SOCIAL SIGNAL PROCESSING

Gergely Magyar, PhD.
Center for Intelligent Technologies
Department of Cybernetics and Artificial Intelligence
Technical University of Košice

DCAI
Department of Cybernetics and Artificial Intelligence

CIT
Center for Intelligent Technologies

WHAT ARE SOCIAL SIGNALS?


Emotions
Face identification
Gaze
Proxemics
Gestures
Prosody
Verbal communication
Body posture

WHAT ARE SOCIAL SIGNALS?

- social signals are ***observable*** behaviors that people display during social interactions

WHAT ARE SOCIAL SIGNALS?

- social signals are ***observable*** behaviors that people display during social interactions
- social signals from an individual ***produces changes*** in others (like creating a belief about the person, generating an appropriate social response, perform any actions)

WHAT ARE SOCIAL SIGNALS?

- social signals are ***observable*** behaviors that people display during social interactions
- social signals from an individual ***produces changes*** in others (like creating a belief about the person, generating an appropriate social response, perform any actions)
- the changes are not random, they follow ***principles and laws*** (in particular ***social norms***)

WHY DO WE NEED SSP?

- the ability to recognize human social signals and social behaviors like turn taking, politeness, and disagreement is essential when building social robots, human-robot interaction, or interactive systems

WHY DO WE NEED SSP?

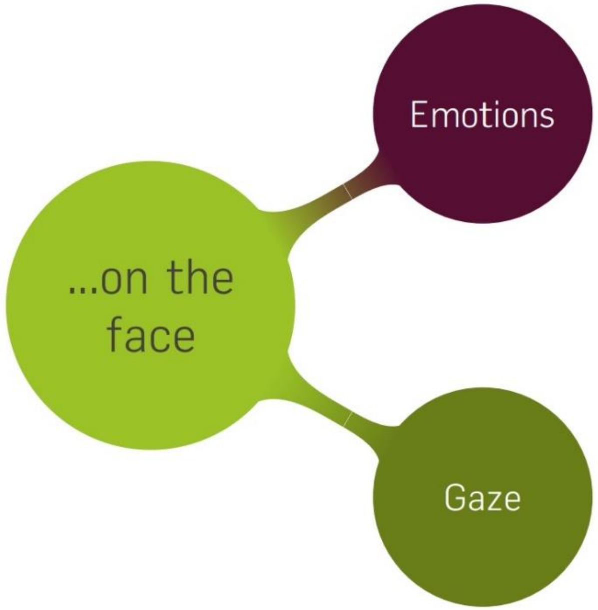
- the ability to recognize human social signals and social behaviors like turn taking, politeness, and disagreement is essential when building social robots, human-robot interaction, or interactive systems
- **3 MAIN PROBLEMS**
  - ***Modeling:*** identification of the principles and laws
  - ***Analysis:*** automatic detection and interpretation
  - ***Synthesis:*** automatic generation of artificial social signals

