Affective computing & Social signal processing:

from the academy to industrial cases

Olga Perepelkina, 2019

http://www.neurodatalab.com/

Founded in 2016.

Neurodata Lab is an Emotion AI Hub, full-range R&D laboratory in affective computing, emotion and social behavior recognition technologies, emotion synthesis and emotion research in the language sciences.

Our company adopts an effective model of integration between fundamental research, applied science and high-tech business. Our technology incorporates a multimodal approach to create highly accurate solutions.

Neurodata Lab is deeply integrated into the global academic community and proud of strong R&D collaborations with universities, centers of knowledge and competence in EU, Russia and US.

Neurodata Lab LLC ®

Miami (Fl), US HQ
Moscow R&D Office
Milan Project Office
Lucerne Project Office

info@neurodatalab.com
www.neurodatalab.com
www.emotionsdemo.com

Our Moscow team

NEURODATA LAB

GEORGE PLIEV
Founder&Managing Partner

MAXIM RYABOV
Managing Director

IGOR LEVIN
Executive Director

OLGA PEREPELKINA
Chief Research Scientist

EVA KAZIMIROVA
Research Scientist

MARIA KONSTANTINOVA
Research Scientist

MIKHAIL GRINENKO
Chief Research Advisor

GRIGORIY STERLING
Deep Learning Expert

MARINA CHURIKOVA
Junior Research Scientist

YAROSLAV LAVRINENKO
Lead Developer

ANDREY BELYAEV
Computer Vision Expert

KRISTINA ASTAKHOVA
Public Relations Officer

GLEB CHERDANTSEV
Project Manager

KONSTANTIN YATKIN
Full Stack Web Developer

MIKHAIL ARTEMYEV
Junior Machine Learning Expert

LEONID SHTANKO
Machine Learning Expert

Scientific advisers

NEURODATA LAB

Corporate partnerships

promobot

Microsoft

SAMSUNG

Research collaborations

University of Glasgow

ITMO UNIVERSITY

MINES ParisTech

UNIVERSITÀ DEGLI STUDI DI GENOVA

UNIVERSITÉ DE GENÈVE

We are open to collaboration!

01

NEURODATA LAB

Multimodal Affective computing

Lorem ipsum dolor sit amet consectetur

Emotions: how do people recognize them?

- People can recognize emotions by separate channels: **faces**, voice, body motion & poses...
- Visual and auditory modalities **affect each other** (e.g. facial movements around mouth region impact vocal acoustics).
- fMRI, ERP studies: compared with unimodal presentations (e.g., face only), multimodal presentations (e.g., face and voice) yield faster and more accurate emotion judgments.

Schirmer, Annett, and Ralph Adolphs. "Emotion perception from face, voice, and touch: comparisons and convergence." Trends in Cognitive Sciences (2017).

Human emotion perception

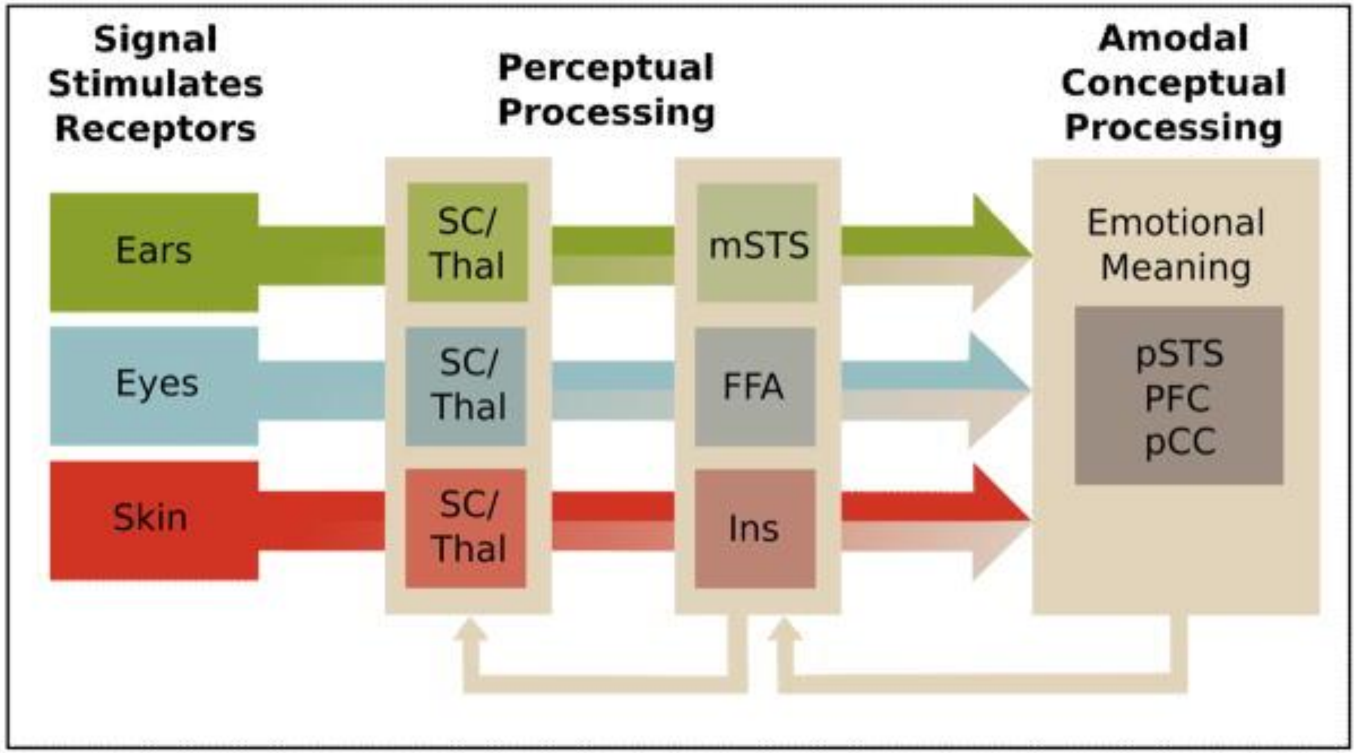
- Multimodal integration is not simply a **late process** that occurs after the individual modalities have been analyzed.
- Instead, it is supported by **multimodal neurons** at some of **the earliest processing stages**, including the superior colliculus (SC) and the thalamus.
- Early aspects of multimodal processing **provide temporal binding** and **cross-modal enhancement** (e.g., reducing a unimodal detection threshold).

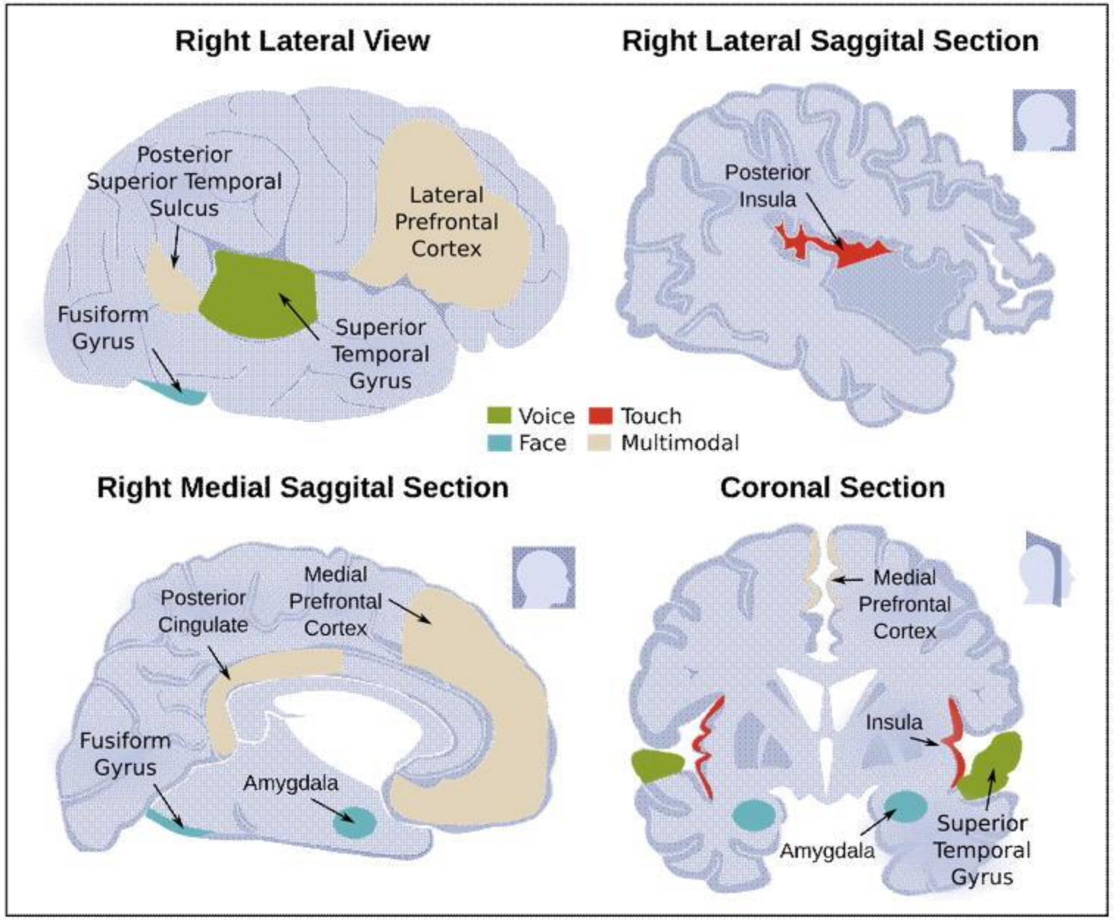
Schirmer, Annett, and Ralph Adolphs. "Emotion perception from face, voice, and touch: comparisons and convergence." Trends in Cognitive Sciences (2017).

Human emotion perception

- At **late stage of multimodal processing**, regions like the PFC, the posterior cingulate, etc. may support **amodal**, **conceptual** representations that involve a modality-nonspecific abstract code .
- Higher-level representations can feed back and **modulate** lower-level representations.

Schirmer, Annett, and Ralph Adolphs. "Emotion perception from face, voice, and touch: comparisons and convergence." Trends in Cognitive Sciences (2017).





SC: Superior Colliculus

Thal: Thalamus

mSTS: medial Superior temporal sulcus

FFA: Fusiform Face Area

Ins: Insula

pSTS: Posterior Superior Temporal Sulcus

PFC: Prefrontal Cortex

pCC: Posterior Cingulate Cortex

It seems that this woman is in pain

Or not?

Serena Williams has won the US Open championship

Challenges in automatic emotion recognition

- Mixed emotions
- Fake emotions
- Hidden emotions
- Missing channels
- Context is important
- “Ground Truth” is noisy

Multimodal affective computing

- Most studies: images (**faces**), audio (**voices**), **texts.** Rarely – body, physiology.
- Accuracy at **Multimodal** data is higher than at **Unimodal** data: by 9,83% in average, for 85% systems.
- We do not know **which type** and **how much "channels"** we need for the best classification.
- The **contribution of individual channels can be different**: for example, models based only on *audio* recognized *fear* better, and models on *visual* signs – recognized *disgust* better (EmotiW challenges).

D'mello et al., 2015; Osman et al., 2017

Multimodal affective computing

Meta analysis of 90 Multimodal systems [2015]:

- Fuse **audio** and **visual** information (55,6%)
- Detect **acted** expressions (52,2%)
- Detect “**basic**” **emotions** and **simple dimensions** of arousal and valence (64,5%)

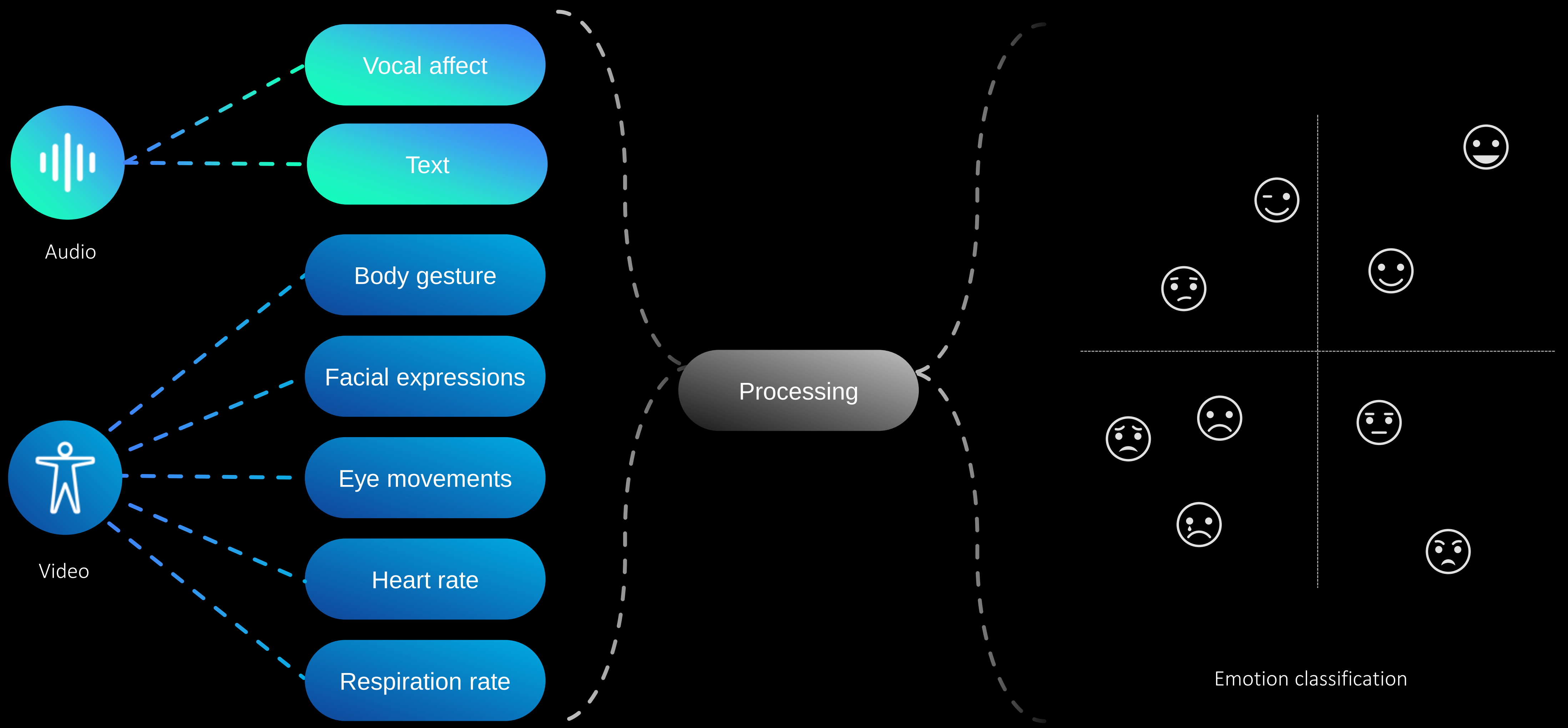
D'mello, Sidney K., and Jacqueline Kory. "A review and meta-analysis of multimodal affect detection systems." ACM Computing Surveys (CSUR) 47.3 (2015): 43

Multimodal fusion

- Relationships between UM and MM Accuracies: there is a very **robust correlation between best UM and MM accuracies** [$r(88) = 0.870$, $p < 0.001$].
- The best **UM accuracy was a significant predictor** 581 ($\beta = 0.795$, $p < 0.001$) but second-best UM accuracy was not ($\beta = 0.104$, $p = 0.174$).
- This indicates that much of **the variance in MM accuracy can be explained by the best UM accuracy**.

