

# There will be Artificial Emotional Intelligence

Professor Björn W. Schuller  
FIEEE, FBCS, FISCA



>Close your eyes.

>Think of AI.

> what does it? ■



Imperial College  
London

**AI.**

**Reasoning  
& Problem  
Solving**

**Knowledge  
Representation**

**Planning**

**Learning**

**A.I.**

**Socioemotional  
Intelligence**

**NLP**

**Motion &  
Manipulation**

**Perception**

**Creativity**

**General  
Intelligence**



Imperial College  
London

## Emotions?

**Survive**

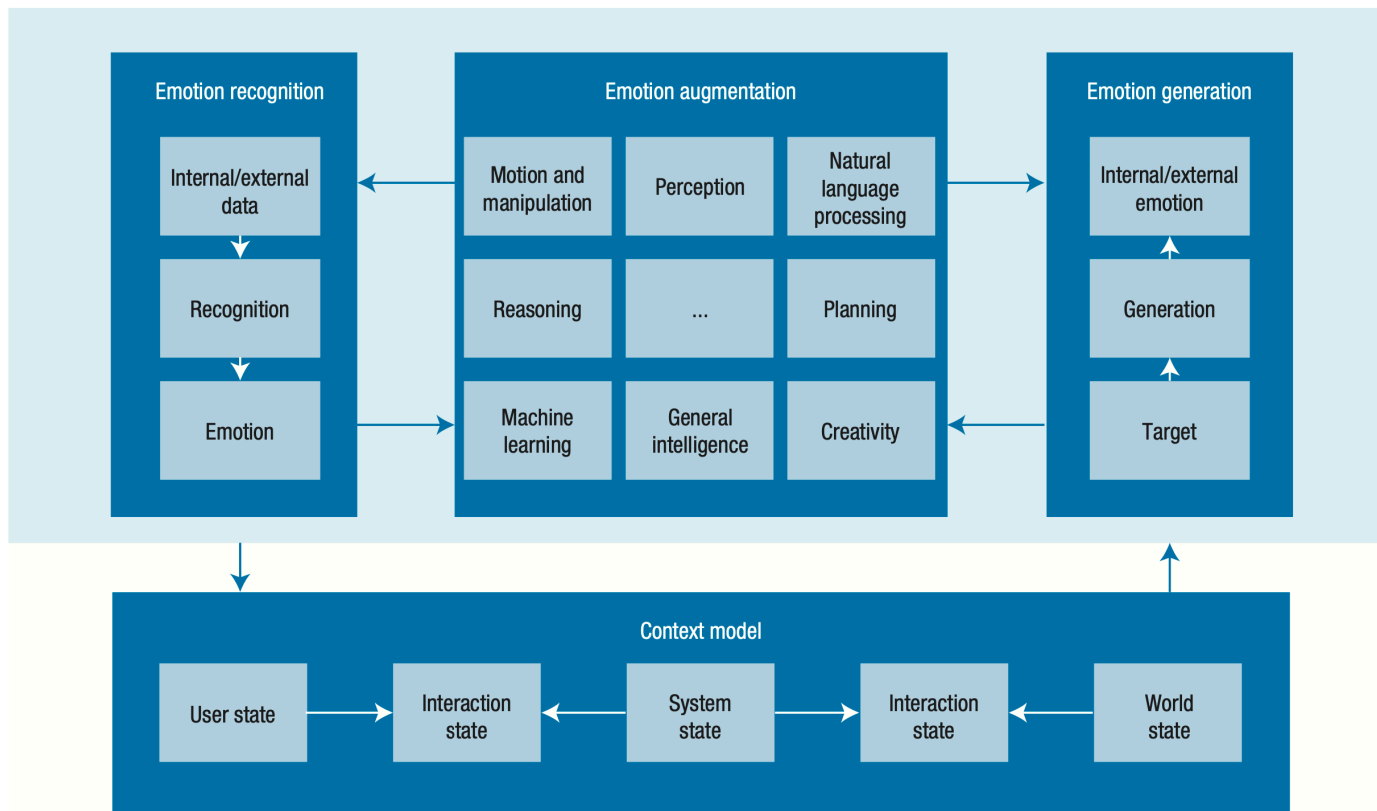
**Motivate**

**Make  
Decisions**

**Inform  
yourself**

**Communicate**

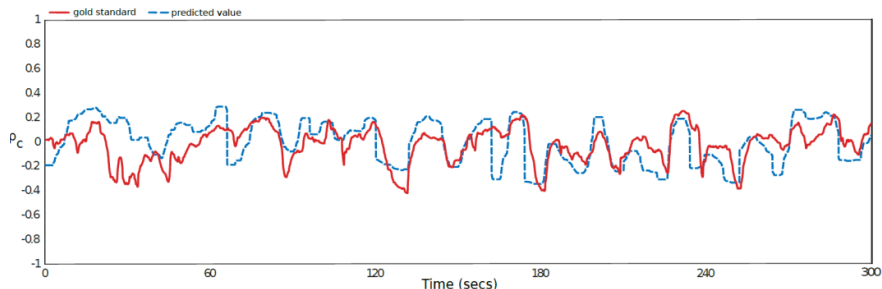
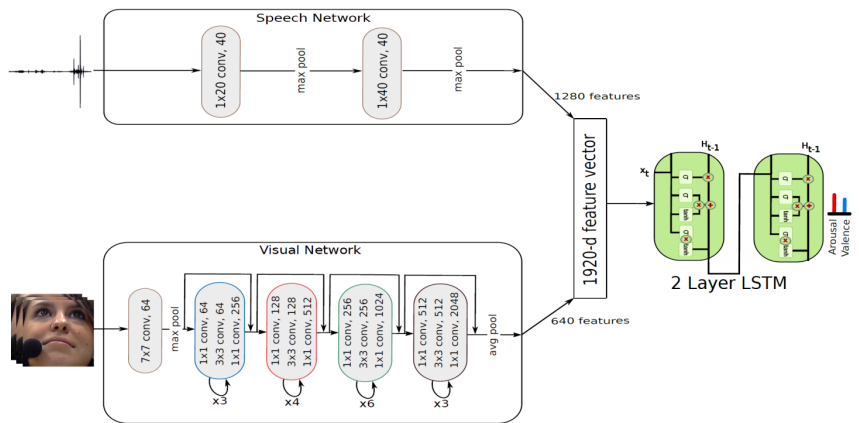




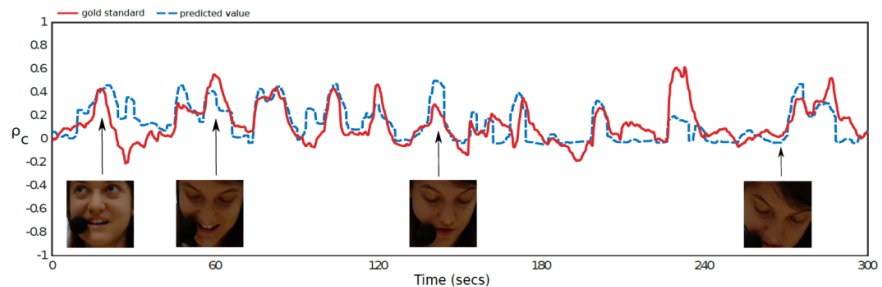
> The tech.



