COMP 3647
Human-AI Interaction Design

Topic 7
Affective Computing: Basics

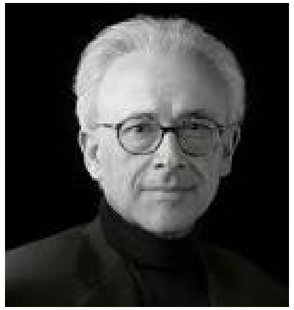
Prof. Effie L-C Law

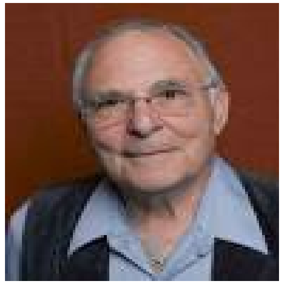
Quiz #1

"Humans are feeling machines that think rather than thinking machines that feel"

Who said that?

A. **Antonio Damasio**, neuroscientist
B. **Rosalind Picard**, computer scientist
C. **Paul Ekman**, psychologist
D. **Lisa Feldman Barret,** psychologist






Durham University

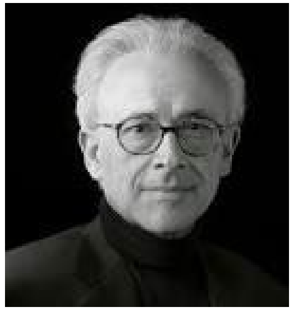
Quiz #2

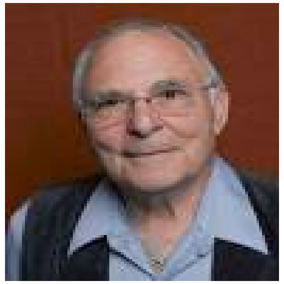
Who proposed six basic emotions:

Angry, Digust, Fear, Happy, Sad, Surprise

Who said that?

A. **Antonio Damasio**, neuroscientist
B. **Rosalind Picard**, computer scientist
C. **Paul Ekman**, psychologist
D. **Lisa Feldman Barret,** psychologist





Quiz #3

Who proposed Theory of Constructed Emotion?

Who said that?

A. **Antonio Damasio**, neuroscientist

B. **Rosalind Picard**, computer scientist

C. **Paul Ekman**, psychologist

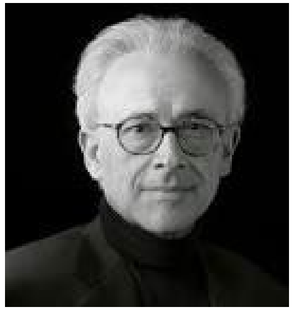
D. **Lisa Feldman Barret,** psychologist

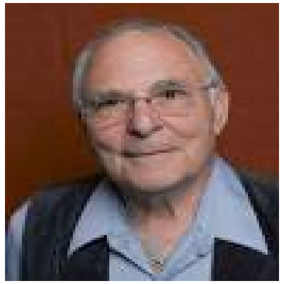
Quiz #4

Who coined the term and found the field "Affective Computing"?

Who said that?

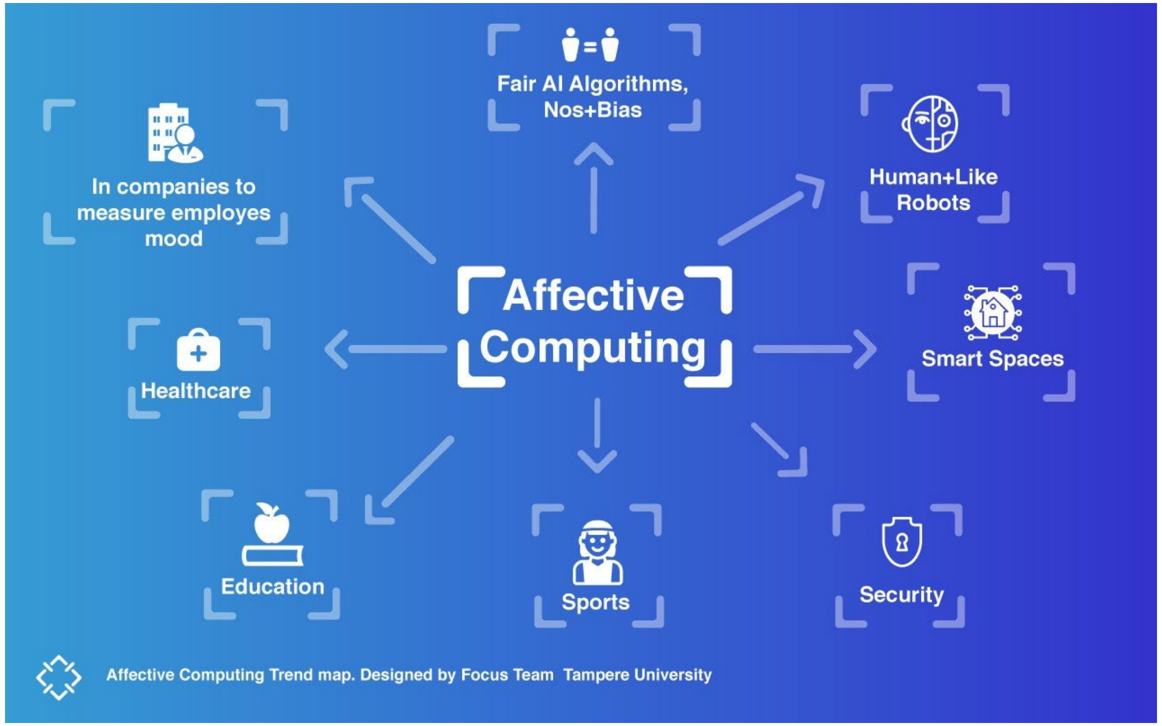
A. **Antonio Damasio**, neuroscientist
B. **Rosalind Picard**, computer scientist
C. **Paul Ekman**, psychologist
D. **Lisa Feldman Barret,** psychologist



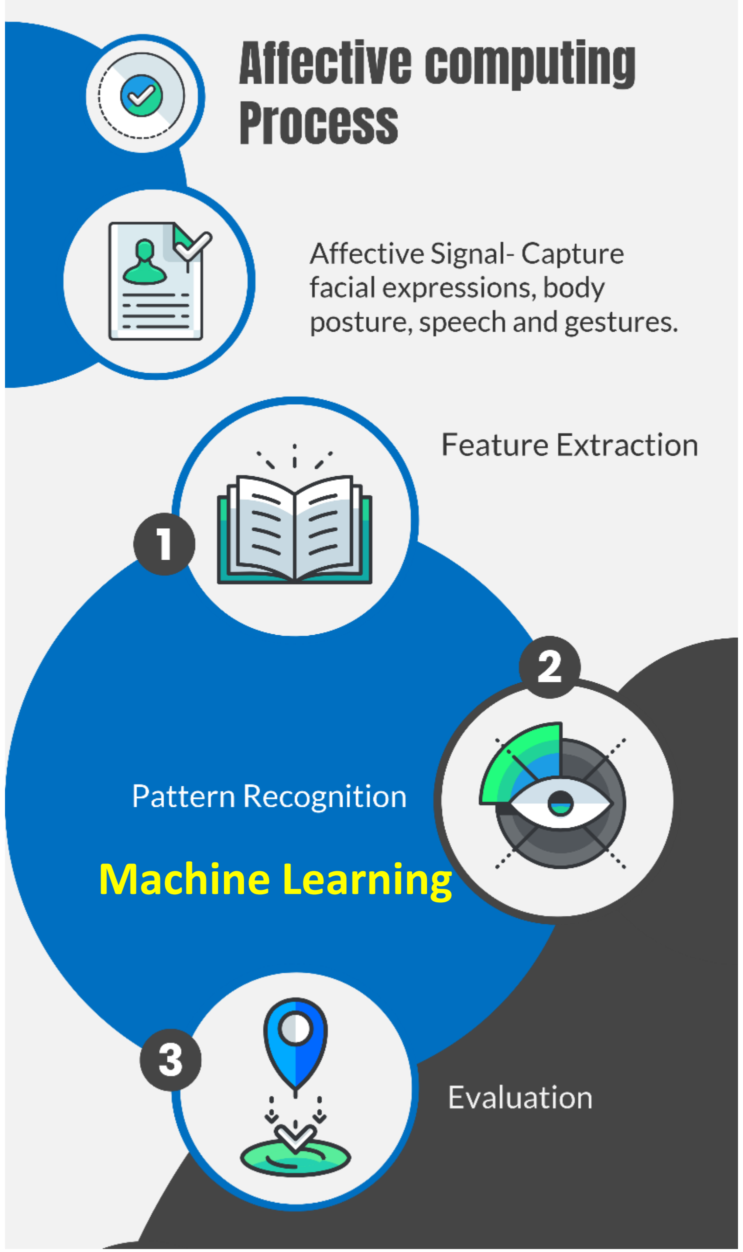



What is Affective Computing?