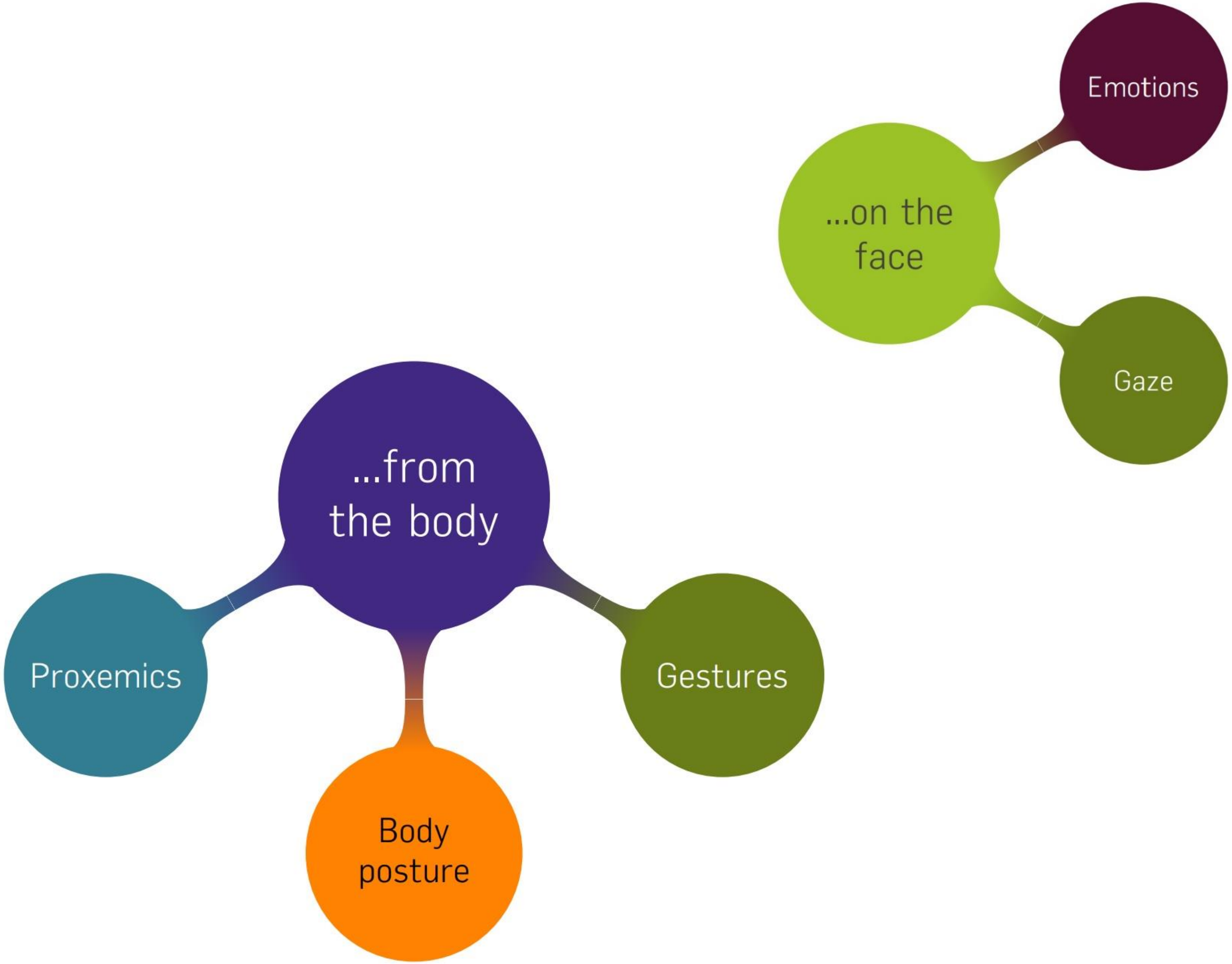
IS IT HARD?