# End-to-End.



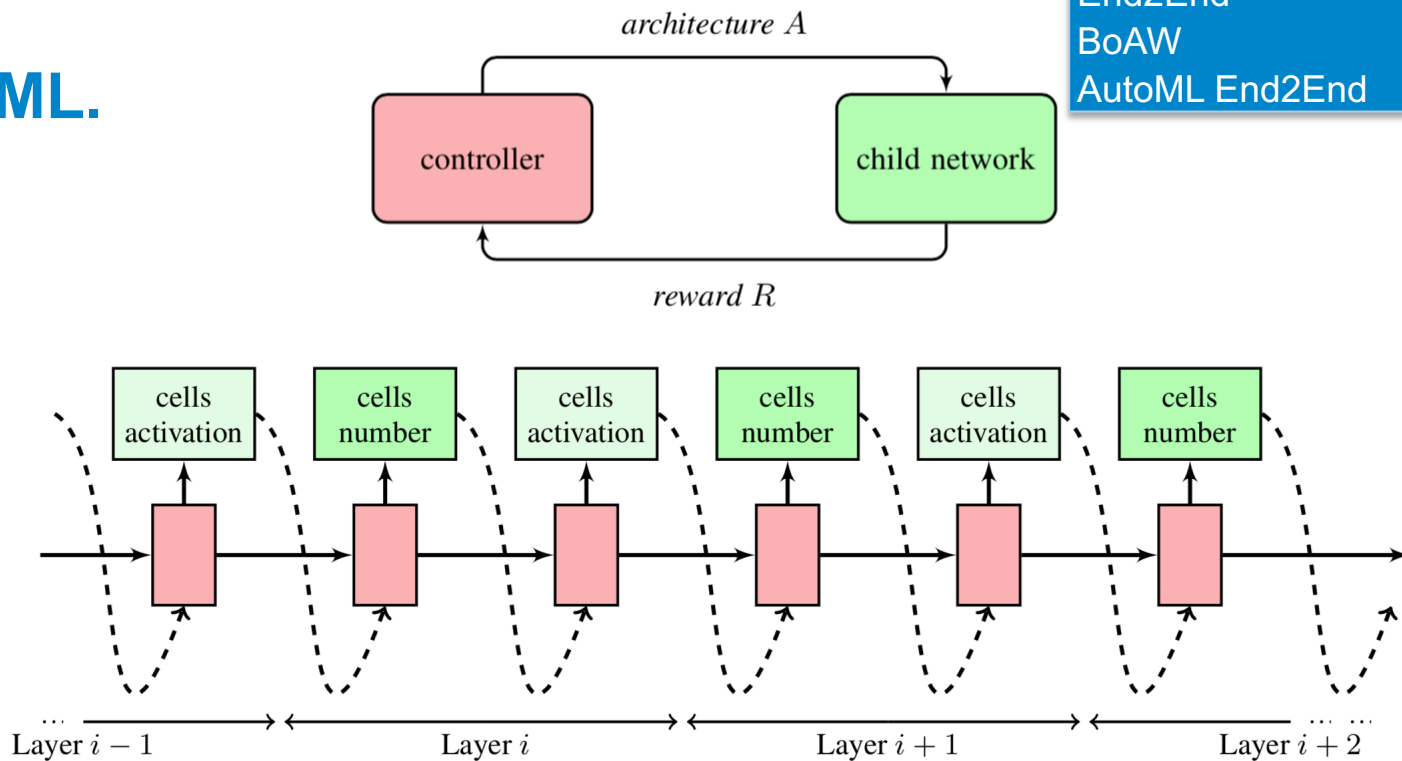
(a) Arousal



(b) Valence



ComParE 18 Crying	%UA
End2End	63.5
BoAW	67.7
AutoML End2End	70.1



## Challenge.

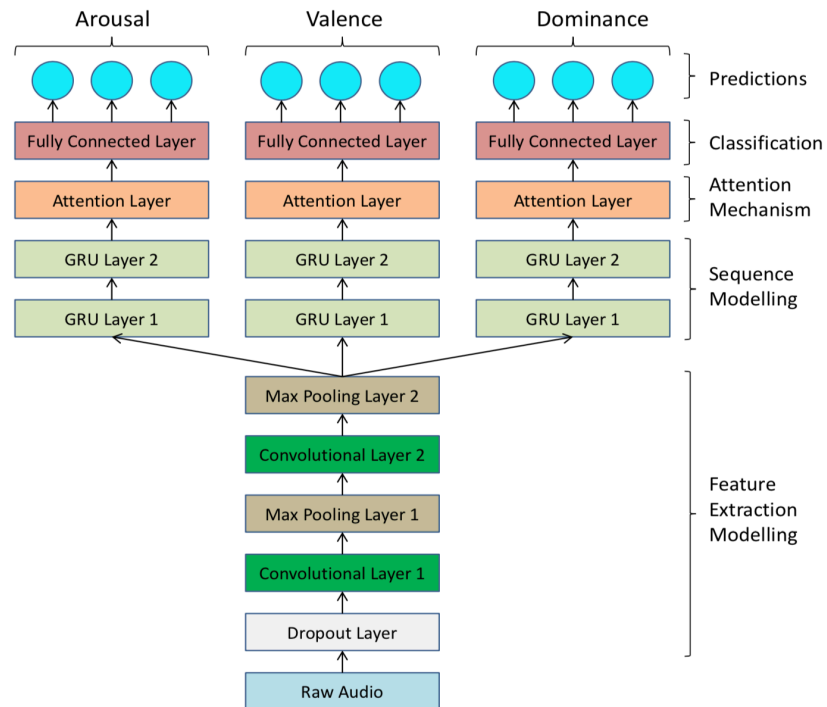
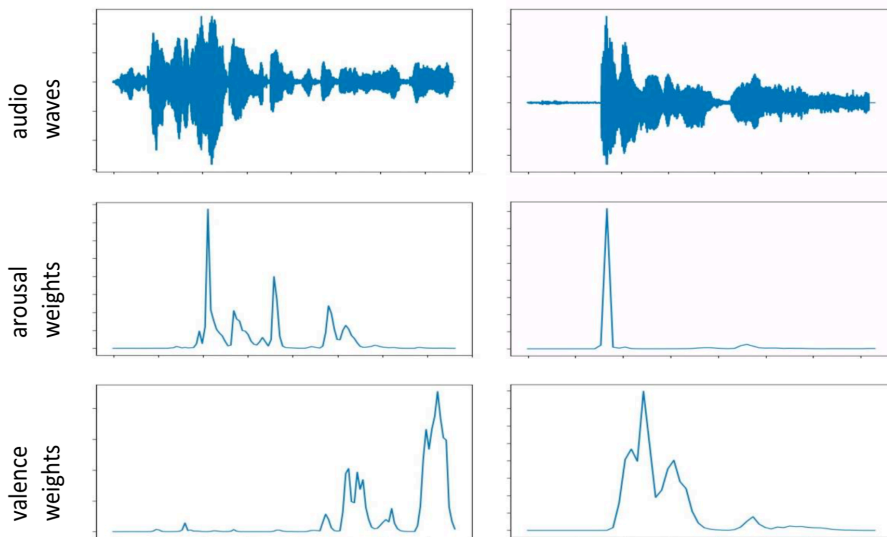
	# Classes	%UA/*AUC/+CC
Interest	[-1,1]	42.8+
Emotion	5	44.0
Negativity	2	71.2