Affective Computing is the study and development of systems and devices that can recognize, interpret, process, and simulate human affects/emotions. It is an interdisciplinary field spanning computer science, psychology, and cognitive science.



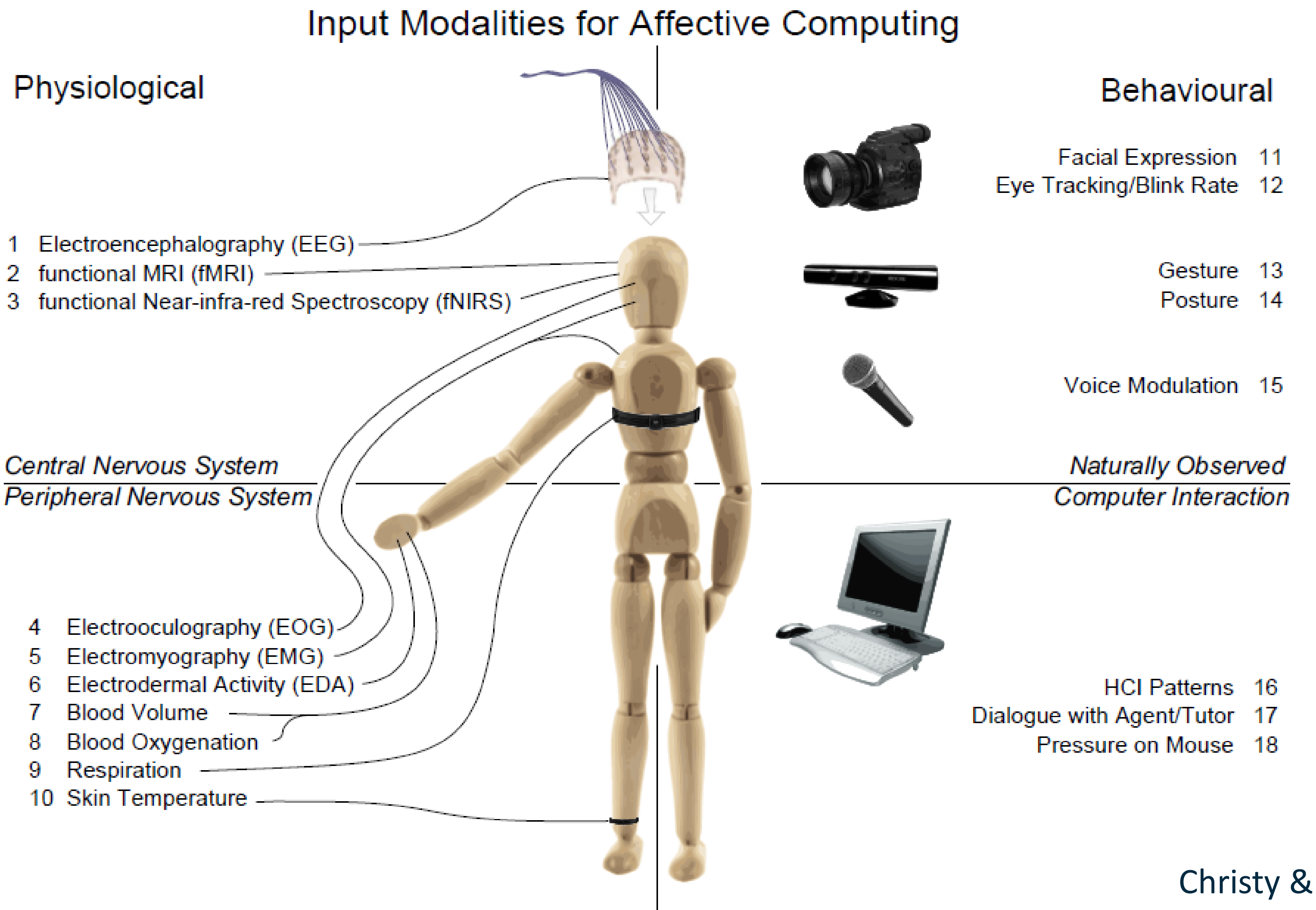
https://medium.com/@focusteamtuni/affective-computing-trends-map-625b554ab0d1



https://www.predictiveanalyticstoday.com/what-is-affective-computing/

Durham University

Automatic Multimodality Emotion Recognition (AMER)



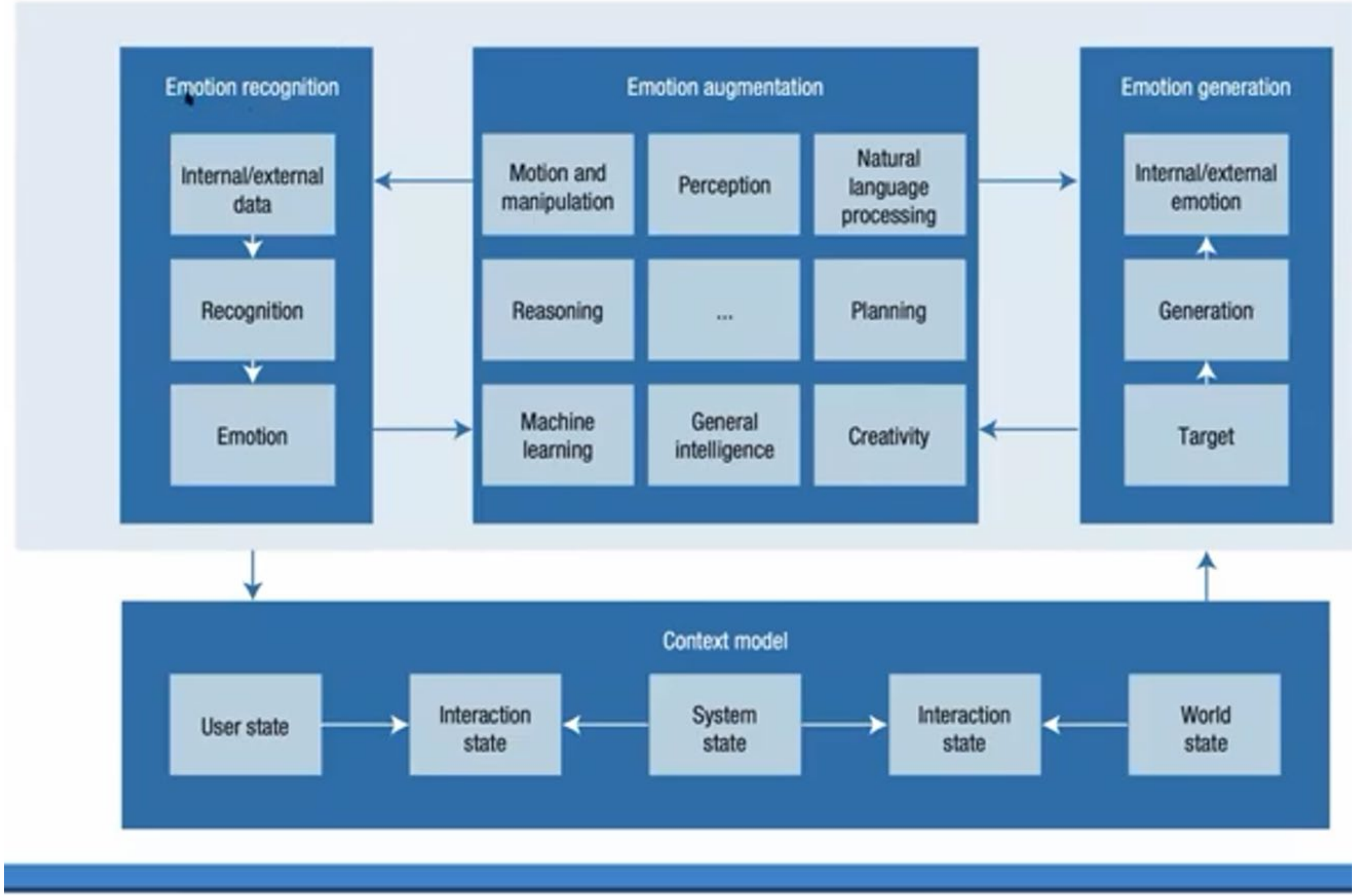
Durham University

Christy & Kuncheva (2014)