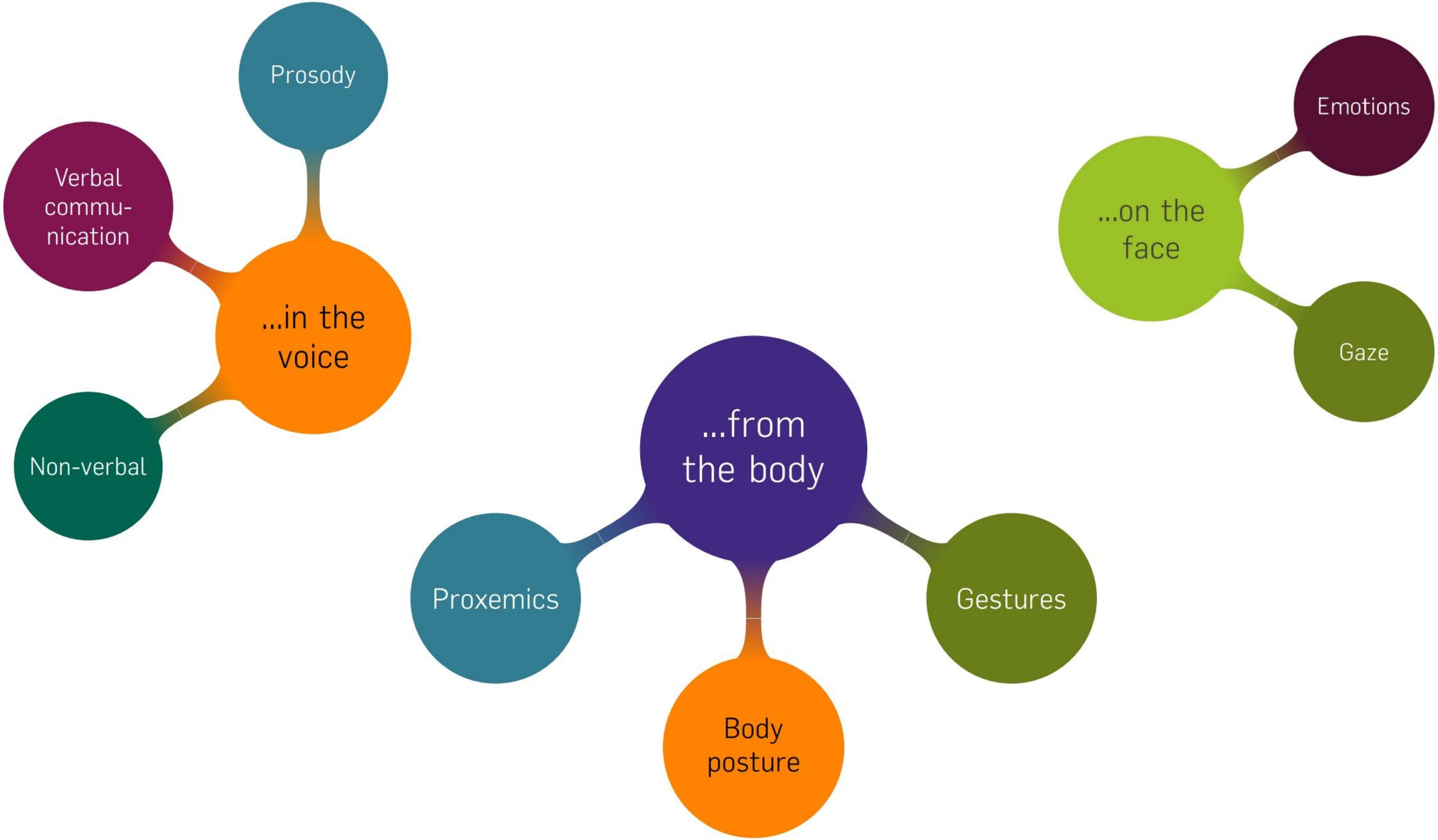
- as with most human activities that seems easy to us, social signal processing is tremendously hard
- often broken down in smaller, sometimes more manageable tasks:

IS IT HARD?

- as with most human activities that seems easy to us, social signal processing is tremendously hard
- often broken down in smaller, sometimes more manageable tasks:
  - people detection
  - face detection
  - face recognition
  - gesture recognition
  - gaze detection
  - facial expression reading (wink, blink, talking, …)
  - detection of social signals from verbal communication
  - emotion recognition (from faces, movement, speech, …)
  - …


...on the face
Emotions
Gaze

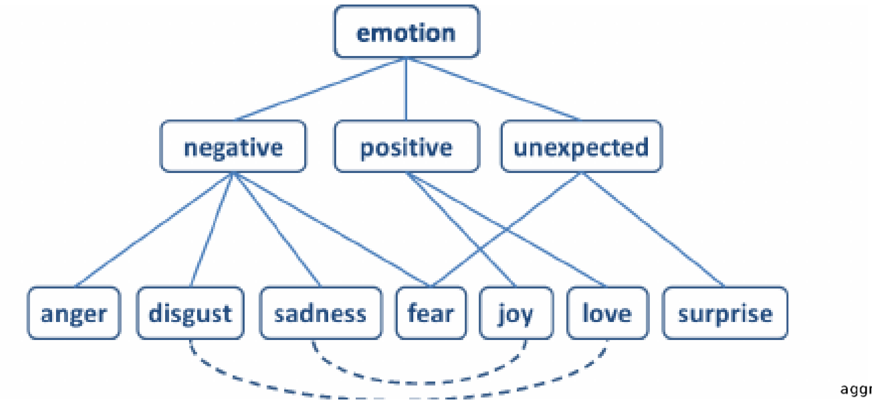

Emotions
...on the face
Gaze
...from the body
Proxemics
Gestures
Body posture


Prosody
Verbal commu-nication
...in the voice
Non-verbal
Emotions
...on the face
Gaze
...from the body
Proxemics
Gestures
Body posture

