Affective Computing

Olga Perepelkina
HSE, 2020

Тема 4. Эмоции и невербальная коммуникация

План лекции

- Проявление эмоций через телесное невербальное поведение
- Влияние невербального поведения на эмоциональное состояние
- Доклад
- Защиты тем групповых проектов
- Системы трекинга движений (motion tracking) и автоматическое распознавание
- Практика OpenPose

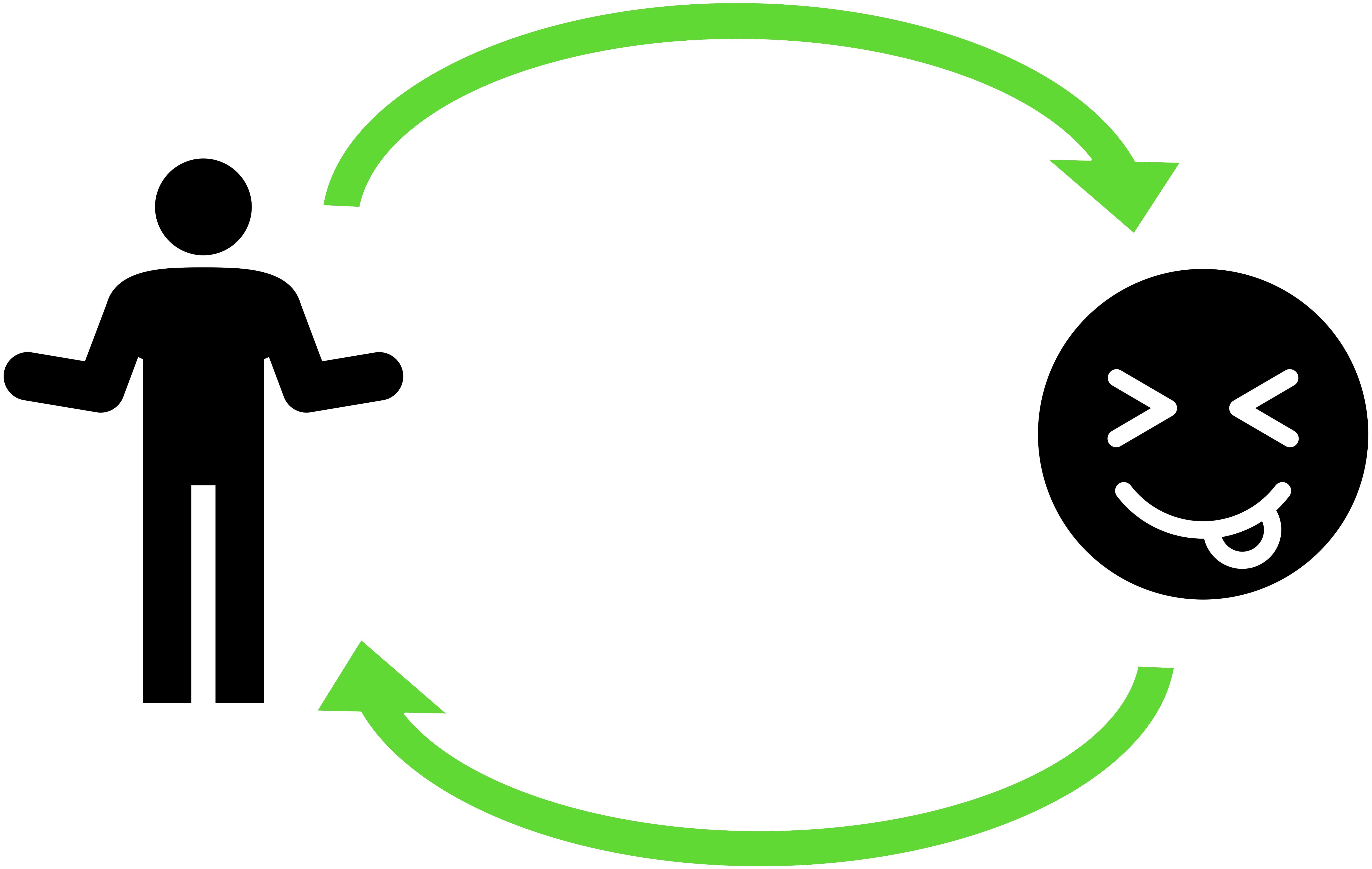
Выражение эмоций с помощью тела
(bodily communication of emotion)

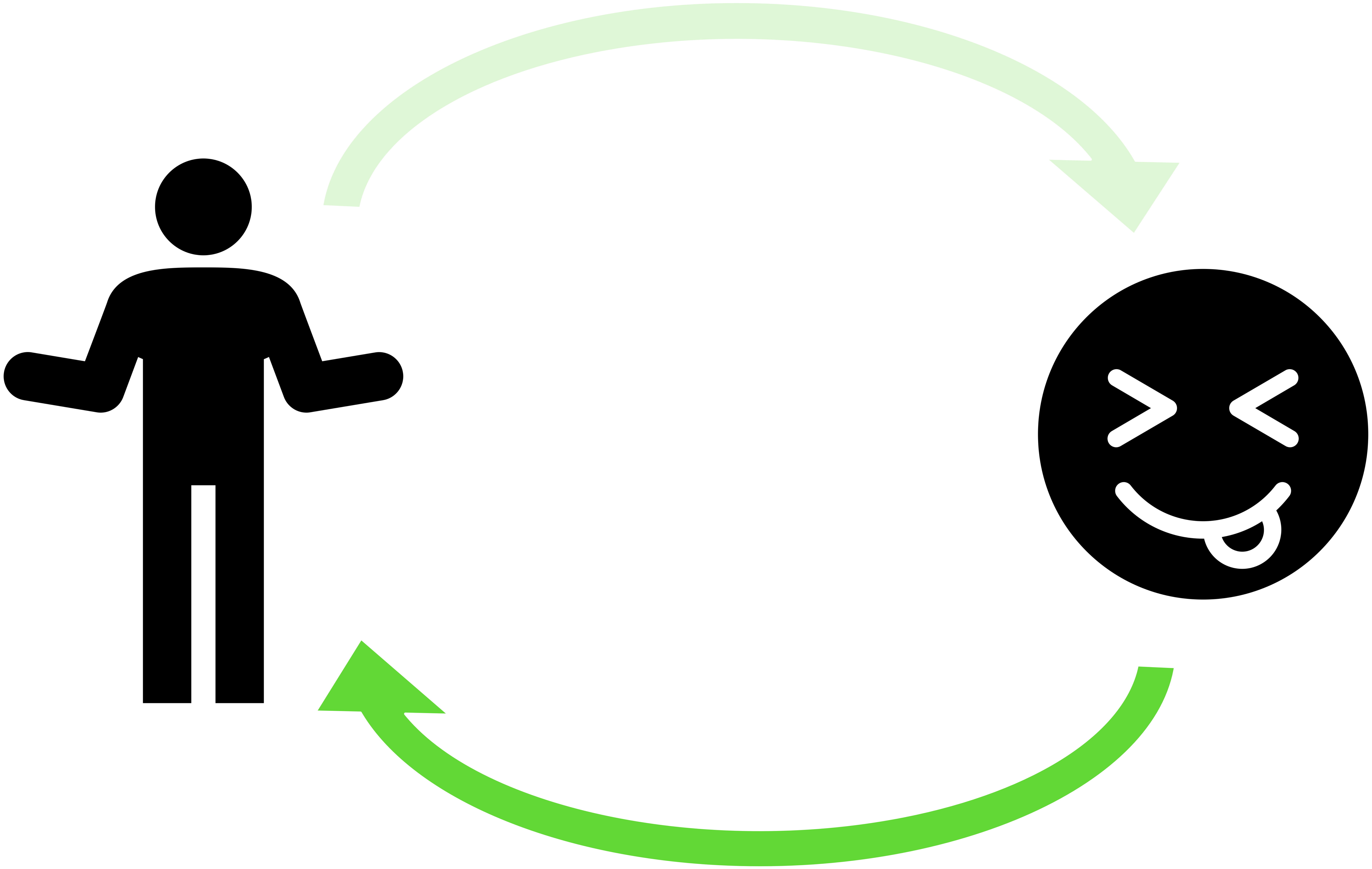
- Тезис о том, что тело участвует в передаче эмоциональной информации, появился еще с работ Дарвина (1872) и Джеймса (1932)

- Затем акцент в изучении эмоций был смещен на выражение лиц

Witkower, Z., & Tracy, J. L. (2018). Bodily Communication of Emotion: Evidence for Extrafacial Behavioral Expressions and Available Coding Systems. *Emotion Review*, 175407391774988O.


RAH SIEGEL-MAGNESS





Bodily communication of emotion

- Важны не менее лицевой экспрессии (facial communication)
- Телесные выражения могут быть визуально различимы из-далека, позволяя выражать эмоции на **дальние дистанции** [de Gelder, 2009; de Gelder, 2016; Martinez, Falvello, Aviezer, & Todorov, 2015]
- Могут быть различимы со спины выражающего, включая моменты, когда лицо **полностью скрыто** от обзора [Coulson, 2004; Sogon & Masutani, 1989]
- Как и в случае с лицевой экспрессией, эмоции по телу распознаются с вероятностью, **выше случайной** (Atkinson et al., 2004; de Gelder & van den Stock, 2011; Tracy & Robins, 2007).

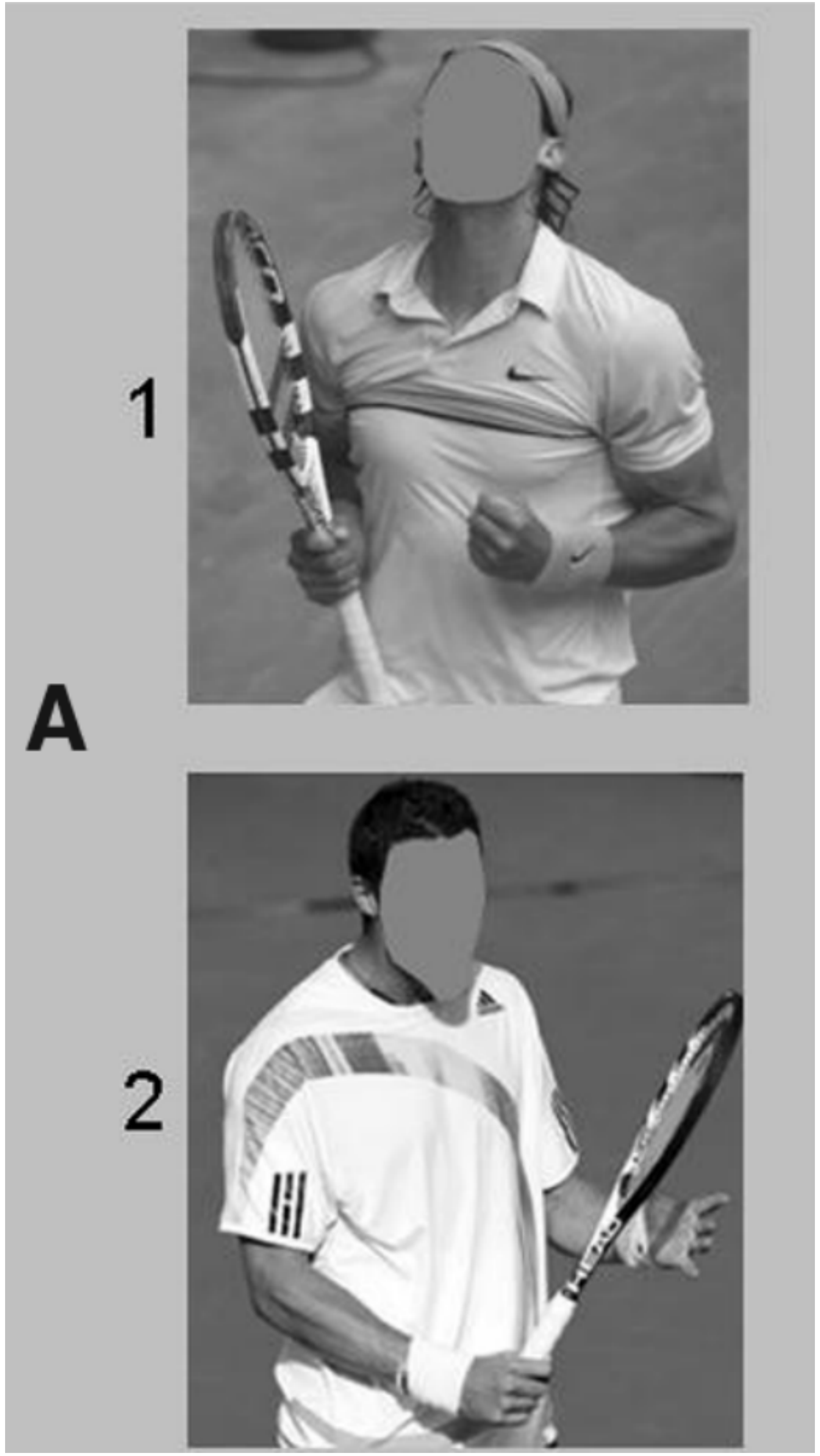
Bodily communication of emotion

- Позы тела и жесты рук обладают значительно большей **степенью свободы** по сравнению с жестами лица. Позы тела имеют неограниченный **словарный запас**, который включает в себя различные комбинации движений частей тела [Gunes, H. et al., 2015].

- Движения тела меньше подвержены **сознательному контролю** и социальной «редакции» по сравнению с лицом [Karg, M., et al., 2013].

- В случае с **сильными эмоциями** распознавание эмоций по телу надежнее, чем по лицу [Aviezer, H., Trope, Y. and Todorov, A. 2012].


1
2
3
B
4
5
6


A
1
2

*
*
ns
2.50
2.00
1.50
1.00
0.50
0.00
-0.50
-1.00
-1.50
-2.00
-2.50
Mean valence
2.0
1.9
-1.0
-0.9
-0.7
-0.5
Lose
Win
Face + Body
Body
Face

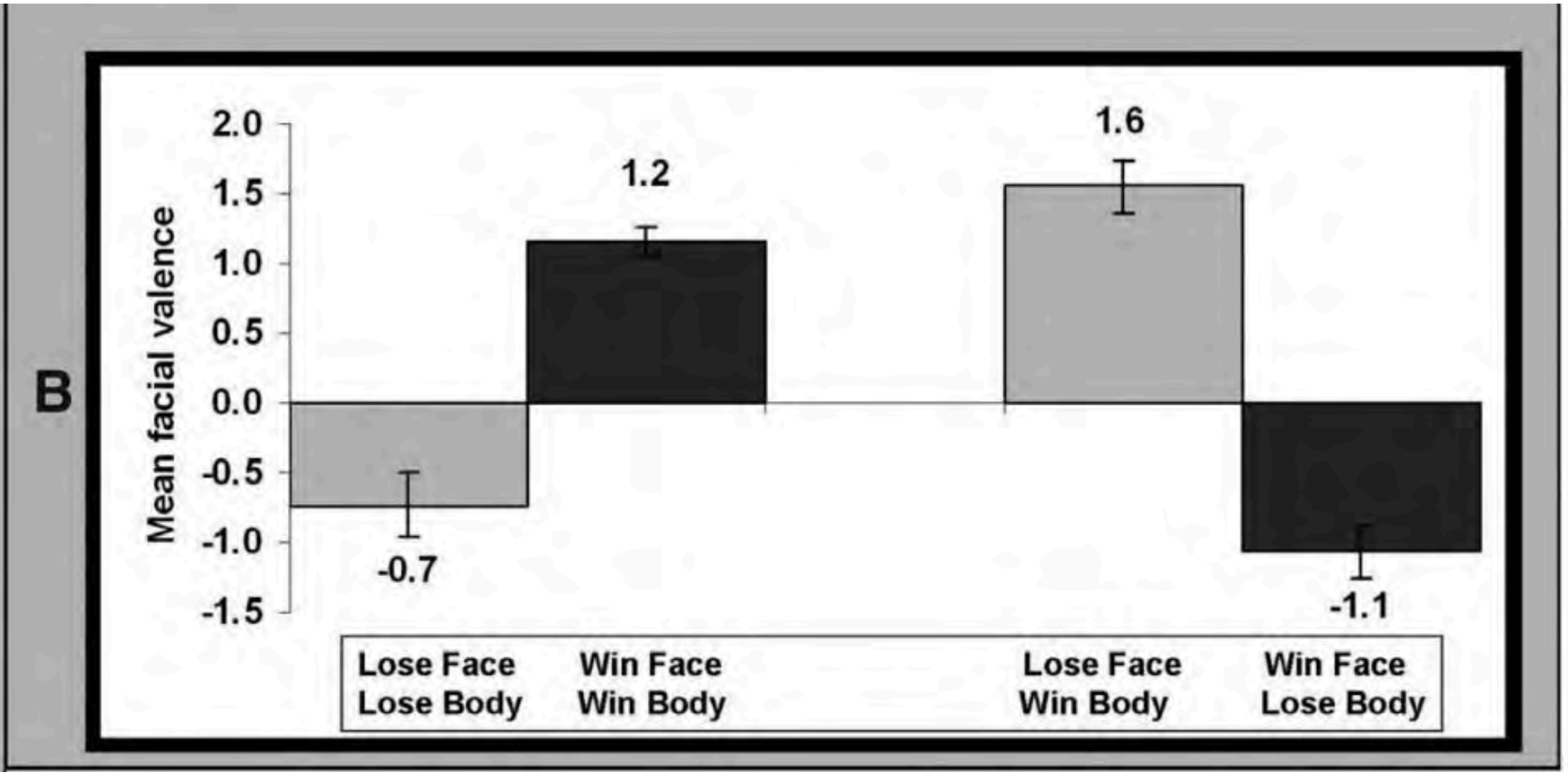


Fig. 2. Experiment 2. (**A**) Examples of original images of players (1) losing or (2) winning a point. The same faces combined with incongruent-valence bodies such as (3) a losing face on a winning body and (4) a winning face on a losing body. [All photos in Fig. 2 credited to a.s.a.p. Creative/Reuters] (**B**) Mean valence ratings of the facial expressions.

Невербальное поведение

- Behavior **types** (including spatial form; leaning back, tilting the head, hands in fists).
- Behavior **qualities** (spatiotemporal form, kinematics, dynamics; e.g., fast, slow, jerky, flowing).

- Динамические стимулы, позволяющие воспринимать и тип поведения, и его качество, получают наиболее высокие значения восприятия (recognition rates) [Atkinson et al., 2004].

- Однако статические изображения тел, показывающие только **тип** поведения, а также динамическое изображение светящихся точек (**качество**) также позволяют распознавать эмоции с вероятностью выше случайной [Atkinson et al., 2004; de Gelder & van den Stock, 2011; Tracy & Robins, 2007]/

Невербальная коммуникация эмоций

- **Encoding**: *выражение* своих эмоций с помощью невербальной информации
- **Decoding**: использование невербальных сигналов для *интерпретации* эмоциональной информации, которую выражают другие

Распознавание эмоций по телесному поведению (bodily behaviors)

- Некоторые эмоции могут проявляться (displayed) и распознаваться (recognized) по определенным типам телесного поведения. Среди этих эмоций:
  - Pride
  - Joy
  - Sadness
  - Shame
  - Embarrassment (смущение)
  - Anger
  - Fear
  - Disgust
- de Gelder & van den Stock, 2011; Keltner, 1995; Tracy, Robins, & Schriber, 2009

Pride

Fig. 3. Pride expression in response to victory shown by a sighted (*left*) and congenitally blind (*right*) athlete.

Joy/Happiness

- Several **bodily behaviors** have been identified as communicating joy or happiness: **upwards bodily movement** (i.e., with the arms, trunk, or shoulders), **upwards head tilt**, **illustrative gestures, opening and closing the hands, expansive bodily displays, and jumping** (e.g., Atkinson et al., 2004; Coulson, 2004; de Meijer, 1989).

- Behavior **quality** appears to be **important for happiness** displays as well; these movements tend to be **fast** and **energetic** (Dael et al., 2013).

- Several bodily behaviors associated with **happiness** (e.g., expansiveness) **overlap with those known to communicate pride**.

Sadness

- Studies have examined both the **type** and **quality** of bodily behaviors associated with sadness.