Emotion + AI Research

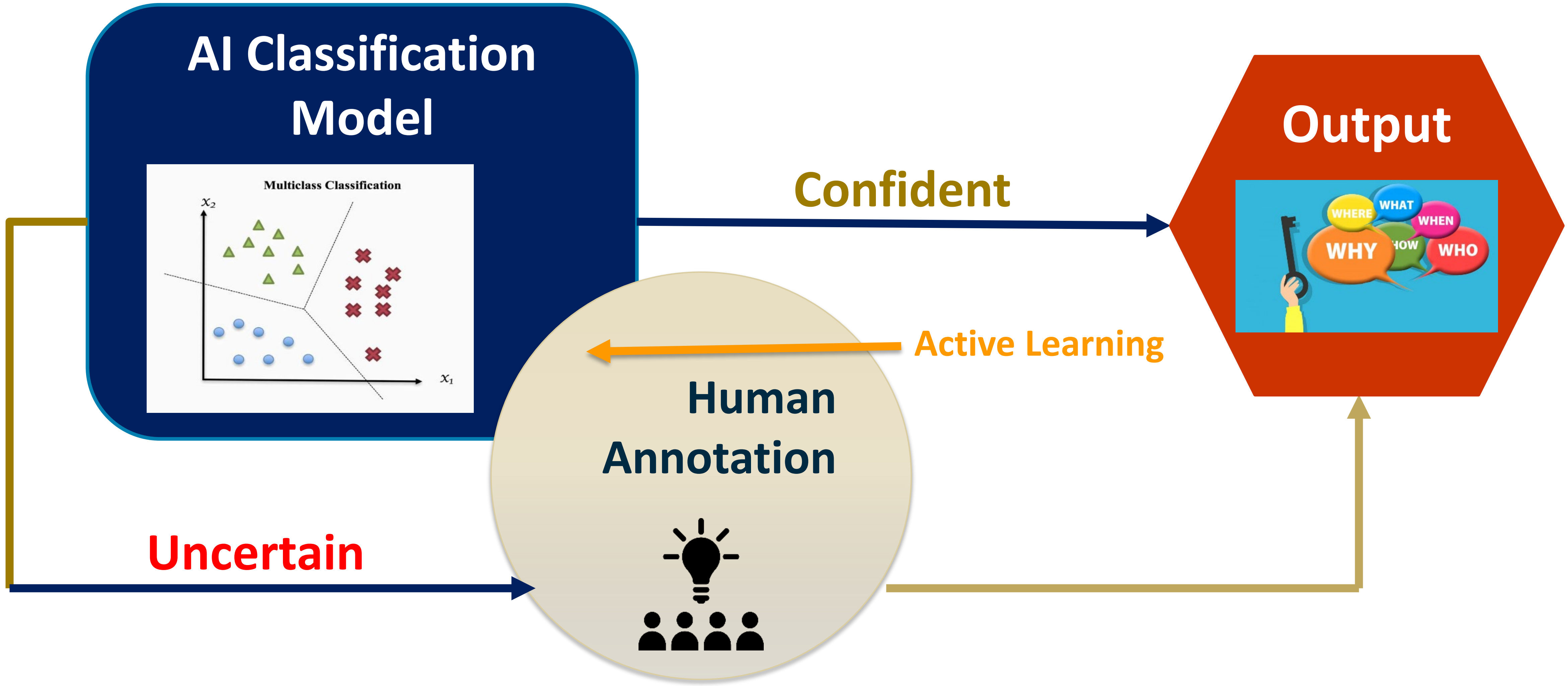
Artificial Emotional Intelligence (AEI)

(Bjorn Schuller, Imperial College London)



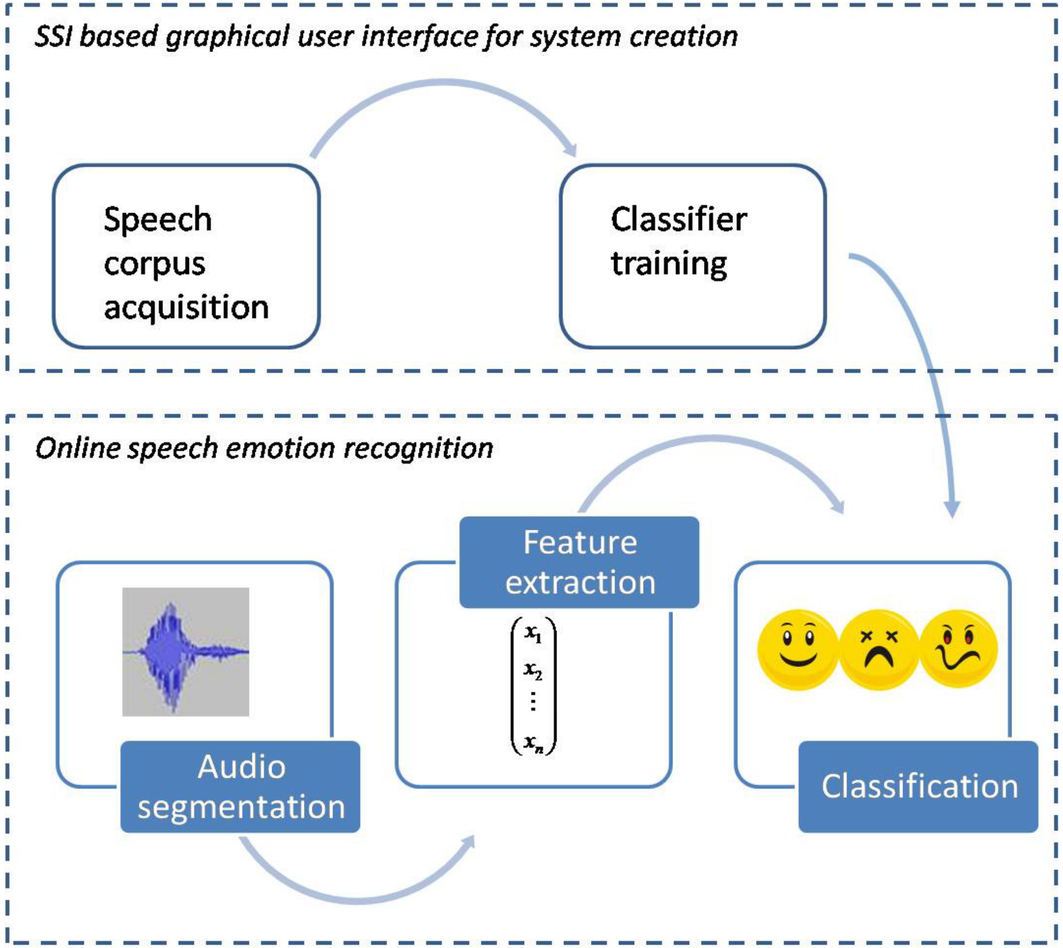
"The Age of Artificial Emotional Intelligence", IEEE Computer, 2018.

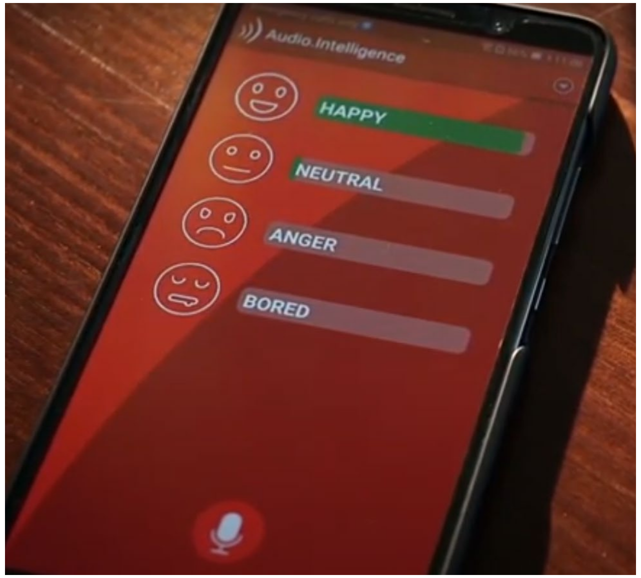
Durham University



Human-in-the-Loop:
(1) Annotate a *subset* of data; (2) Train a model with labelled data; (3) Make predictions on unlabelled data; **(4) Validate *uncertain* predictions**; (5) Revise the model to improve prediction.