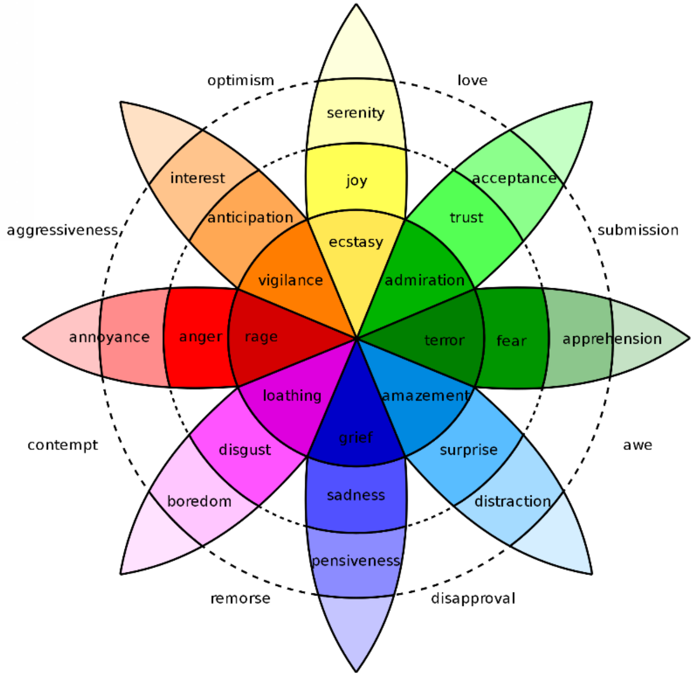
EMOTION RECOGNITION

- Emotions are connected to
  - feeling
  - mood
  - affect
- Components of emotions
  - cognitive
  - physiological
  - behavioral

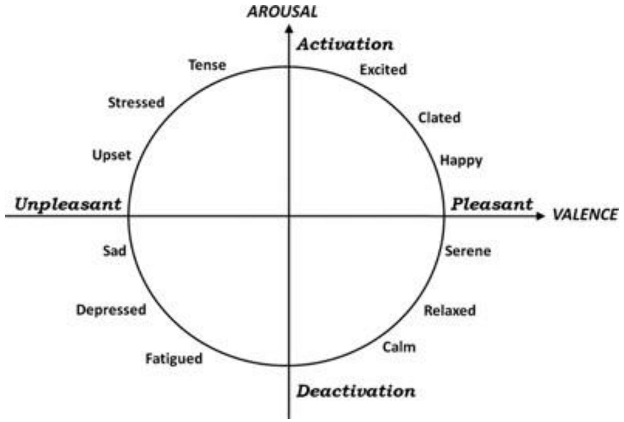
EMOTION MODELS



discrete emotion models



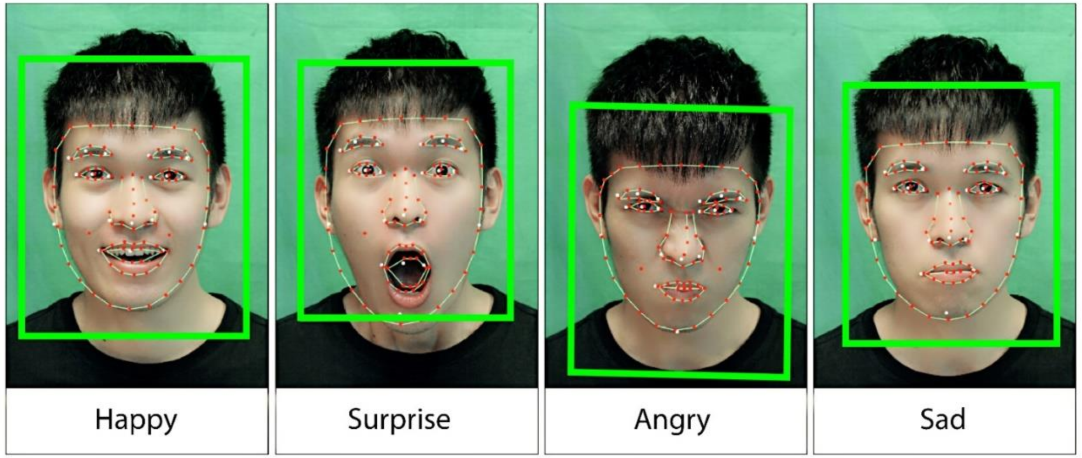
hybrid emotion models



dimensional emotion models

FACIAL EMOTION RECOGNITION

1) face detection
2) landmark identification
3) facial expression classification
4) mapping expressions to emotions