- Behavior **types** include **slumped shoulders** and a **collapsed upper body**, **downwards head tilt**, **arms in front of the body**, and the **head cradled in hands** (e.g., Parkinson et al., 2017; Sawada et al., 2003; Wallbott & Scherer, 1986; see Table 1).

- Behaviors tend to be **slow** and include **less overall movement**.

- These behaviors have been documented in both **encoding** and **decoding** studies.

Shame and embarrassment

- As is the case for pride, the bodily expressions of shame and embarrassment are particularly important when compared to their facial expressions.
- **Facial expressions alone cannot capture the complex message sent by shame expressions** (Tracy & Robins, 2007), and both **encoders and decoders utilize bodily cues more than facial cues when communicating shame** (App et al., 2011; Tracy & Matsumoto, 2008; Tracy et al., 2009).

Shame and embarrassment

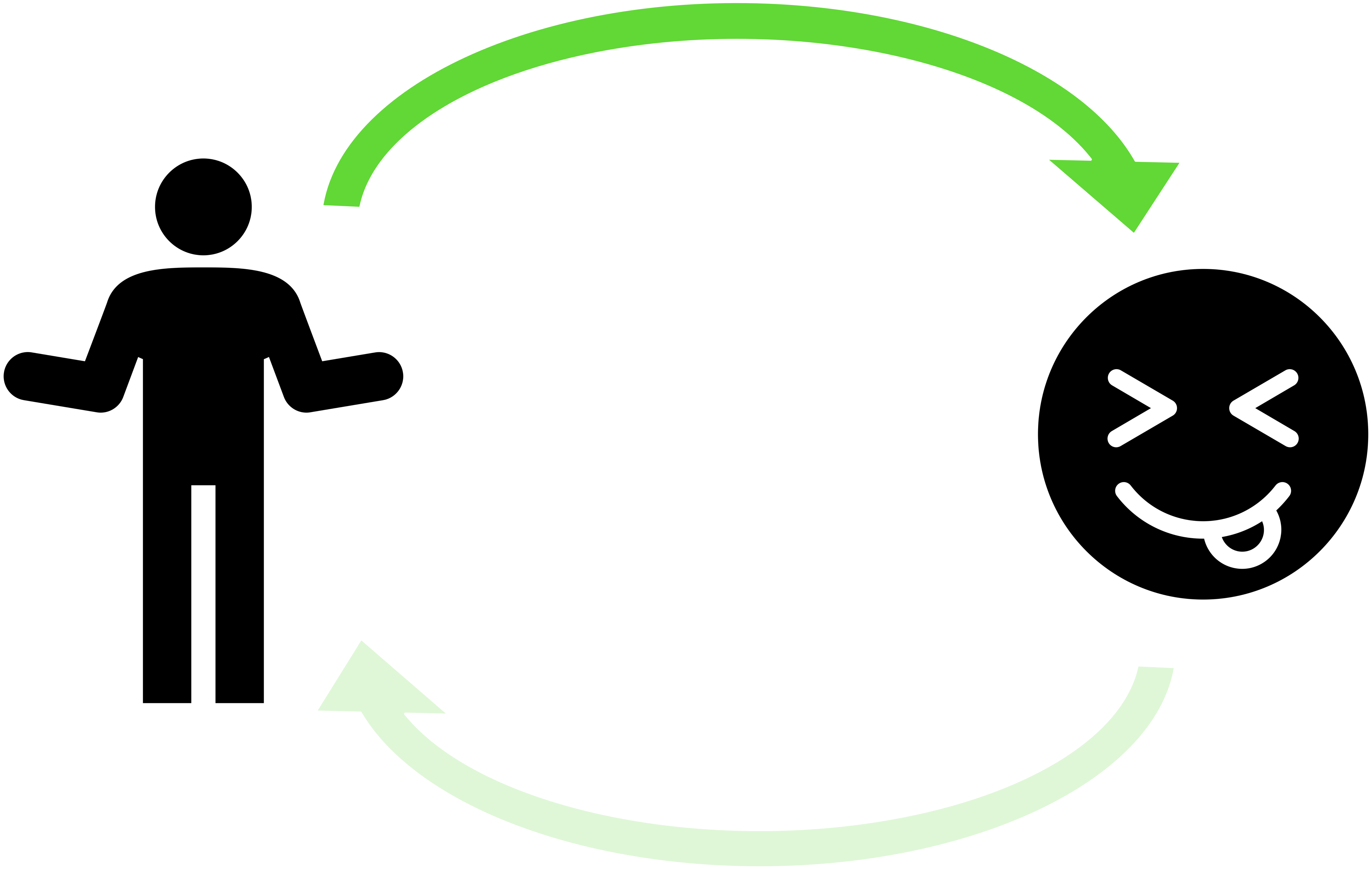
- Olympic-level **judo athletes** from numerous countries spontaneously displayed a head tilted down, shoulders slumped, and chest narrowed after losing a match—presumably an intense shame experience.
- Narrowed chest and slumped shoulders were displayed even by **congenitally blind athletes**, making it unlikely that the shame expression is a learned response to failure.
- However, these blind athletes **did not display a downwards head tilt**, raising questions about the innateness or universality of that particular behavior in encoding **shame**.

Limitations in Research Identifying Bodily Expressions

- although many of the emotions found to have bodily expressions are theoretically and empirically distinct (in terms of self-report, facial expression, eliciting cognitions, etc.), their bodily expressions **overlap**.
- Distinct bodily expressions—often including facial components—have been identified for only certain emotions.

Limitations in Research Identifying Bodily Expressions

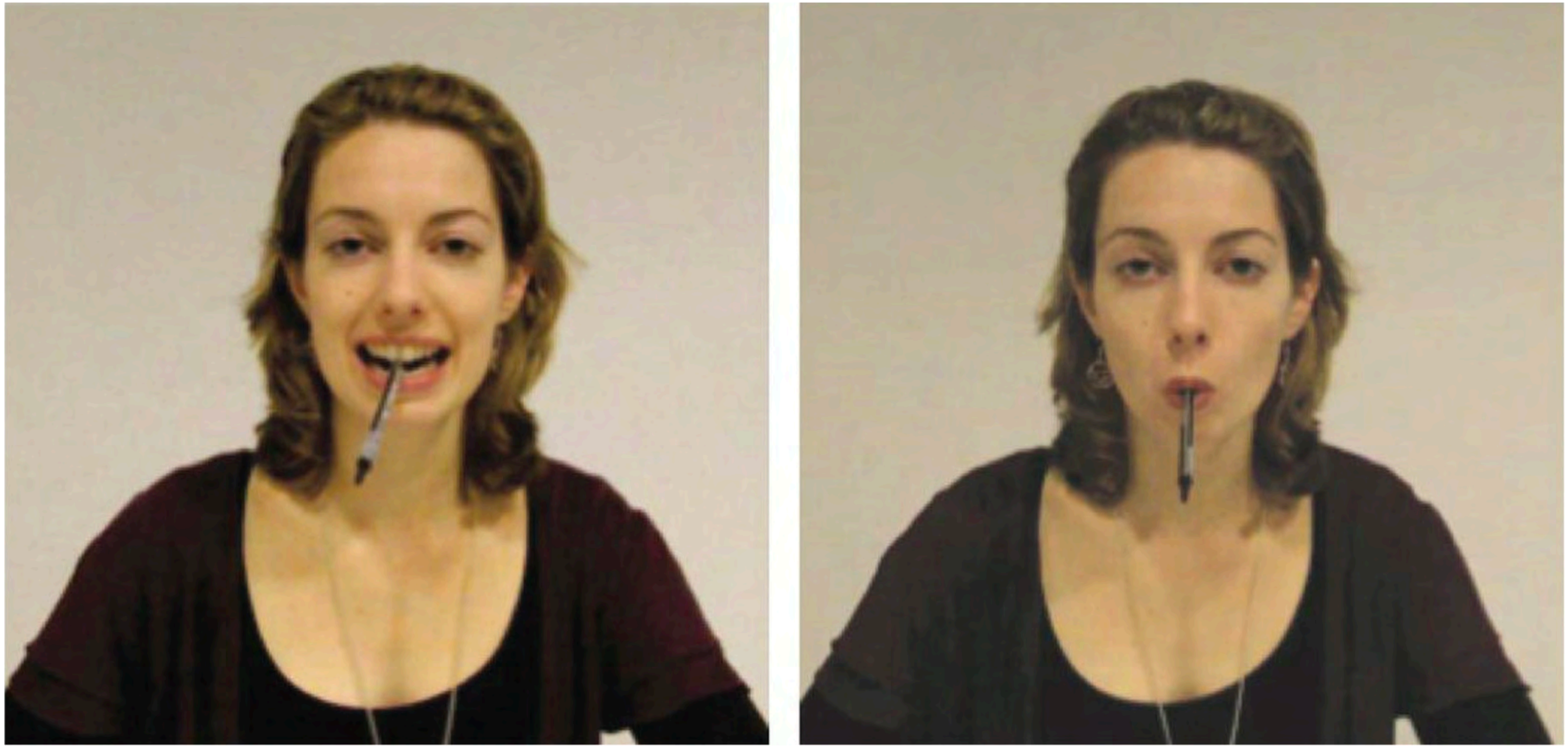
- A second limitation is the imprecise way in which bodily behaviors tend to be measured, resulting in the loss of valuable information. For example, several studies identify head tilt as a behavior that communicates emotion, but the head is a complex body part that can move in many ways, including turning (head yaw/rotation around the vertical axis/“no” gesture), tilting (head pitch/rotation around the frontal axis/“yes” gesture), rolling (rotation around the sagittal axis; e.g., raising an ear to the sky), and backward versus forward (e.g., neck extension forward and backwards)



Facial Feedback гипотеза

- **Дарвин**: выражение эмоции усиливает ее; подавление эмоции – ослабляет.
- **Джеймс**: не выражай страсть, и она умрет. Паника усиливается бегством. Досчитай до 10 и причина гнева покажется смешной.
- **Теория Джеймса-Ланге**: стимул -> физиологическая реакция -> эмоциональное переживание

Позы лица и эмоции



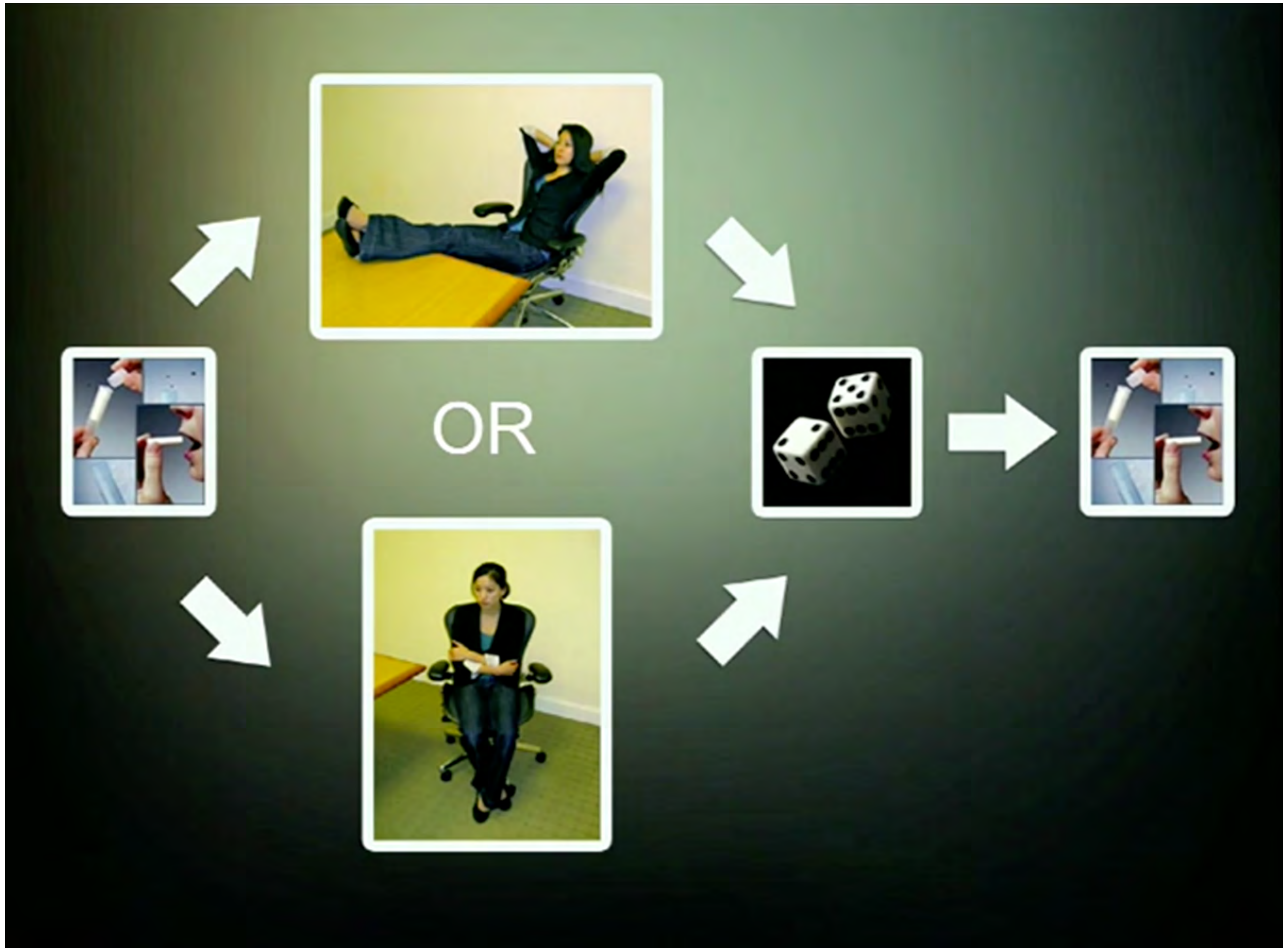
Strack, Martin & Stepper, 1988

Позы тела и эмоции

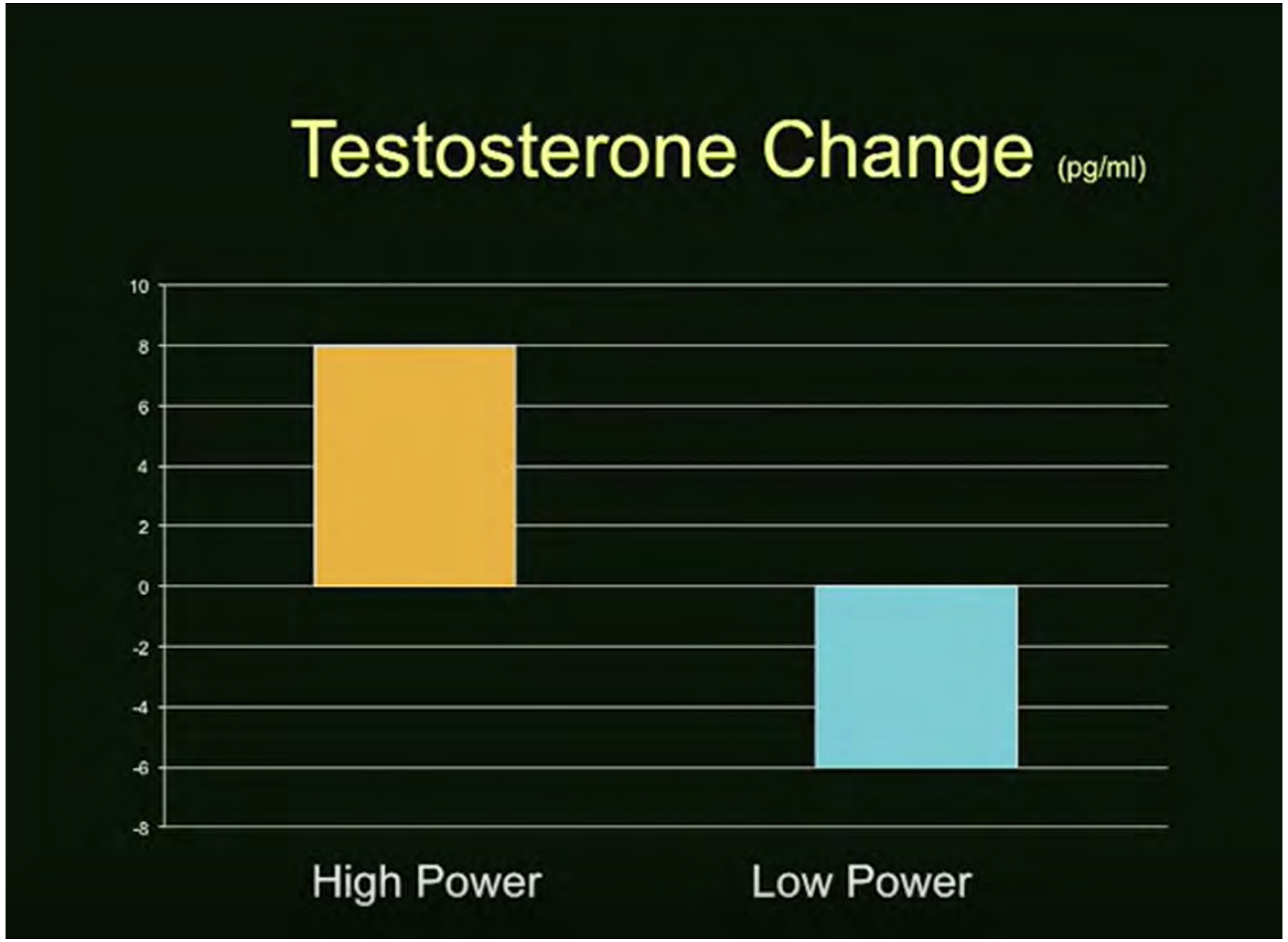
D. R. Carney , A. J.C. Cuddy , A. J. Yap, 2010

Позы тела и эмоции

Позы тела и эмоции



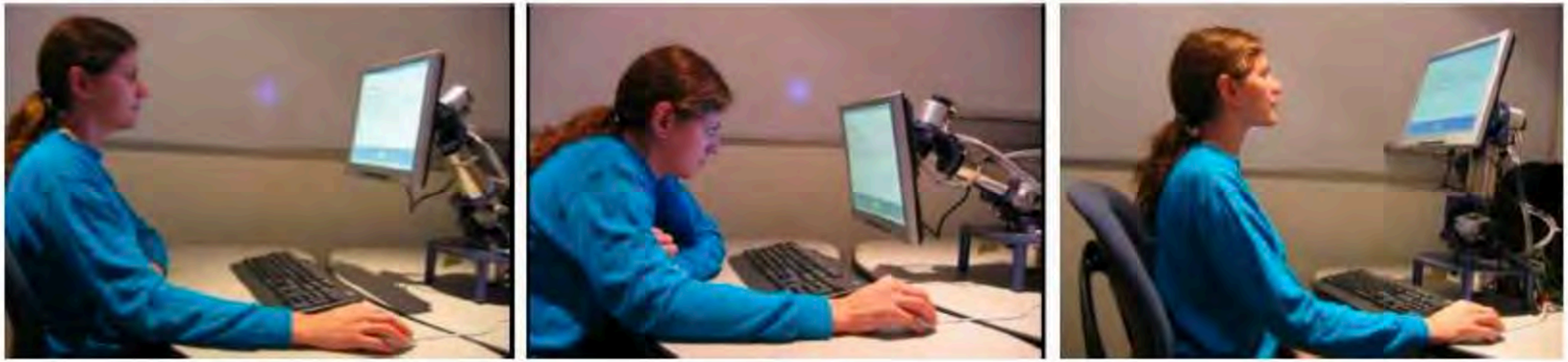
Позы тела и эмоции



Приложение в Affective Computing

Idea: shape the body to shape emotion

- Explore if computer's "posture" influences its user's posture
- Determine if interaction of user's posture with their affective state leads to improved persistence in problem solving



- Users adjusted posture as they performed the tasks
- And solved more problems when they were upright

Stoop to Conquer: Posture and Affect Interact to Influence Computer Users' Persistence
Hyung-il Ahn, Alea Teeters, Andrew Wang, Cynthia Breazeal, Rosalind Picard

Относиться к исследованиям осторожно!