Speech Emotion Recognition: Emotion + AI



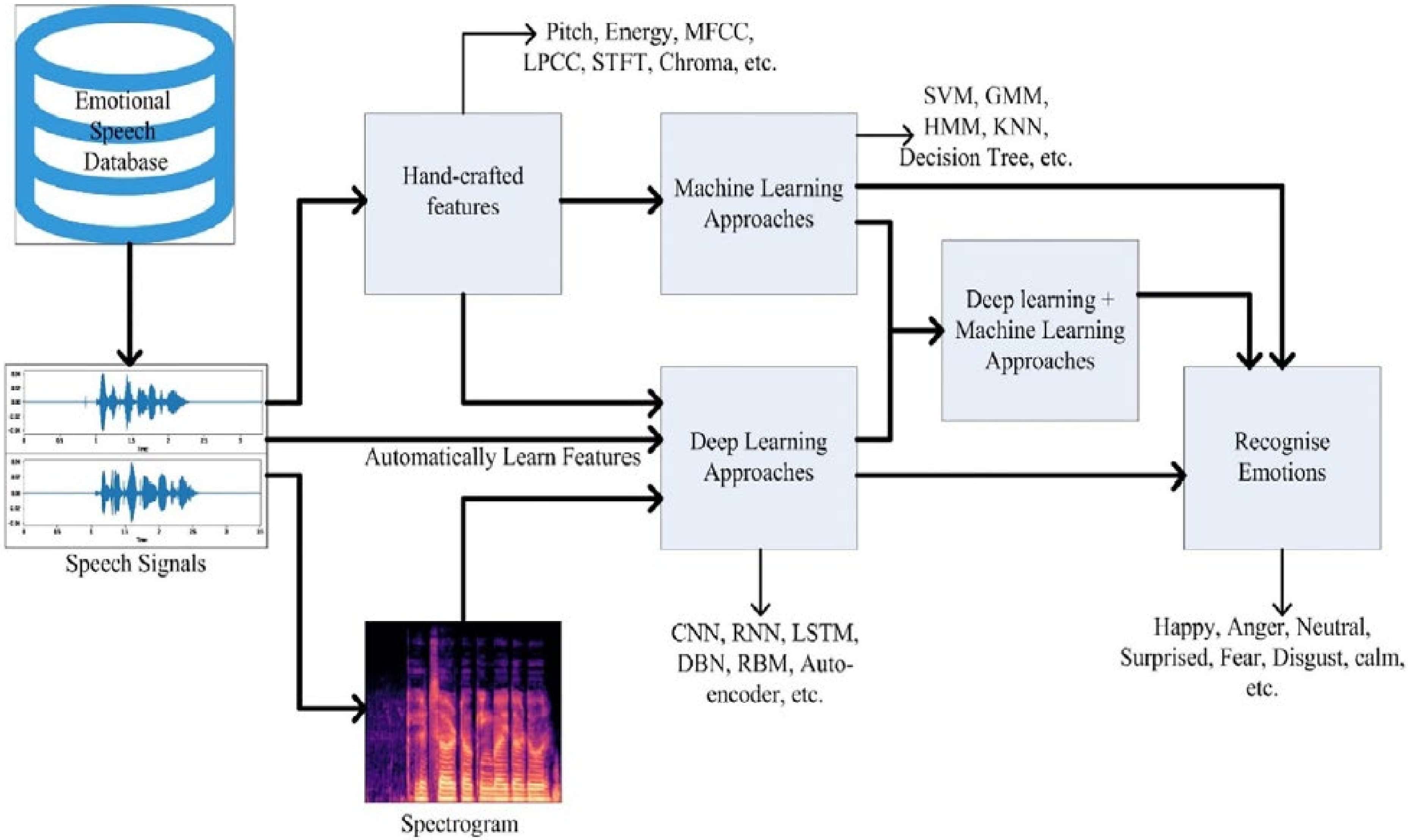


audEERING/
University of Augsburg

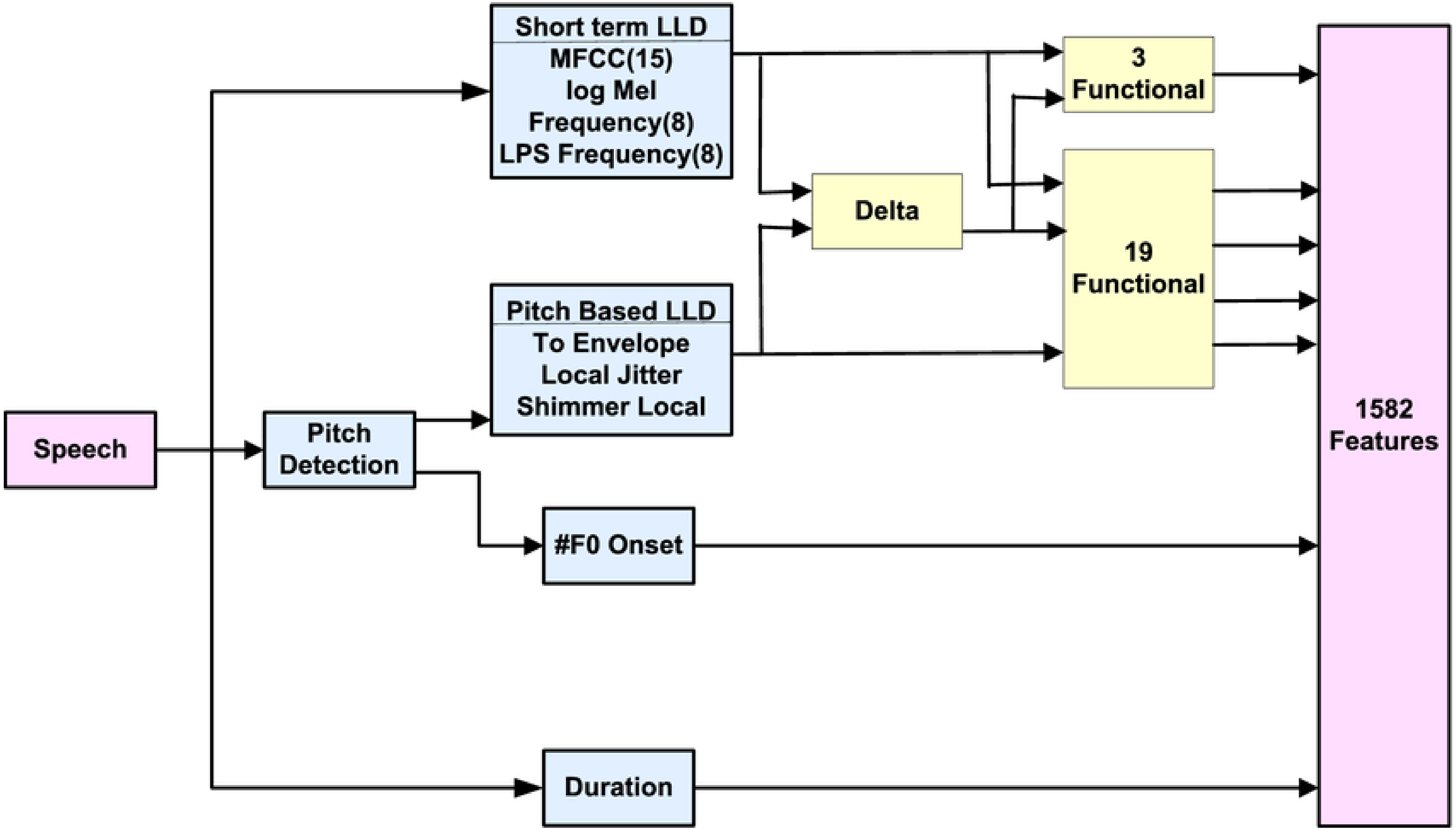
https://www.informatik.uni-augsburg.de/lehrstuehle/hcm/projects/tools/emovoice/

Law, E. L. C., Soleimani, S., Watkins, D., & Barwick, J. (2020). Automatic voice emotion recognition of child-parent conversations in natural settings. *Behaviour & Information Technology*, 1-18.

Speech Emotion Recognition (SER)