D'mello, Sidney K., and Jacqueline Kory. "A review and meta-analysis of multimodal affect detection systems." *ACM Computing Surveys (CSUR)* 47.3 (2015): 43

Multimodal Human Behavior Processing Unit



02

Databases

Lorem ipsum dolor sit amet consectetur

RAMAS

The Russian Acted Multimodal Affective Set [2017-2018]

RAMAS: motivation

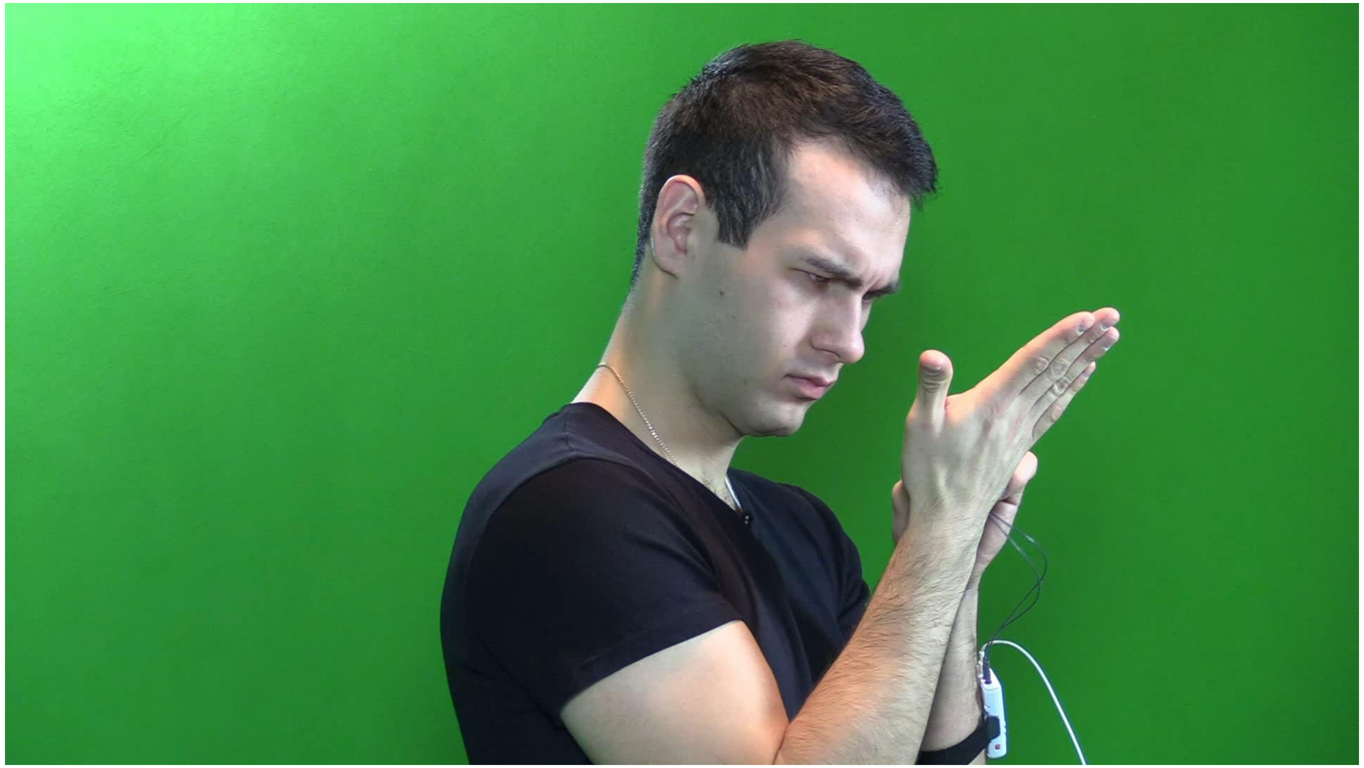
- There were no multimodal emotional databases in Russian
- The database with synchronously recorded video, audio, full body motion capture and vital signals (PPG, EDA).

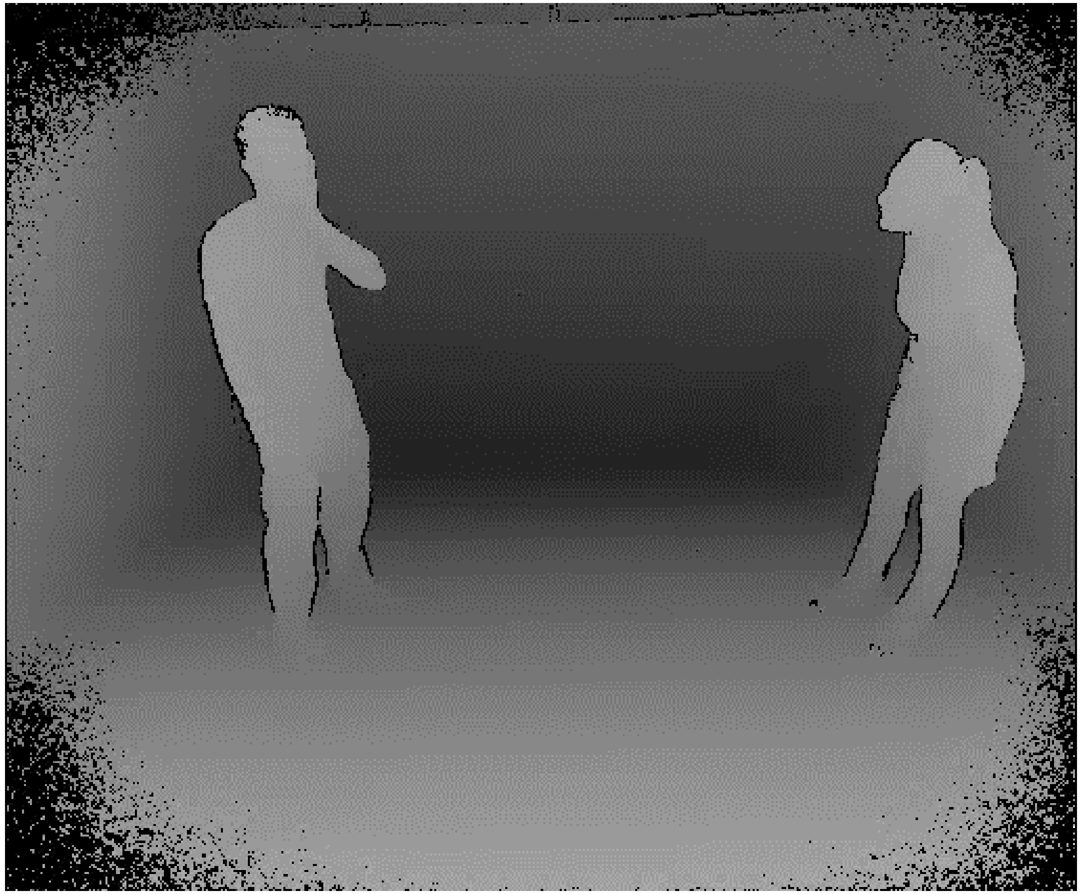
RAMAS: structure

- 10 actors (5 females и 5 males), 18-28 years old
- 6 play acted emotions: **Disgust**, **Anger**, **Happiness**, **Fear**, **Surprise**, **Sadness + Neutral**
- 2 social categories: **Domination**, **Submission**.
- 5 modalities: face, eyes, body (Kinect), voice, physiology (EDA, PPG)
- RAMAS contains 150 videos in the Russian language.
- 21 annotators (15 females, 6 males)
- Inter-rater agreement (Krippendorff's alpha): **0.44**
- Total length: ~**7 hours**

TABLE I
MEAN AND MEDIAN KRIPPENDORFFS ALPHA FOR EACH SCALE

Scale	**Mean alpha**	**Median alpha**
Disgusted	0.54	0.66
Happy	0.58	0.6
Angry	0.5	0.56
Scared	0.47	0.48
Domination	0.45	0.46
Submission	0.46	0.44
Surprised	0.41	0.38
Sad	0.35	0.31





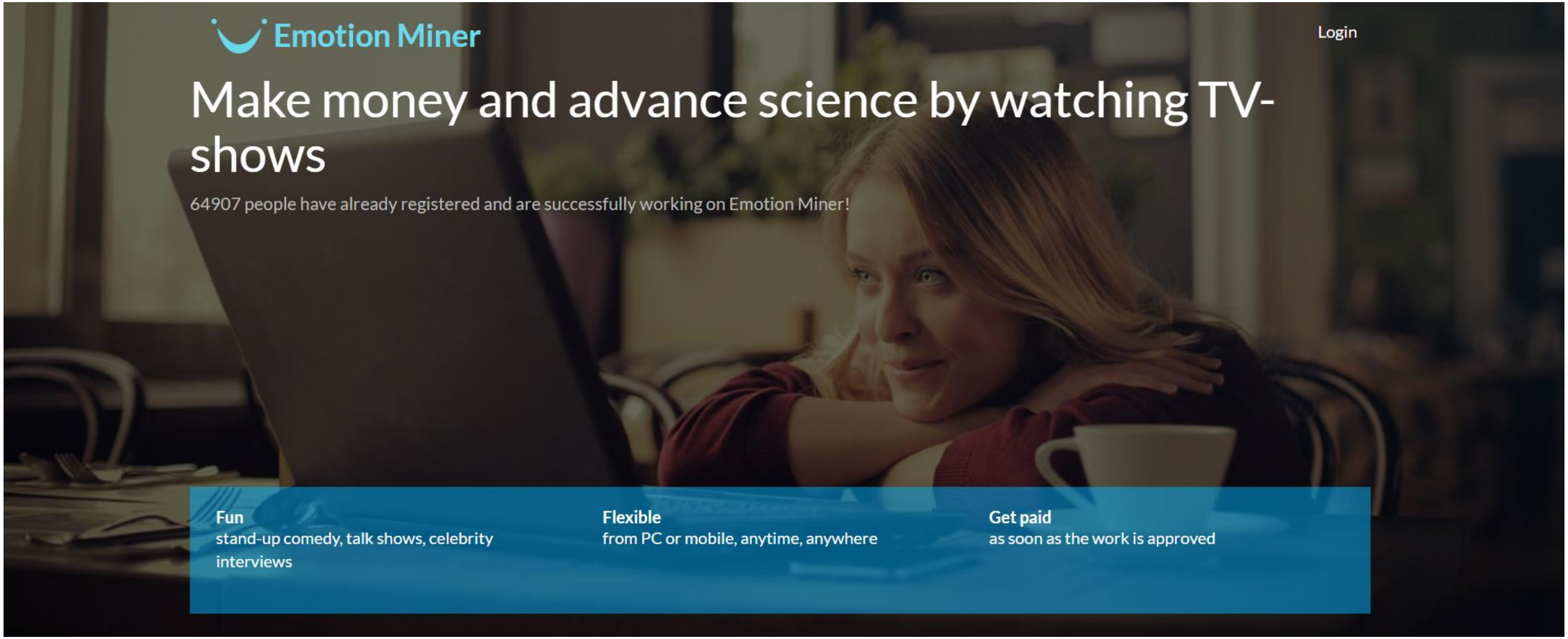
RAMAS_PPG

- 353 videos close-up videos in naturalistic settings (actors are moving, speaking, experiencing emotions, etc.) with duration 57 ± 31 sec.
- Total length: ~ 6 hours
- Ground truth: contact PPG by Shimmer Consensys GSR Development Kit.
- Ground truth heart rate: 94 ± 12 bits per sec.

RAMAS: links & access

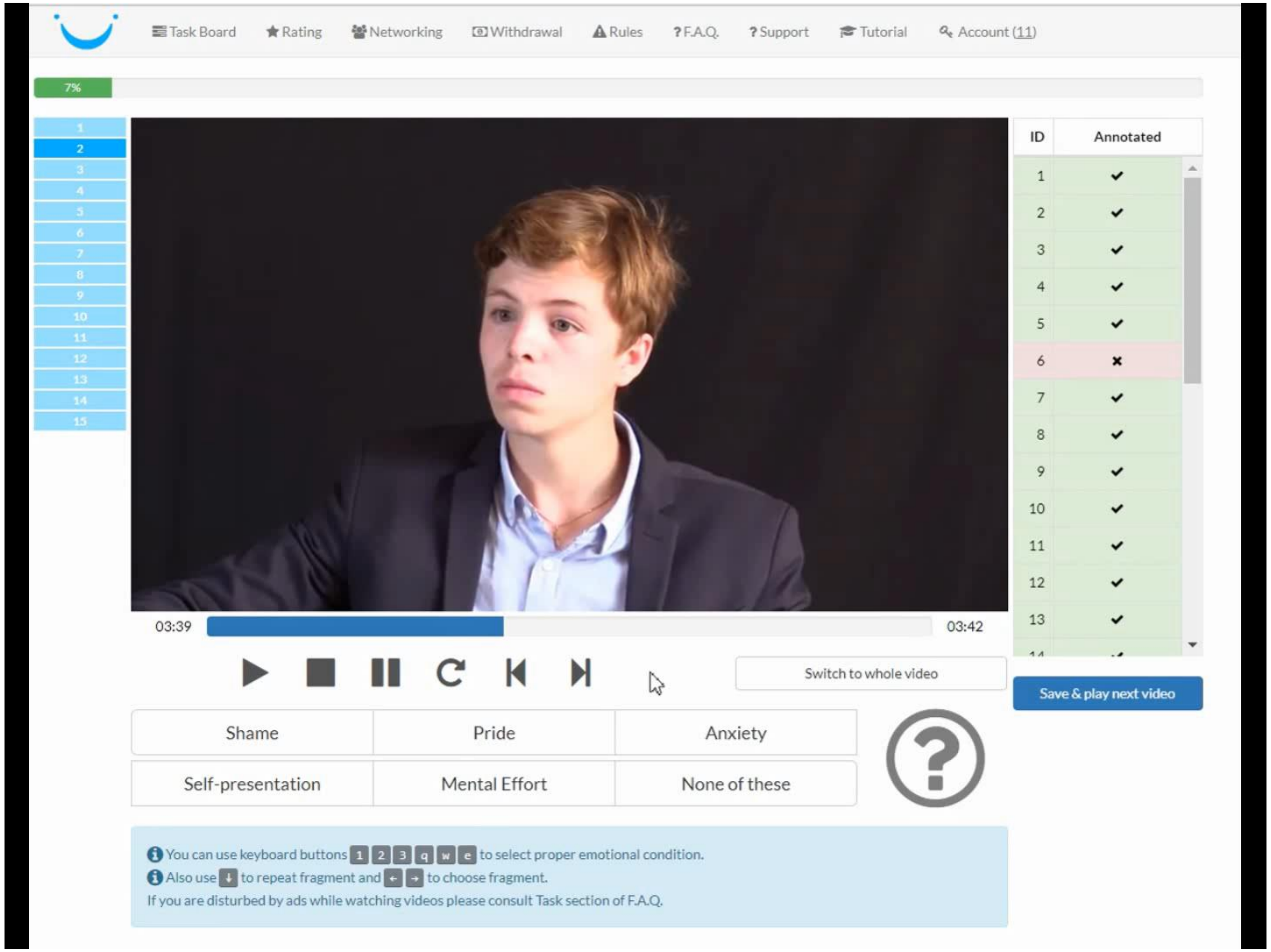
- https://neurodatalab.com/category/ramas/
- https://link.springer.com/chapter/10.1007/978-3-319-99579-3_52
- Perepelkina, O., Kazimirova, E., & Konstantinova, M. (2018, September). RAMAS: Russian Multimodal Corpus of Dyadic Interaction for Affective Computing. In *International Conference on Speech and Computer* (pp. 501-510). Springer, Cham.