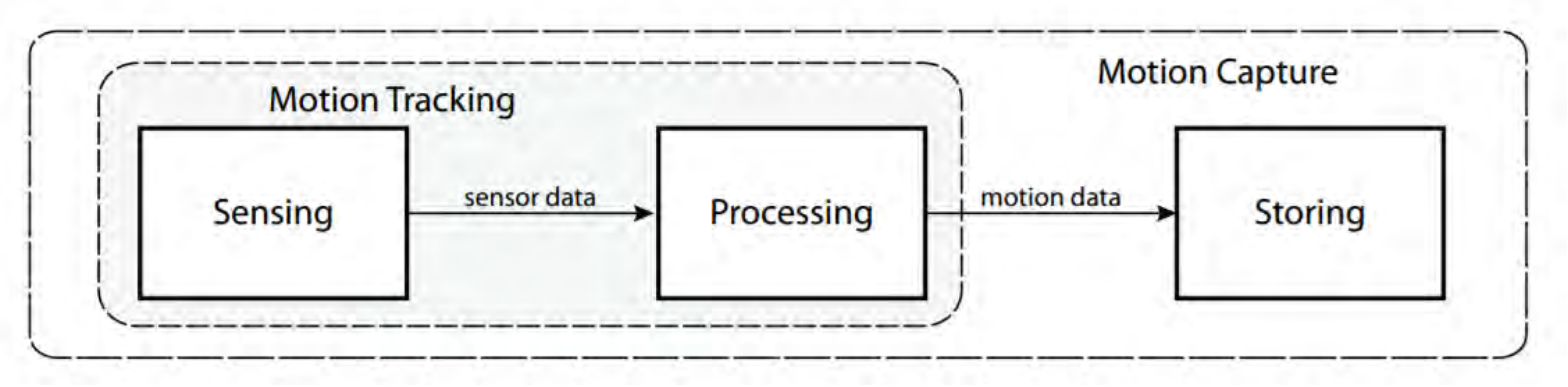
???

Registered Replication Report: Strack, Martin, & Stepper (1988) E.-J. Wagenmakers*, T. Beek*, L. Dijkhoff*L. Dijkhoff, **2016**

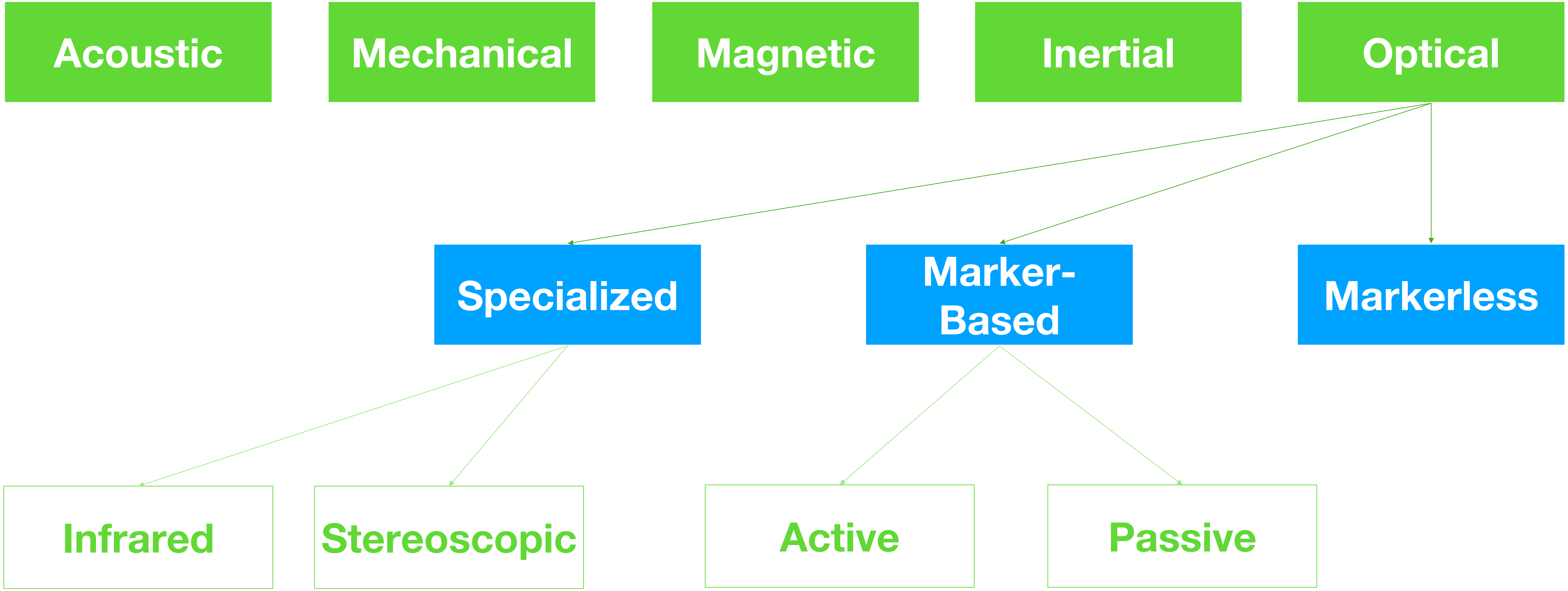
Доклад

Motion Tracking

Motion capture (mocap) & Motion tracking

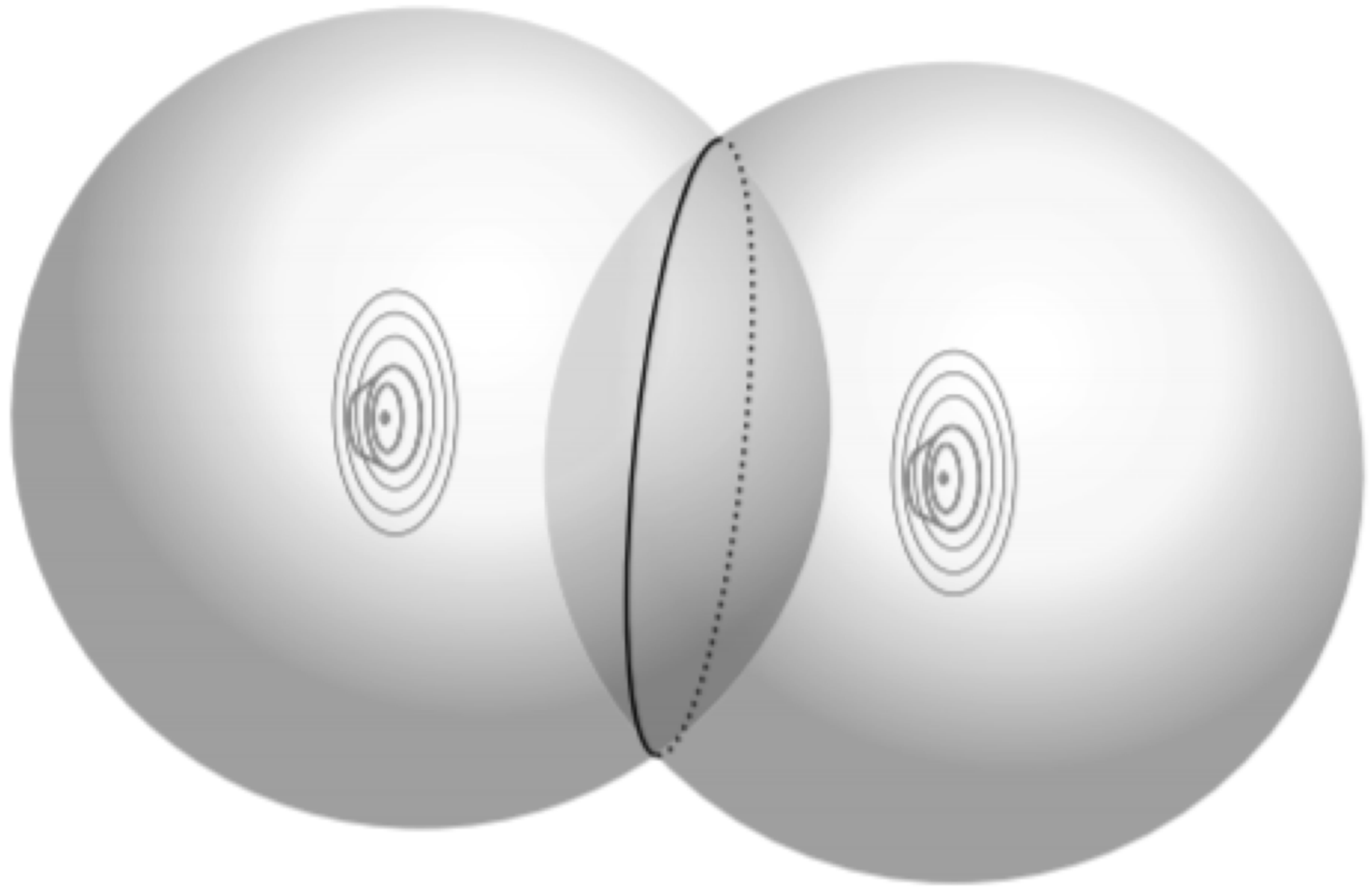
- 

Tracking Types



Acoustic Tracking

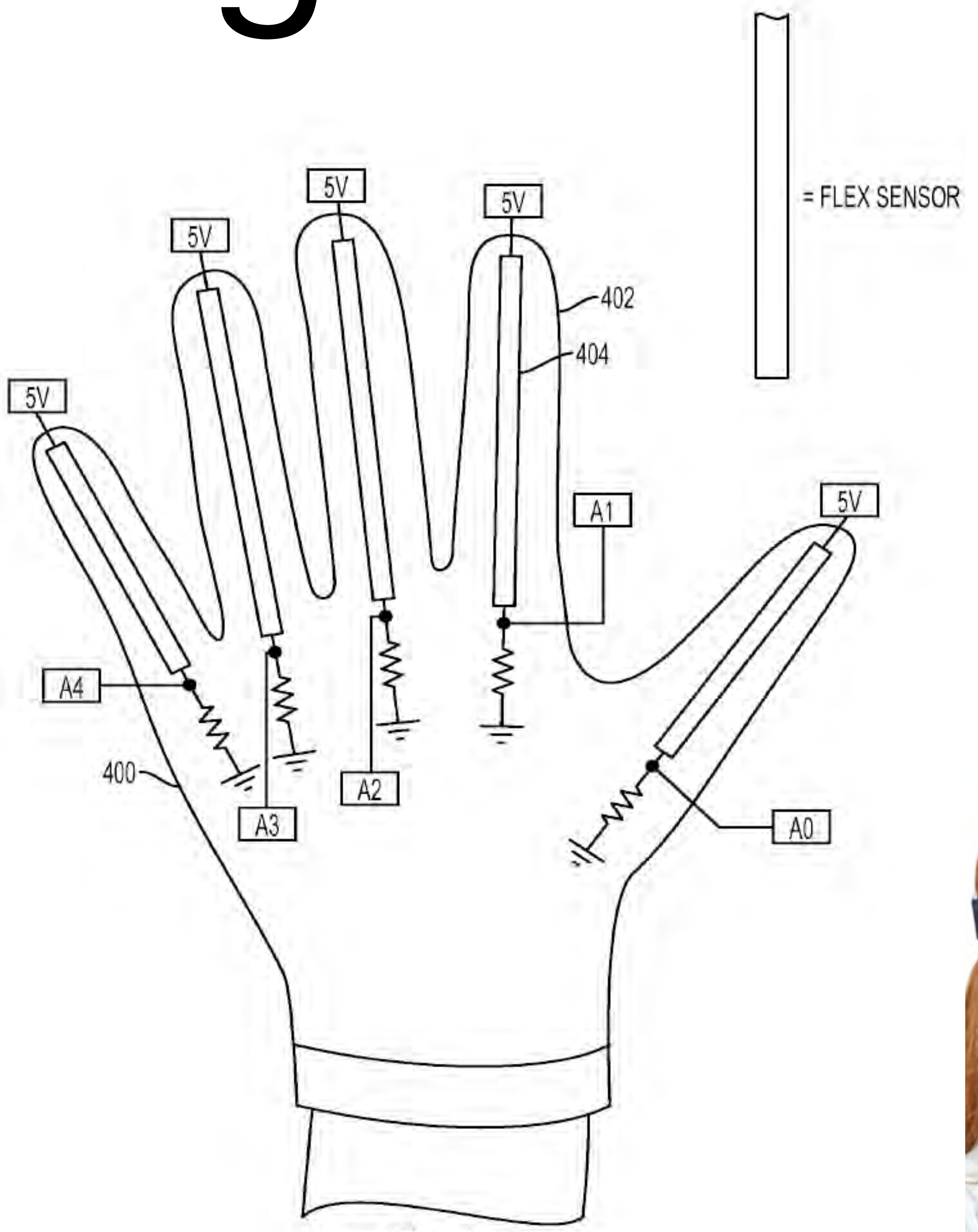
- = Ultrasonic Tracking
- Acoustic tracking systems calculate position upon the wavelength of an acoustic signal and the speed of sound.
- Systems based on time of flight measure the time between the sending of a signal from a transmitter and its being picked up by a receiver.



Distance measurements from two acoustic transmitters can determine the position of a receiver to be somewhere along a circle.

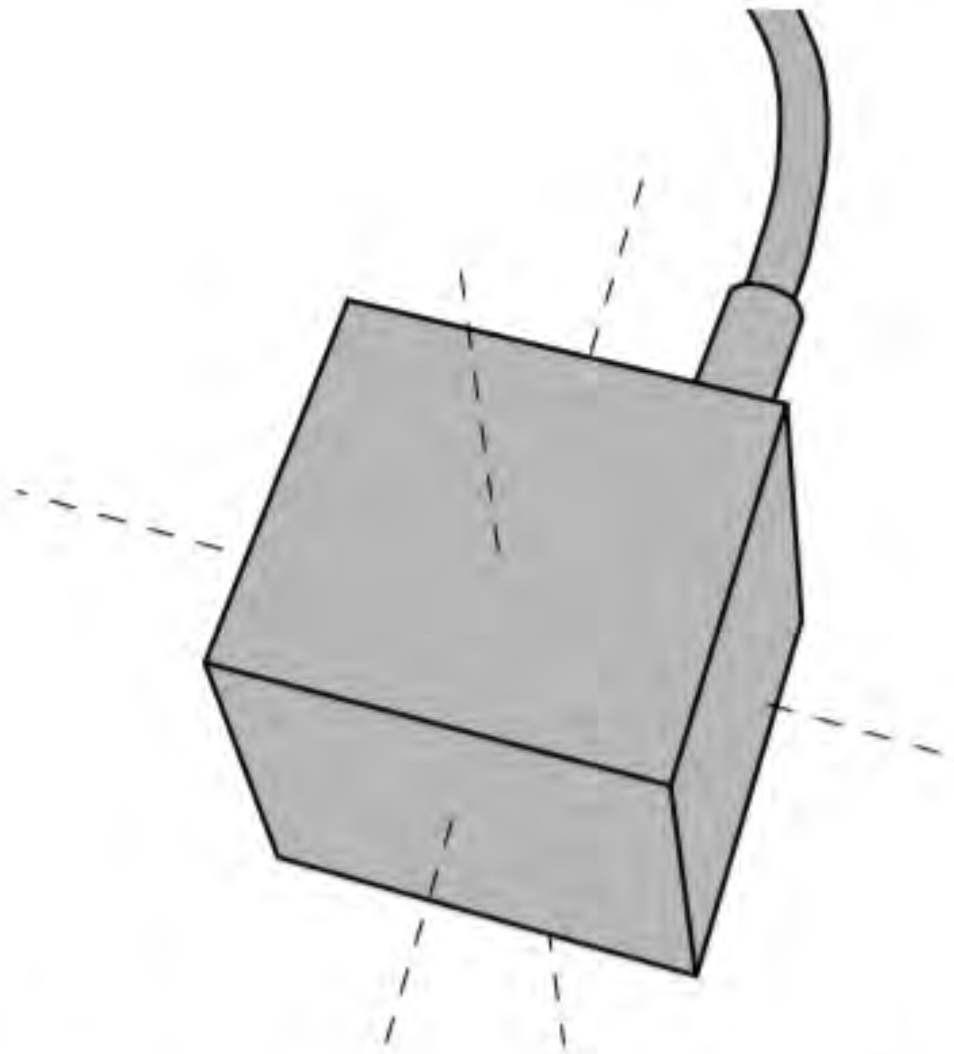
Mechanical Tracking

- Bend sensors: measure angles or lengths between the mechanical parts

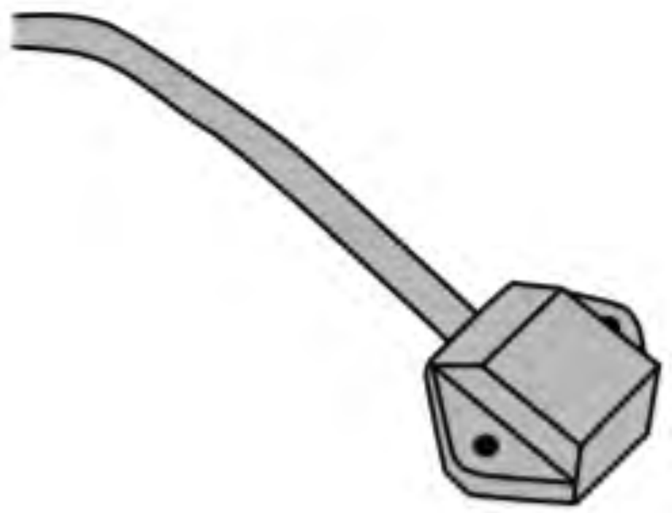


Magnetic Tracking

- Is based on the principle of induction, which explains how an electric current is induced in a coil when it is moved in a magnetic field



Source, sequentially setting up three perpendicular magnetic fields



Sensor, containing three perpendicular coils where voltages are induced by the magnetic fields from the source

Inertial Tracking

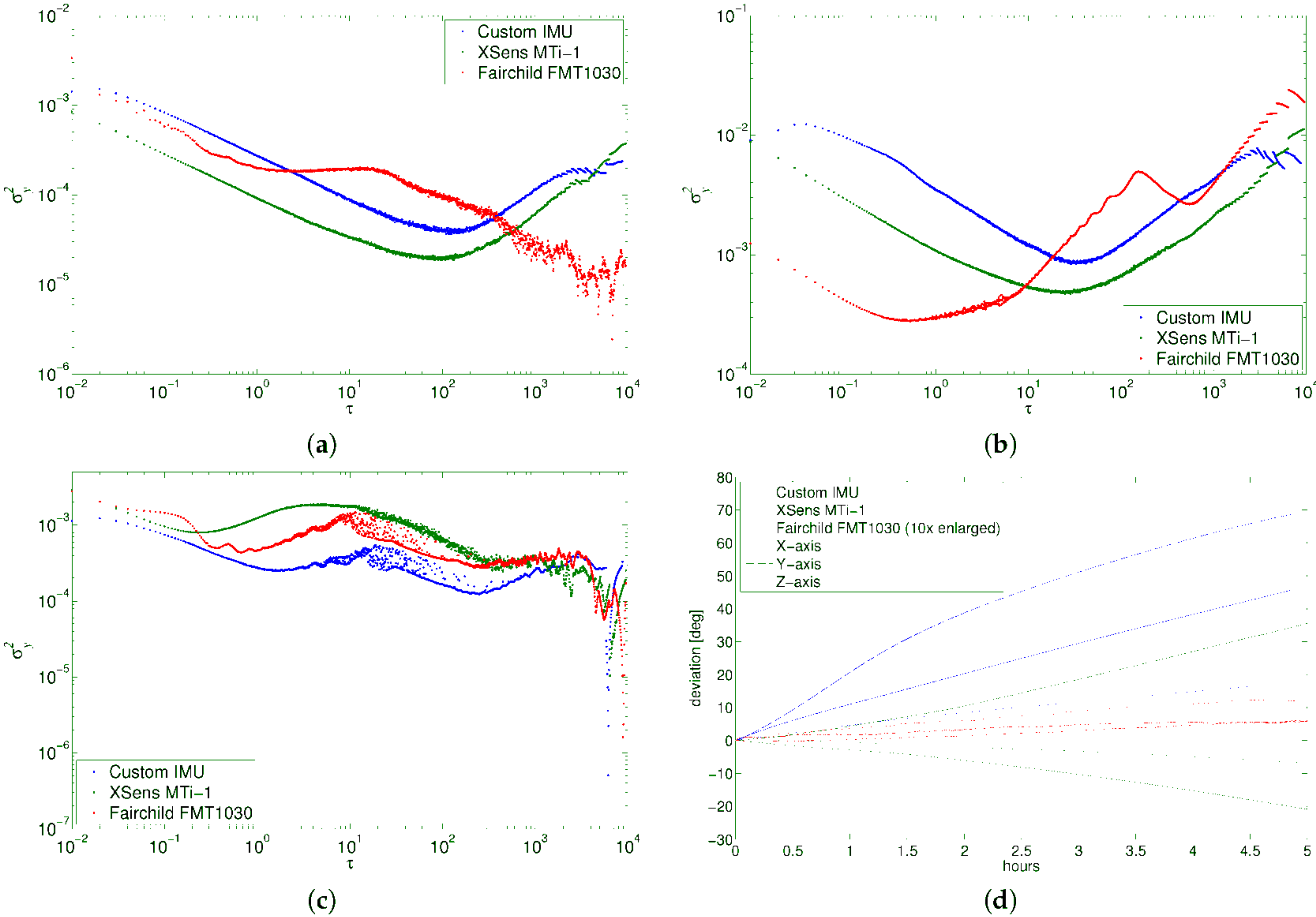
- Inertial tracking systems include those based on accelerometers and gyroscopes. These sensors are based on the physical principle of inertia.