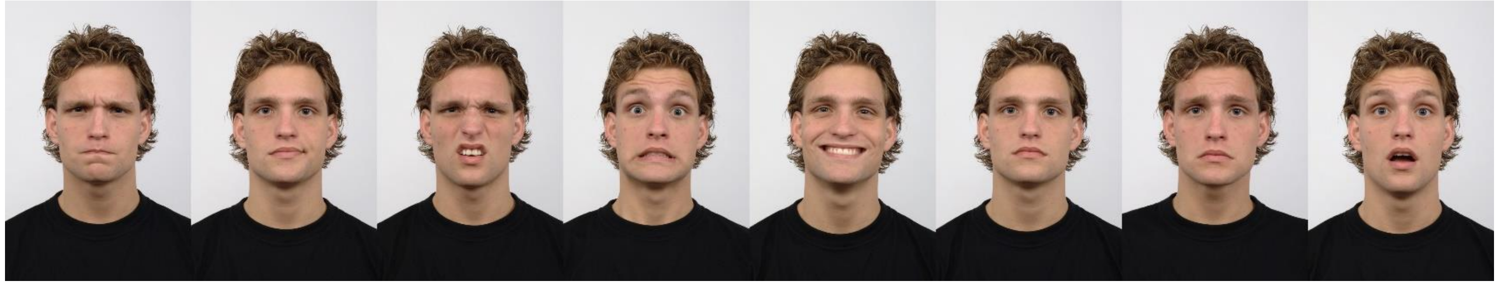
EMOTION RECOGNITION SERVICES

- Affectiva (anger, contempt, disgust, fear, joy, sadness, surprise)
- Amazon Rekognition (anger, calmness, confusion, disgust, happiness, sadness, surprise)
- Face++ (anger, disgust, fear, happiness, neutral, sadness, surprise)
- Google Vision (anger, joy, sorrow, surprise)
- Kairos (anger, disgust, fear, joy, sadness, surprise)
- Microsoft Face (anger, contempt, disgust, fear, happiness, neutral, sadness, surprise)
- Sightcorp F.A.C.E. (anger, disgust, fear, happiness, sadness, surprise)
- Sighthound Cloud (anger, disgust, fear, happiness, neutral, sadness, surprise)

IMAGE DATASET

- Karolinska Directed Emotional Faces
  - 973 frontal images with 2 sets of 70 individuals expressing 7 emotions

- Radboud Faces Database
  - 1,608 frontal images with 3 gaze directions from 67 models expressing 8 emotions



EVALUATION RESULTS

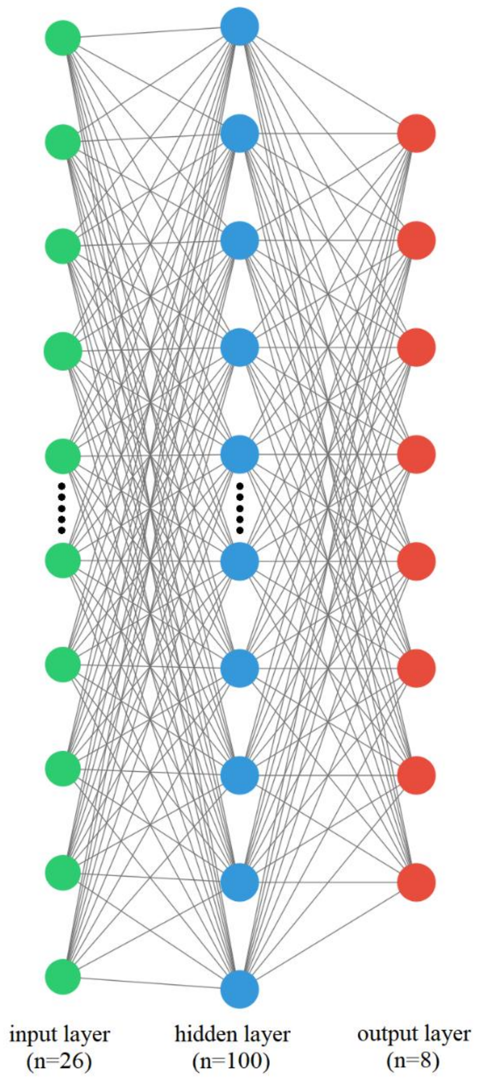
API	KDEF (%)	RaFD (%)	Overall (%)
Affectiva	54.57	64.68	60.87
F.A.C.E.	63.21	61.32	62.03
Kairos	45.74	26.99	34.06
Rekognition	52.62	39.74	44.60
Face++	77.08	71.33	73.50
Google	43.47	36.63	39.21
MS Face	75.33	76.24	75.90
Sighthound	62.18	72.33	68.50