EMDC: Emotion Miner Data Corpus



~ 65000 registered users at www.emotionminer.com

EMDC: annotation

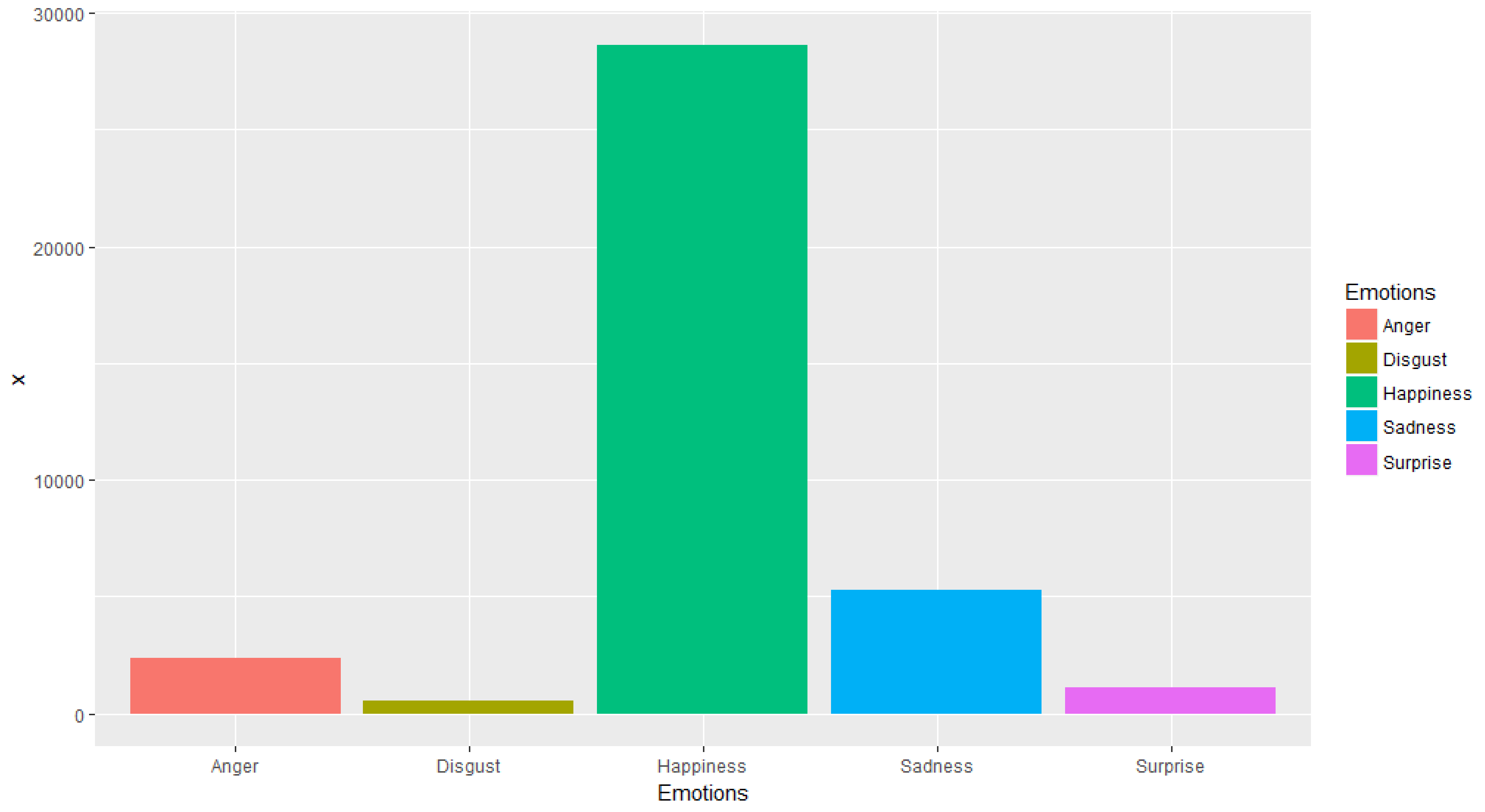


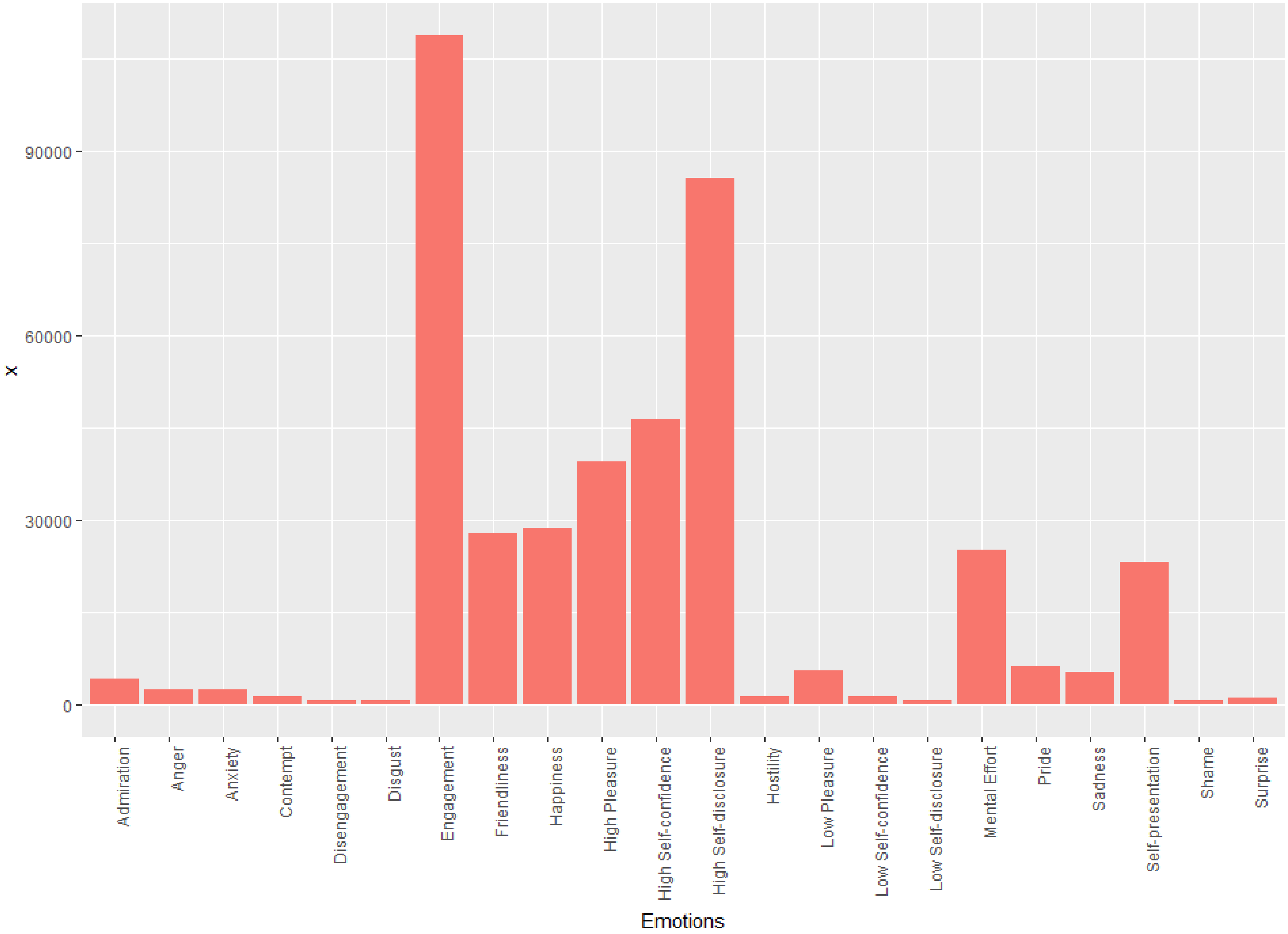
EMDC: structure

- **160 hours** of labelled public audiovisual content in English (more than **110 000 annotated video** ~5 sec fragments)
- 10+ annotators for each fragment
- 22 emotional states and social categories
- ~ 1500 annotators in total (29 ± 10 years old)

EMDC: data collection

- **Video content**: conversations & monologues
  - TV-shows, monologues (biographies), interviews, talks, “trash-show”
- Content selection:
  - 1-2 people in the frame
  - English language
  - Without loud noises, background music, etc.


30000
20000
10000
0
x
Anger
Disgust
Happiness
Sadness
Surprise
Emotions
Emotions
Anger
Disgust
Happiness
Sadness
Surprise



90000
60000
30000
0
x
Admiration
Anger
Anxiety
Contempt
Disengagement
Disgust
Engagement
Friendliness
Happiness
High Pleasure
High Self-confidence
High Self-disclosure
Hostility
Low Pleasure
Low Self-confidence
Low Self-disclosure
Mental Effort
Pride
Sadness
Self-presentation
Shame
Surprise
Emotions

Correlations between scales (>0.5)

- Friendliness – Happiness: 0.65
- Hostility – Anger: 0.56
- Happiness – High Pleasure: 0.54
- Friendliness – High Pleasure: 0.50

EMDC: link

- https://neurodatalab.com/category/emotion-miner-data-corpus/
- Database is partly open for our collaborators
- In future we will share it with the research community

03

Engagement & Disengagement detection

[ULM University & Neurodata Lab]

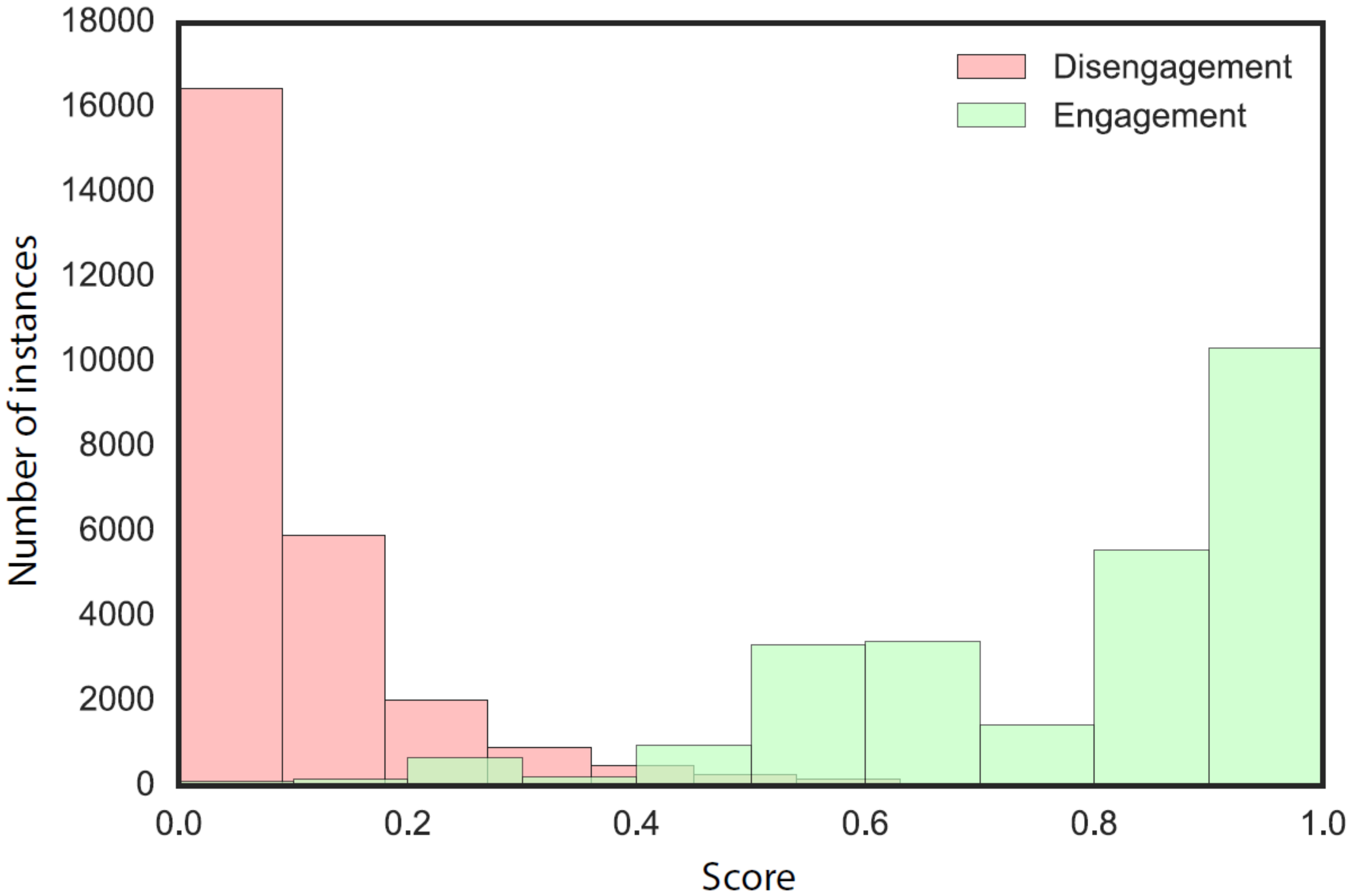
Lorem ipsum dolor sit amet consectetur

EMDC: Engagement & Disengagement detection

- Automatic detection of disengagement and engagement states is a challenge in a wide range of contexts.
- Engagement is "**the value that a participant in an interaction attributes to the goal of being together with the other participant(s) and of continuing the interaction**" [Poggi, 2007].
- "**The process where two (or more) participants establish, maintain and end their perceived connection**" [Nakano, Ishii, 2010; Turney, 1994] and reflects how much the subject is **interested in and willing to continue the current dialogue** [Inoue, Lala et al., 2018].

For the study we used **26 446 fragments** from 981 videos (~ 36 hours). This part of the EM corpus contained all the available disengagement fragments and randomly chosen videos with engagement.

98.2% of fragments were labeled as "engagement" and only **1.8%** were labeled as "disengagement".