The Xsens suit consists of 17 MTx sensors combining inertial sensors and magnetometers. Full body motion capture is obtained through the use of a kinematic model.

Inertial Tracking: Advantages

- Self-contained: they do not rely on external sources such as acoustic ultrasound sensors or cameras which require line-of-sight.
- The sensors rely on physical laws that are not affected by external factors such as ferromagnetic objects or light conditions.
- The sensors are very small and lightweight (useful in portable devices).
- The systems have low latencies and can be sampled at very high sampling rates.

Inertial Tracking: Drift

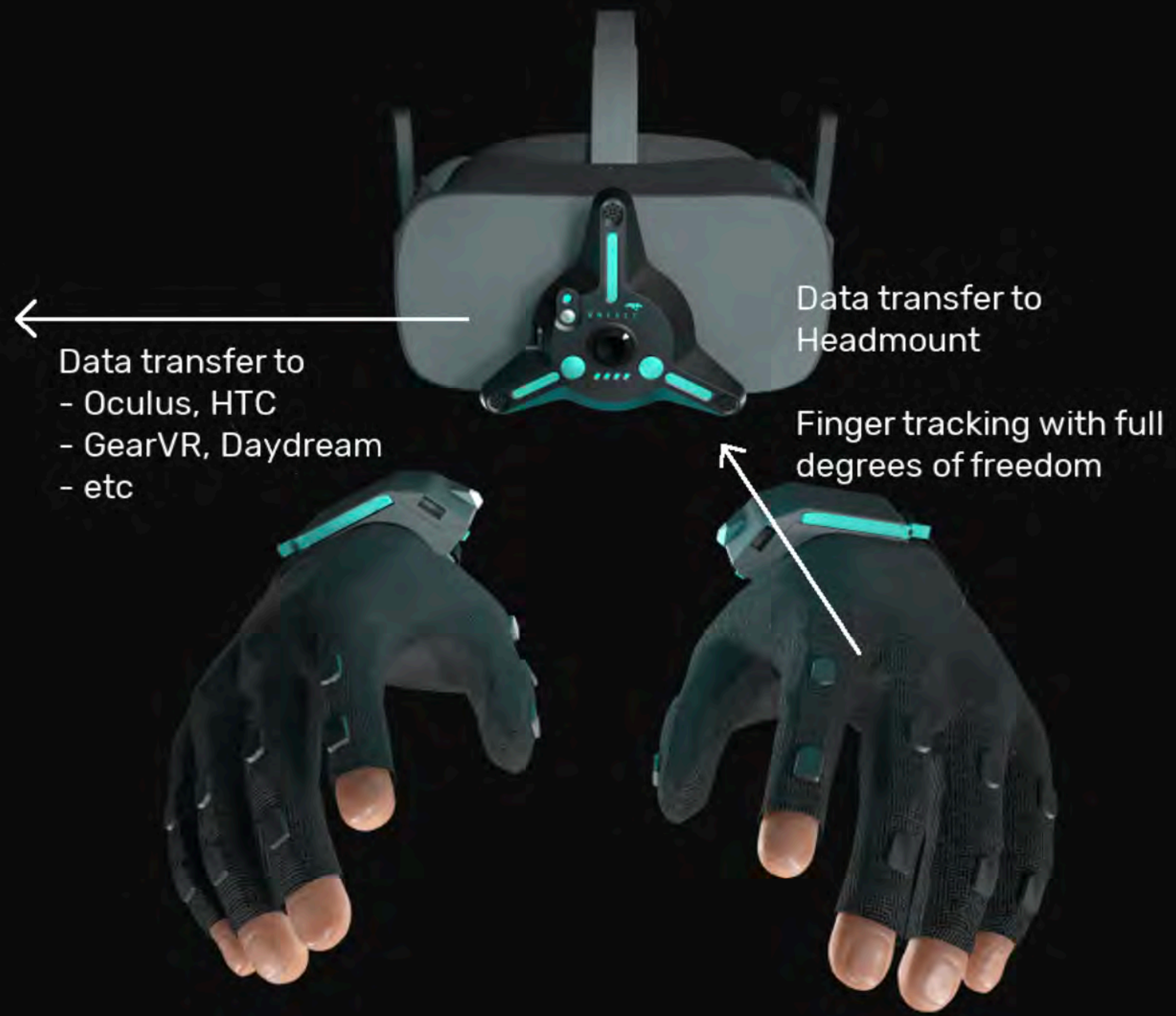


- Static experiment results: representative Allan variances for the tested IMUs: (a) gyroscopes; (b) accelerometers; (c) magnetometer (for yaw axis-Z); (d) drift resulting from gyroscope integration.
- Szczęsna A. et al. Inertial Motion Capture Costume Design Study //Sensors. – 2017. – T. 17. – №. 3. – С. 612.

Axis Neuron

- 3-axis accelerometer, 3-axis gyroscope sensor and 3-axis magnetometer
- Calculated orientation (quaternion)
- Dimension: 12.5mm*13.1mm*4.3mm
- Weight: 1.2 gram

Sensoryx

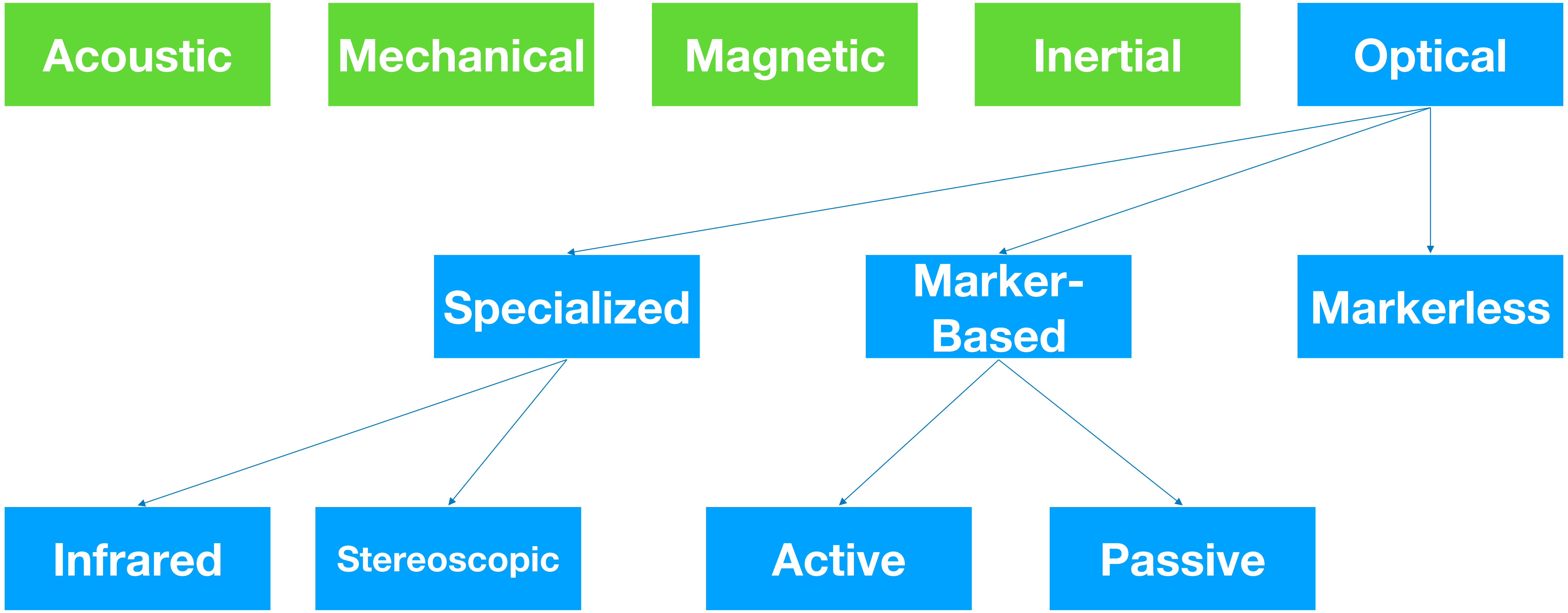


The VRfree glove system in contrast deploys multiple inertial measurement units (IMU) on finger segments, capable to measure their full 3D orientation.

Using in-house developed, hardware accelerated sensor fusion algorithms, the VRfree glove is capable of precisely tracking full hand gestures at high frame rates and low latency.

Optical Tracking & Computer Vision

- Regular video cameras
- Infrared (IR) video cameras
- Depth cameras

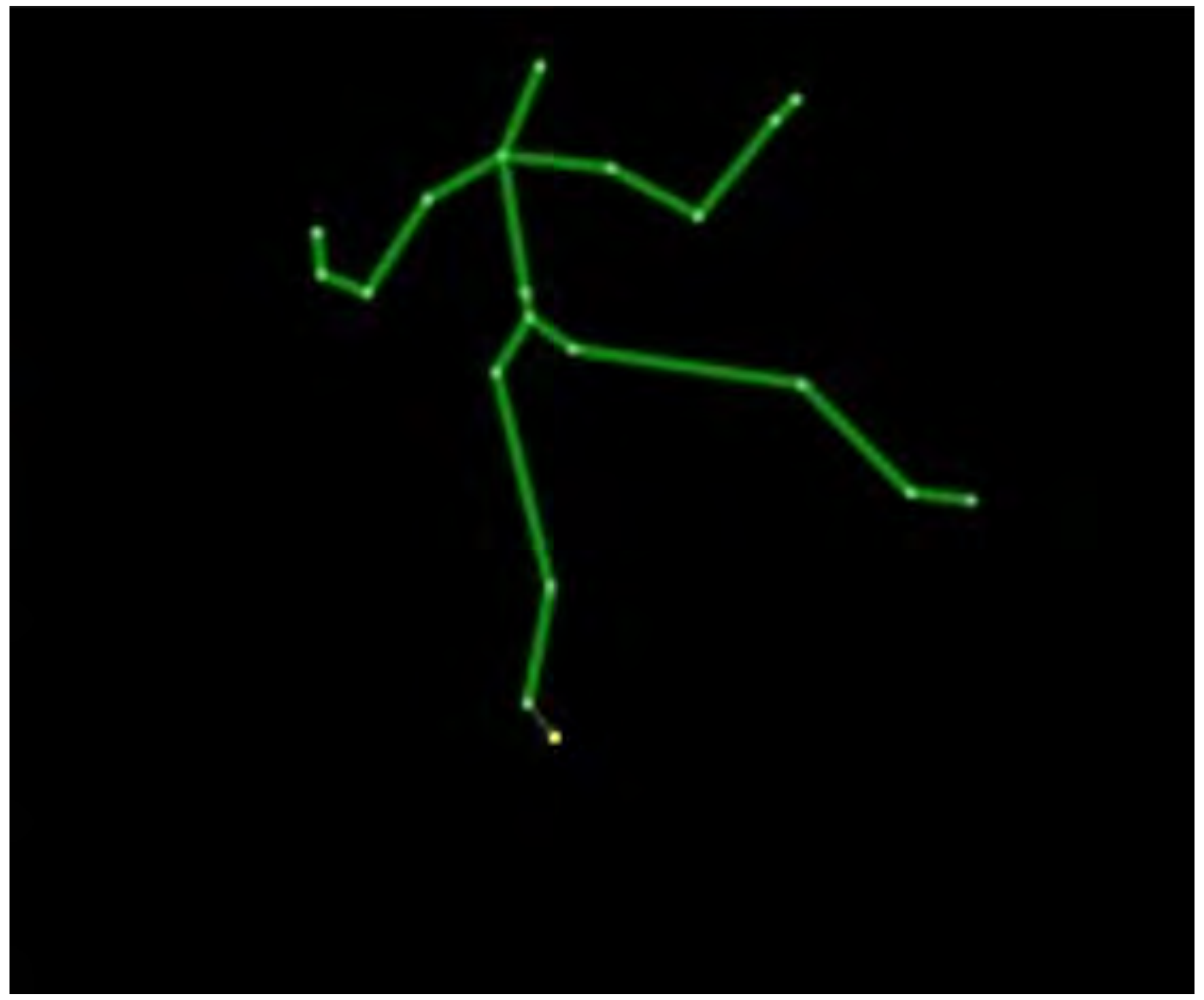


Optical Tracking: Depth (IR) Camera

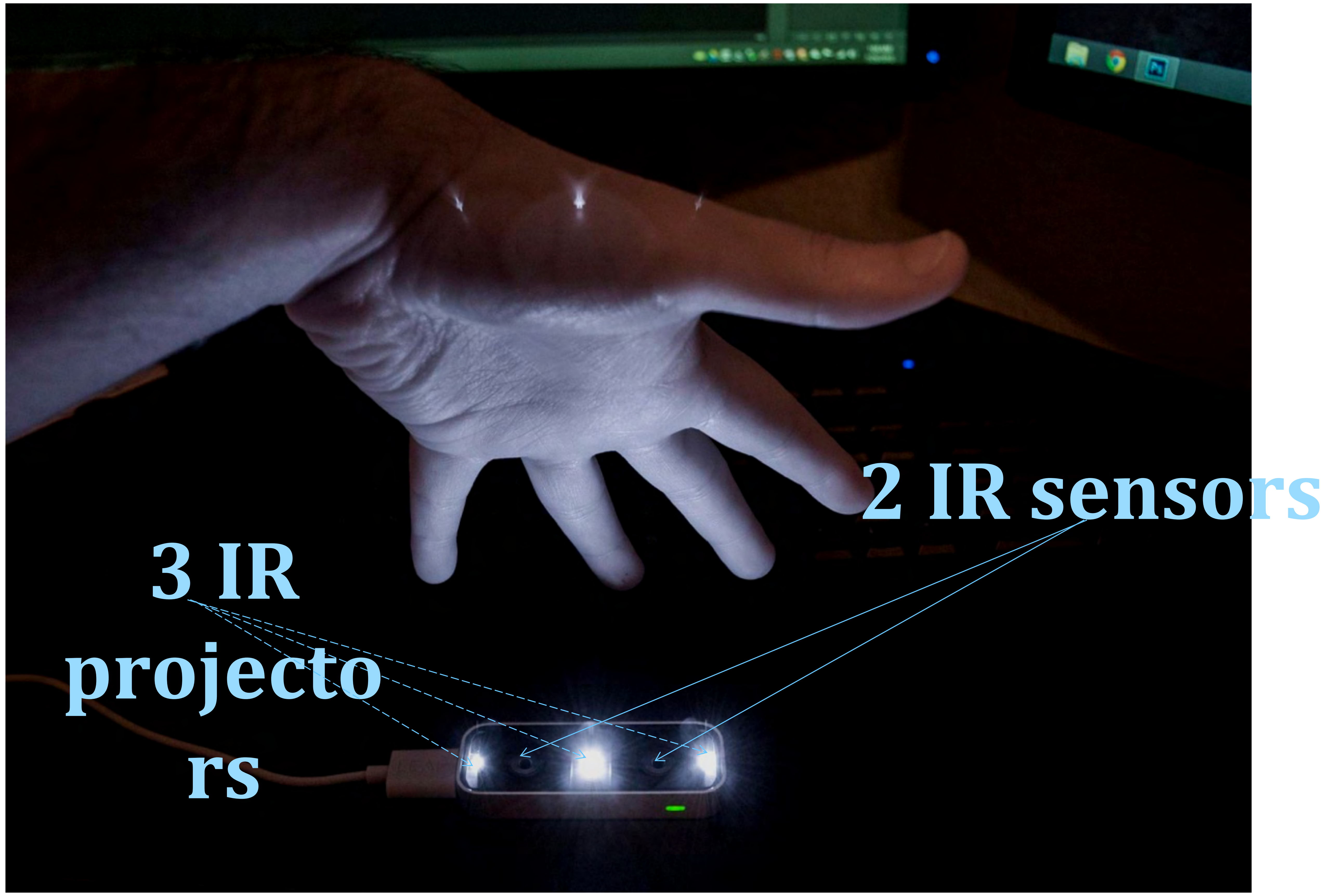
- Microsoft Kinect

Optical Tracking: Depth (IR) Camera

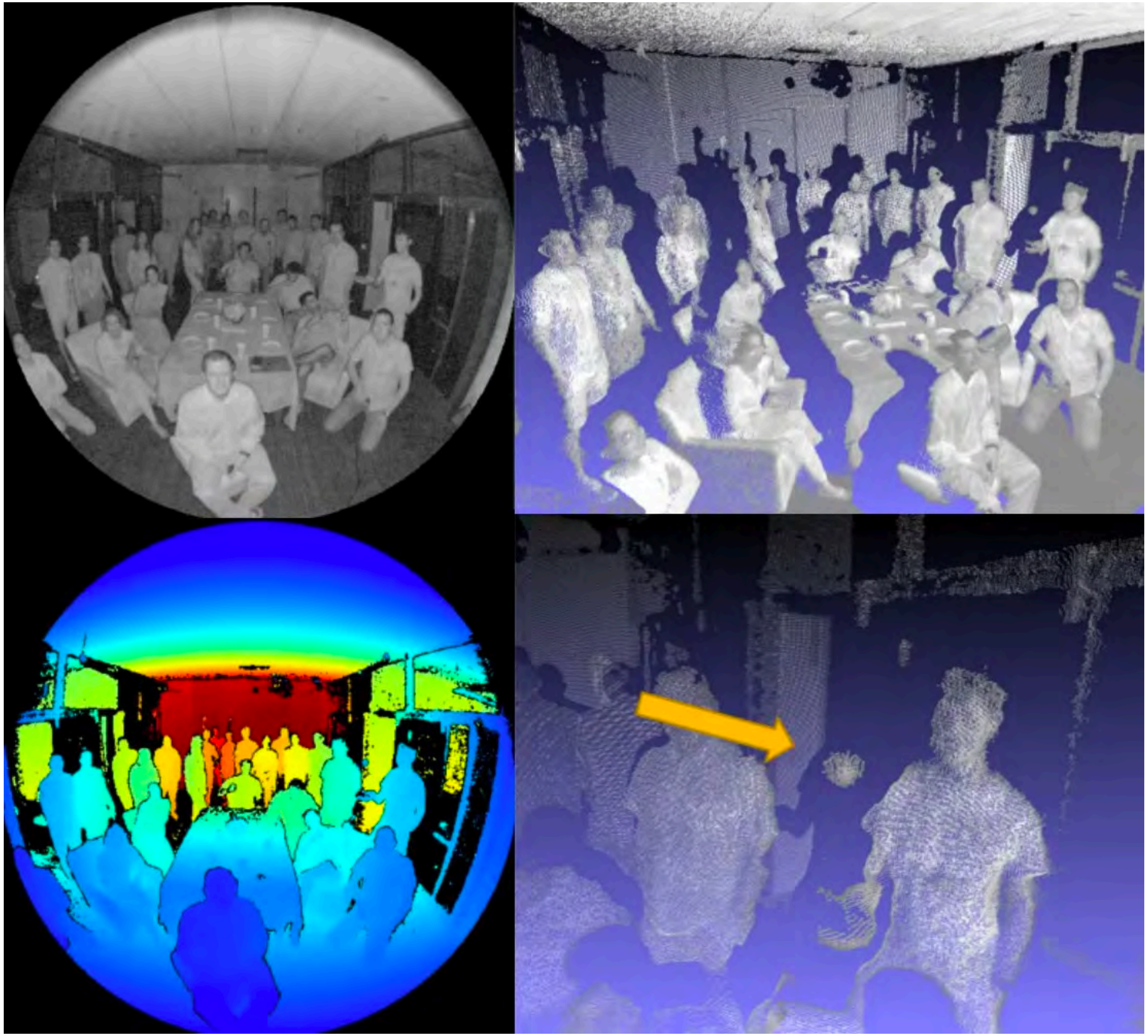
- Kinect uses structured light and machine learning: first it computes a depth map using structured light, then it applies machine learning to infer body position (skeleton).