	# Classes	%UA/*AUC/+CC
Deception	2	72.1
Sincerity	[0,1]	65.4+
Social Signals	2x2	92.7*
Conflict	2	85.9
Emotion	12	46.1

## INTERSPEECH COMPARE

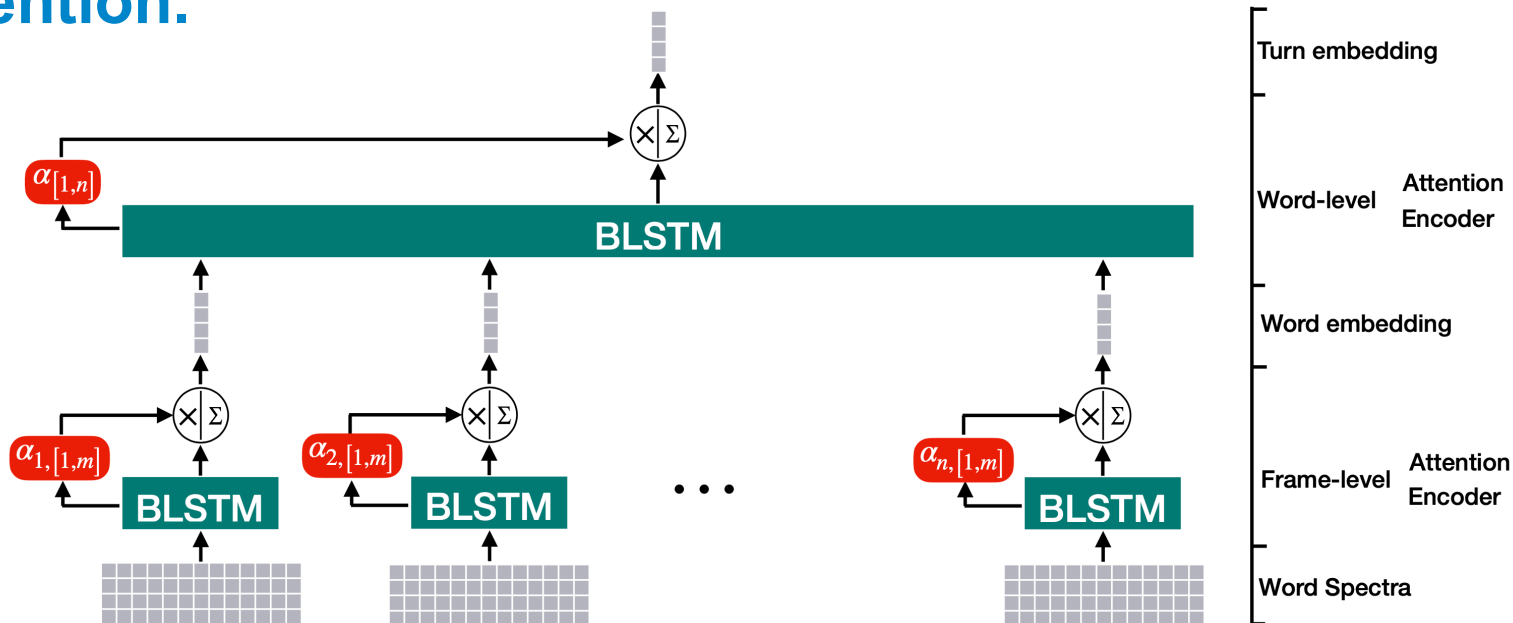
	# Classes	%UA/*AUC/+CC
Affect: Atypical	4	45.0
Affect: Self-Ass.	3	68.4
Crying	3	78.6
Heart Beats	3	56.2
Elderly A/V	2x3	63.8
Escalation	3	(59.8)

# Attention.



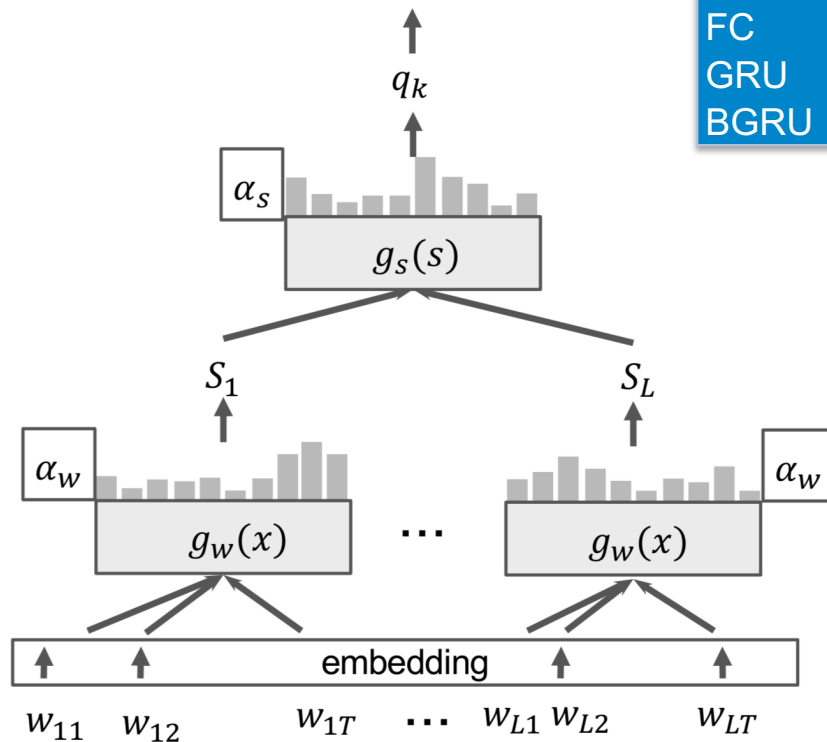


## Attention.



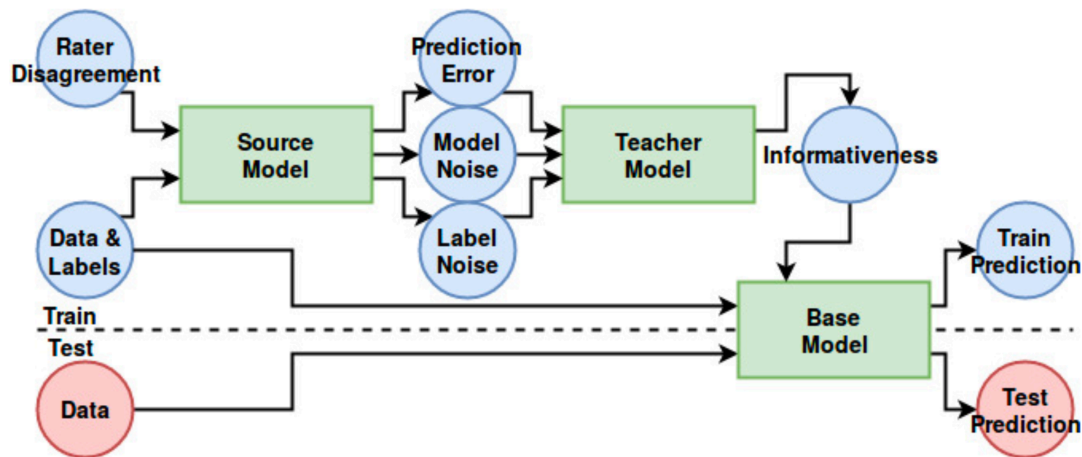
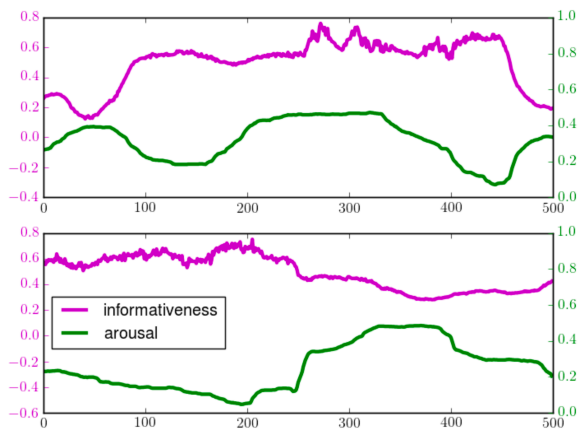
# Attention.