OpenSMILE: Open-source Speech and Music Interpretation by Large-space Extraction

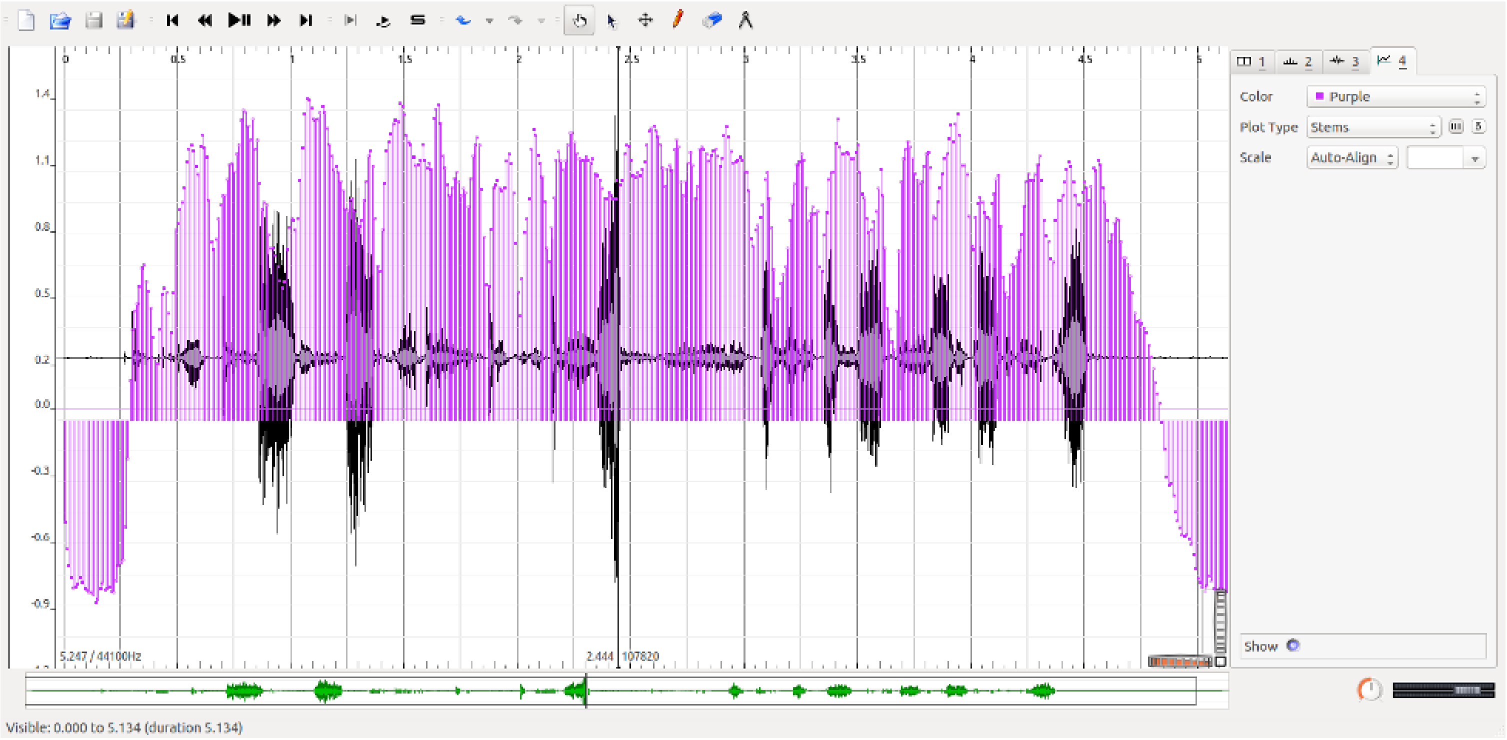


Durham University

OpenSMILE 3.0: Speech Features

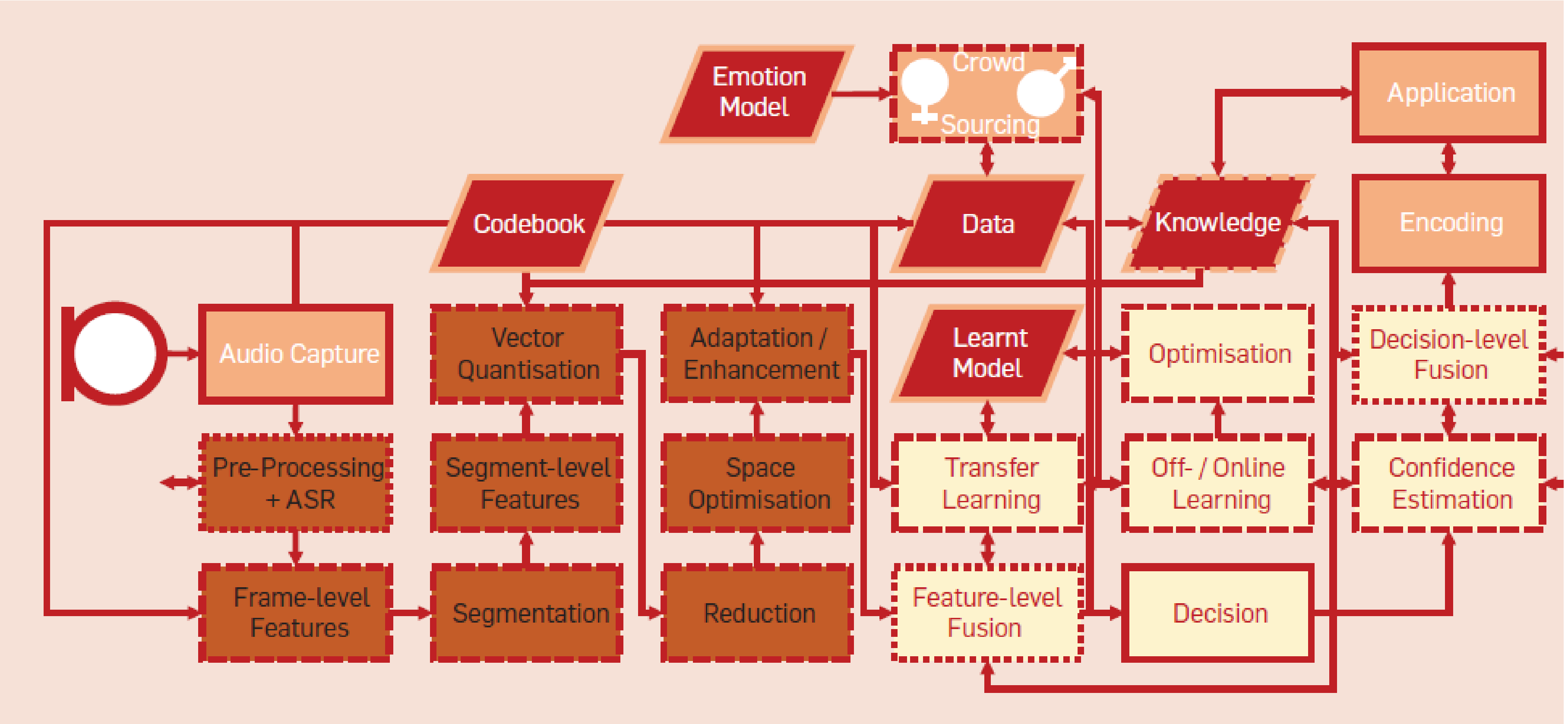
https://www.audeering.com/research/opensmile/

- Signal energy
- Loudness
- Mel-/Bark-/Octave-spectra
- MFCC (Mel-frequency cepstral coefficient)
- PLP-CC (perceptual linear prediction cepstral coefficient)
- Pitch
- Voice quality (Jitter, Shimmer)
- Formants
- LPC (linear predictive coding)
- Line Spectral Pairs (LSP)
- Spectral Shape descriptors

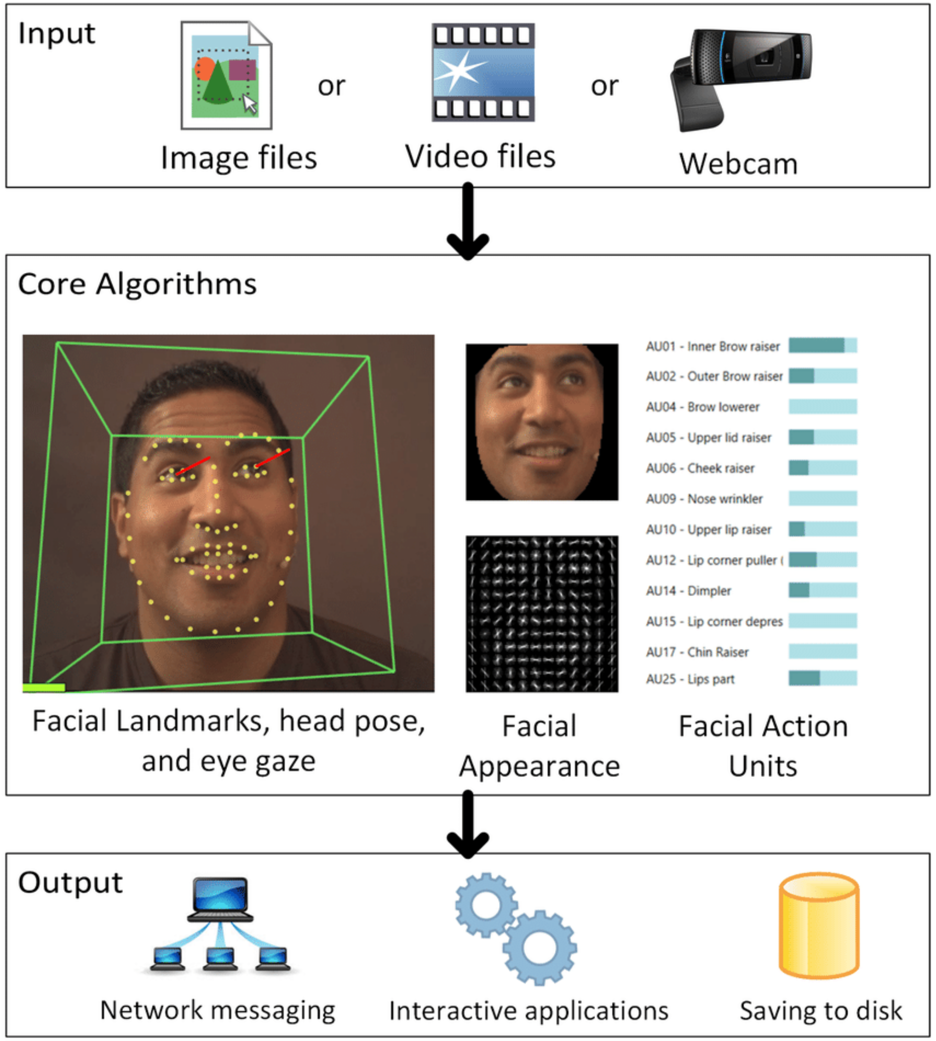


SER: Challenges (Schuller, 2018)