Leap Motion

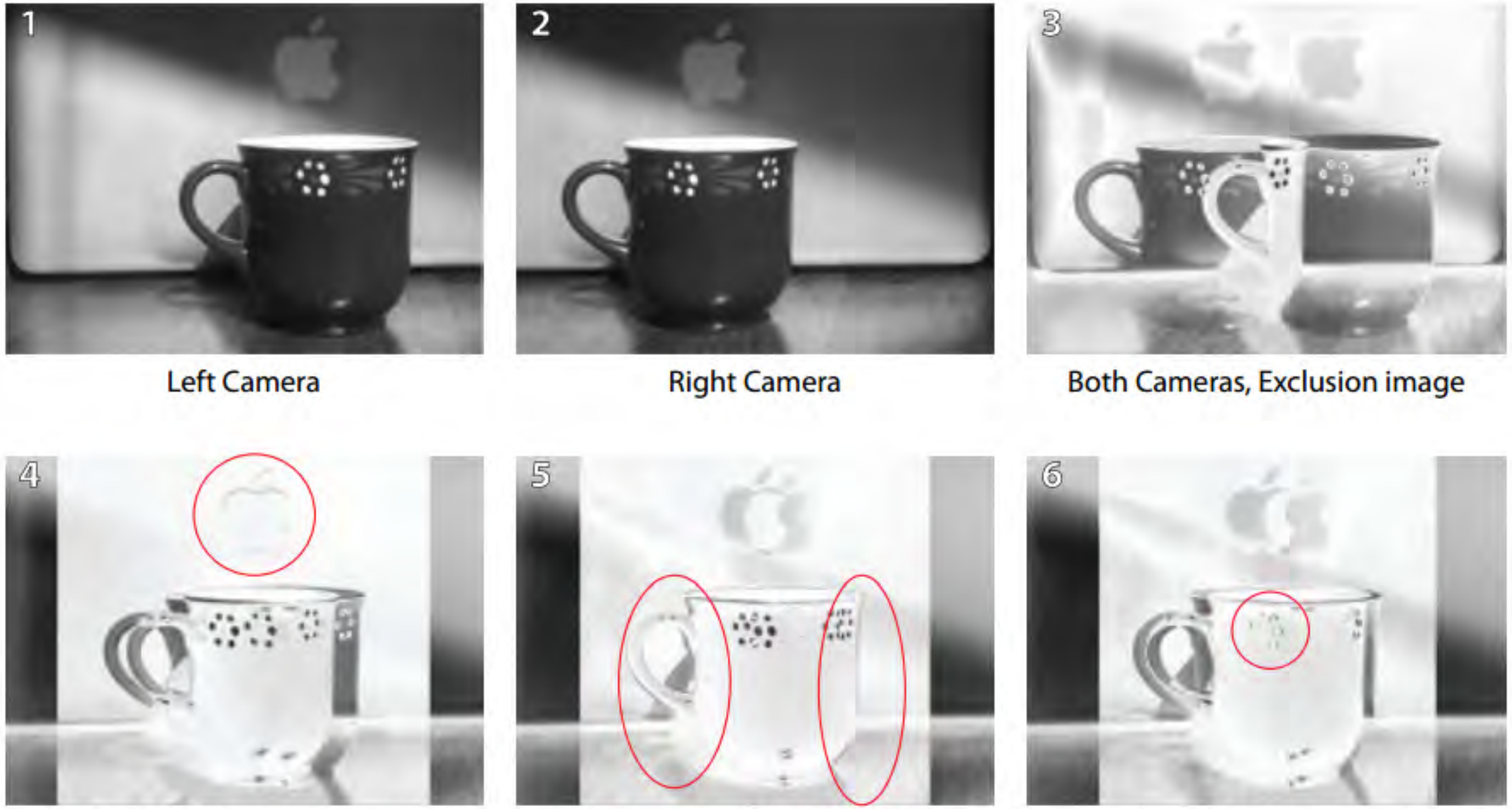


Azure Kinect



	Ultrasound	Infrared triangulation	Laser	TeraRanger Time-of-Flight ToF
High reading frequency			•	•
Long range			•	•
Minimal weight	•	•		•
Small form factor	•	•		•
Eye safety	•	•	Class 1 laser only	•
Use with multiple sensors				•

Optical Tracking: Depth (Stereo) Camera



Left Camera

Right Camera

Both Cameras, Exclusion image

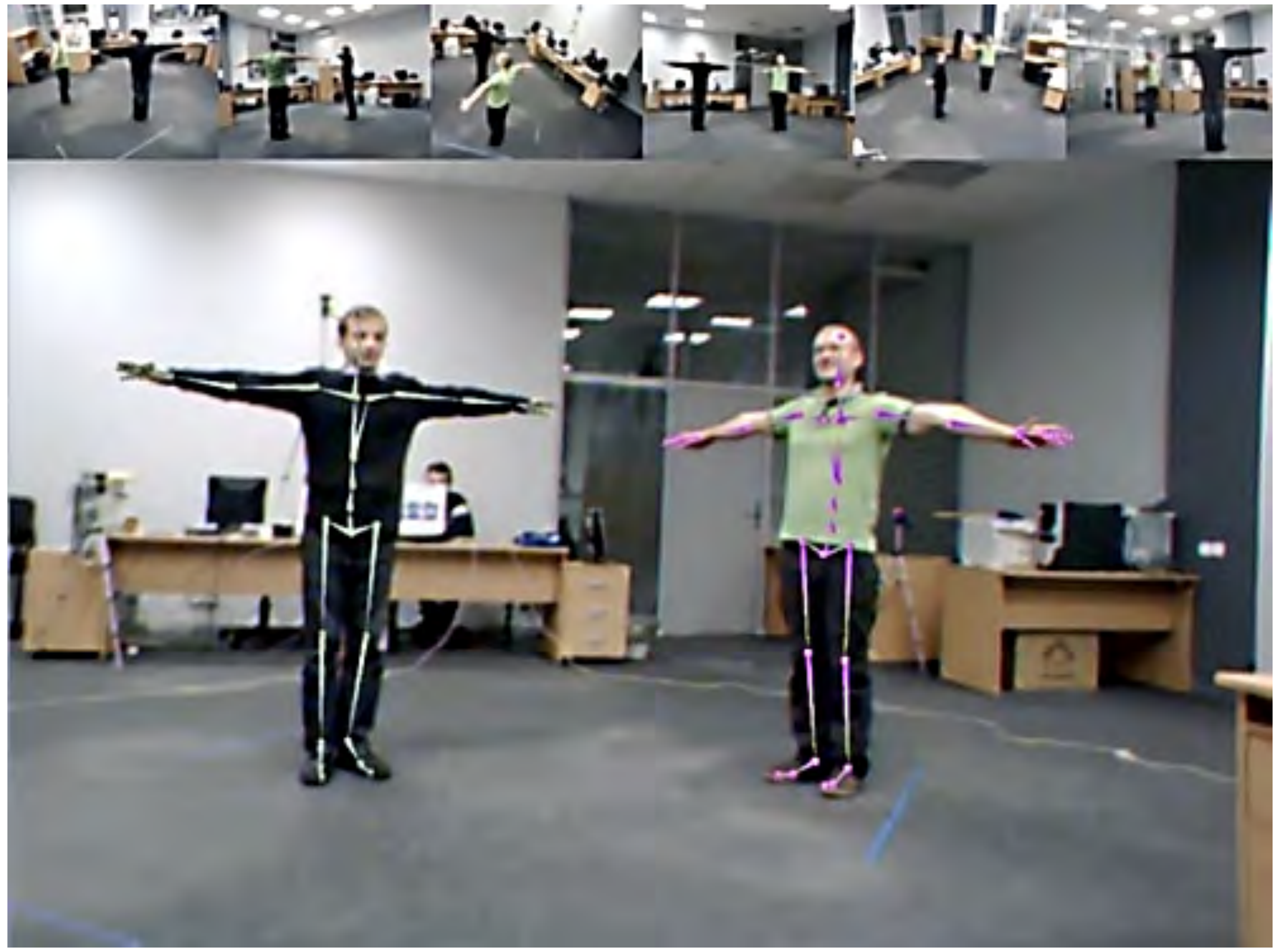
By shifting the images away from each other, the Apple logos in the images overlap

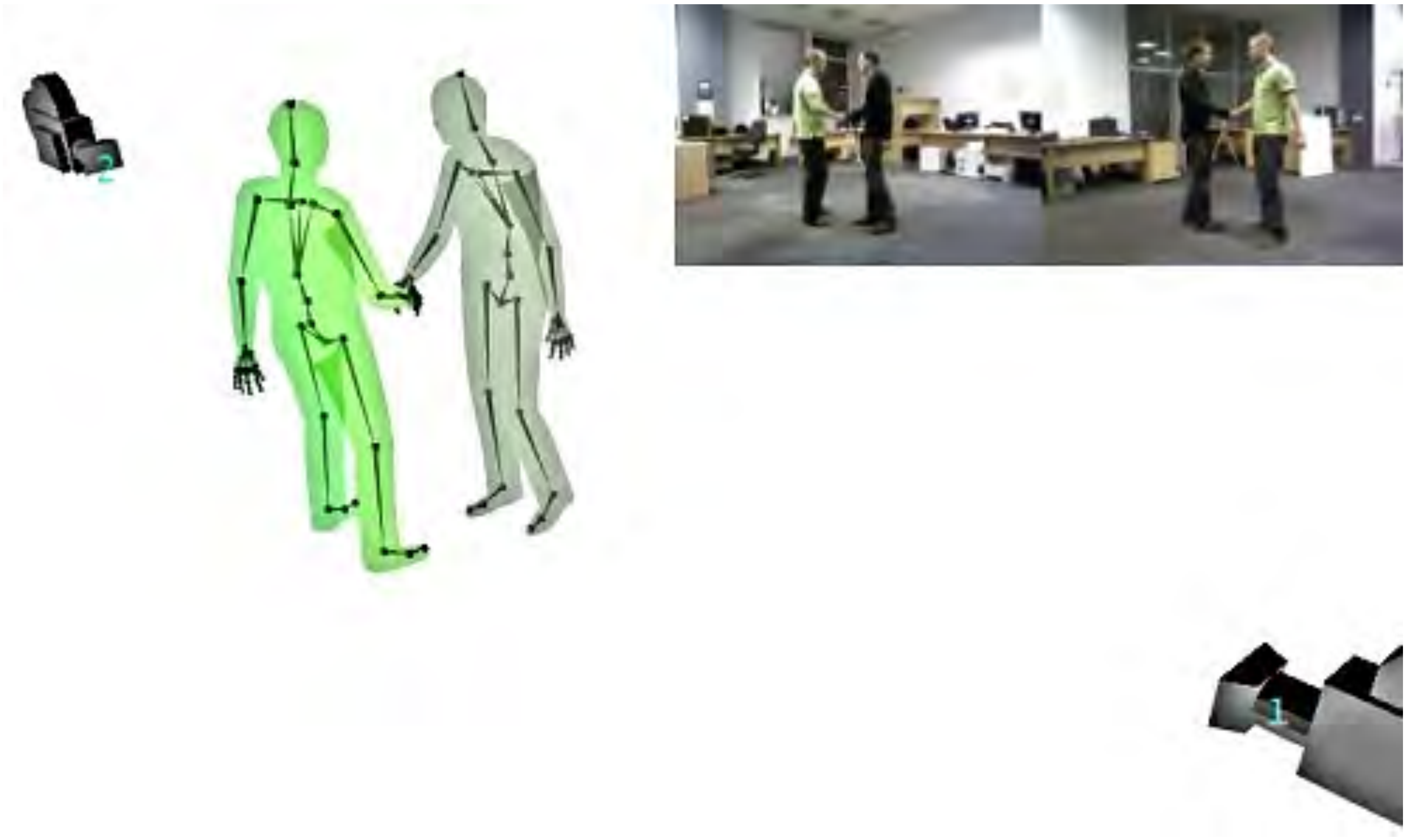
By shifting further, the edges of the mug overlap

Even more, and the flower pattern closest on the mug overlaps

Depth (Stereo) Tracking:

iPi Soft





http://ipisoft.com

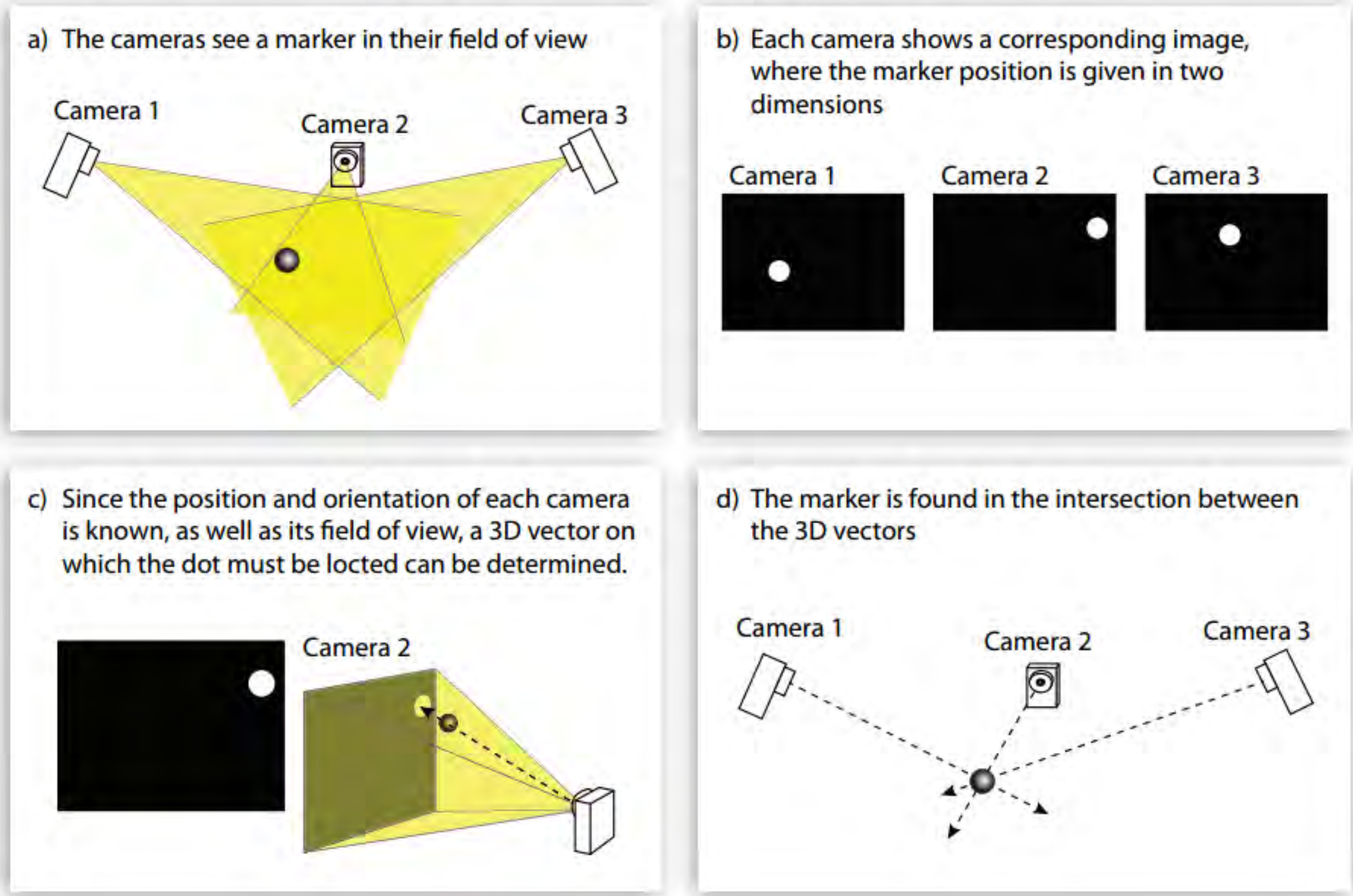
Optical Tracking: Marker-Based

- Passive reflective markers: reflect light from an external source (IR LEDs).

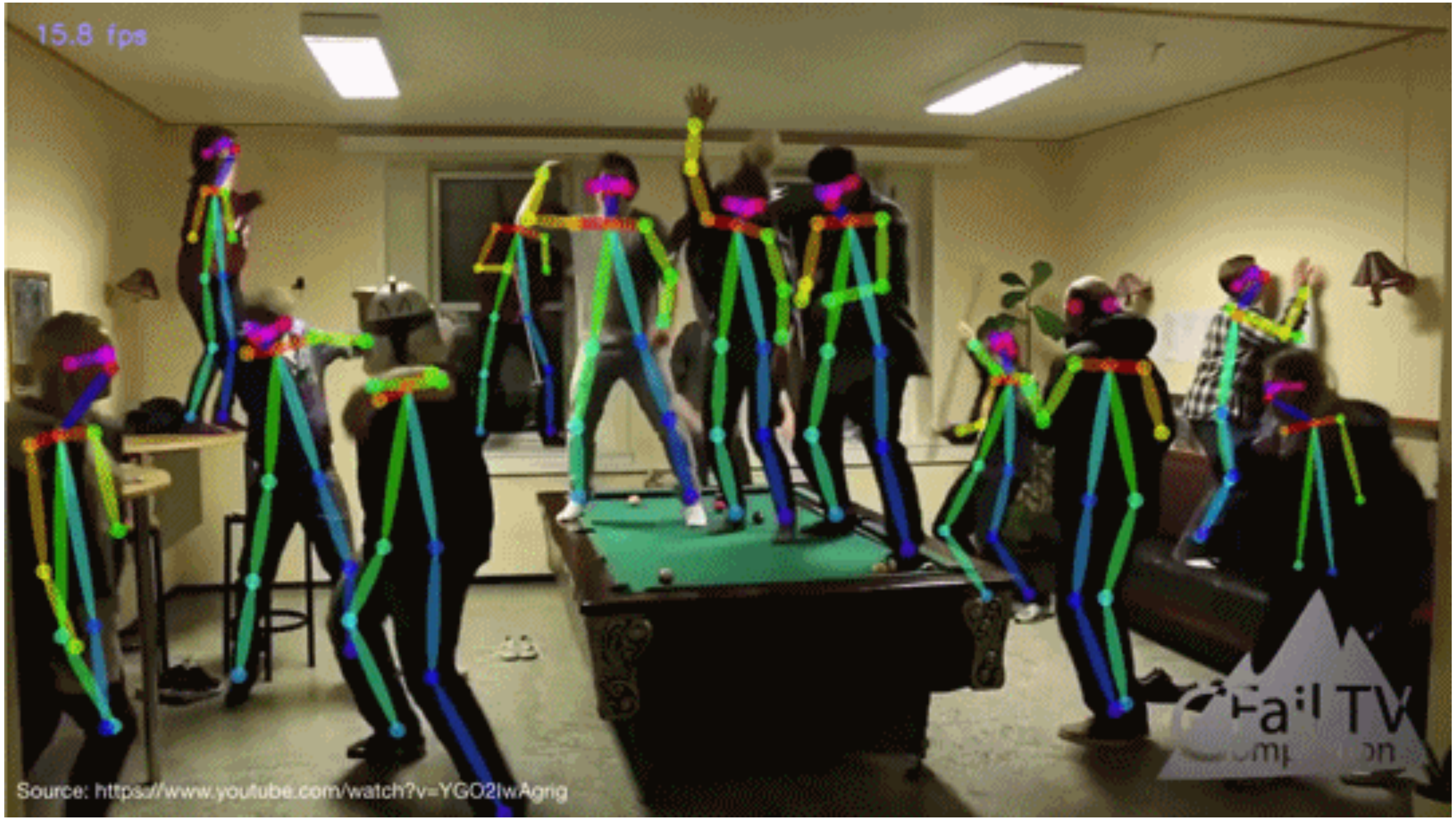
- Active light/IR-emitters.