UAR [%]	USoMS	w/o	w/ Att.
+/-		Att.	
FC		81.4	85.2
GRU		86.5	88.5
BGRU		87.8	91.0

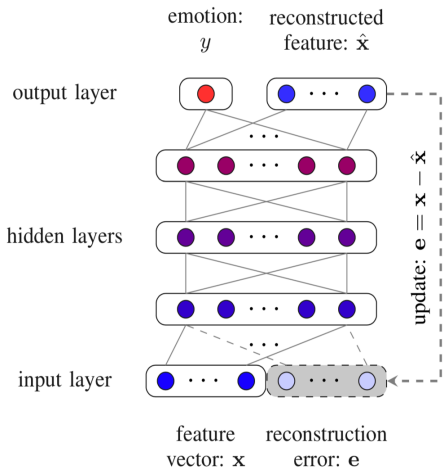


# Uncertainty.

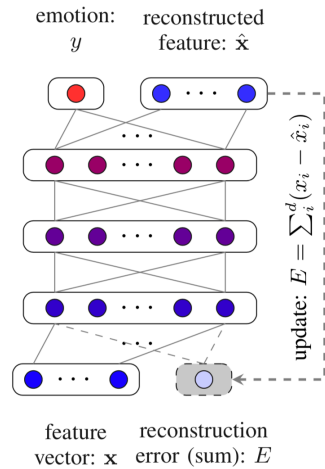
CCC RECOLA	Audio- Va	Video- Ar
e2e	.607	.322
... + model uncertainty	.691	.324
... + label uncertainty	.680	.319
... + teacher model	.693	.352



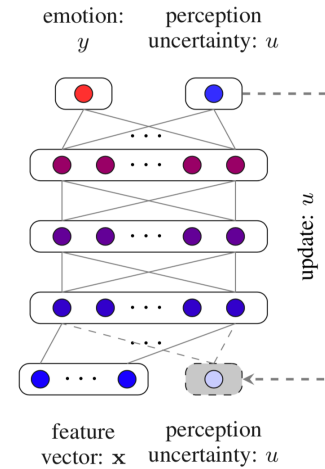
# Uncertainty.



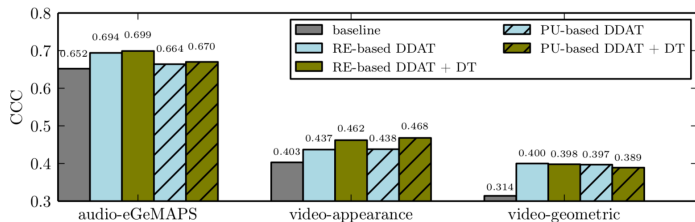
(a) based on reconstruction error (vector)



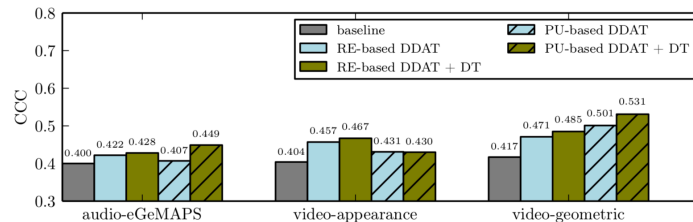
(b) based on reconstruction error (scalar)



(c) based on perception uncertainty



(a) arousal



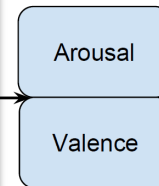
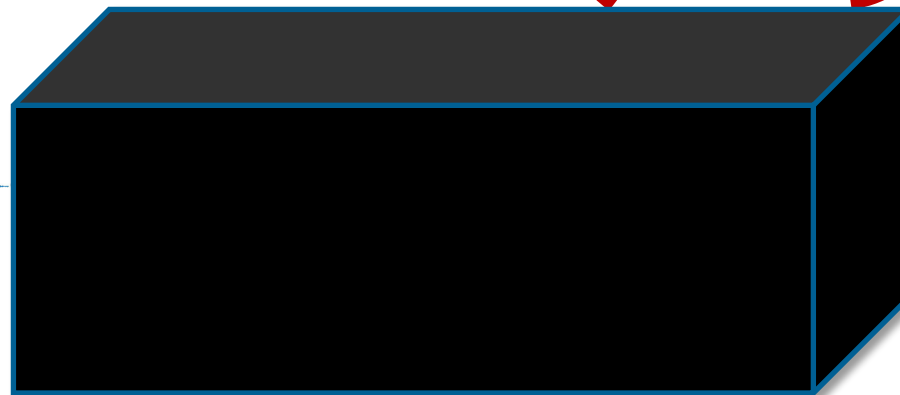
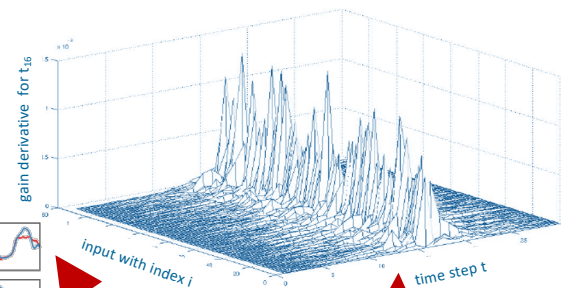
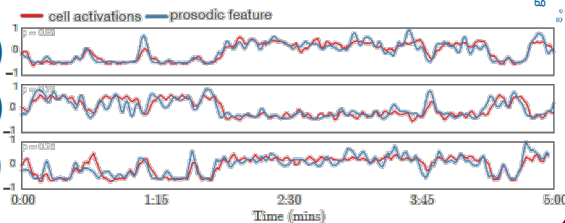
(b) valence

# Explainable.

loudness range (.77)

loudness mean (.73)

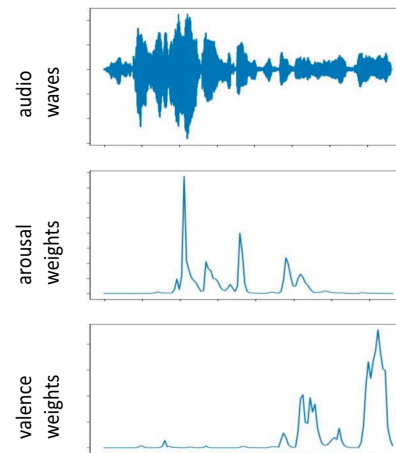
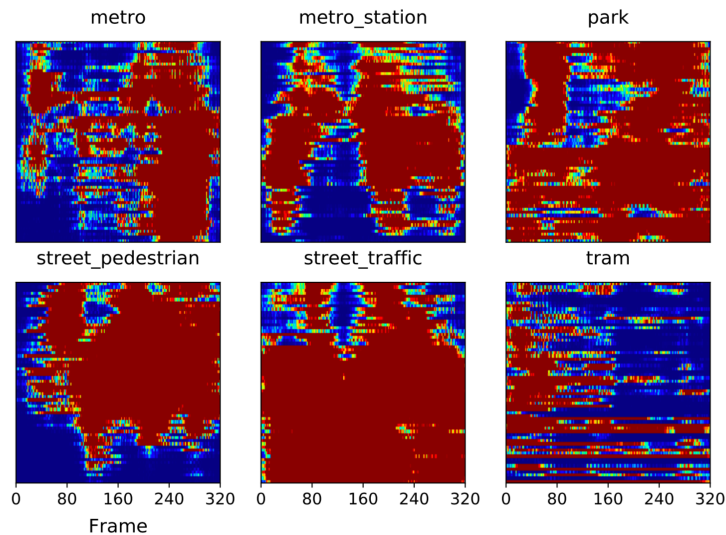
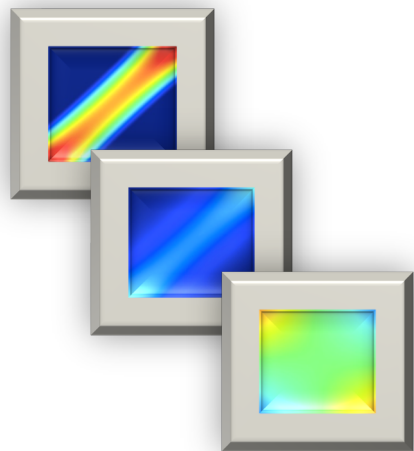
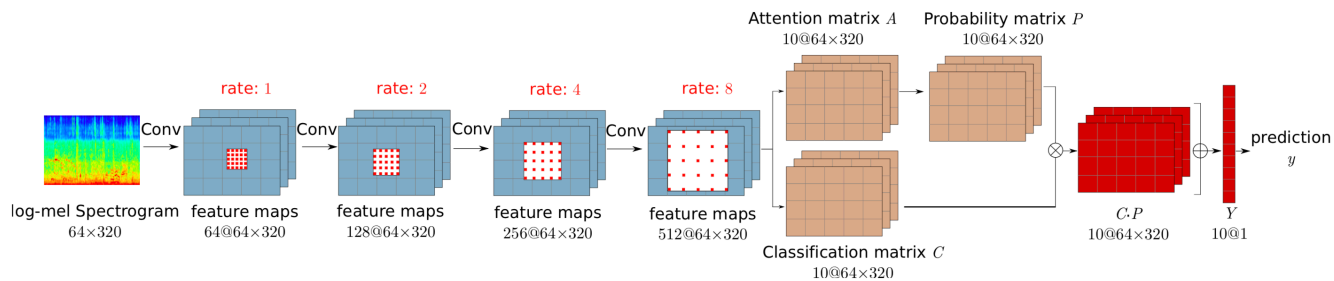
pitch mean (.71)



