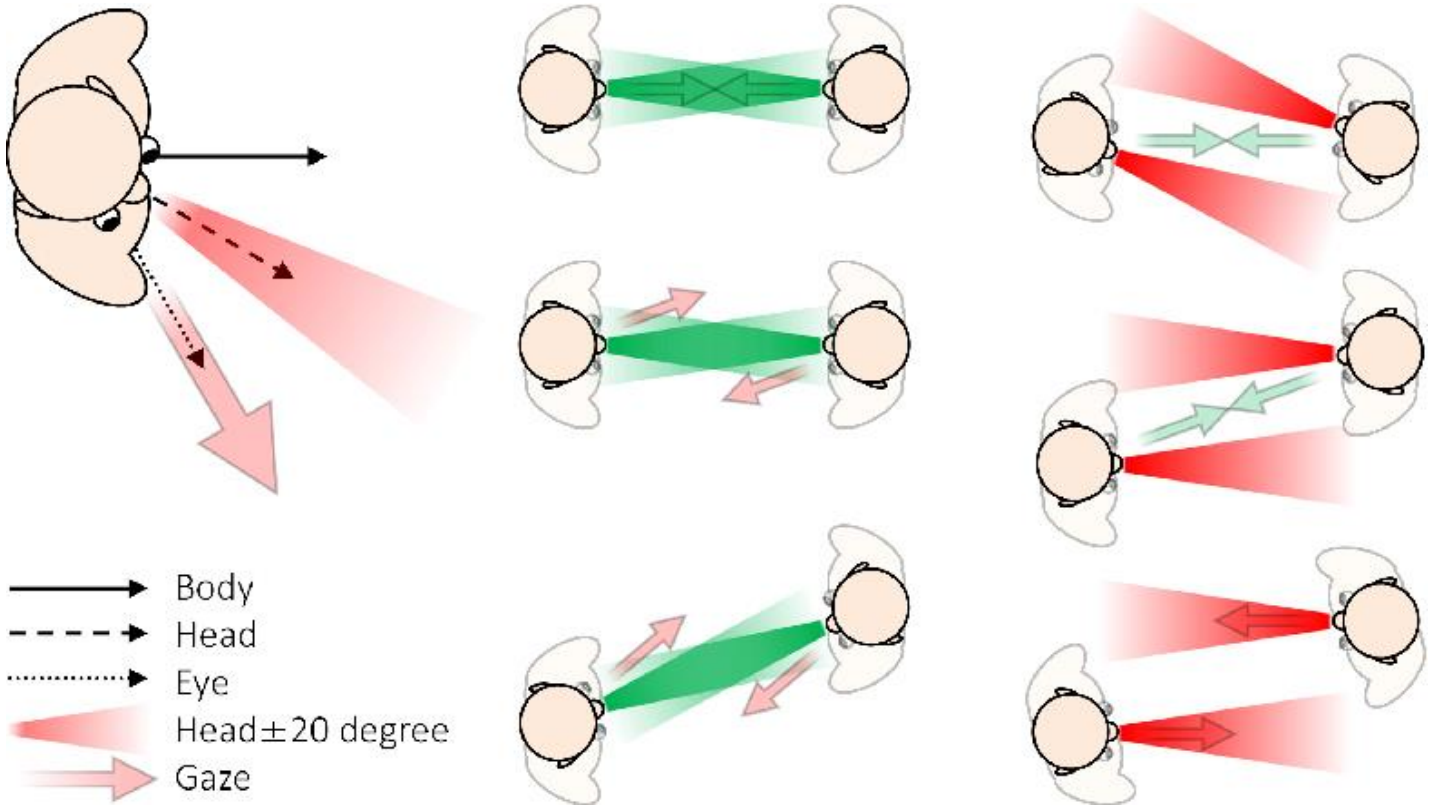
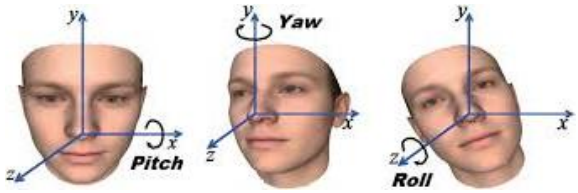


Image from Hachisu et al (2018). FaceLooks: A Smart Headband for Signaling Face-to-Face Behavior. *Sensors*.