Features

- **“Audio”**: 65 low-level descriptors (LLDs) and their derivatives from the ComParE feature set were obtained, 130 features in total. For feature extraction open-source toolkit openSMILE was used.
- **“Face”**: As face features embeddings from person identification model were used. To obtain the embeddings we trained the ResNet-50 neural network (pretrained on ImageNet) on a dataset with more than 10000 faces.
- **“Body”**: 2D body poses from each frame of the videos were extracted. 18 keypoints were detected for each person within the frame.

Features

- **“Eyes”**. Coordinates of pupil with respect to nose obtained with a Convolutional Neural Network (CNN) were used.
- **“Lips”**: The speaking lips technology identifies the probability of a person speaking at a particular moment [ResNet-50 + LSTM].

Model

- The **threshold** for considering particular label was 0.5. This means that at least 50% of annotators had to agree on it.
- We used **logistic regression with L2 regularization** to build uni- and multimodal models.
- Modalities fusion is performed at **feature-level fusion** and **decision-level fusion**.
- Metric: **UAR** (unweighted average recall)

1 modality	2 modalities	3 modalities	4 modalities
	Audio + Eyes 0.6932 (0.00)		
Audio 0.6709	Audio + Lips 0.6919 (0.00)	A + L + B 0.7038 (0.03)	A + L + B + E 0.7041 (0.84)
Lips 0.6621	Audio + Body 0.6864 (0.02)	A + L + E 0.7037 (0.00)	
	Audio + Face 0.6746 (0.45)		

Figure 6: Performance (UAR) of different modality combinations. Each box contains the names of the modalities used, the performance of the system and t-test p-value of the current result (green color for significant results, red color for non-significant), compared to the best one from the previous level.

The best scores among all combinations are achieved by **FLF** except Audio+Face (**DLF**).

Conclusions

- Most contributing modalities in engagement/disengagement detection: **audio** and **speaking lips**
- Engagement is highly connected with speech parameters
- Highest performance: logistic regression with class weighting, four modalities (audio, lips, eyes, body), feature-level-fusion

Link

Fedotov, D., Perepelkina, O., Kazimirova, E., Konstantinova, M., & Minker, W. (2018, October). Multimodal approach to engagement and disengagement detection with highly imbalanced in-the-wild data. In *Proceedings of the Workshop on Modeling Cognitive Processes from Multimodal Data* (p. 2). ACM.

04

Automatic emotion recognition from expressive body gestures

Lorem ipsum dolor sit amet consectetur

Bodily communication of emotion

- Bodily behaviors are **important** for emotional communication. They are likely to be as important as facial communication.
- Bodily displays can be visually salient from afar, allowing for successful emotion communication across **long distances** [de Gelder, 2009; de Gelder, 2016; Martinez, Falvello, Aviezer, & Todorov, 2015]
- Bodily expressions of emotion often can be interpreted from behind an expresser, including when the **face is completely hidden** [Coulson, 2004; Sogon & Masutani, 1989]
- It was shown that in the case of **strong emotions**, body expressions are even more revealing and **reliable** than facial expressions [Aviezer, H., Trope, Y. and Todorov, A. 2012. Body Cues, Not Facial Expressions, Discriminate Between Intense Positive and Negative Emotions. Science, 30 (Nov 2012), 1225-1229].

- Behavior **types** (including spatial form, poses): leaning back, tilting the head, hands in fists.

- Behavior **qualities** (spatiotemporal form, kinematics, dynamics): fast, slow, jerky, flowing.

- Dynamic stimuli that allow for the perception of **both type and quality** typically receive the highest recognition rates [e.g., Atkinson et al., 2004].
- Static photographs of bodily displays that show only **type**, and dynamic point light displays that focus on **qualities** are also recognized well above chance [Atkinson et al., 2004; de Gelder & Van den Stock, 2011; Tracy & Robins, 2007]

Automatic emotion recognition from 2D video

Bodily information: feature extraction [2D]

Hand-craft approach

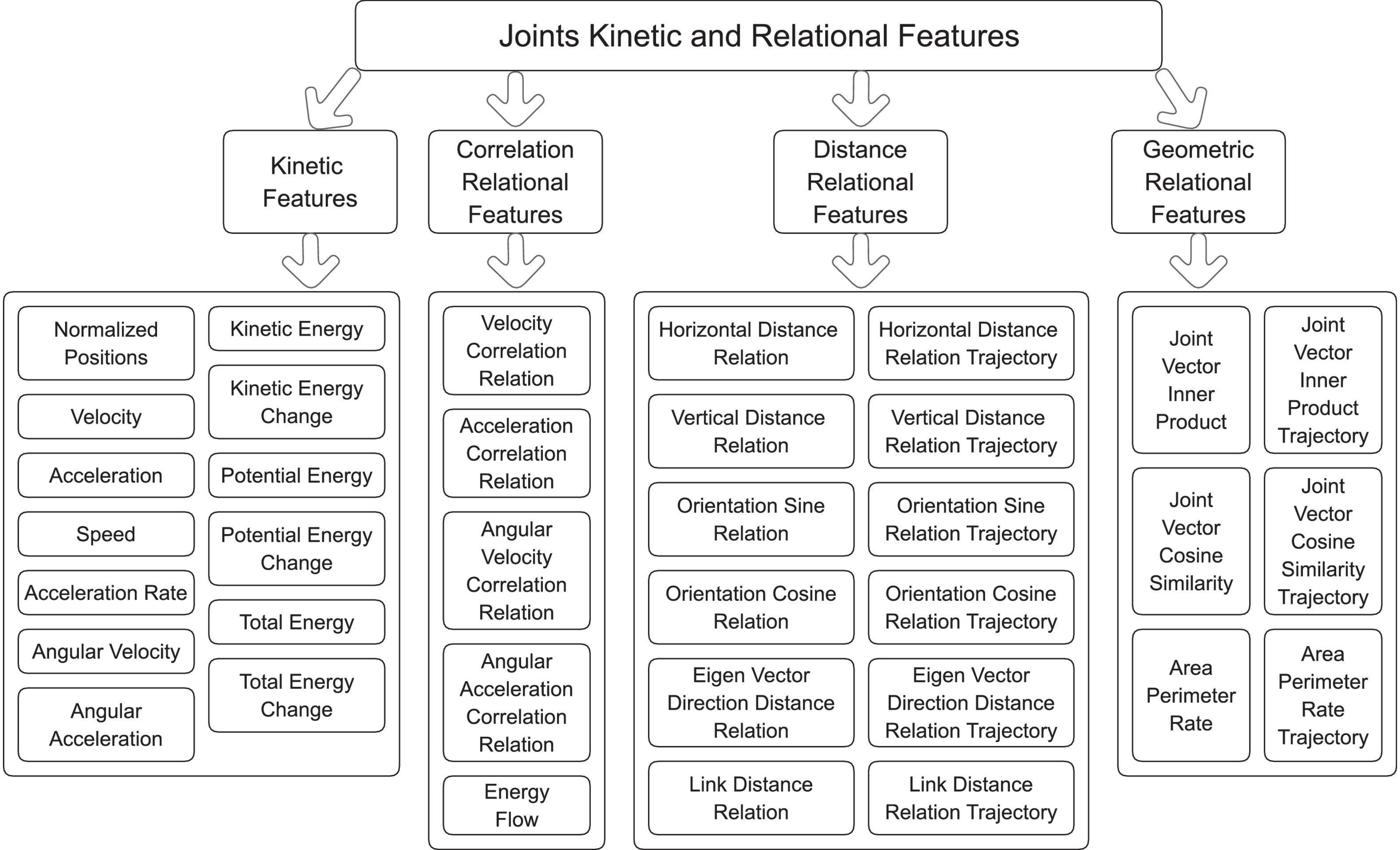
- Extract body poses from each frame
- Calculate features: kinematic features, expressive features, etc.
- Select features or reduce dimensionality [e.g., PCA, t-SNE]

Deep-learned approach

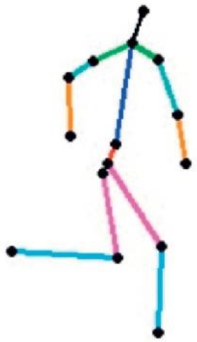
- Use neural network (NN) for feature extraction
- Transfer learning: train NN at huge dataset [AVA, Kinetics datasets] to classify images
- Use pretrained neural network in the final database [e.g., FABO]

Hand-crafted features

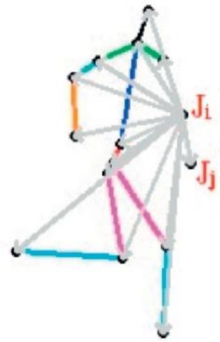
Get skeleton in each frame [2D coordinates of body parts] => calculate some features


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

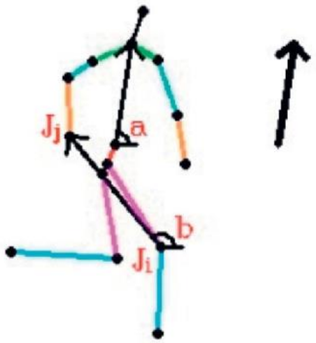
(a) A video frame



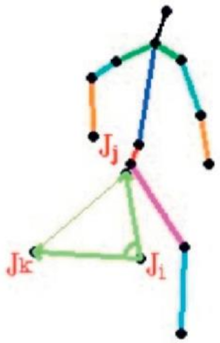
(b) Joints and skeleton



(c) Distance relation



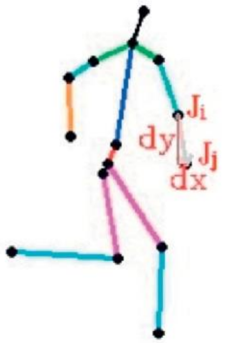
(d) Orientation (cosine/sine) relation



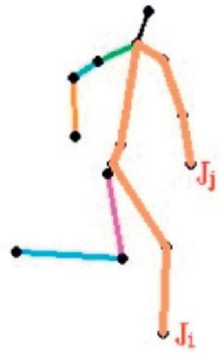
(e) Angle (geometric) relation

(f) Trajectories of joints

(g) Velocity of joints

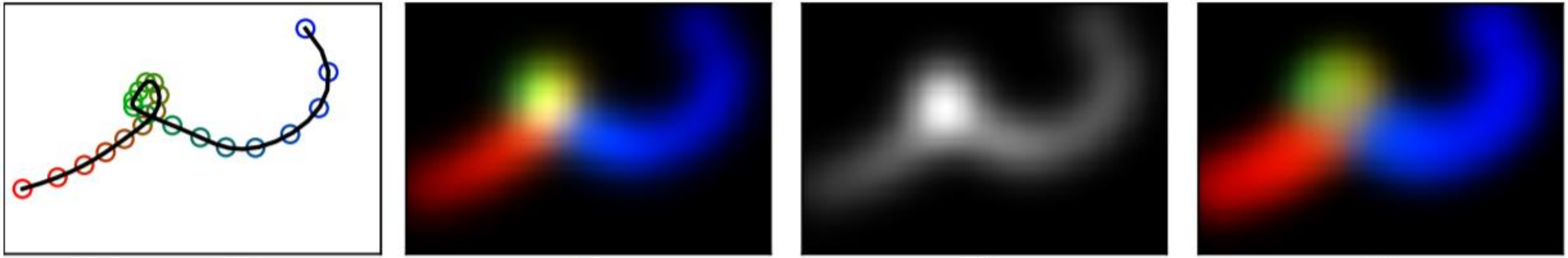


(h) Horizontal (vertical) distance relation



(i) Link distance relation

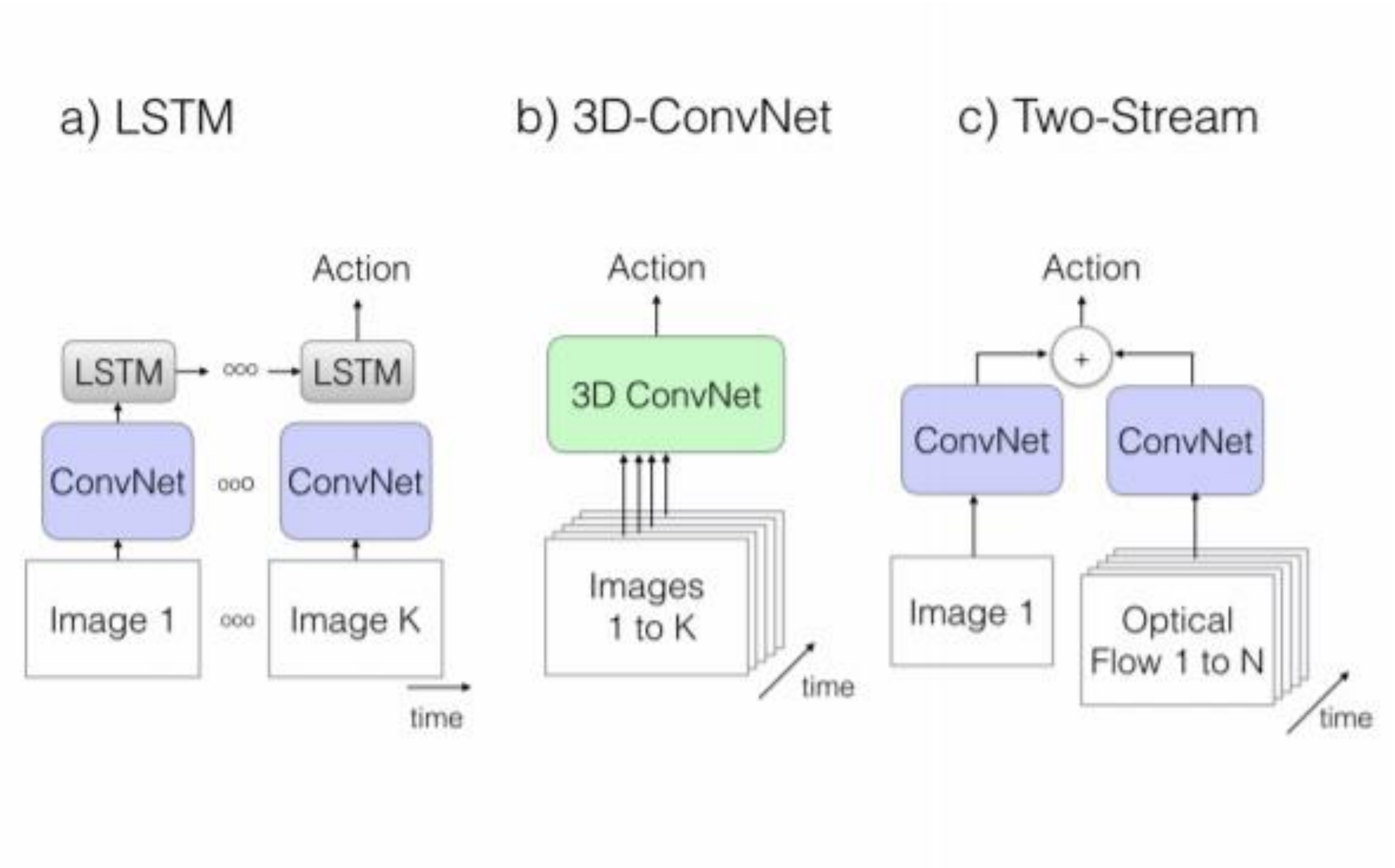
PoTion [Pose moTion representation]