Input raw waveform at 16kHz

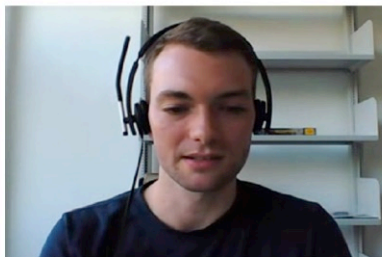
Output label at 25Hz

# Explainable.

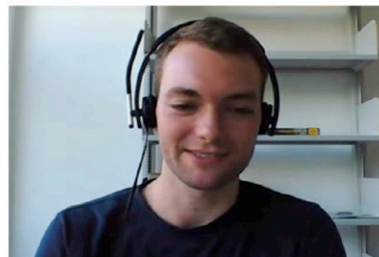


## Confidence.

CCC	Arousal	Valence
Baseline	.386	.478
+Confidence	.450	.515



(a)  $E^{(A)} = 0.01, \sigma^{(A)} = 0.01$



(b)  $E^{(A)} = 0.01, \sigma^{(A)} = 0.11$



(c)  $E^{(V)} = 0.79, \sigma^{(V)} = 0.47$

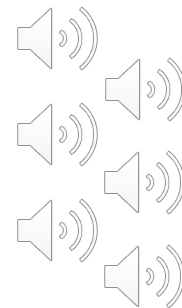
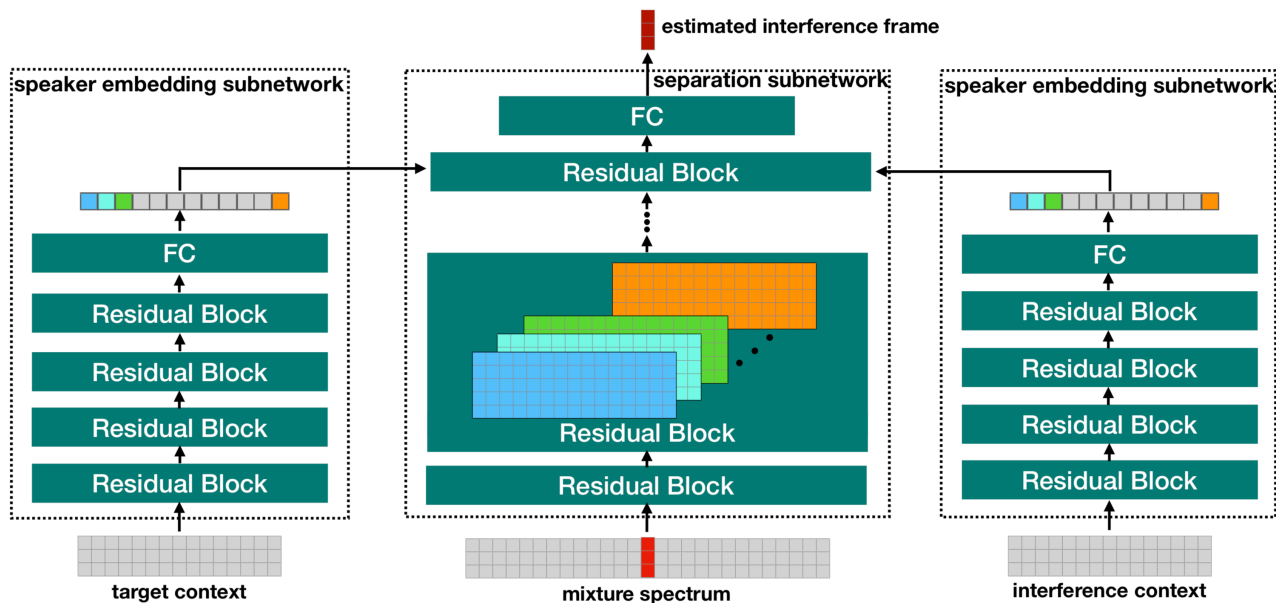


(d)  $E^{(V)} = 0.69, \sigma^{(V)} = 0.70$

Reliable.

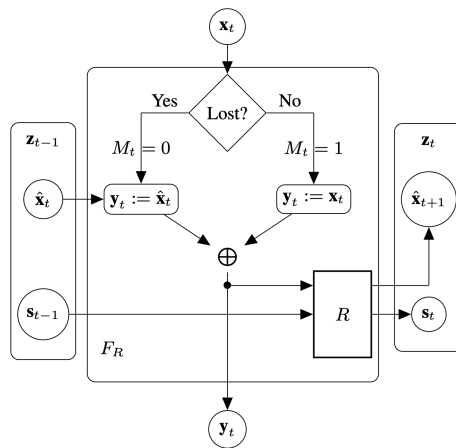
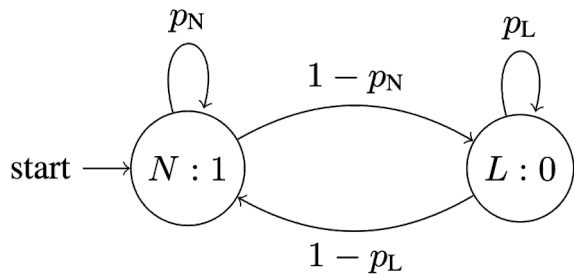
—))) + N-HANS

Separation	SDR	SAR
Deep Clustering	0.84	2.09
Deep Attractor	1.81	3.29
N-HANS	4.79	8.44