VOTING SYSTEM

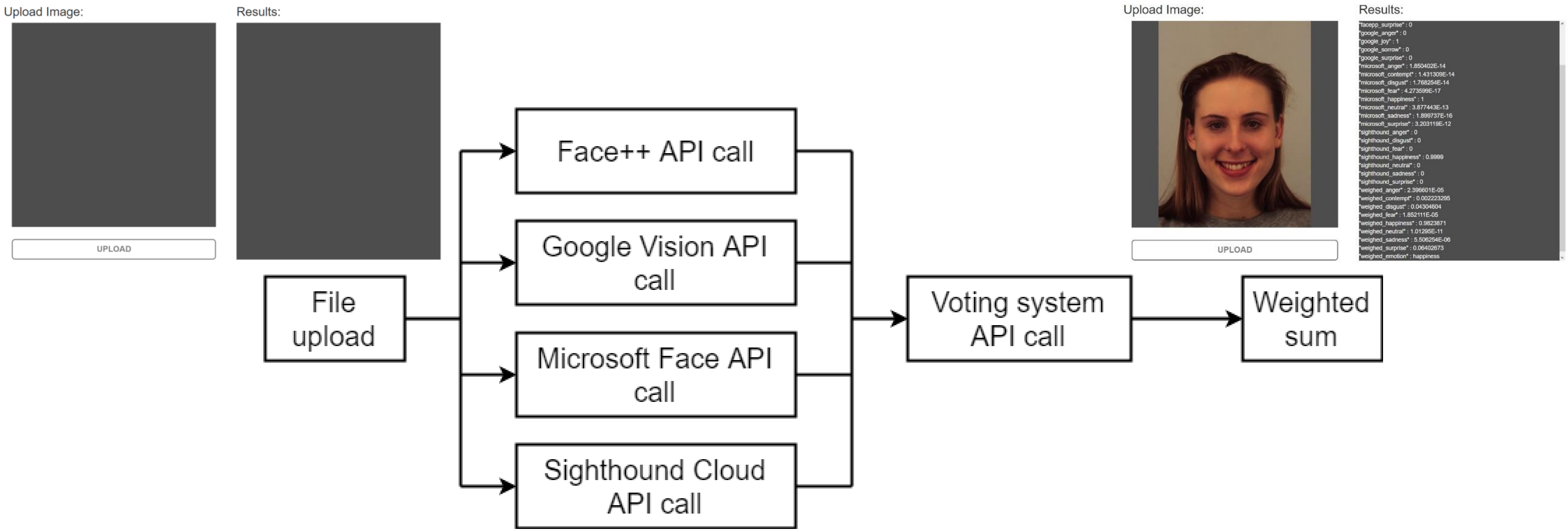
- Input is the result from multiple emotion recognition services
- Neural network computes weighted average of confidence values
- Output are confidence values for each emotion category

VOTING SYSTEM

- Online service created in MS Machine Learning Studio
  - multiclass neural network
  - 100 hidden nodes
  - learning rates: 0.01, 0.02, 0.04
  - number of iterations: 164-500, number of points: 3
  - initial learning weight: 0.1
  - momentum: 0
  - min-max normalizer
- 75% training set, 25% testing set (non-stratified sampling)