Trajectory of joint j — $\mathcal{U}_j$ — $\mathcal{I}_j$ — $\mathcal{N}_j$

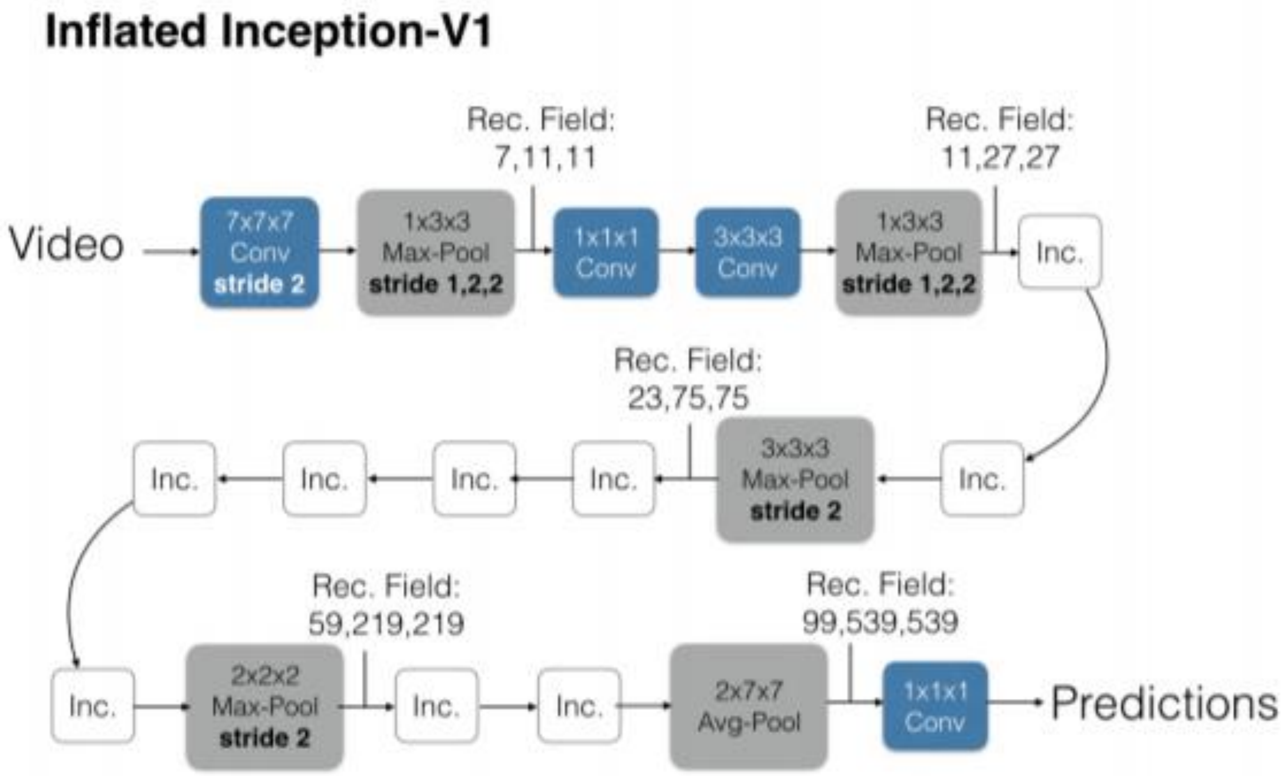
https://hal.inria.fr/hal-01764222/document

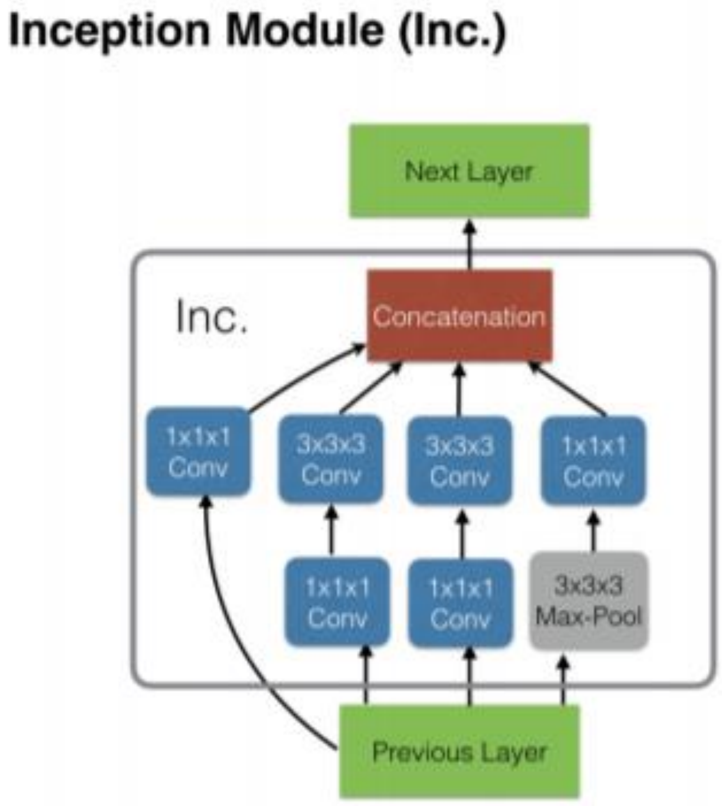
Deep-learned approaches: examples



Approaches are usually used for action recognition from videos

I3D [Two-Stream Inflated 3D ConvNet]



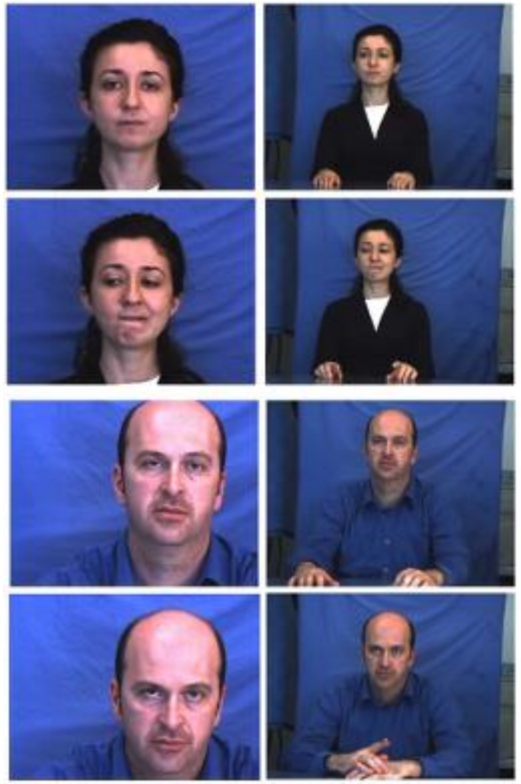


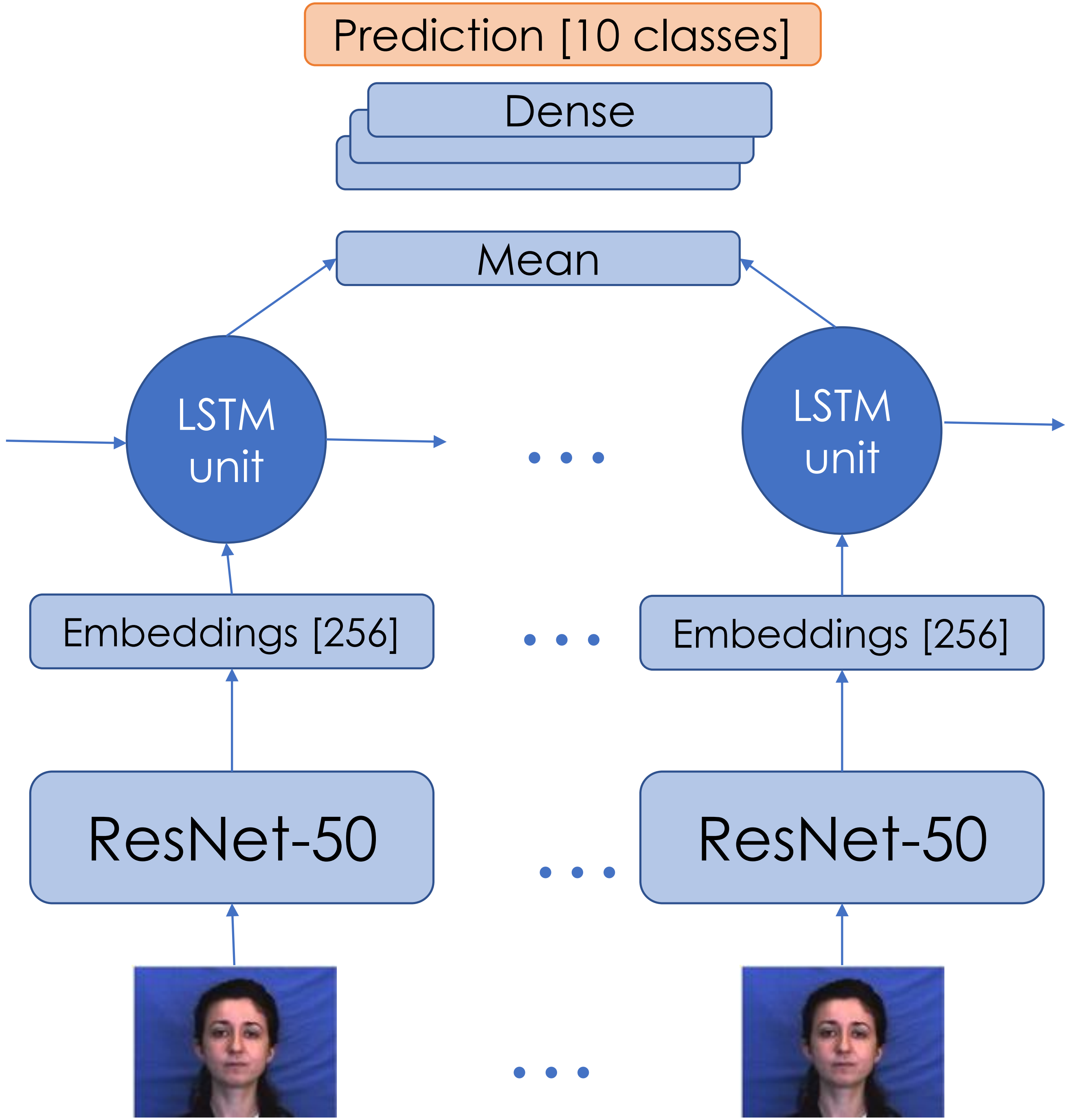
https://arxiv.org/pdf/1705.07750.pdf

Experiments on FABO database

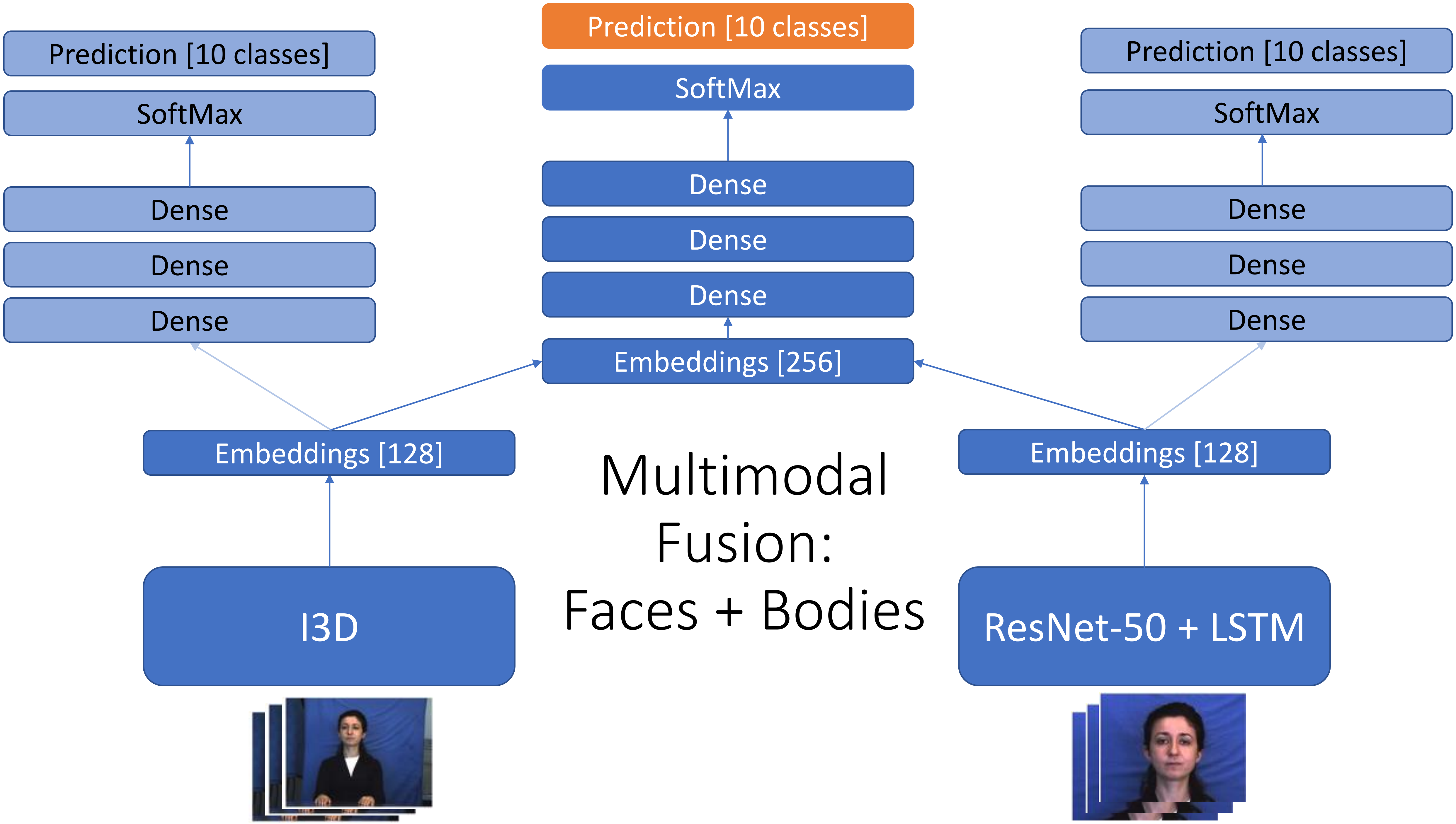
- Anger
- Anxiety
- Boredom
- Disgust
- Fear
- Happiness
- Puzzlement
- Sadness
- Surprise
- Uncertainty

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





Face: Architecture


Prediction [10 classes]
SoftMax
Dense
Dense
Dense
Prediction [10 classes]
SoftMax
Dense
Dense
Dense
Embeddings [256]
Prediction [10 classes]
SoftMax
Dense
Dense
Dense
Embeddings [128]
Embeddings [128]
Multimodal Fusion: Faces + Bodies
I3D
ResNet-50 + LSTM

Results: FABO

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

Leave one subject out cross validation

Experiments on EmotionMiner database

- For the experiment we selected data labelled for 4 classes:
- Happiness, Anger, Sadness, Neutral
- [4400 fragments, ~ 6 hours]

Emotion Miner

Results: EmotionMiner

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

Conclusions

- In our experiments on a small play acted database (**FABO**) there were **no gain in performance from multimodal fusion**.
- **FABO** database: the **body modality** showed the highest performance [database specificity].
- Naturalistic database (**EmotionMiner**): there is **a small gain in performance (~2%) from multimodal fusion**.
- Previous works have shown that improvement from using multimodal fusion is usually stronger in played databases [D'mello, Sidney K., and Jacqueline Kory. "A review and meta-analysis of multimodal affect detection systems." *ACM Computing Surveys (CSUR)* 47.3 (2015): 43].