Mel Slater's lab, Barcelona

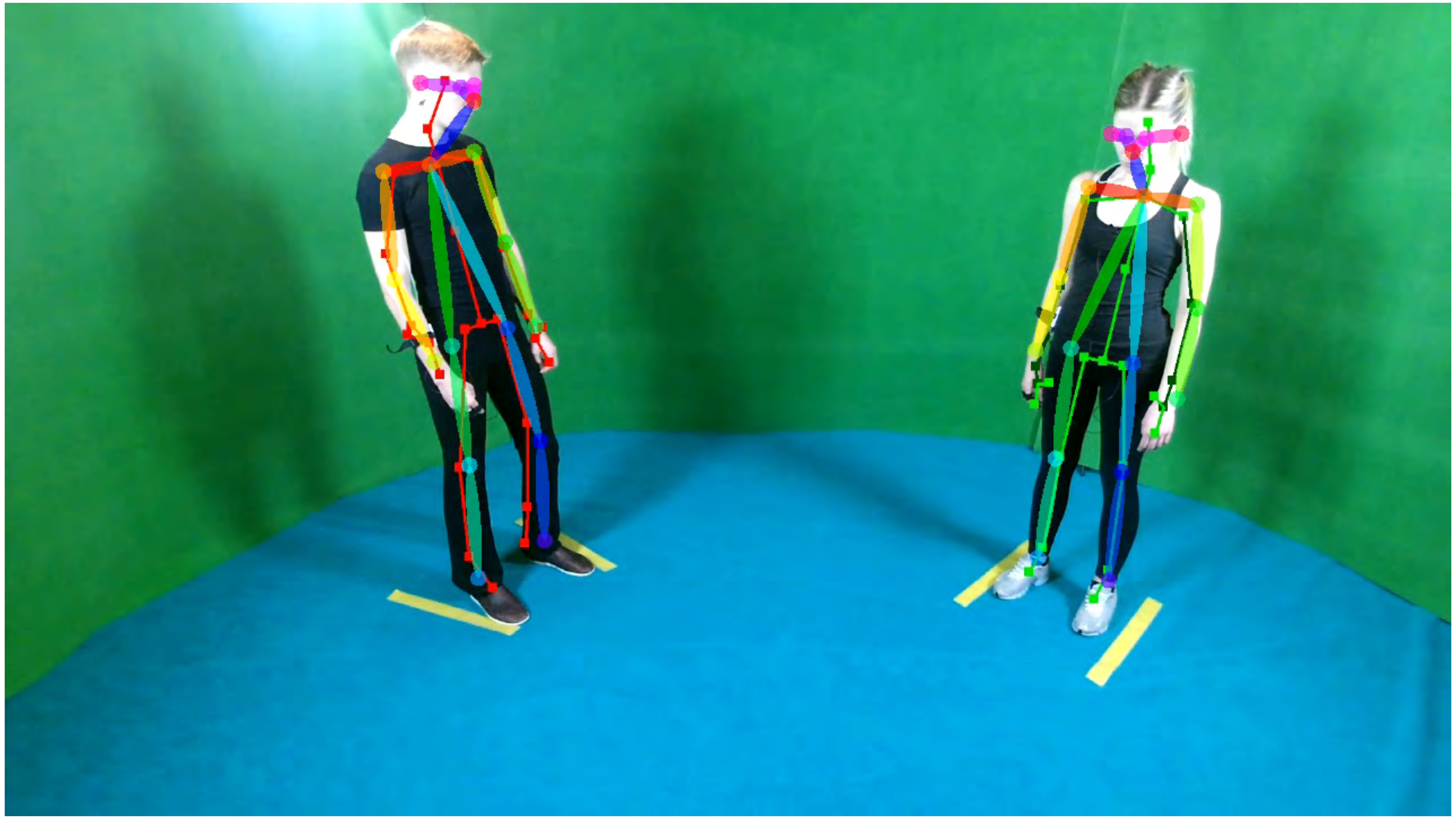
Optical Tracking: Marker-Based



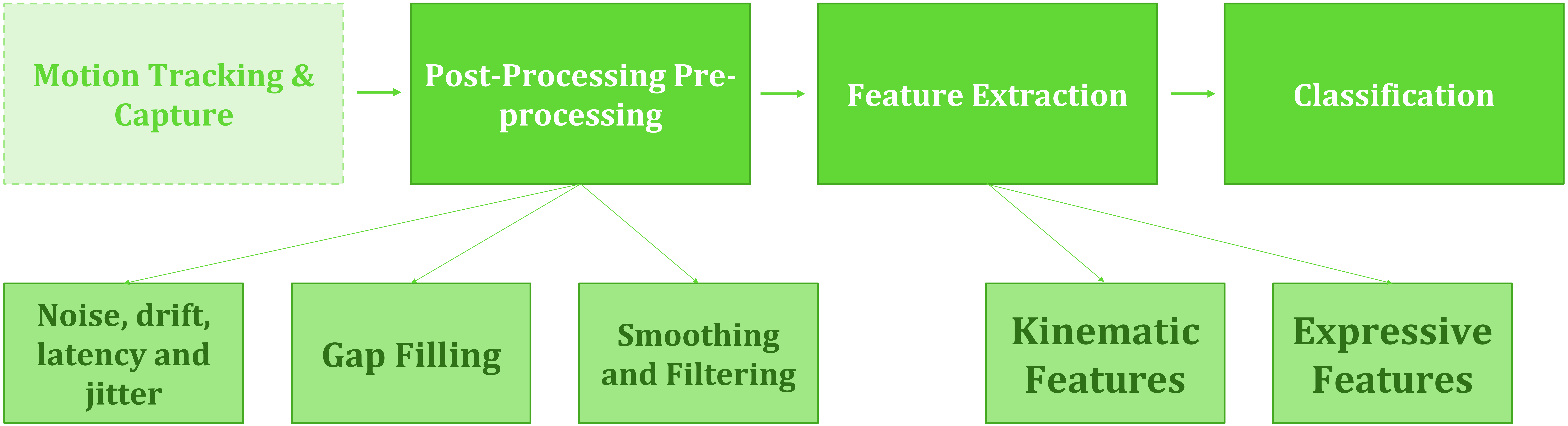
2D video analysis: Computer Vision



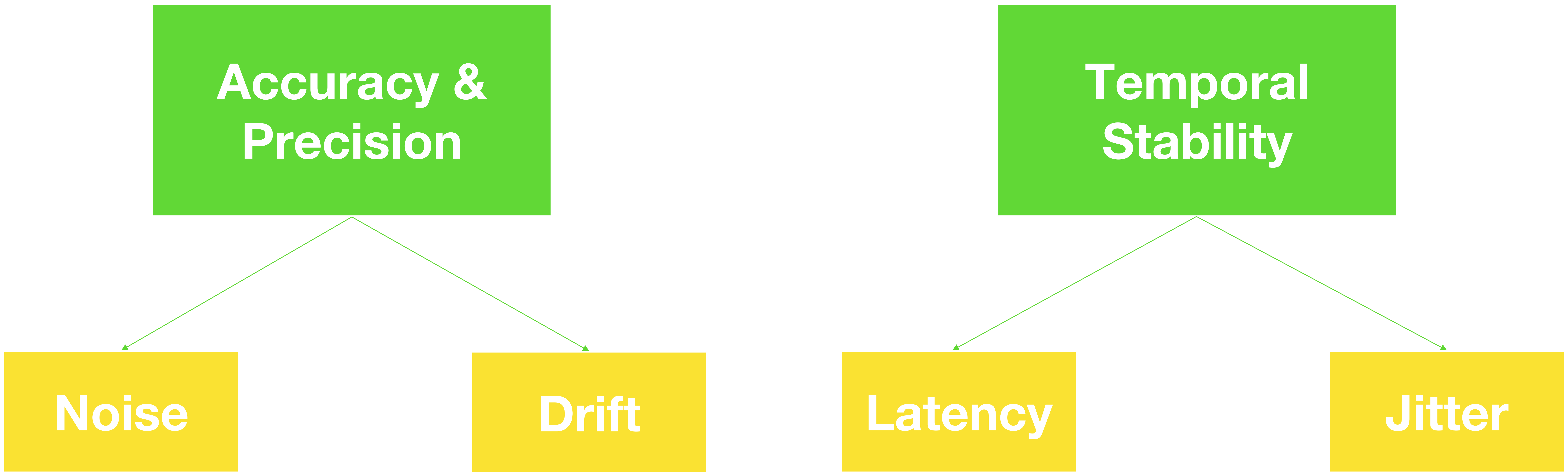
Realtime Multi-Person Pose Estimation



Data analysis



Post-Processing: tracking performance



Post-Processing: Noise

- The level of noise can be measured by the standard deviation (SD) of a static (i.e. without motion) measurement over a time period.

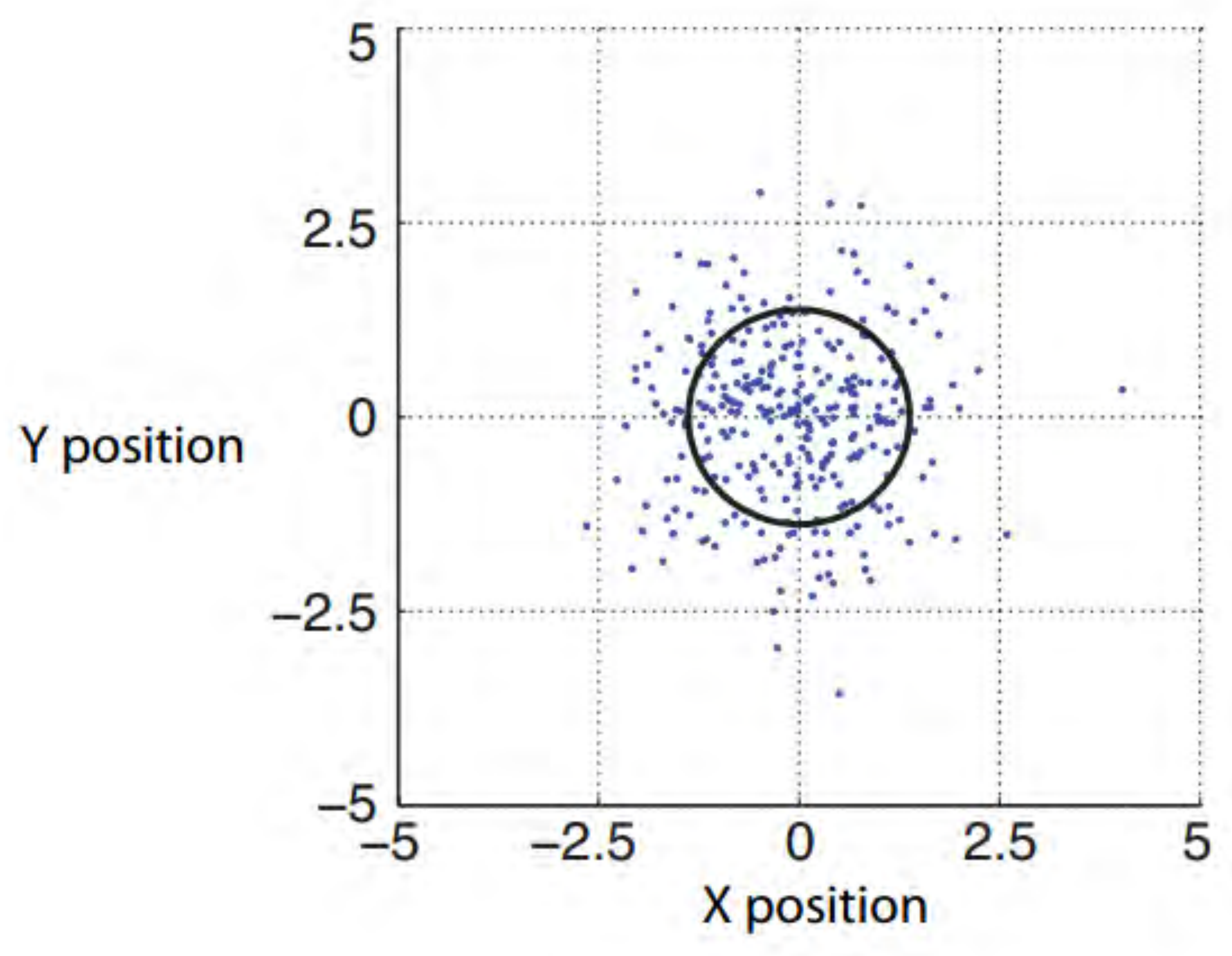
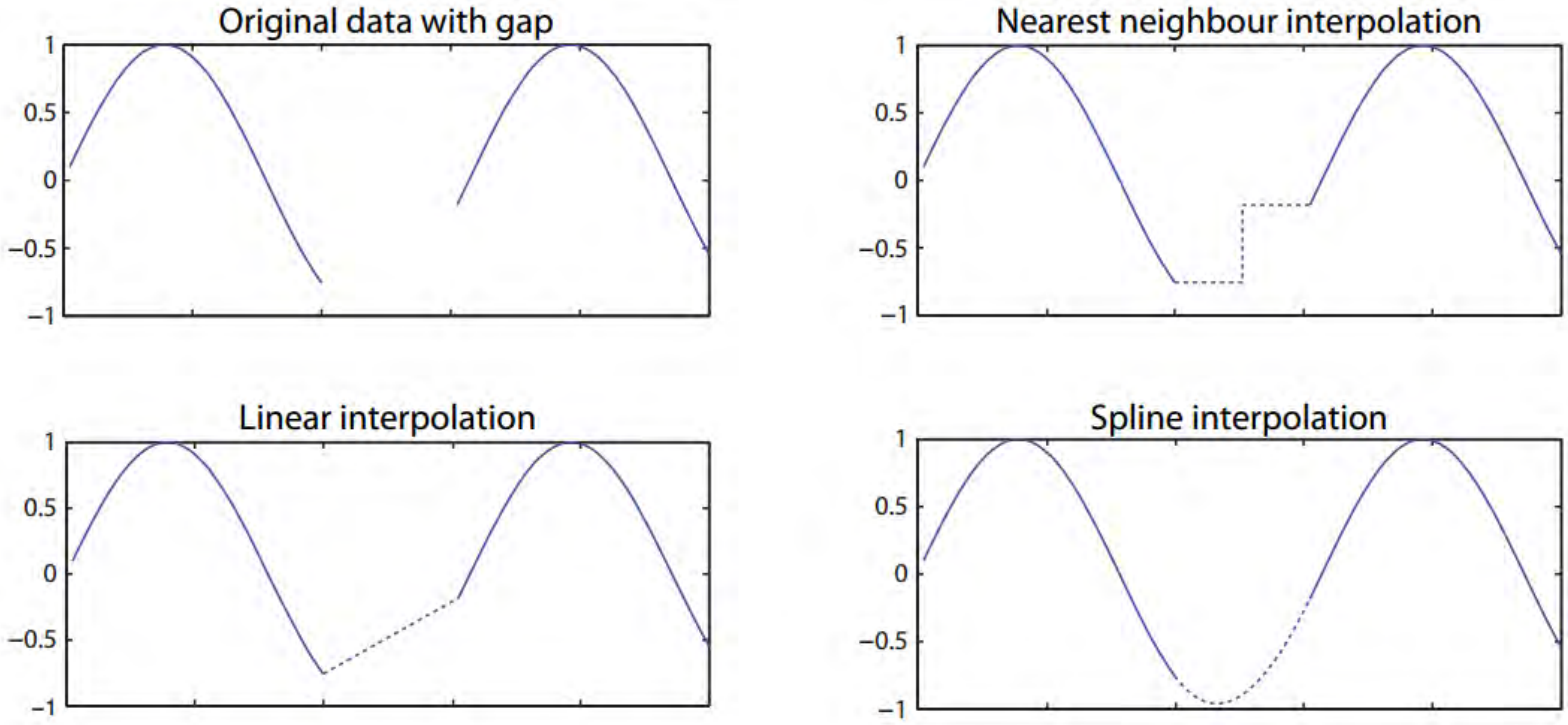


Illustration of how noise can be calculated as the standard deviation of a static position recording. The individual dots display 300 position samples (randomly generated for this example), and the circle has a radius equal to the standard deviation of the position samples.

Post-Processing: Temporal Stability

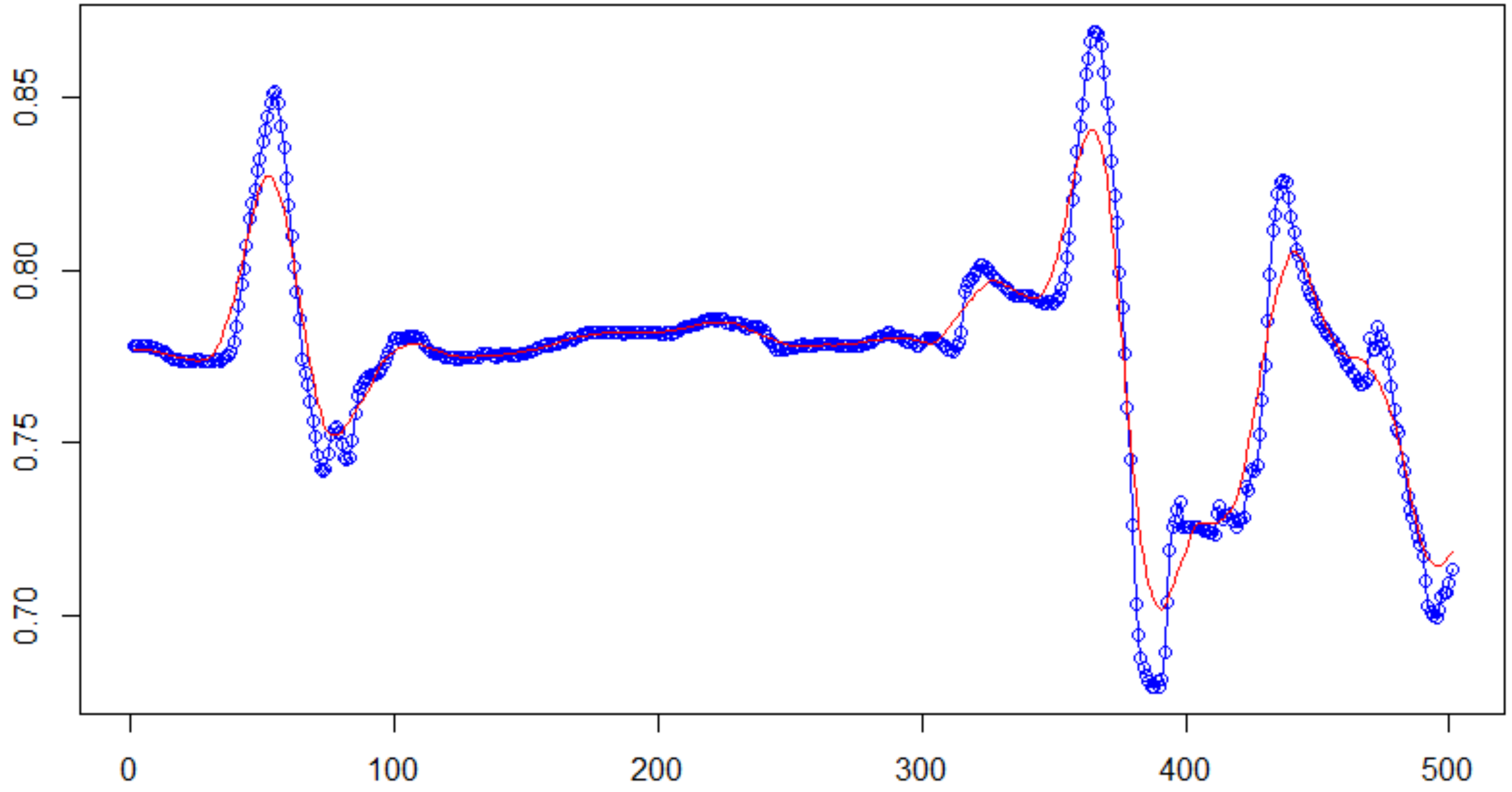
- The latency of an interactive system is the time delay from when a control action occurs until the system responds with some feedback.
- Jitter means any temporal instability in the time interval between data frames. In other words, absence of jitter would mean that the data samples are perfectly periodic.