WORKFLOW


Upload Image:
Results:
UPLOAD
File upload
Face++ API call
Google Vision API call
Microsoft Face API call
Sighthound Cloud API call
Voting system API call
Weighted sum
Upload Image:
UPLOAD
Results:
"facepp_surprise" : 0
"google_anger" : 0
"google_joy" : 1
"google_sorrow" : 0
"google_surprise" : 0
"microsoft_anger" : 1.850402E-14
"microsoft_contempt" : 1.431309E-14
"microsoft_disgust" : 1.768254E-14
"microsoft_fear" : 4.273599E-17
"microsoft_happiness" : 1
"microsoft_neutral" : 3.877443E-13
"microsoft_sadness" : 1.899737E-16
"microsoft_surprise" : 3.203119E-12
"sighthound_anger" : 0
"sighthound_disgust" : 0
"sighthound_fear" : 0
"sighthound_happiness" : 0.9999
"sighthound_neutral" : 0
"sighthound_sadness" : 0
"sighthound_surprise" : 0
"weighed_anger" : 2.396601E-05
"weighed_contempt" : 0.002223295
"weighed_disgust" : 0.04304604
"weighed_fear" : 1.852111E-05
"weighed_happiness" : 0.9823871
"weighed_neutral" : 1.01295E-11
"weighed_sadness" : 5.506254E-06
"weighed_surprise" : 0.06402673
"weighed_emotion" : happiness

WORKFLOW

Upload Image:

UPLOAD

Results:

"facepp_surprise" : 0
"google_anger" : 0
"google_joy" : 1
"google_sorrow" : 0
"google_surprise" : 0
"microsoft_anger" : 1.850402E-14
"microsoft_contempt" : 1.431309E-14
"microsoft_disgust" : 1.768254E-14
"microsoft_fear" : 4.273599E-17
"microsoft_happiness" : 1
"microsoft_neutral" : 3.877443E-13
"microsoft_sadness" : 1.899737E-16
"microsoft_surprise" : 3.203119E-12
"sighthound_anger" : 0
"sighthound_disgust" : 0
"sighthound_fear" : 0
"sighthound_happiness" : 0.9999
"sighthound_neutral" : 0
"sighthound_sadness" : 0
"sighthound_surprise" : 0
"weighed_anger" : 2.396601E-05
"weighed_contempt" : 0.002223295
"weighed_disgust" : 0.04304604
"weighed_fear" : 1.852111E-05
"weighed_happiness" : 0.9823871
"weighed_neutral" : 1.01295E-11
"weighed_sadness" : 5.506254E-06
"weighed_surprise" : 0.06402673
"weighed_emotion" : happiness

VOTING SYSTEM ACCURACY

API	KDEF (%)	RaFD (%)	Overall (%)
Affectiva	54.57	64.68	60.87
F.A.C.E.	63.21	61.32	62.03
Kairos	45.74	26.99	34.06
Rekognition	52.62	39.74	44.60
Face++	77.08	71.33	73.50
Google	43.47	36.63	39.21
MS Face	75.33	76.24	75.90
Sighthound	62.18	72.33	68.50
Voting system	**90.75**	**97.08**	**94.69**

VOTING SYSTEM ACCURACY

		Actual class								Precision (%)
		A	C	D	F	H	N	Sa	Su	
Predicted class	A	303	0	3	3	0	0	6	0	96.19
	C	3	188	0	0	0	0	1	0	97.92
	D	7	0	327	5	0	0	3	0	95.61
	F	4	0	2	298	0	0	7	6	94.01
	H	0	0	0	1	340	0	0	0	99.71
	N	10	13	0	1	0	340	7	2	91.15
	Sa	13	0	8	12	0	0	316	0	90.54
	Su	0	0	0	20	0	0	0	332	94.32
Recall (%)		89.12	93.53	96.18	87.65	100	100	92.94	97.65	

HOW TO CLASSIFY SOCIAL SIGNALS?

- raw signals will in most cases require pre-processing to extract features
- the raw social signal (audio or video) requires pre-processing to extract between 10 and over 1000 **features**
  - a raw signal contains too much data, and cannot be fed to the classifier immediately
  - pre-processing extracts feature data which is relevant for the information which we are after (pitch, volume/energy, duration, …)
  - these features then form the input for the classifier

EXAMPLE: RECOGNIZING GENDER FROM SPEECH

- can we automatically recognize someone’s gender from speech?
- 3168 recorded voice samples, collected from male and female speakers

- the voice samples are pre-processed by acoustic analysis and 20 features are extracted

EXAMPLE: RECOGNIZING GENDER FROM SPEECH

- performance
  - kNN (k=7): 97,8% classified correctly
  - SVM: 97,5% classified correctly
- recognizing gender from speech is easy and robust
- all classification algorithms can deal with this problem