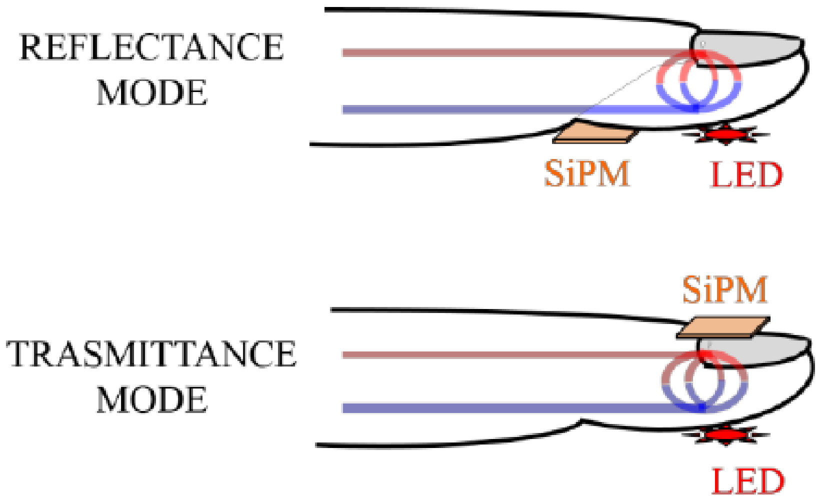
05

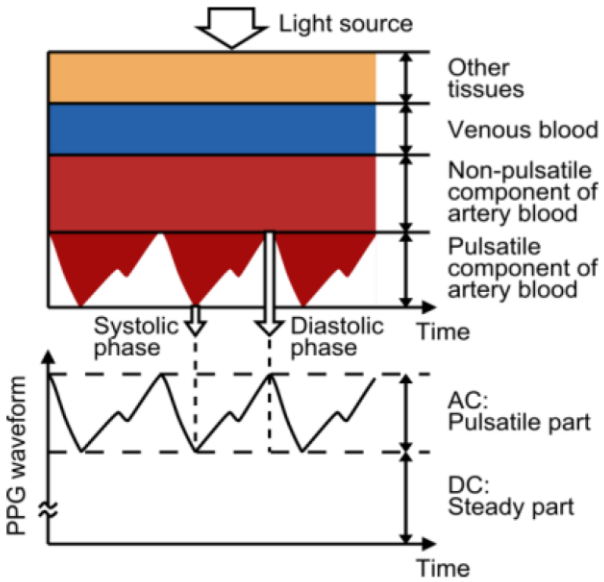
Heart rate estimation from video

Lorem ipsum dolor sit amet consectetur

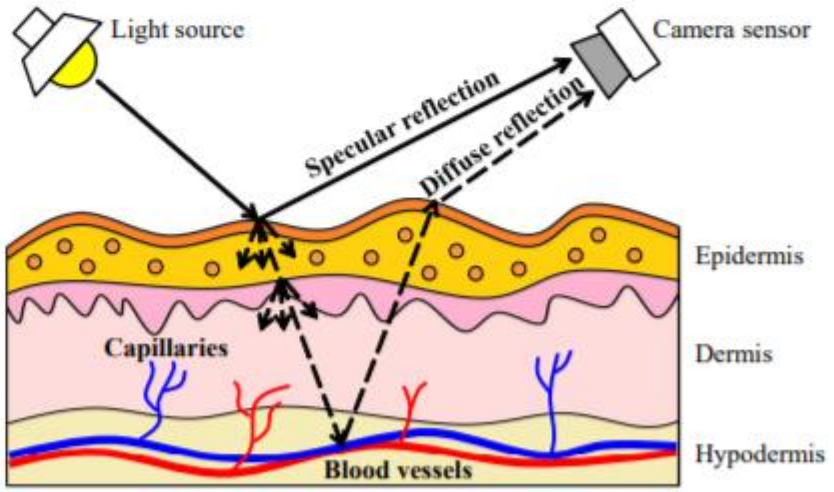
Photoplethysmography

Photoplethysmography (PPG) is an optical method to measure skin light reflectance or transmission intensity variations. This technology is widely used in medicine, sport and other applications.





Imaging photoplethysmography

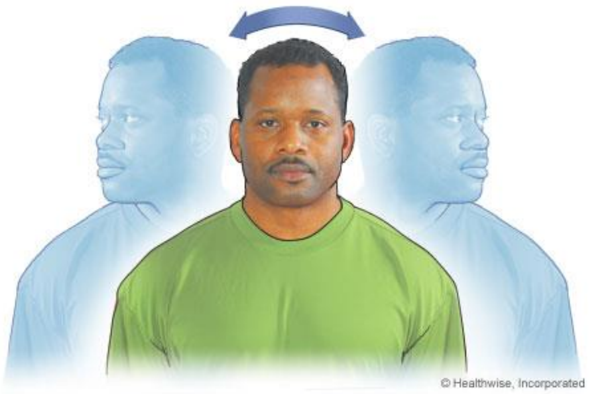


Remote video photoplethysmography (or Imaging photoplethysmography, **iPPG**) requires only **ambient light** and digital **camera** to achieve person vital signals. The digital camera detects **skin color changes** that represent a light absorption by hemoglobin.

iPPG is an appealing method that can be potentially useful for obtaining person heart rate data in realistic settings.