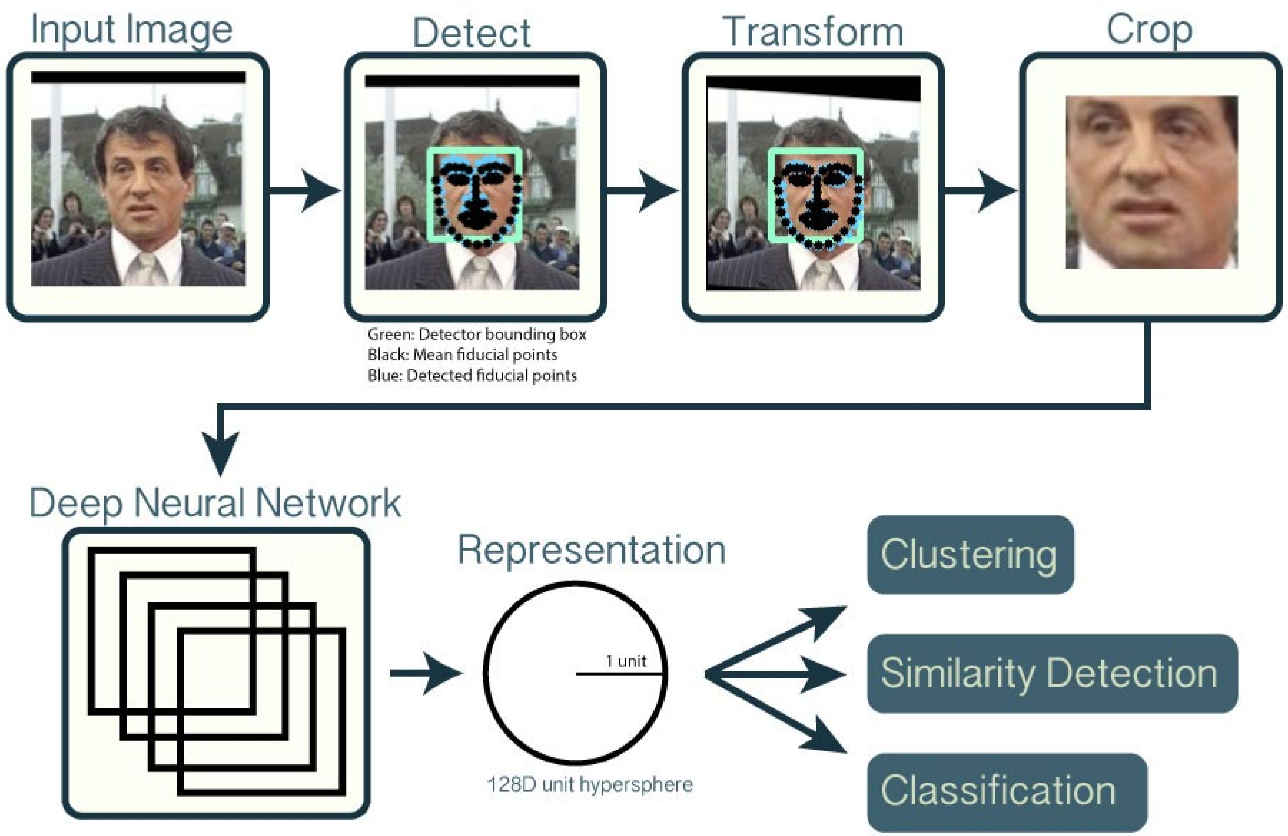
- Automatic SER requires an appropriate **emotion representation** (**modelling**)
- Robustness of prediction requires accurate data labelling **(annotation)** considering **states and traits** (i.e. context-awareness)



OpenFace

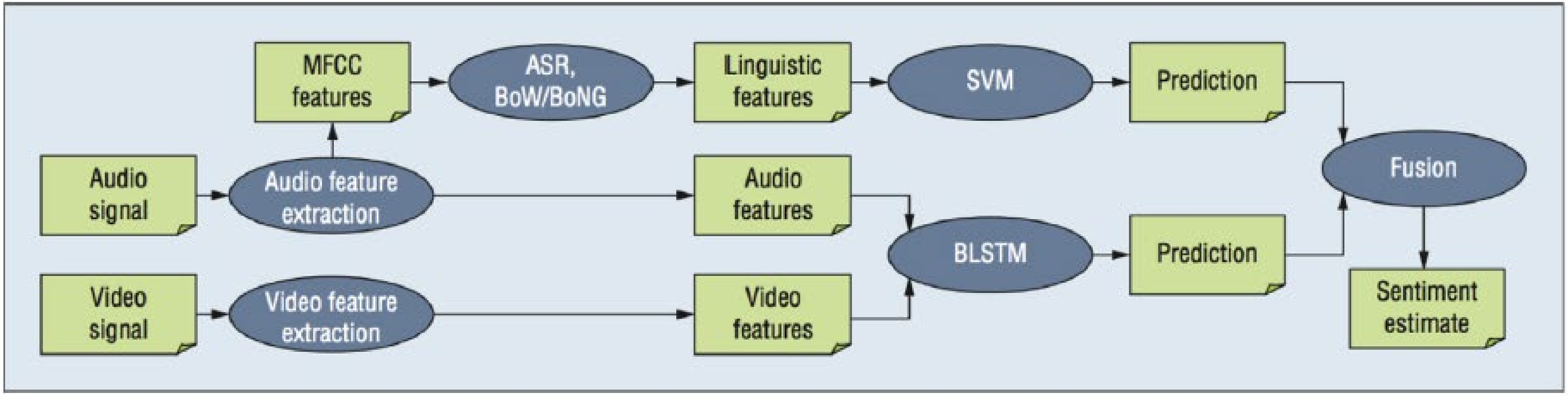


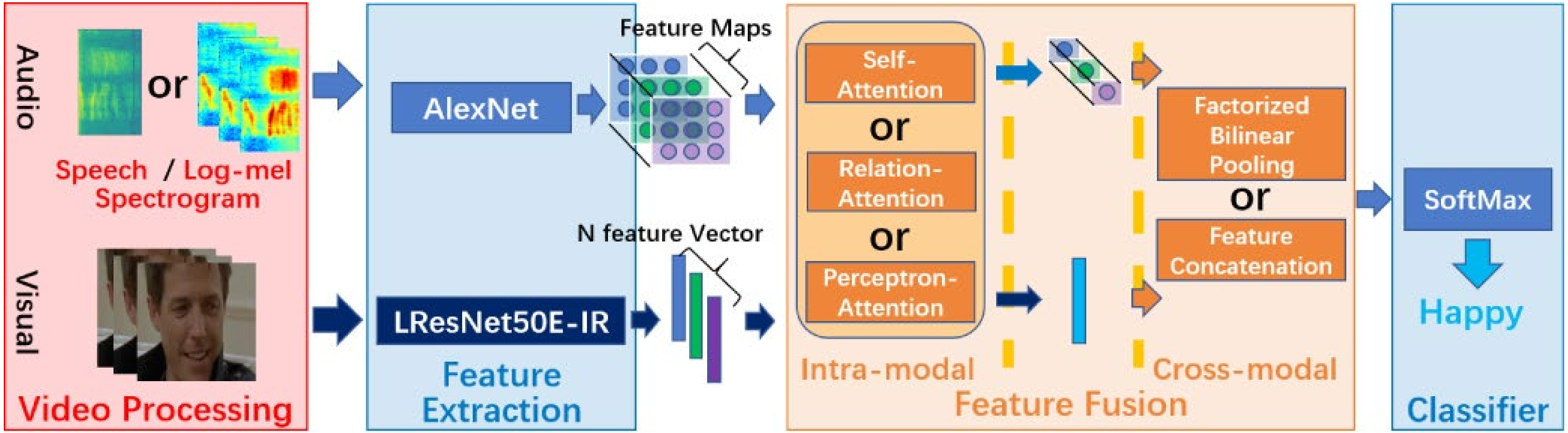
Input
Image files
or
Video files
or
Webcam
Core Algorithms
AU01 - Inner Brow raiser
AU02 - Outer Brow raiser
AU04 - Brow lowerer
AU05 - Upper lid raiser
AU06 - Cheek raiser
AU09 - Nose wrinkler
AU10 - Upper lip raiser
AU12 - Lip corner puller (
AU14 - Dimpler
AU15 - Lip corner depres
AU17 - Chin Raiser
AU25 - Lips part
Facial Landmarks, head pose, and eye gaze
Facial Appearance
Facial Action Units
Output
Network messaging
Interactive applications
Saving to disk



Input Image
Detect
Transform
Crop
Green: Detector bounding box
Black: Mean fiducial points
Blue: Detected fiducial points
Deep Neural Network
Representation
1 unit
128D unit hypersphere
Clustering
Similarity Detection
Classification