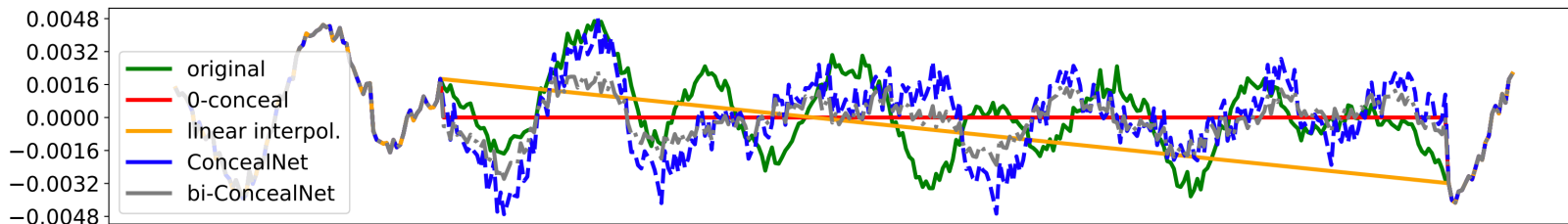


# Package Loss?



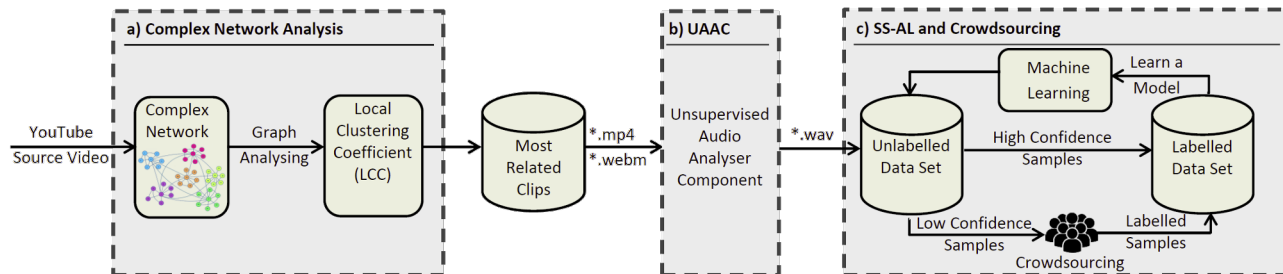
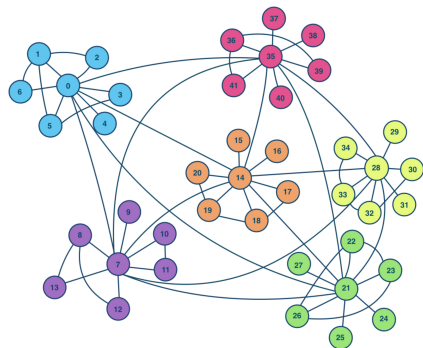
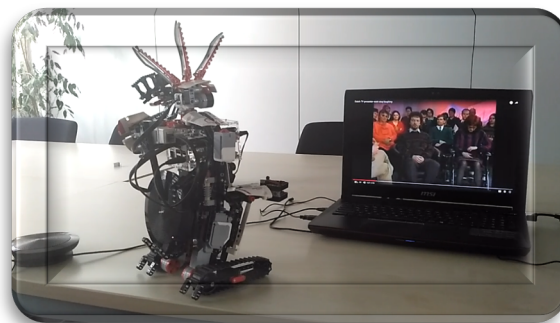
CCC	w/o	CNet
Arousal	.769	-
Original	.769	-
35% drop	.681	.769
50% drop	.583	.754
85% drop	.564	.676

CCC	w/o	CNet
Valence	.432	-
Original	.432	-
35% drop	.267	.420
50% drop	.140	.399
85% drop	.104	.266



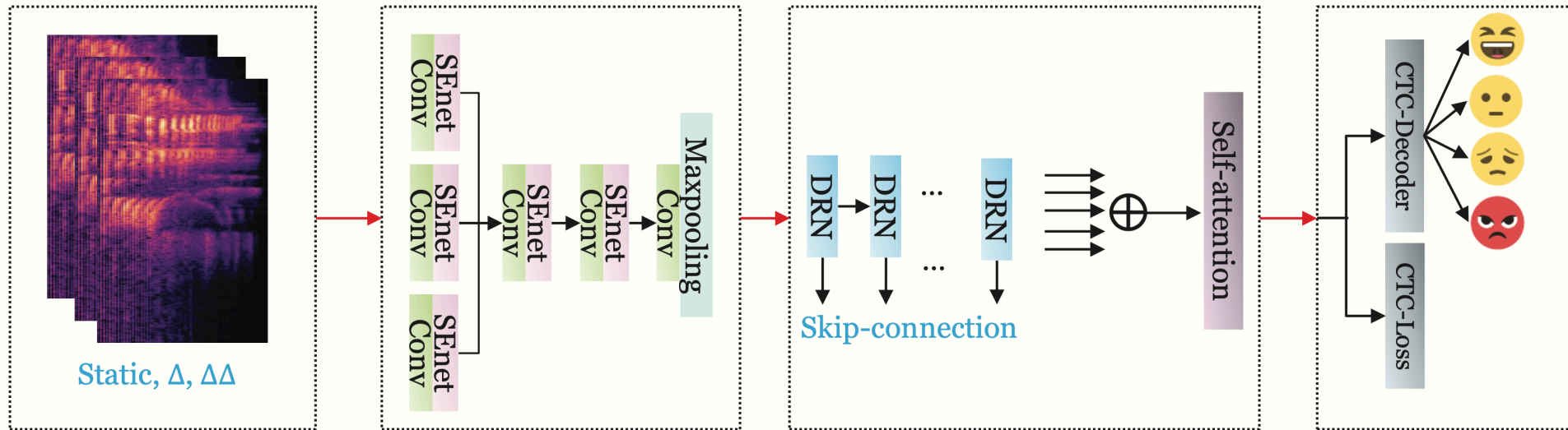
# Data.

%UA	3 hrs
Freezing	70.2
Intoxication	72.6
Screaming	97.0
Threatening	73.8
Coughing	97.6
Sneezing	85.2

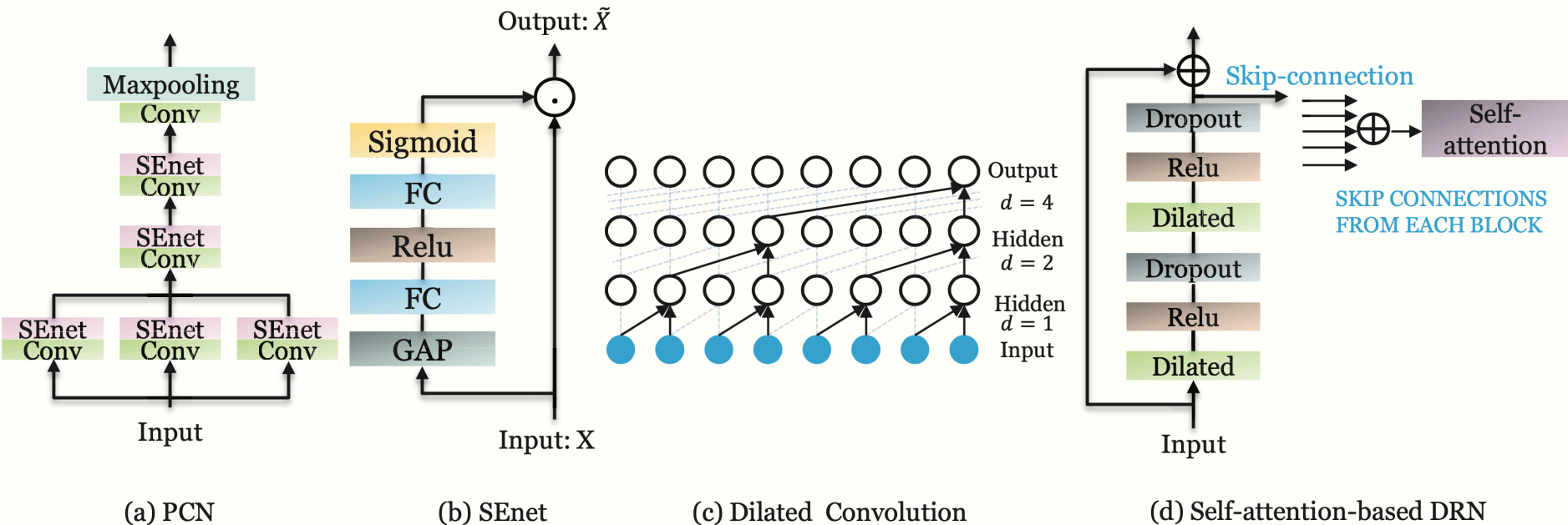


## Timing.

Combining a Parallel 2D CNN with a Self-Attention-DRN-CTC for Discrete Speech Emotion Recognition