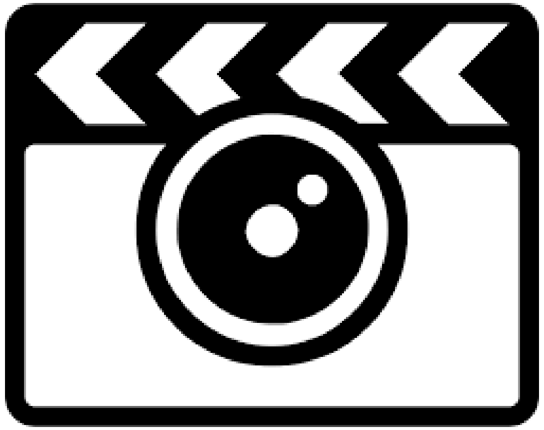
iPPG: challanges

Motion artifacts

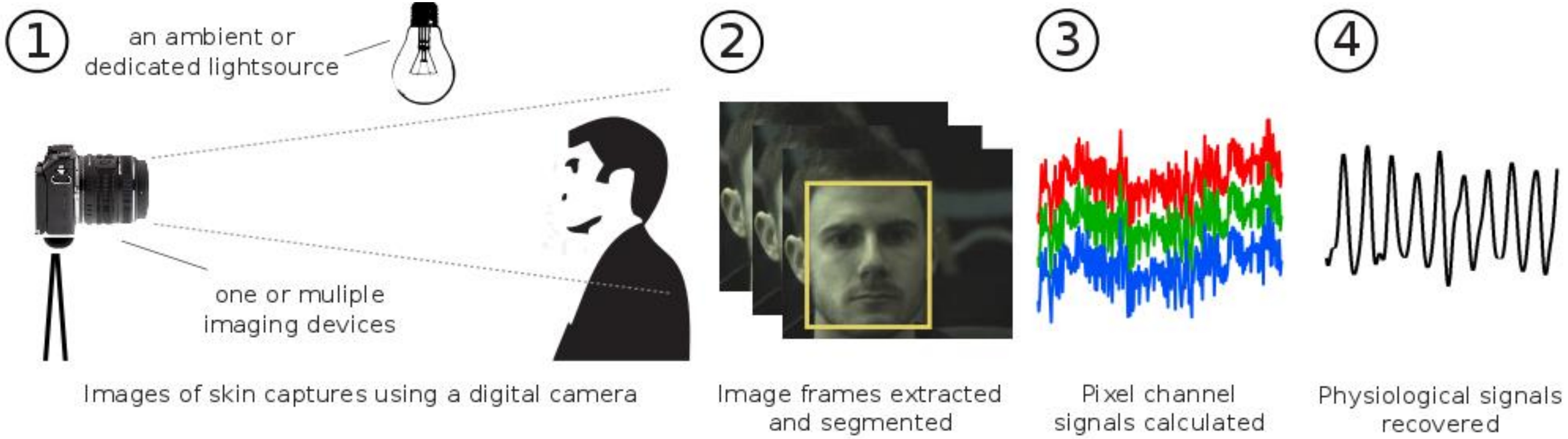


Changing lighting parameters

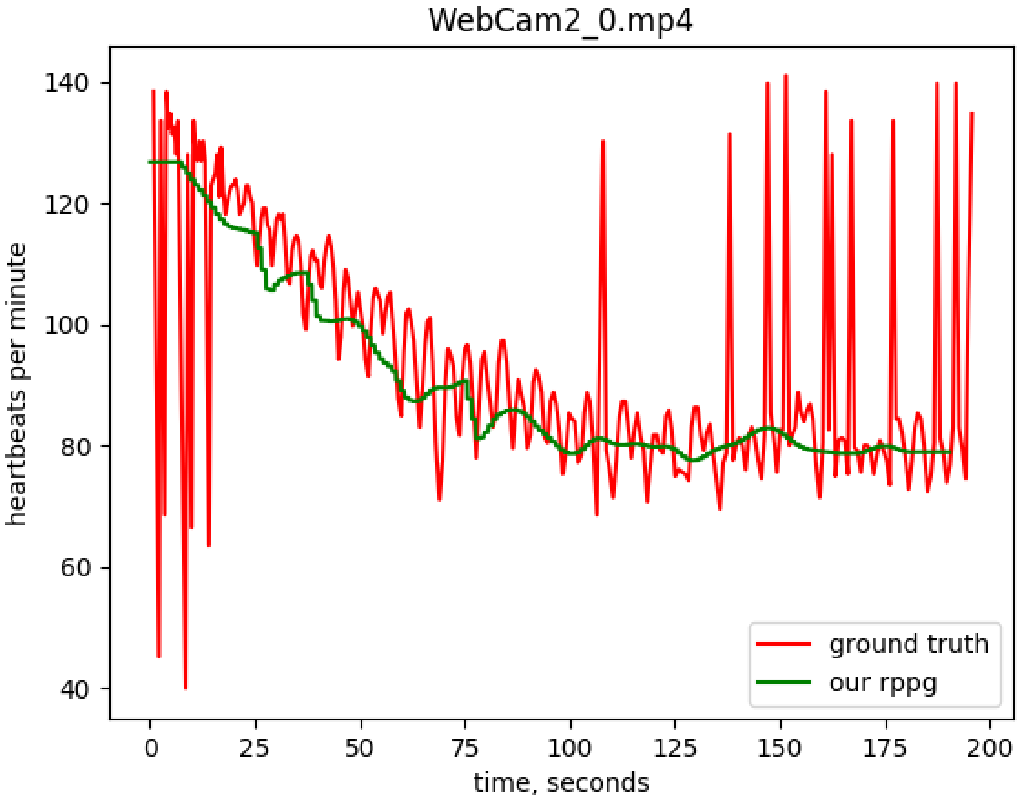
Low quality of video



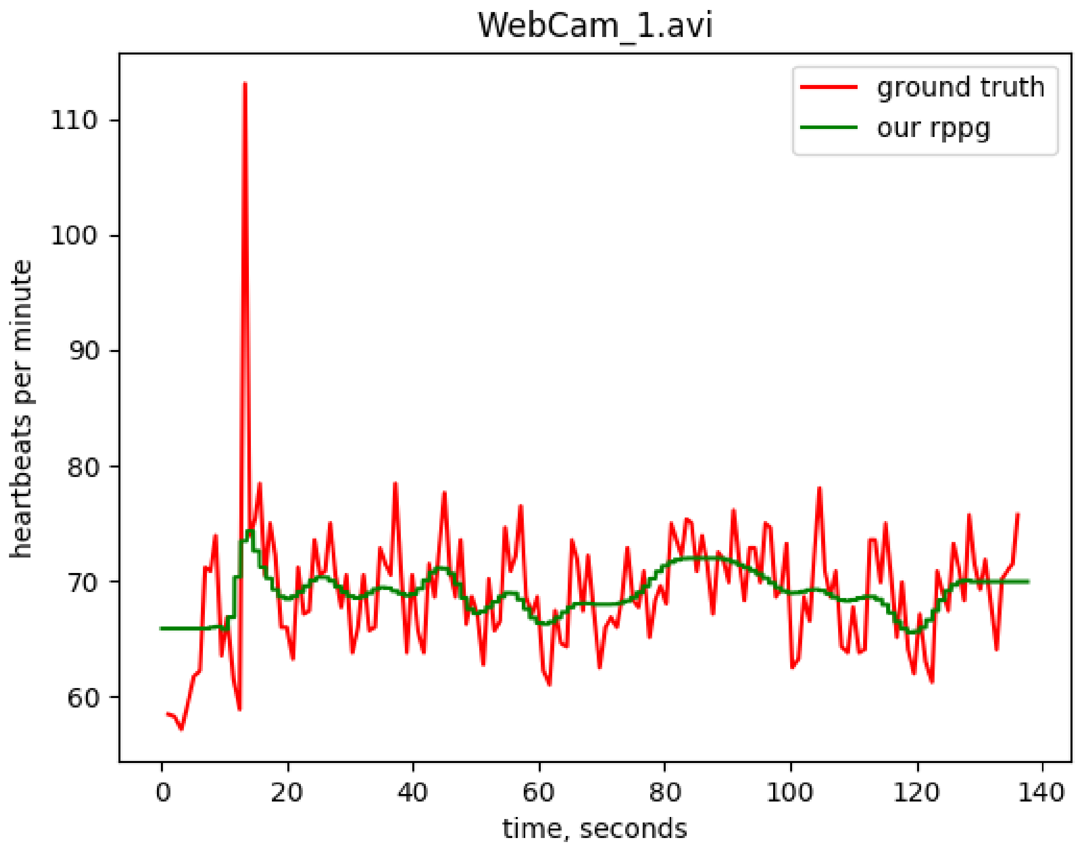
iPPG general scheme



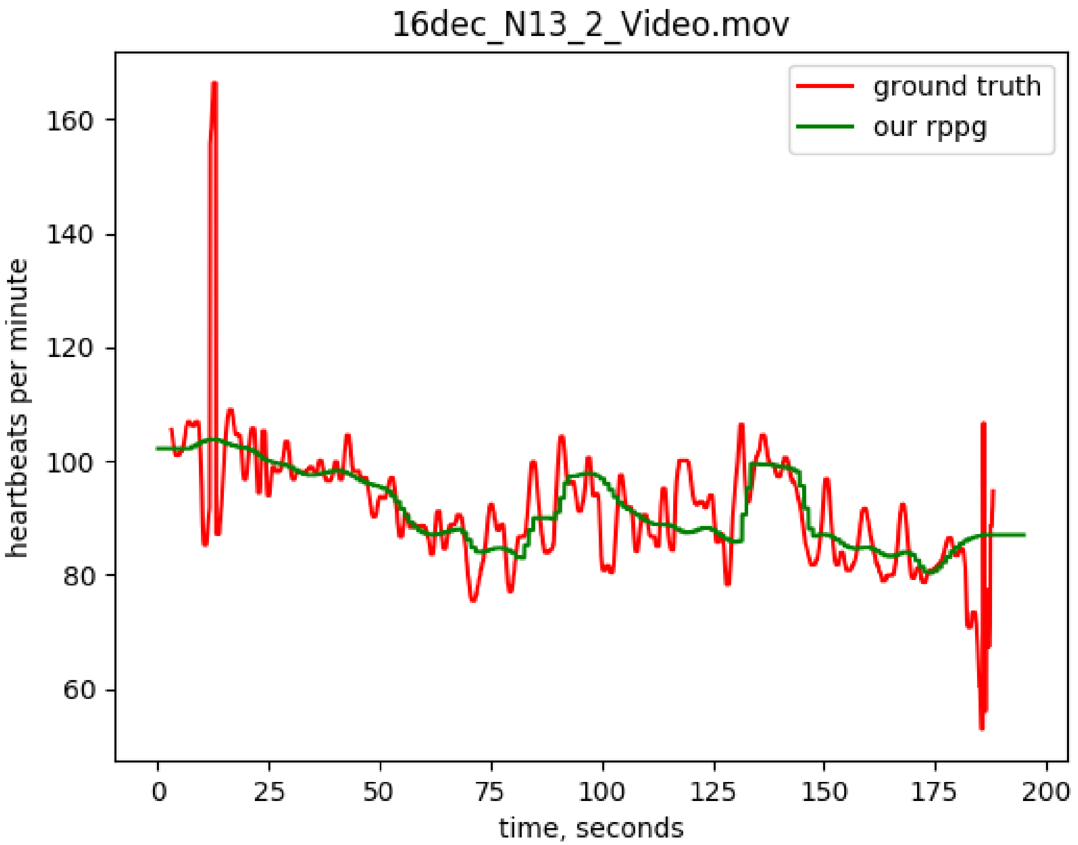
After physical exercises [trend in heart rate]



Blinking illumination [from the screen]



Movements and speech [naturalistic data]



UBFC-RPPG database

- Logitech C920 HD Pro, 30fps, 640x480.
- A CMS50E **transmissive pulse oximeter** was used to obtain the **ground truth PPG**
- The subject sits in front of the camera (about 1m away from the camera) with his/her face visible and is required to play a time sensitive mathematical game that aimed at augmenting their heart rate while simultaneously emulating a normal human-computer interaction scenario
- **41 videos**, duration of each video is ~ 1 min (~40 min in total)
- Bobbia et al., 2017

iPPG algorithms comparison

	Our approach	“Heartbeat”	“HR measurement using camera”
UBFC	3.032	14.586	26.024
RAMAS	20.167	20.744	29.641

Bobbia et al., 2017: RMSE for UBFC **2.4**

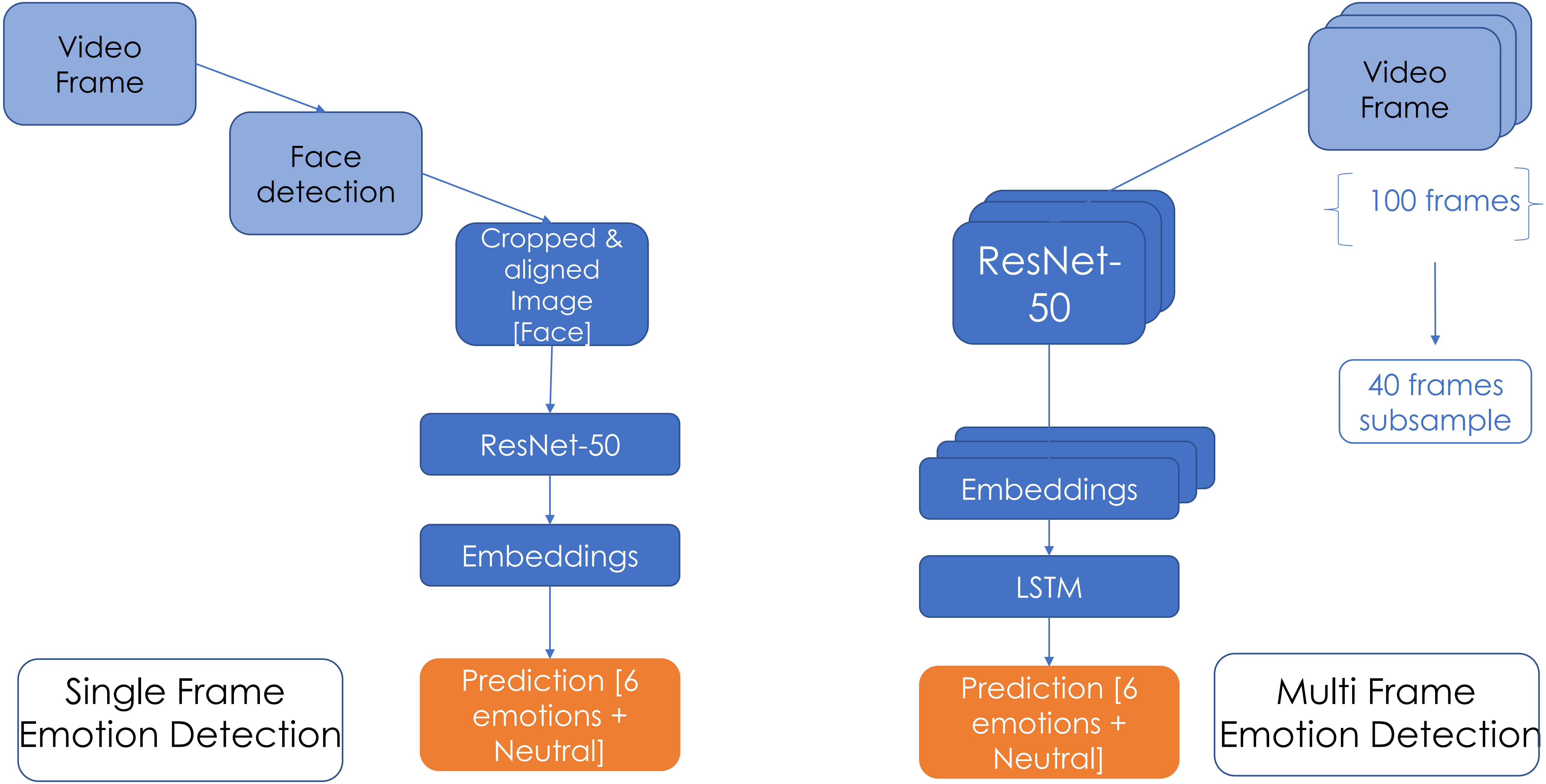
Metric: RMSE

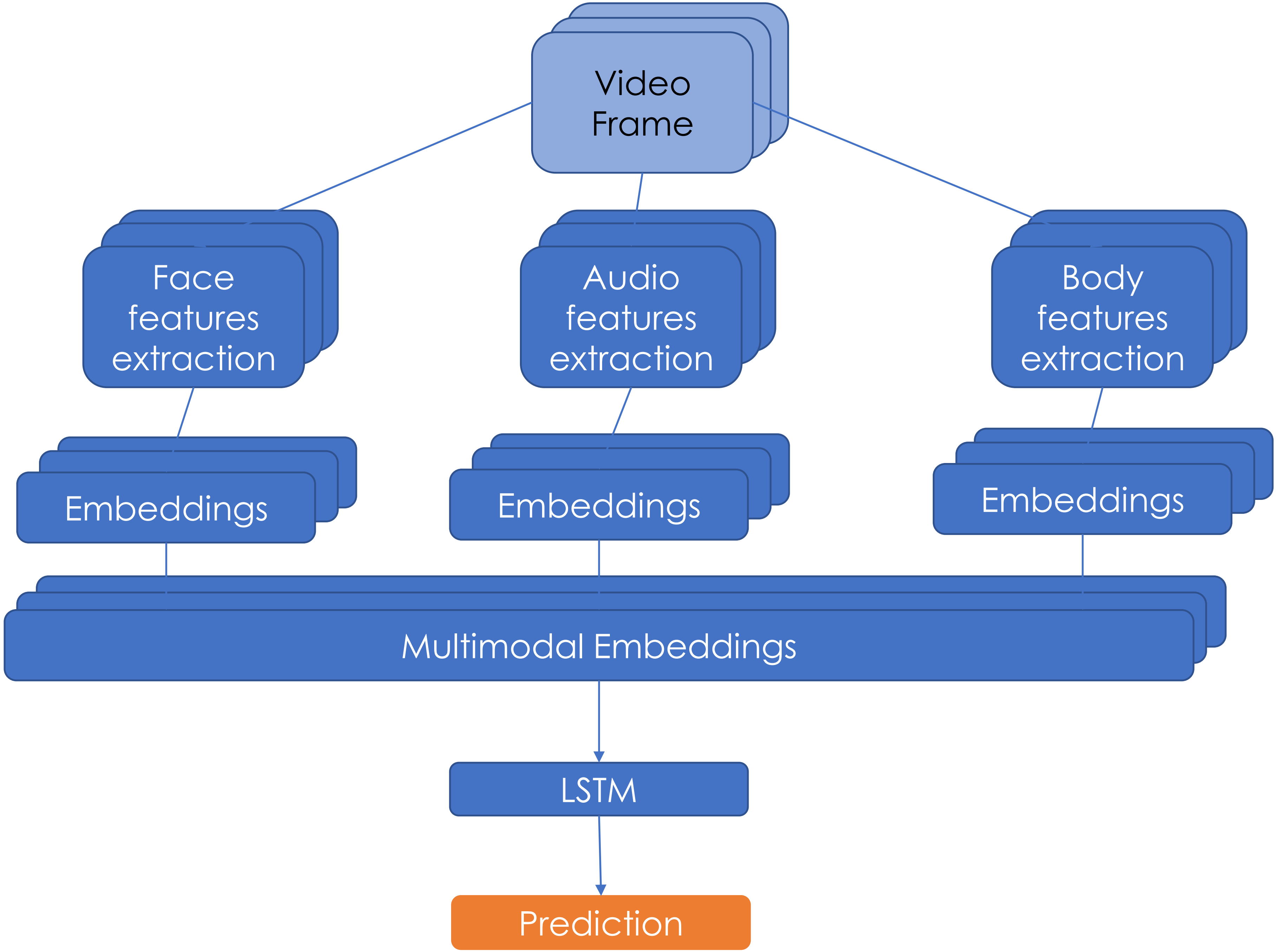
$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}$$

05

Commercial APIs comparison

Lorem ipsum dolor sit amet consectetur


Video Frame
Face detection
Cropped & aligned Image [Face]
ResNet-50
Embeddings
Prediction [6 emotions + Neutral]
Single Frame Emotion Detection
Video Frame
100 frames
40 frames subsample
ResNet-50
Embeddings
LSTM
Prediction [6 emotions + Neutral]
Multi Frame Emotion Detection


Video Frame
Face features extraction
Audio features extraction
Body features extraction
Embeddings
Embeddings
Embeddings
Multimodal Embeddings
LSTM
Prediction

Emotion Recognition algorithms comparison

API

- Affectiva
- Microsoft
- Amazon
- Neurodata Lab

Databases

- SAVEE
- AFEW
- RAVDESS

SAVEE

- Play acted emotions. 4 male actors in 7 emotions [Anger, Disgust, Fear, Happiness, Sadness, Surprise, Neutral]. 480 British English utterances in total. Audio + Video [close-up].

S. Haq and P.J.B. Jackson (2010).

	Neurodata Lab	Affectiva	Microsoft	Amazon
Precision	0.61	0.22	0.57	0.43
Recall	0.37	0.35	0.48	0.34
F1	0.32	0.24	0.42	0.25
Accuracy	0.32	0.30	0.56	0.30

AFEW

- Video fragments extracted from movies. 1156 videos in total. Emotions: angry, disgust, fear, happy, sad, surprise and the neutral. Audio + Video.

A. Dhall, R. Goecke, S. Lucey, and T. Gedeon (2011).

	Neurodata Lab	Affectiva	Microsoft	Amazon
Precision	0,39	0,13	0,67	0,32
Recall	0,32	0,15	0,39	0,3
F1	0,26	0,09	0,36	0,28
Accuracy	0,3	0,15	0,45	0,15

RAVDESS

- Acted emotions. 24 actors x 60 short videos (1440 videos in total). Audio + Video (close up). Emotions: neutral, calm [was combined], happy, sad, angry, fearful, disgust, surprised).

Livingstone SR, Russo FA (2018)

	Neurodata Lab	Affectiva	Microsoft	Amazon
Precision	0,58	0,32	0,57	0,39
Recall	0,52	0,33	0,37	0,44
F1	0,49	0,26	0,33	0,39
Accuracy	0,51	0,31	0,39	0,44

AFEW	Neurodata Lab	Affectiva	Amazon	Microsoft
Precision	0,39	0,13	0,32	0,67
Recall	0,32	0,15	0,3	0,39
F1	0,26	0,09	0,28	0,36
Accuracy	0,3	0,15	0,15	0,45

SAVEE	Neurodata Lab	Affectiva	Amazon	Microsoft
Precision	0,61	0,22	0,43	0,57
Recall	0,37	0,35	0,34	0,48
F1	0,32	0,24	0,25	0,42
Accuracy	0,32	0,3	0,3	0,56

RAVDESS	Neurodata Lab	Affectiva	Amazon	Microsoft
Precision	0,58	0,32	0,39	0,57
Recall	0,52	0,33	0,44	0,37
F1	0,49	0,26	0,39	0,33
Accuracy	0,51	0,31	0,44	0,39

06

Applications & Industrial Cases

Lorem ipsum dolor sit amet consectetur

NEUR DATA LAB
Microsoft
empatica
HNL
heart never lies
amazon.com
tobii
EYETRACKING
Affective Computing
Apple
BRAIQ
soul machines
human
Fraunhofer
synapbox
eyeris
:) Affectiva
sightcorp
Affective Sciences
SWISS NATIONAL CENTER OF COMPETENCE IN RESEARCH
KAIROS
IBM
BEYONDVERBAL
THE EMOTIONS ANALYTICS COMPANY
SKY BIOMETRY
Google
EMRAYS
GOOD VIBRATIONS
TUNING YOUR VIBES
VOKATURI
emotion recognition by speech
ENTROPIK TECH
MOTISTA™
SPEECH GRAPHICS
Noldus
Information Technology
IMOTIONS
BIOMETRIC RESEARCH PLATFORM
audEERING™
intelligent Audio Engineering
SENSUM
TAWNY
EMOTION RESEARCH LAB
NVISO
crowdemotion
EMOSHAPE
EMOTIONS NEXT FRONTIER
Panasonic
EMOTIBOT
ADOREBOARD
Opsis
HEARTBEAT Ai
TECHNOLOGIES
real eyes
emotional intelligence
feel
cogito
nuralogix
SilverLogic
LABS
Receptiviti
emteq
cognoviLabs
Q°emotion

Challenges for Emotion AI in Industry

- The market of Emotion AI is very young.
- **Not so many successful business cases**. Sometimes companies do not understand how Emotional AI technologies can help to solve their problems.
- On the other hand, some huge companies open R&D centers and develop their own Emotion AI technologies [Samsung, Microsoft, Automotive companies].
- When we are integrating Emotion AI technology in one business process, it usually concerns many other business processes. And the solution may be **too expensive** in comparison with the value that the technology may provide.