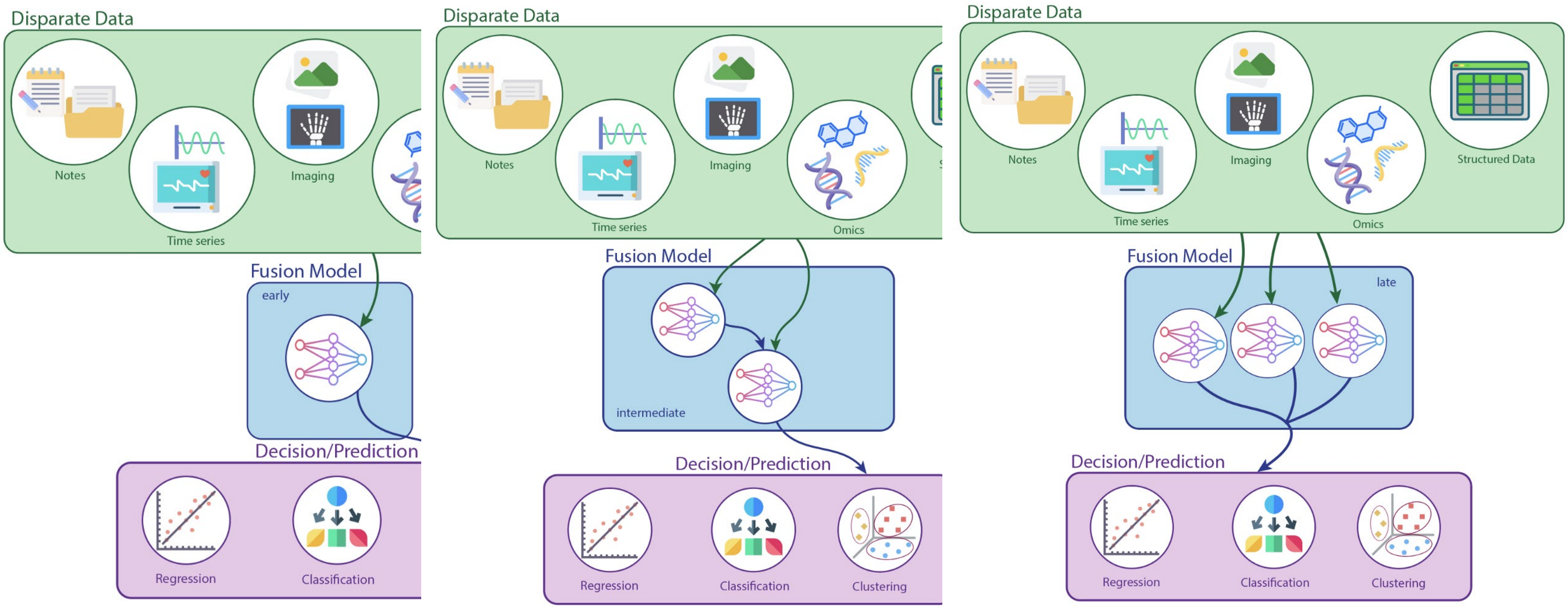
Data Fusion





Durham University

Three Types of Fusion:
Early (Feature), Intermediate (Joint), Late (Decision)



Durham University

https://pub.towardsai.net/multimodal-machine-learning-data-fusion-d1d8776e2cb0

Björn Schuller's Invited Talk (2022)

Björn Schuller

Imperial College London

SER

- **Intro** Patent in 1970ies
- **Chapter 1** First real papers in mid 1990ies
  Expert Features, mostly acted,
  mostly categories, noise robustness,
  low comparibility
- **Chapter 2** Late Naughties
  Systematic Brute Forcing & standardised
  features (openSMILE), spontanoues
  realistic data, challenges (Interspeech '09)
- **Chapter 3** Deep Learning – 2010s
  End-2-End Learning (2016), AutoML (2018),
  Dimensions
- **Chapter 4** 2020ies… Several Companies, Application

Audio Transcript | Chat Messages

Search transcript

interesting. But then, after um that,
it's really degrading for most of the
most challenging database on the

Emotion Corpora/ Databases/ Datasets

Database	Type	Size [hrs]	Speakers
eNTERFACE'05 [23]	Acted	-	42
CREMA-D [10]	Acted	-	91
RAVDESS [19]	Acted	-	24
IEMOCAP [5]	Acted	≈ 12	10
MSP-IMPROV [8]	Acted	≈ 9.5	12
CreativeIT [25, 26]	Acted	≈ 8	16
SEMAINE [24]	Natural	≈ 75	150
MAHNOB-HCI [41]	Natural	-	27
RECOLA [35]	Natural	≈ 3.75	46
SEWA [18]	Natural	44	398
CMU-MOSEI [45]	Natural	≈ 65	1,000
MSP-Face	Natural	≈ 24.7 (+46)	302

Durham University

Vidal et al. (2020).

CREMA-D (Cao et al. 2014, IEEE Transaction on Affective Computing)

Code	Sentence
DFA	Don't forget a jacket.
IEO	It's eleven o'clock,
IOM	I'm on my way to the meeting.
ITH	I think I have a doctor's appointment.
ITS	I think I've seen this before.
IWL	I would like a new alarm clock.
IWW	I wonder what this is about.
MTI	Maybe tomorrow it will be cold.
TAI	The airplane is almost full.
TIE	That is exactly what happened.
TSI	The surface is slick.
WSI	We'll stop in a couple of minutes.

12 sentences of emotionally neutral content

6 emotions: Anger, Disgust, Fear, Happy, Sad +
Neutral

IEO: 3 intensity levels (low, medium, high) for each emotions but NOT neutral: 3*5 + 1 = 16

Each of 11 sentences: 11 * 6 Emotions = 66

91 Actors * (16 + 66) = 7462 -> 7442 video

Each video split: audio-only, visual-only, mixed

7442*3 clips. Each annotated by 10 raters

223,269 clips / ~ 90 clips = 2443 raters

Durham
University

CREMA-D: Inclusivity and Diversity of Database (Bias in AI)

Actors' Age Distribution

Age	# actors
20-29 YRS	34
30-39 YRS	23
40-49 YRS	16
50-59 YRS	12
60-69 YRS	5
OVER 70 YRS	1