Existing Software (head gaze)



- OpenFace
  - https://github.com/TadasBaltrusaitis/OpenFace
- Chehra face tracking
  - https://sites.google.com/site/chehrahome/
- Watson: head pose estimation
  - http://sourceforge.net/projects/watson/
- Random forests
  - http://www.vision.ee.ethz.ch/~gfanelli/head_pose/head_forest.html
  - (requires a Kinect)
- IntraFace
  - http://www.humansensing.cs.cmu.edu/intraface/

Existing Software (eye gaze)

- OpenFace: gaze from a webcam
  - https://github.com/TadasBaltrusaitis/OpenFace
- EyeAPI: eye pupil detection
  - http://staff.science.uva.nl/~rvalenti/
- EyeTab
  - https://www.cl.cam.ac.uk/research/rainbow/projects/eyetab/
- OKAO Vision: Gaze estimation, facial expression
  - http://www.omron.com/ecb/products/mobile/okao03.html (Commercial software)

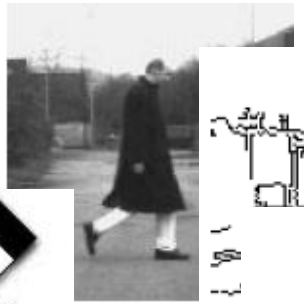
Articulated Body Tracking: OpenPose

Existing Software (body tracking)

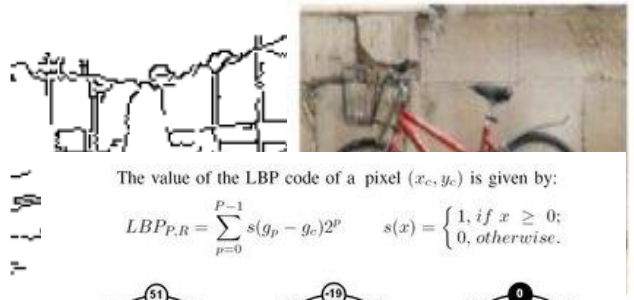
- OpenPose
  - https://github.com/CMU-Perceptual-Computing-Lab/openpose
- Microsoft Kinect
  - http://www.microsoft.com/en-us/kinectforwindows/
- OpenNI
  - http://openni.org/
- Convolutional Pose Machines
  - https://github.com/shihenw/convolutional-pose-machines-release

Visual Descriptors

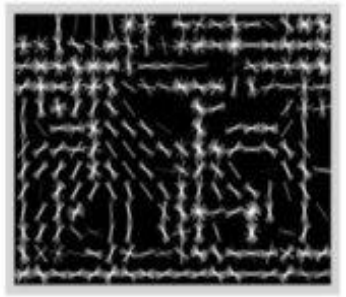
I



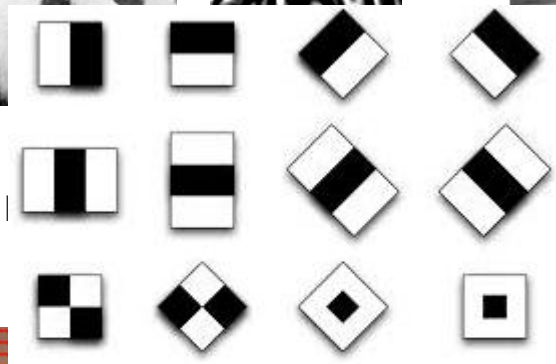
Edge

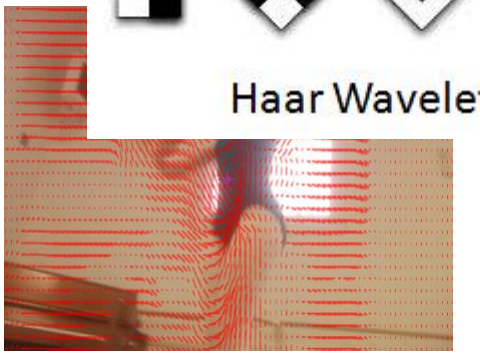


LBP

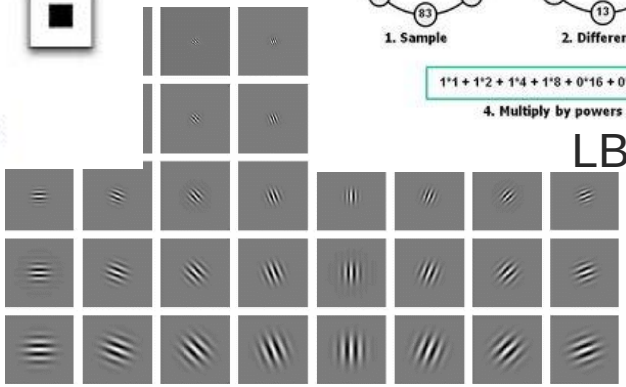


Oriented Gradients

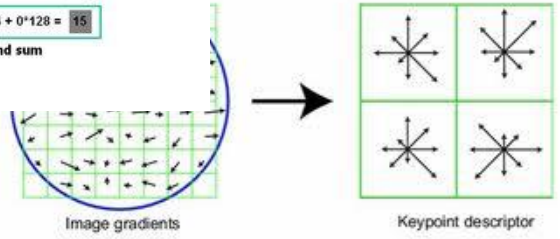




Optical Flow



Gabor Jets



SIFT descriptors

Existing Software (visual descriptors)

- OpenCV: optical flow, gradient, Haar filters…
- SIFT descriptors
  - http://blogs.oregonstate.edu/hess/code/sift/
- dlib – HoG
  - http://dlib.net/
- OpenFace: Aligned HoG for faces
  - https://github.com/TadasBaltrusaitis/CLM-framework