Neurodata Lab Business Solutions

Customer Experience Management

CX Management solution for customer emotion analysis piloting at Societe Generale's Rosbank.

Link

Robotics & Virtual assistants

Emotion Recognition and CX Management system for Promobot's service robot. Presented at CES 2019.

Link

API

Tools and instruments for automatic emotion recognition for everyone

Link

CX Management

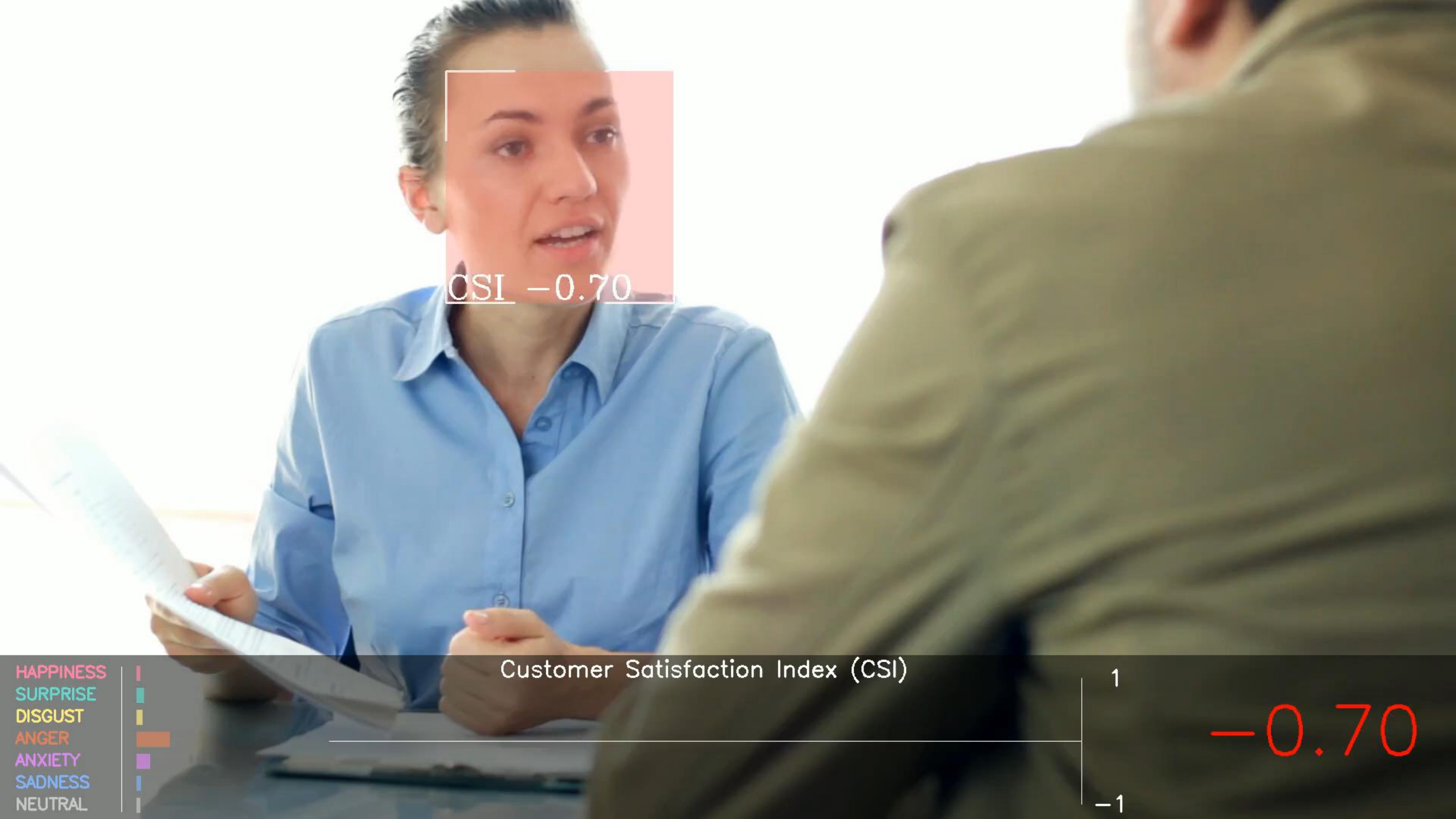
Existing methods of customer service evaluation

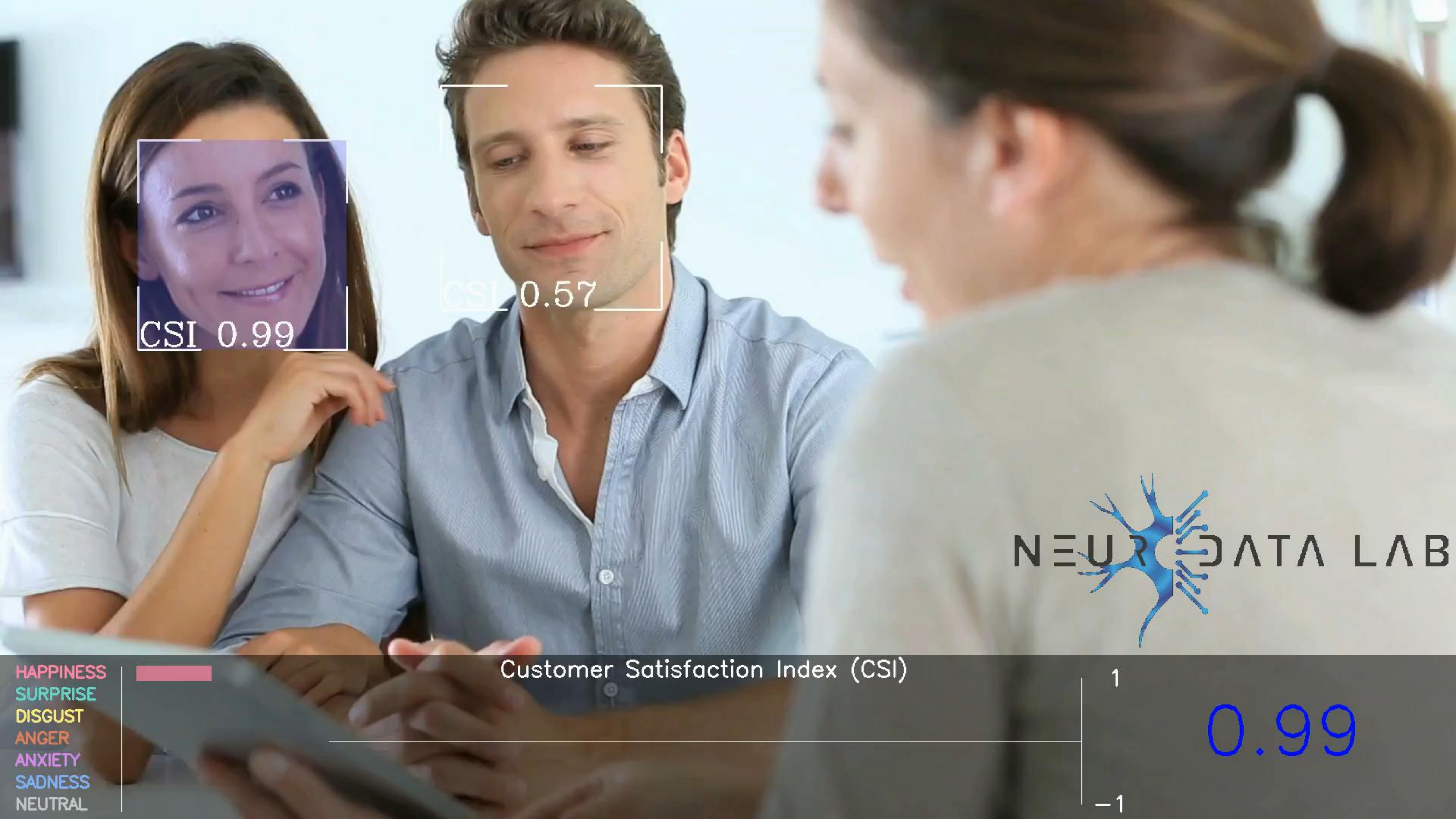
NEURODATA LAB

	Coverage	Integrity	Feedback	Cost	Cause analysis
E-mail/SMS survey	average	low ☹	none ☹	affordable	sometimes ☹
Phone survey	low ☹	average	none ☹	expensive ☹	sometimes ☹
Special commission	extremely low ☹	low ☹	sometimes ☹	Extremely expensive ☹	present
«Mystery customer»	extremely low ☹	average	sometimes ☹	Extremely expensive ☹	present
«rate the service» button	low ☹	extremely low ☹	sometimes ☹	average	none ☹

Situation analysis indicates a lack of a good solution among the existing methods of customer service efficiency evaluation:

1. Most solutions doesn`t provide full coverage, besides the sampling is usually not representative (doesn`t show the full picture of the actual audience) and doesn`t provide any conclusions of value for the company
2. That`s why the integrity of the results is low. The delay in survey and an opportunity to fake the results (for instance, it is not always clear how many times "rate the service" button) also influence the quality of the results.
3. The company does not receive the real account of the situation when the spending are excessive


CSI −0.70
HAPPINESS
SURPRISE
DISGUST
ANGER
ANXIETY
SADNESS
NEUTRAL
Customer Satisfaction Index (CSI)
1
−1
−0.70


CSI 0.99
CSI 0.57
NEUR DATA LAB
HAPPINESS
SURPRISE
DISGUST
ANGER
ANXIETY
SADNESS
NEUTRAL
Customer Satisfaction Index (CSI)
1
−1
0.99

Promobot- emotional robot

Neurodata Lab API

Emotion AI

Multimodal Emotion Recognition

Face Emotion Recognition

Customer Satisfaction Index

Voice Emotion Recognition

Smart trackers

Person Identification

Face Clustering

Face Detection

Heart Rate Estimation [video PPG]

Single Frame Age Estimation

Speaker Diarization

Body Pose Estimation

Eye Tracking

Coming in 2019

Available for everyone

Special conditions for non-profit organizations

NEURODATA LAB TOOLS

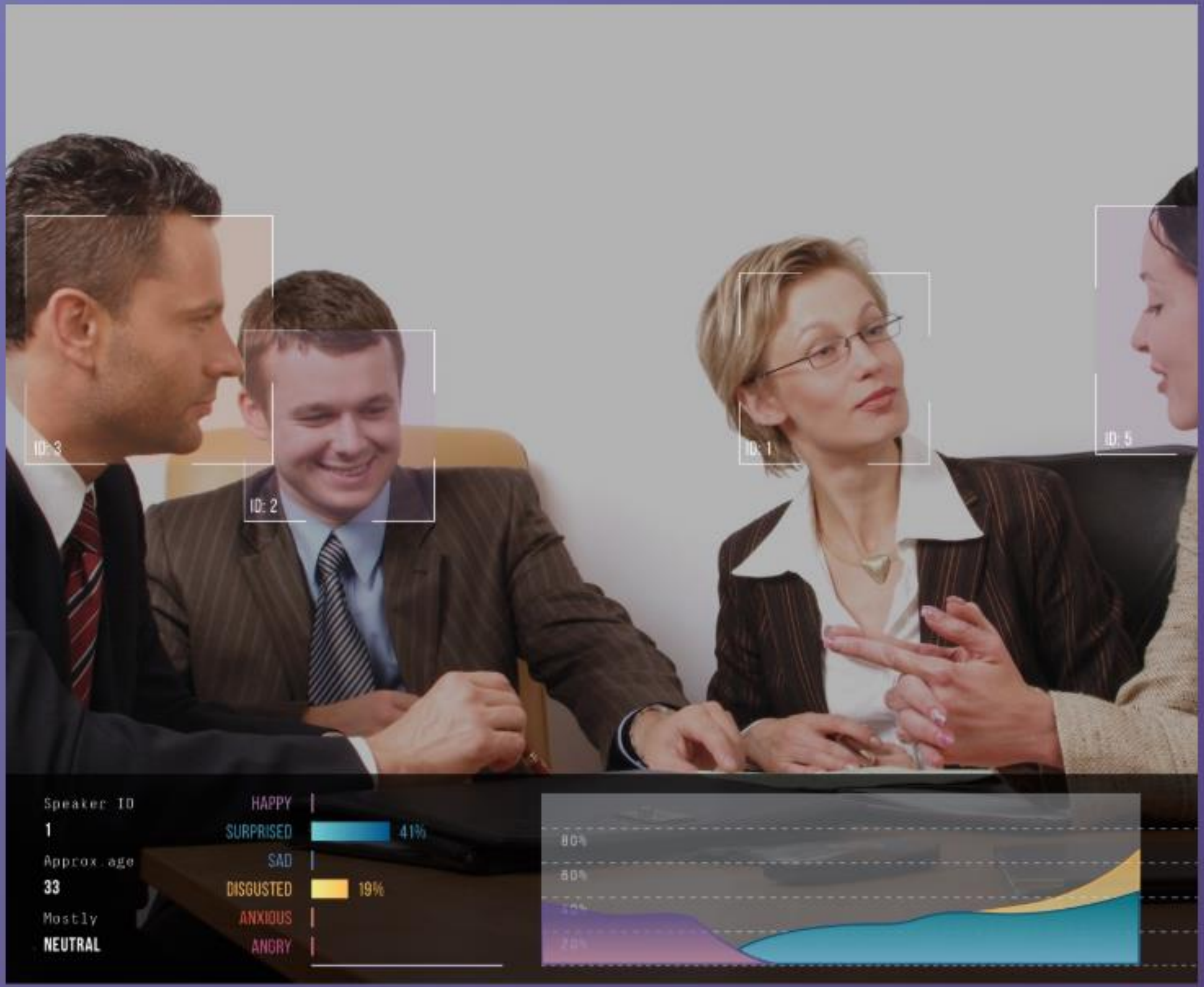
LOGIN SIGN UP

EMOTION RECOGNITION TOOL

Multimodal Human Behavior Processing Unit

- recognizes 7 emotional states
- in real time and on pre-recorded video
- up to 4 people in the frame

Demo



Terms of service Contact Sales Contact Support

Contacts

- www.neurodatalab.com
- www.emotionsdemo.com

- Olga Perepelkina: o.perepelkina@neurodatalab.com

NEURODATA LAB

Thank you