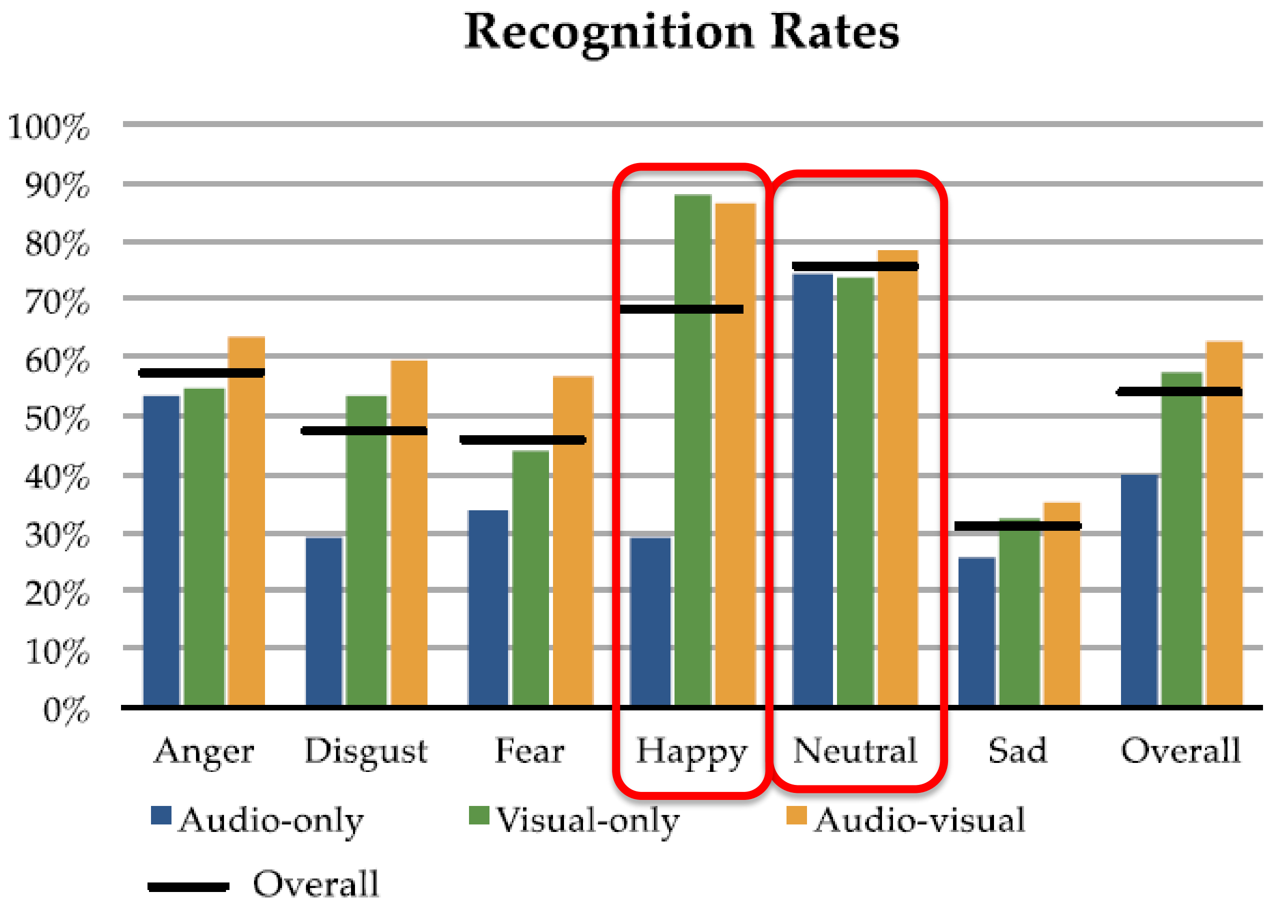
Race/Ethnicity Distribution

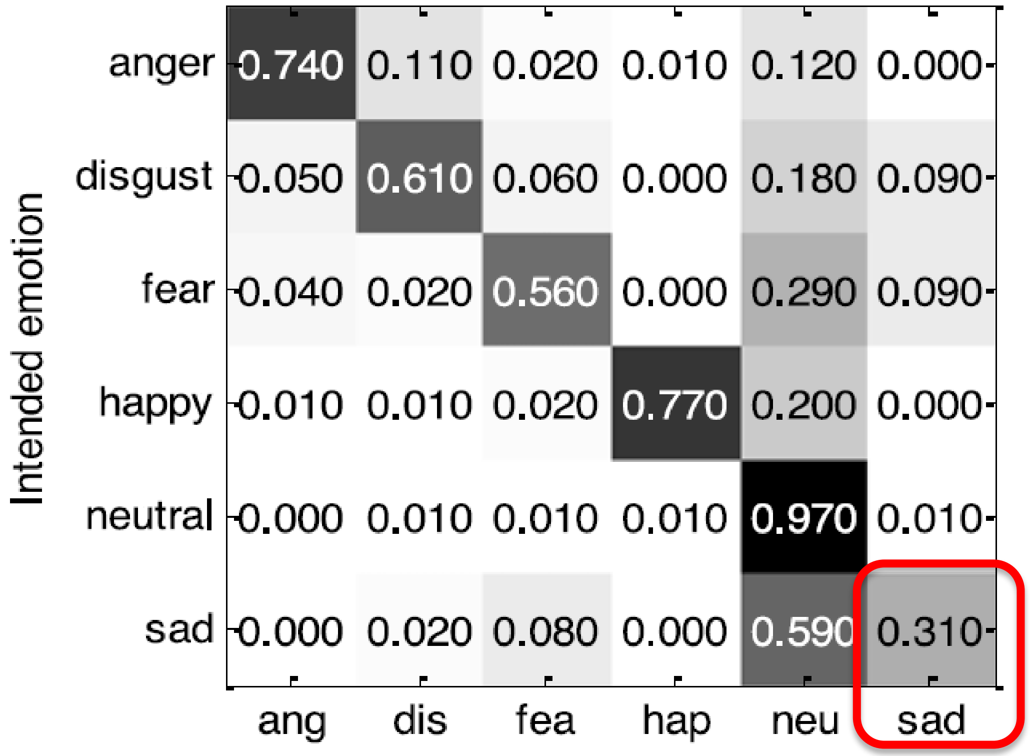
Race/Ethnicity	Raters	Actors
Caucasian	73.60%	58.24%
Hispanic	10.80%	10.99%
African American	8.10%	23.08%
Asian	4.50%	7.69%
Other/No Answer	3.00%	0.00%

Test your Emotional Intelligence!

Google Form

CREMA-D: Emotion Recognition Rates



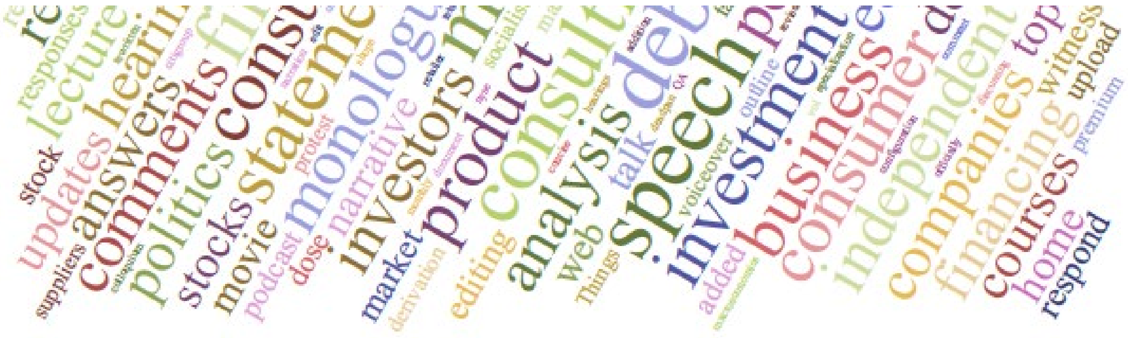


CMU-MOSEI (Multimodal Opinion Sentiment and Emotion Intensity)

Dataset	# S	# Sp	Mod	Sent	Emo	TL (hh:mm:ss)
CMU-MOSEI	**23,453**	**1,000**	$\{l, v, a\}$	✓	✓	**65:53:36**
CMU-MOSI	2,199	98	$\{l, v, a\}$	✓	✗	02:36:17
ICT-MMMO	340	200	$\{l, v, a\}$	✓	✗	13:58:29
YouTube	300	50	$\{l, v, a\}$	✓	✗	00:29:41
MOUD	400	101	$\{l, v, a\}$	✓	✗	00:59:00
SST	11,855	–	$\{l\}$	✓	✗	–
Cornell	2,000	–	$\{l\}$	✓	✗	–
Large Movie	25,000	–	$\{l\}$	✓	✗	–
STS	5,513	–	$\{l\}$	✓	✗	–
IEMOCAP	10,000	10	$\{l, v, a\}$	✗	✓	11:28:12
SAL	23	4	$\{v, a\}$	✗	✓	11:00:00
VAM	499	20	$\{v, a\}$	✗	✓	12:00:00
VAM-faces	1,867	20	$\{v\}$	✗	✓	–
HUMAINE	50	4	$\{v, a\}$	✗	✓	04:11:00
RECOLA	46	46	$\{v, a\}$	✗	✓	03:50:00
SEWA	538	408	$\{v, a\}$	✗	✓	04:39:00
SEMAINE	80	20	$\{v, a\}$	✗	✓	06:30:00
AFEW	1,645	330	$\{v, a\}$	✗	✓	02:28:03
AM-FED	242	242	$\{v\}$	✗	✓	03:20:25
Mimicry	48	48	$\{v, a\}$	✗	✓	11:00:00
AFEW-VA	600	240	$\{v, a\}$	✗	✓	00:40:00

Total number of sentences	23453
Total number of videos	3228
Total number of distinct speakers	1000
Total number of distinct topics	250
Average number of sentences in a video	7.3
Average length of sentences in seconds	7.28
Total number of words in sentences	447143

Topics: Reviews (16%), Debate (3%), Consulting (2%)



CMU-MOSEI: Rationale & Sources

- Diversity in the training samples
- Variety in the topics
- Diversity in speakers

Sources:

- Social multimedia: monologue videos of opinions
  - language in the form of spoken text
  - visual via perceived gestures and facial expressions
  - acoustic through intonations and prosody
- 5000 videos, 14 experts quality check → 3228 videos
- Automatic check: facial feature extraction confidence
- 57% male vs. 43% female
- Tokenisation: Punctuation marks rather than Stanford CoreNLP tokenizer

Summary

- Humans are NOT good emotion recognisers
- AI-powered emotion recognition applications (ERA) may (not) be better in terms of accuracy, which in general remains moderate.
- The ground truth of ERA is based on human annotators → paradox?!
- More research on Affective Computing is MUCH needed!

Suggested Reading

Christy, T., & Kuncheva, L. I. (2014). Technological advancements in affective gaming: A historical survey. *GSTF Journal on Computing (JoC)*, *3*(4), 1-10.

Gandhi, A., Adhvaryu, K., Poria, S., Cambria, E., & Hussain, A. (2022). Multimodal sentiment analysis: A systematic review of history, datasets, multimodal fusion methods, applications, challenges and future directions. *Information Fusion*.

Picard, R. W. (2000). *Affective computing*. MIT press.

Poria, S., Cambria, E., Bajpai, R., & Hussain, A. (2017). A review of affective computing: From unimodal analysis to multimodal fusion. *Information Fusion*, *37*, 98-125.

Schuller, B. W. (2018). Speech emotion recognition: Two decades in a nutshell, benchmarks, and ongoing trends. *Communications of the ACM*, *61*(5), 90-99.

Tomar, P. S., Mathur, K., & Suman, U. (2022). Unimodal approaches for emotion recognition: A systematic review. *Cognitive Systems Research*.

Zhao, S., Yao, X., Yang, J., Jia, G., Ding, G., Chua, T. S., ... & Keutzer, K. (2021). Affective image content analysis: Two decades review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.