# Timing.

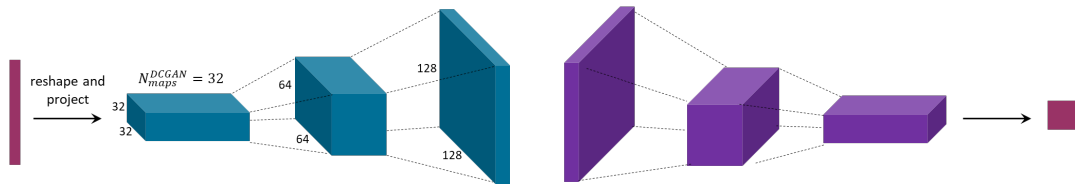
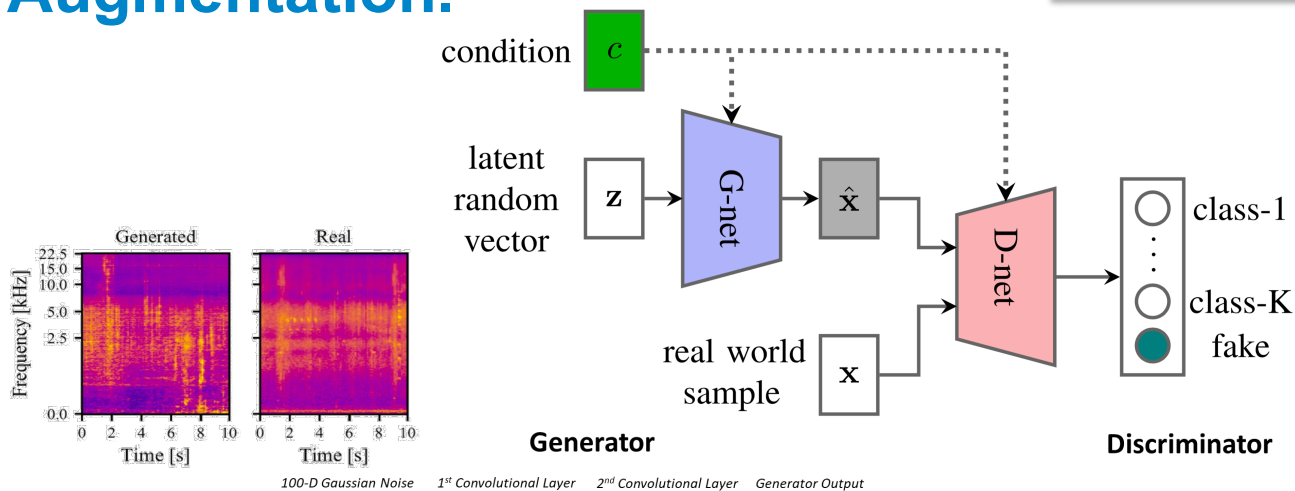


## Timing.

Methods [%]	IEMOCAP		FAU-AEC
	WA	UA	UA
CNN-BLSTM [31]	68.8	58.4	–
CNN-GRU [20]	71.5	64.2	–
BLSTM-CTC [10]	66.9	65.1	41.4
PCN	68.6	56.8	38.4
PCNSE	69.8	58.5	38.8
PCN-SABLSTM	70.8	62.7	39.8
PCNSE-SABLSTM	72.1	65.4	40.5
PCN-SADRN	71.1	62.5	41.1
PCNSE-DRN w/ Global max-pooling	71.5	62.0	38.1
PCNSE-DRN w/ Global average-pooling	71.2	62.5	39.5
PCNSE-SADRN	72.5	65.0	40.4
PCNSE-SADRN-CTC	<b>73.1</b>	<b>66.3</b>	41.1

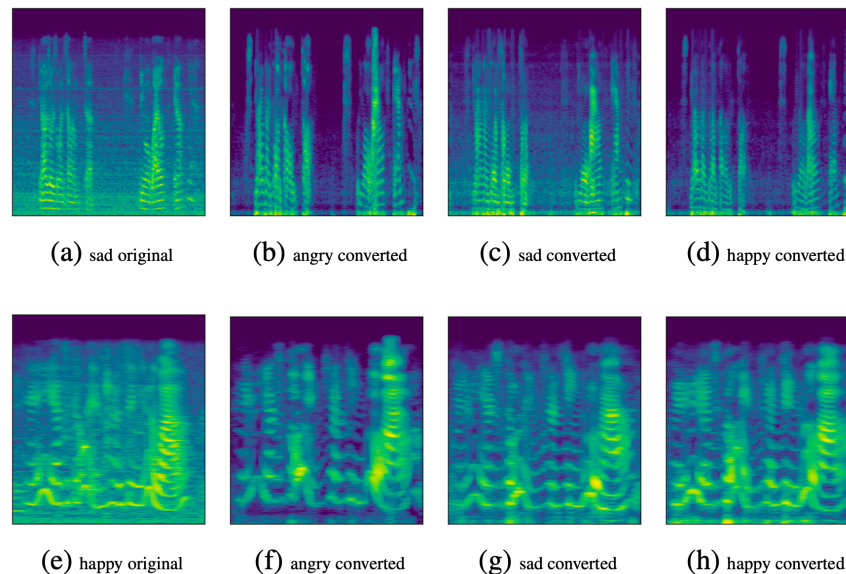
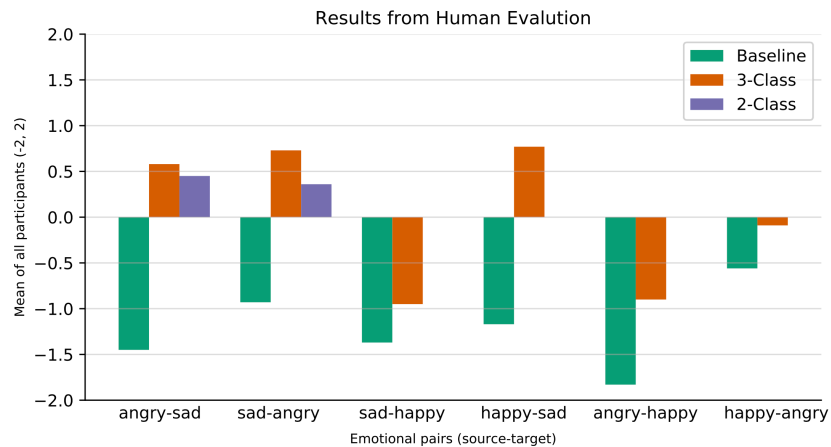
# Augmentation.

CCC RECOLA	Arousal	Valence
ComParE+LSTM	.382	.187
e2e (2016)	.686	.261
CAN	.737	.455



# Conversion.

Train on	%Micro F1	%Macro F1
IEMOCAP	61.6	50.4
Converted	60.0	51.7
Both	63.7	56.4



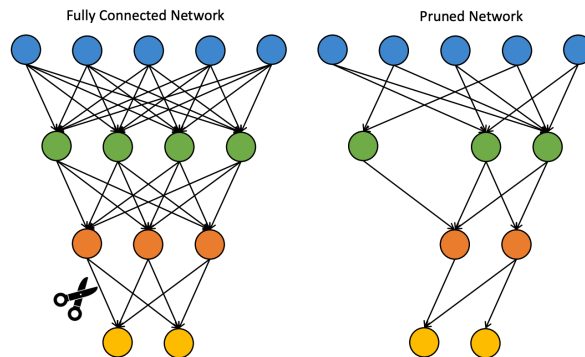
# Embedded.