Post-Processing: gap-filling



Post-Processing: smoothing and filtering

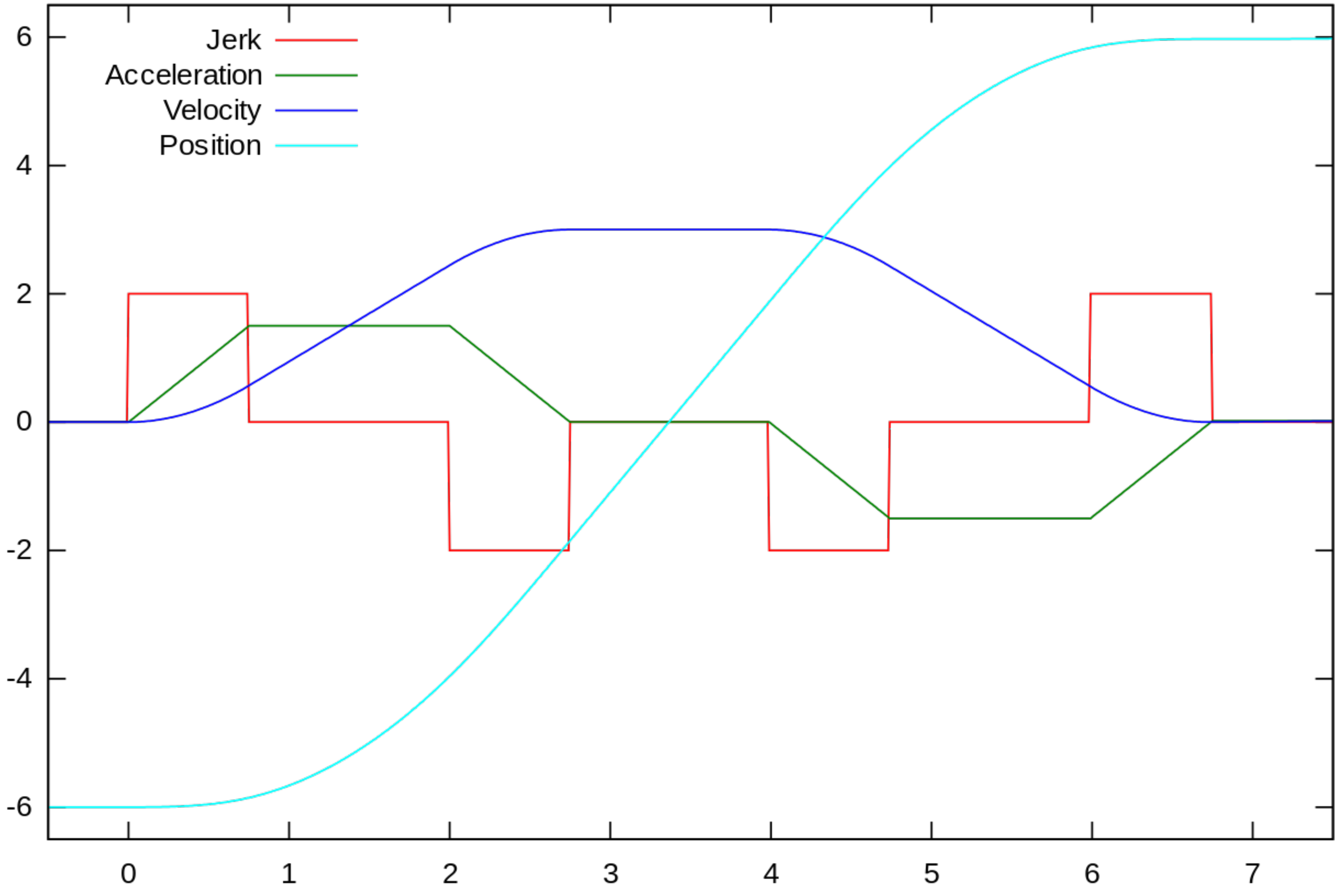
- Moving Average
- Savitzky-Golay filter
- Low-pass filters
- … etc



Feature Extraction: Kinematic Features

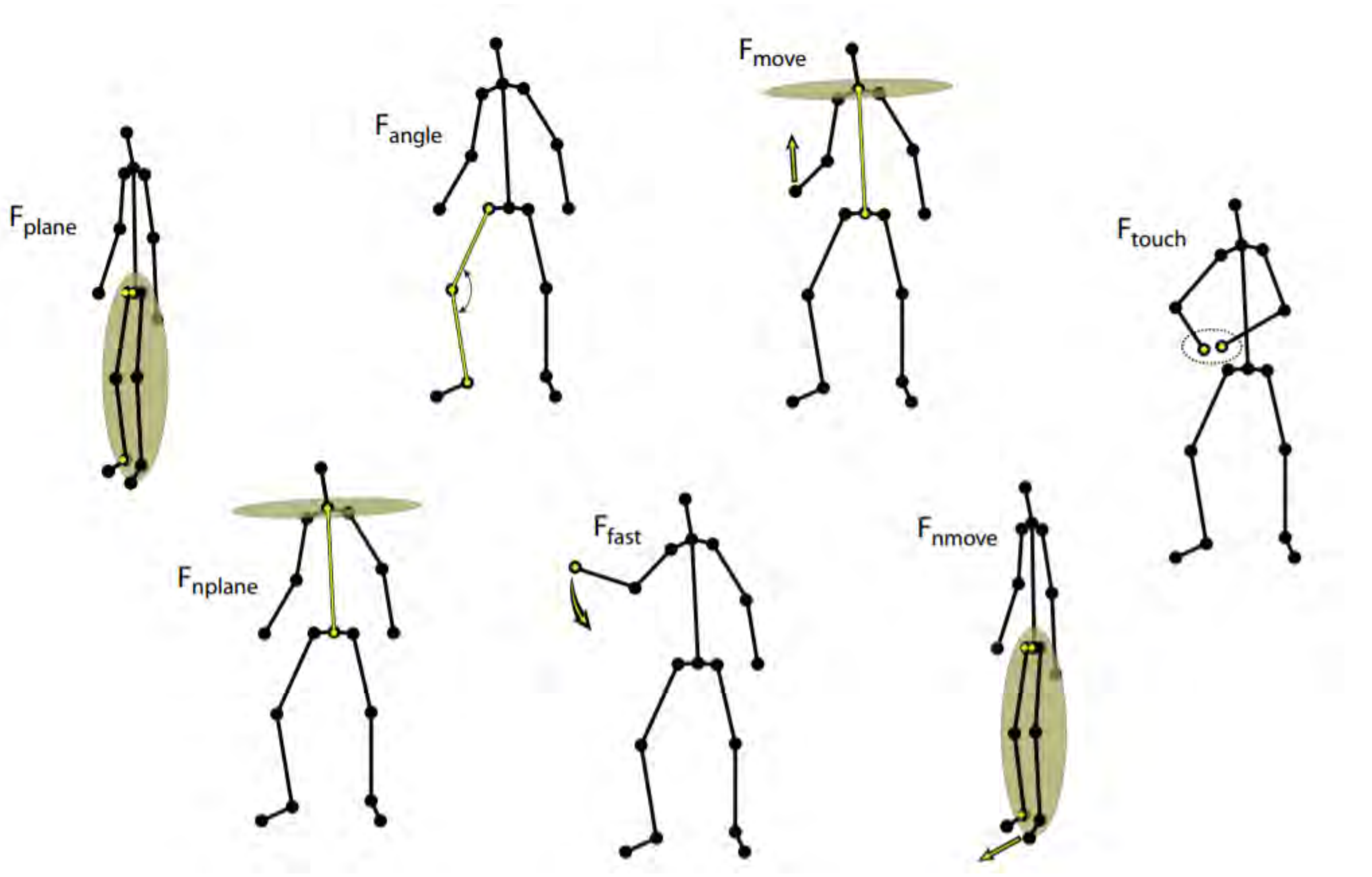
$$v_i = \frac{s_{i+1} - s_{i-1}}{2\Delta t}$$

- Positions and Trajectories
- Velocity:
- Acceleration: $(v_i)'$
- Jerk: $(v_i)''$



Feature Extraction: Kinematic Features

- F_{plane} defines a plane by the position of three joints and determines whether a fourth joint is in front of or behind this plane.
- F_{angle} specifies two vectors given by four joints and tests whether the angle between them is within a given range.
- F_{fast} specifies a single joint and assumes a value of 1 if the velocity of the joint is above a chosen threshold.
- F_{touch} measures the distance between two joints or body segments and assumes a value of 1 if the distance is below a certain threshold.



Feature Extraction: Expressive Features

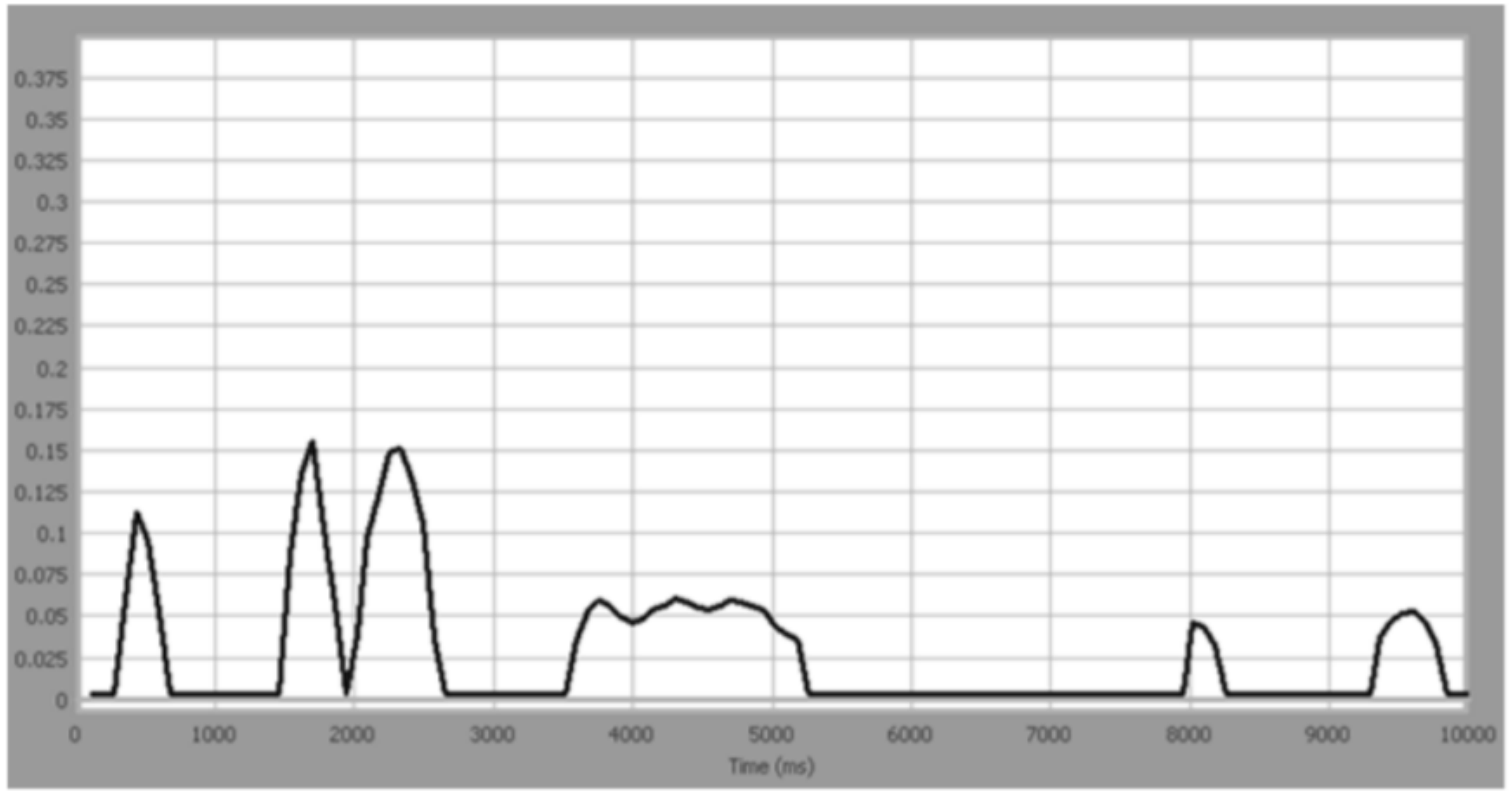
- Kinetic Energy: $KE(f) = \frac{1}{2}\sum_{i=1}^{n} m_i v_i^2(f)$
- Contraction Index is a measure of how the user's body uses the space surrounding it.

Feature Extraction: Expressive Features

- Symmetry Index
- Periodicity
- Directness
- Impulsivity
- etc.

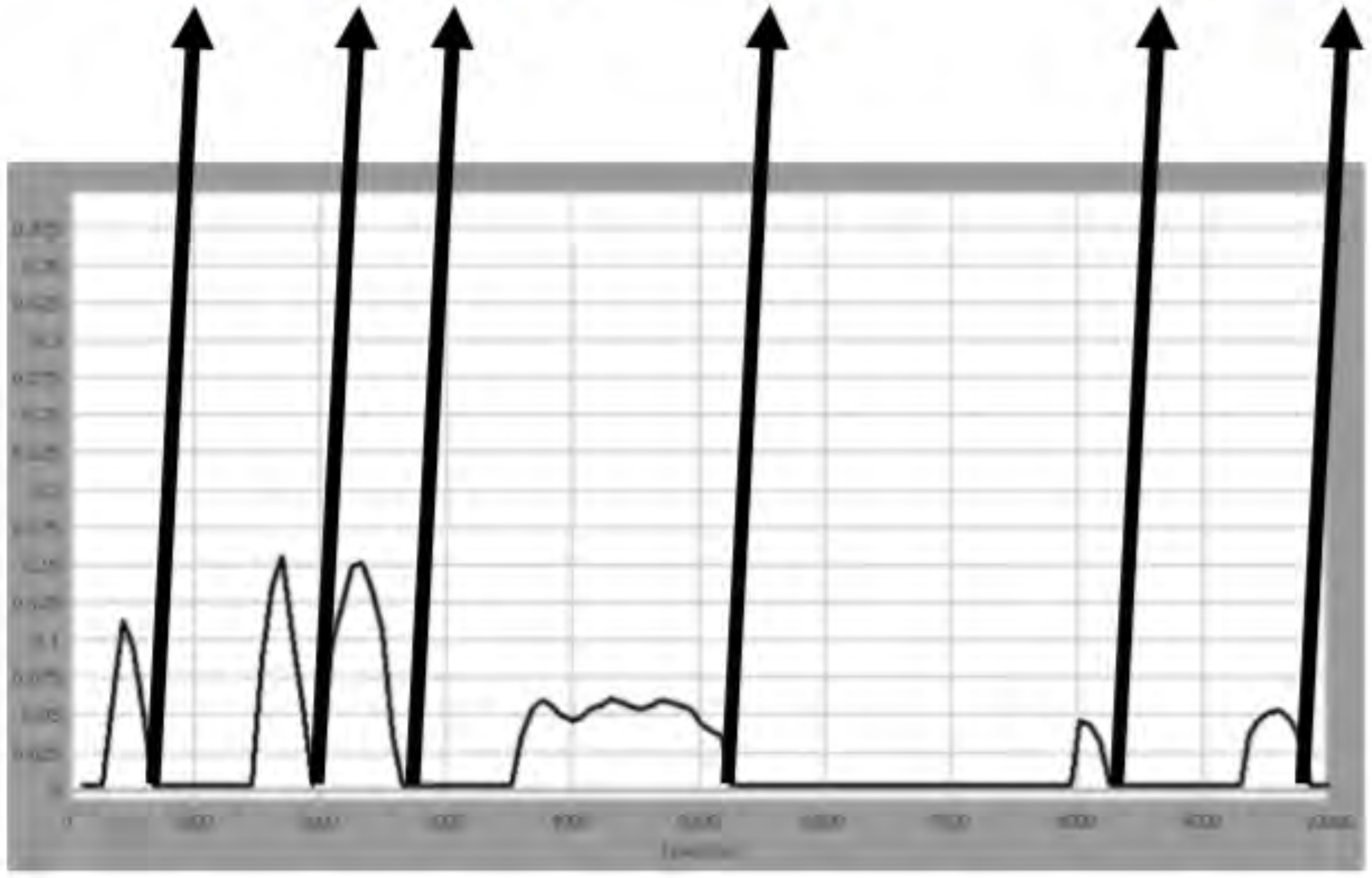
Movement Segmentation

- Motion Phases: the user is moving.
- Pause Phases: the user does not appear to move
- A threshold on the Motion Index / Kinetic Energy.



Movement Segmentation

Motion	0.1	0.15	0.15	0.05	0.05	0.07
Contraction	0.3	0.9	0.3	0.5	0.7	0.1
Symmetry	0.8	0.9	0.8	0.1	0.2	0.3



Bodily information: feature extraction [2D]

Hand-craft approach

- Extract body poses from each frame
- Calculate features: kinematic, expressive features, etc.
- Select features or reduce dimensionality [e.g., PCA]

Deep-learned approach

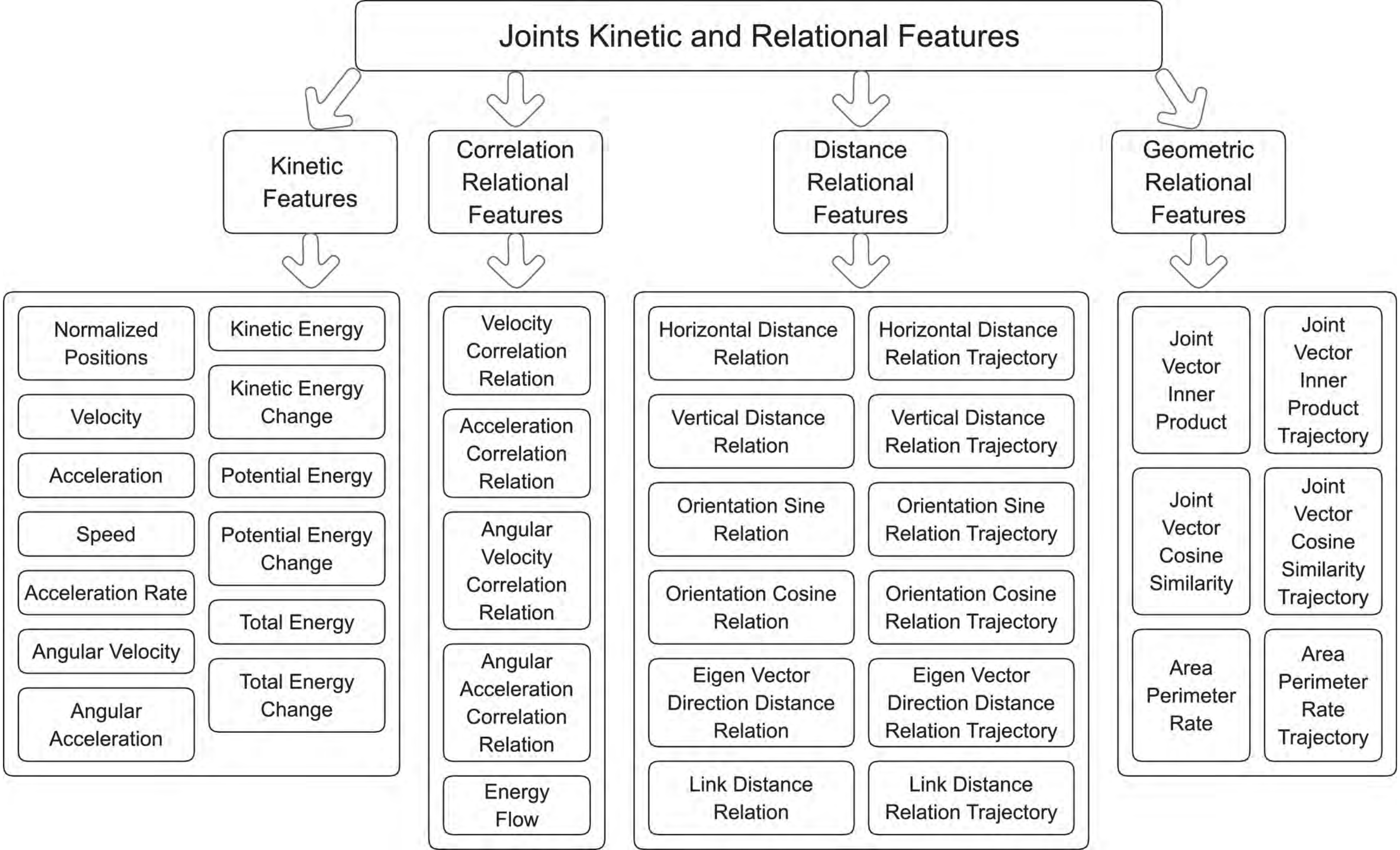
- Use neural network (NN) for feature extraction
- Train NN at huge dataset [AVA, Kinetics datasets] to classify actions
- Use pretrained neural network in the final emotional database [e.g., FABO]

Hand-crafted features

Get skeleton in each frame [2D coordinates of body parts] => calculate some features

High-level features (HLPF)

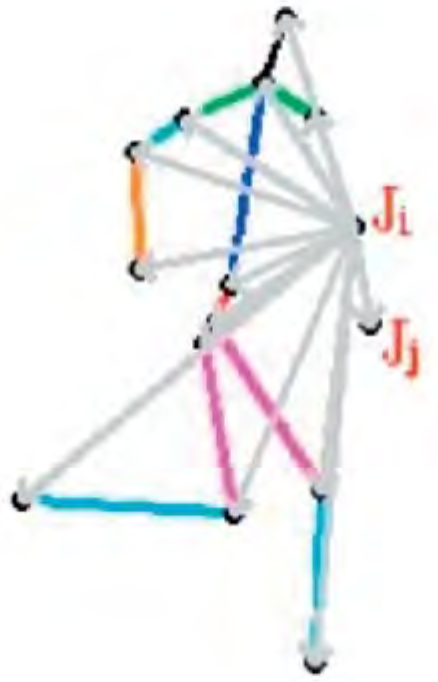
- HLRF is a pose-based description method [calculated features from body poses]
- Features are extracted <u>based on the positions (*x* and *y* coordinates) of joints</u>
- Three features in HLPF describe **single joint information**: normalized joint positions, the trajectories in Cartesian coordinates and the trajectories in polar coordinates.
- Four features in HLPF describe **pairwise joints relations**: distance relation, orientation relation and the trajectories of these relations.
- The other two features, i.e., the angle relation and its trajectory, describe **triplet joints relations.**



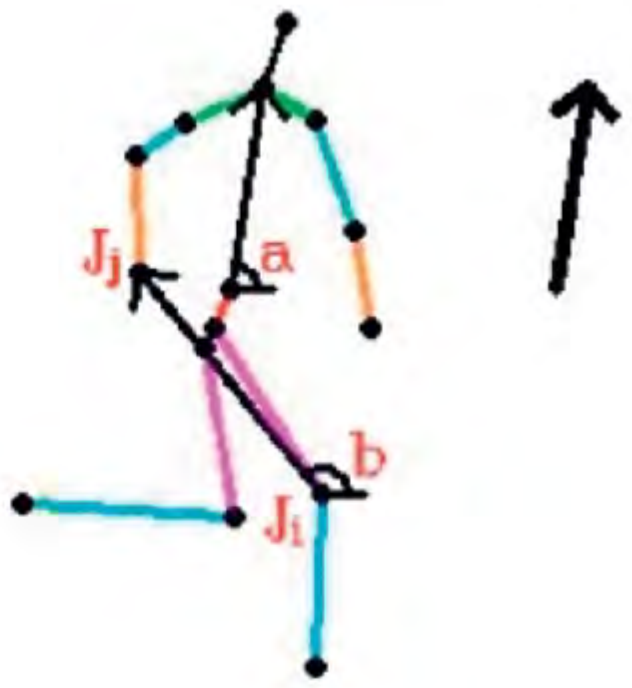
Joints Kinetic and Relational Features
Kinetic Features
Correlation Relational Features
Distance Relational Features
Geometric Relational Features
Normalized Positions
Kinetic Energy
Velocity
Kinetic Energy Change
Acceleration
Potential Energy
Speed
Potential Energy Change
Acceleration Rate
Total Energy
Angular Velocity
Total Energy Change
Angular Acceleration
Velocity Correlation Relation
Acceleration Correlation Relation
Angular Velocity Correlation Relation
Angular Acceleration Correlation Relation
Energy Flow
Horizontal Distance Relation
Horizontal Distance Relation Trajectory
Vertical Distance Relation
Vertical Distance Relation Trajectory
Orientation Sine Relation
Orientation Sine Relation Trajectory
Orientation Cosine Relation
Orientation Cosine Relation Trajectory
Eigen Vector Direction Distance Relation
Eigen Vector Direction Distance Relation Trajectory
Link Distance Relation
Link Distance Relation Trajectory
Joint Vector Inner Product
Joint Vector Inner Product Trajectory
Joint Vector Cosine Similarity
Joint Vector Cosine Similarity Trajectory
Area Perimeter Rate
Area Perimeter Rate Trajectory

(a) A video frame

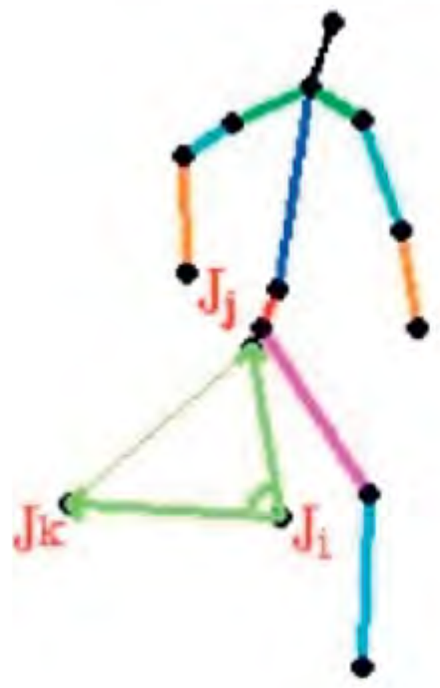
(b) Joints and skeleton



(c) Distance relation



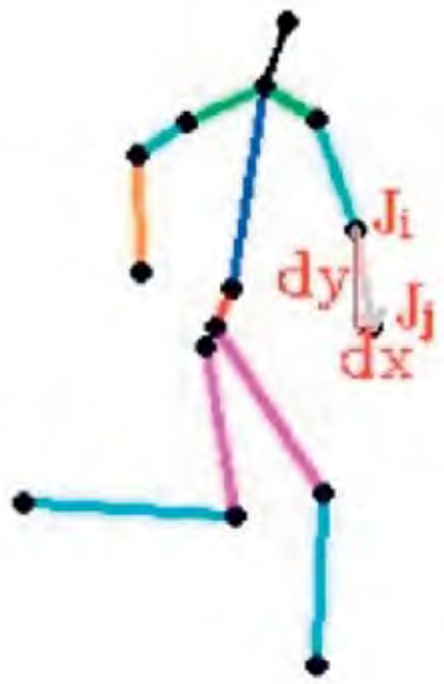
(d) Orientation (cosine/sine) relation



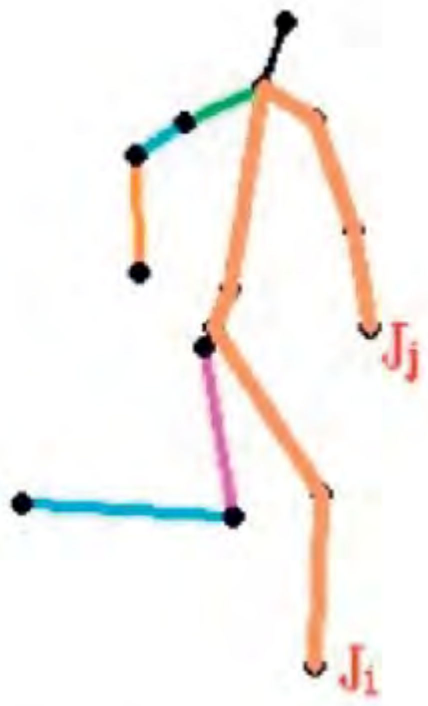
(e) Angle (geometric) relation

(f) Trajectories of joints

(g) Velocity of joints



(h) Horizontal (vertical) distance relation

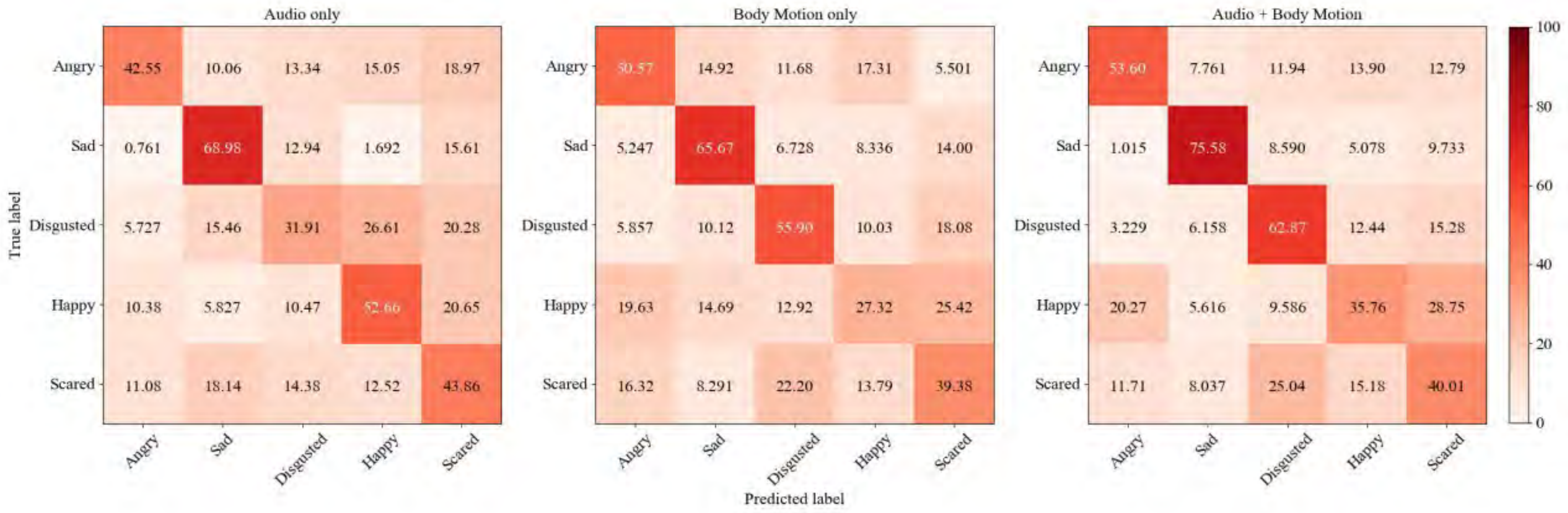


(i) Link distance relation

Experiments on RAMAS dataset

- 6 basic emotions: angry, disgust, happy, sad, scare, surprise.
- Multimodal: video, audio, motion, physiology.
- Play-acted

Results on RAMAS dataset



5 classes (random choice 20%): RNN

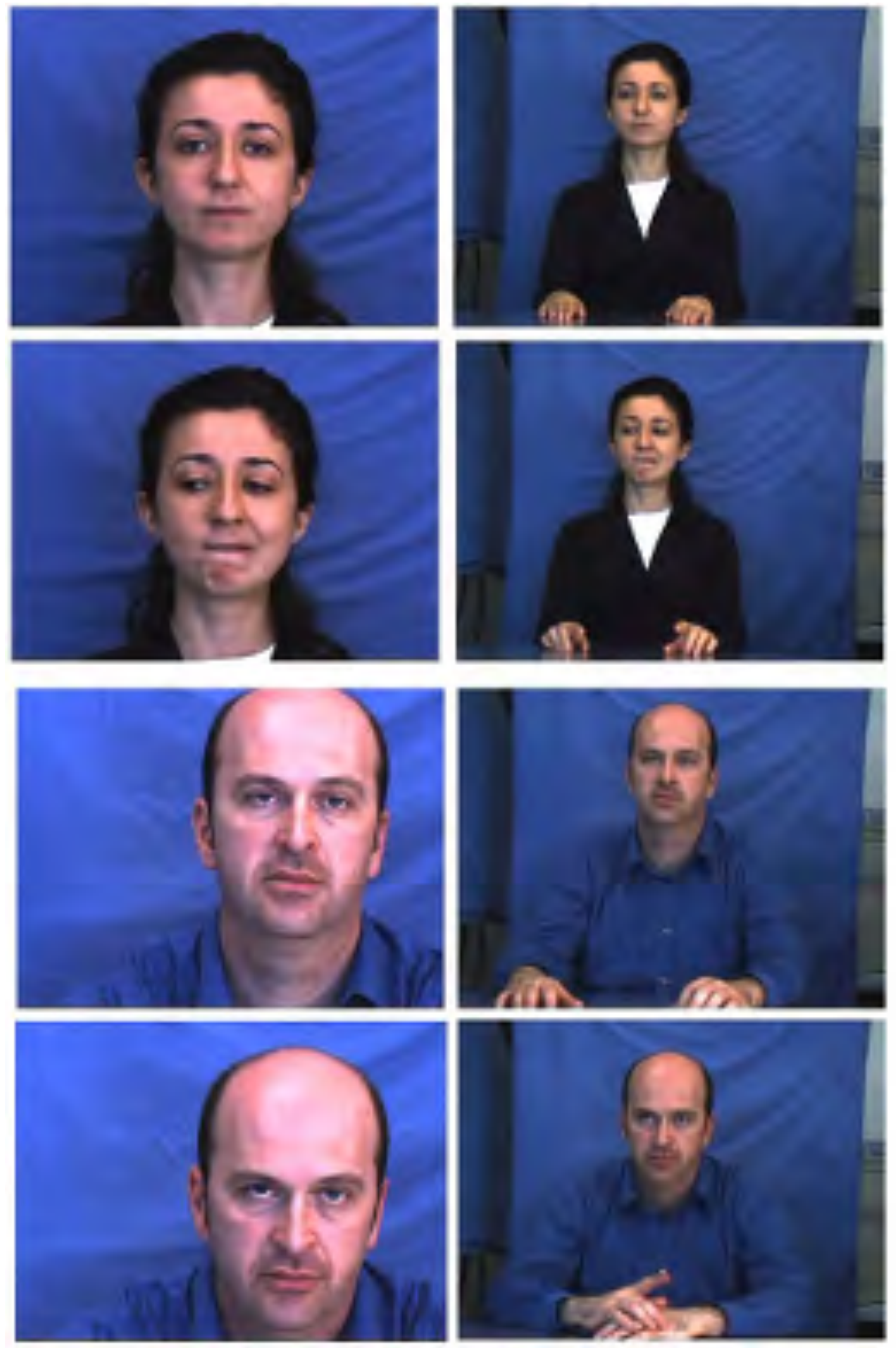
49% for audio and motion

52 % for audio + motion

Experiments on FABO database

- Anger
- Anxiety
- Boredom
- Disgust
- Fear
- Happiness
- Puzzlement
- Sadness
- Surprise
- Uncertainty

- 10 subjects
- Play acted (posed) emotions
- Collected by Hatice Gunes and Massimo Piccardi in 2005
- Onset & Offset of Emotions (annotated)

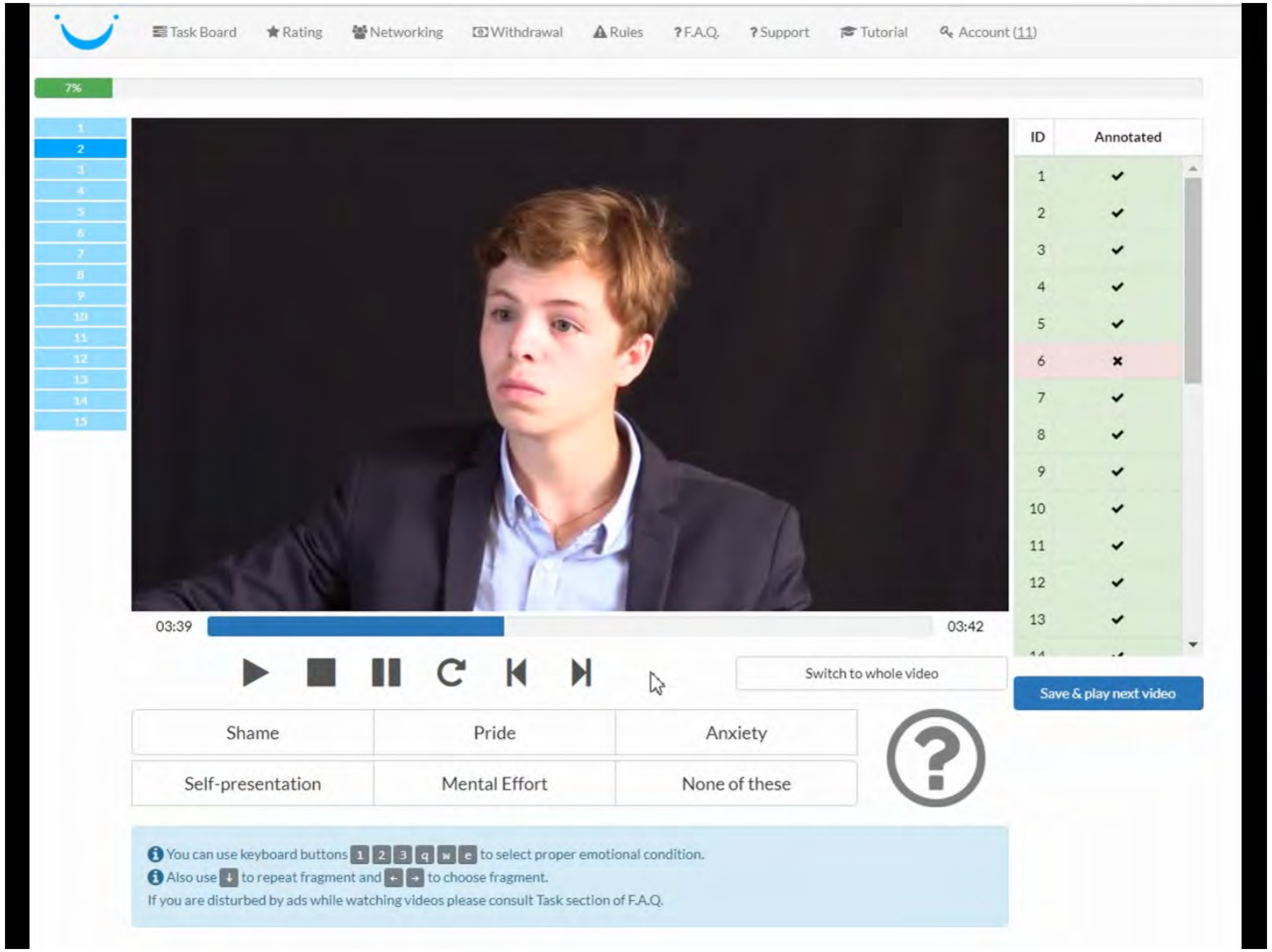


Results: FABO

	Face	Body	Face + Body
Precision	0.51	**0.56**	0.45
Recall	0.52	**0.56**	0.46

Experiments on EmotionMiner database

- 160 hours of marked up pubic audiovisual content in English (more than 110 000 annotated video ~5 sec fragments)
- 10+ annotators for each fragment
- 22 emotional states and social categories
- For the experiment selected: Happiness, Anger, Sadness, Neutral [4400 fragments, ~ 6 hours]



Results: EmotionMiner

	Face	Body	Face + Body
Precision	0.63	0.55	**0.65**
Recall	0.62	0.54	**0.64**

Практика

http://bit.ly/2GJeKz2

HW_3

- Доделать практическое задание
- Прислать отчет на почту o.perepelkina@neurodatalab.com
- Дедлайн: 7 февраля
- Заполнить форму по проектам: http://bit.ly/3b5fJaL
  Прислать до 4 февраля