Equal UAR	Size/MB	%Rdtcn
Uncompressed	229	-
Quantised	55	76
Pruned	48	79
Quantised + pruned	12	95

2.7831237	-2.8960736	4.000346
1.1066585	-4.8258014	-2.8570464
-3.9000013	-1.7565128	3.135015

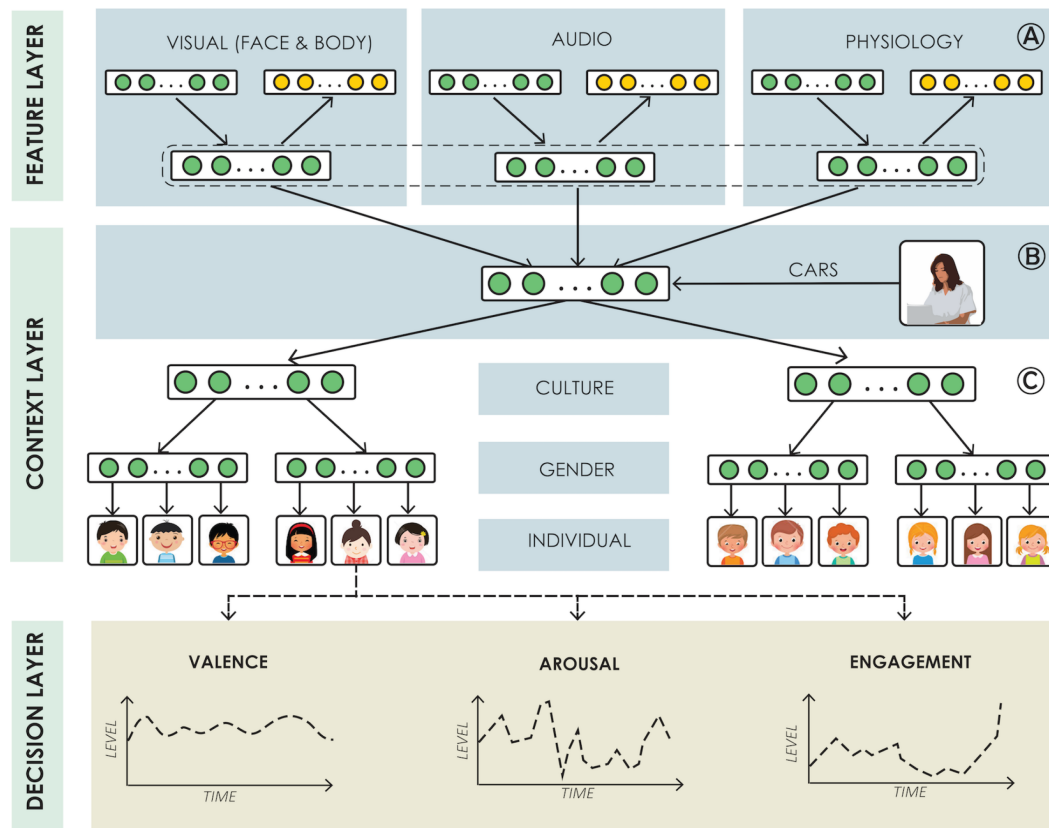


2	-2	4
1	4	-2
-3	-1	3

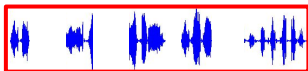
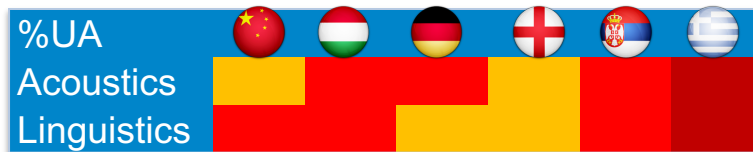




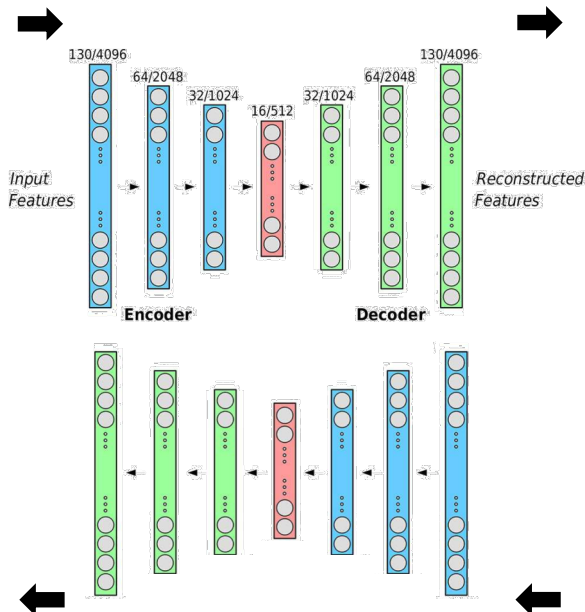
# Holism.



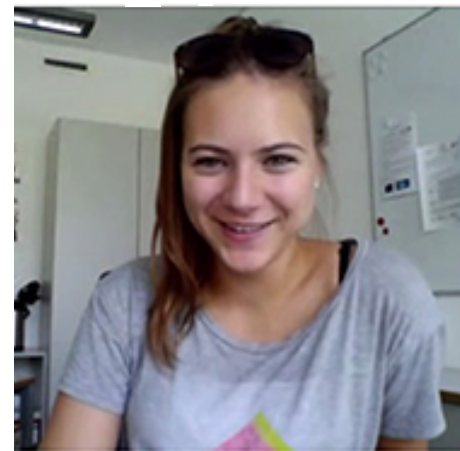
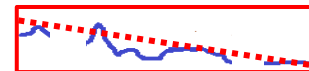
# Adaptive.



{RMSE<sub>2</sub>}

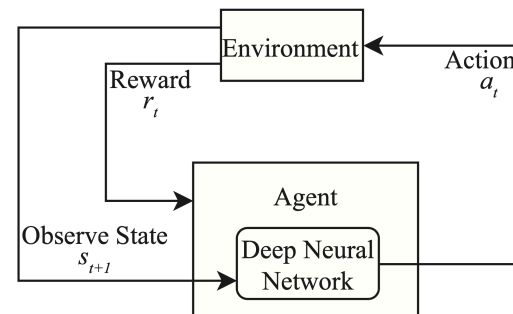
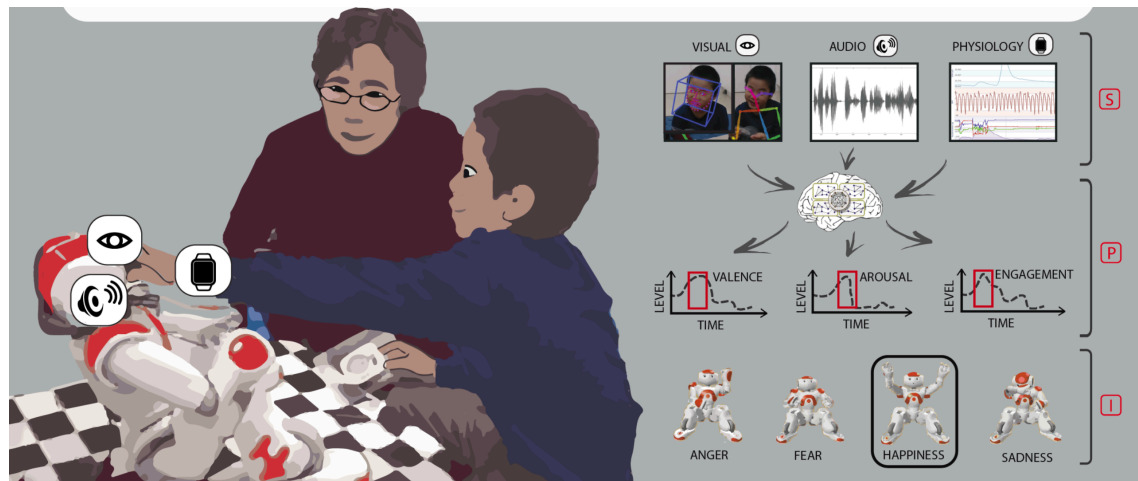


{RMSE<sub>1</sub>}



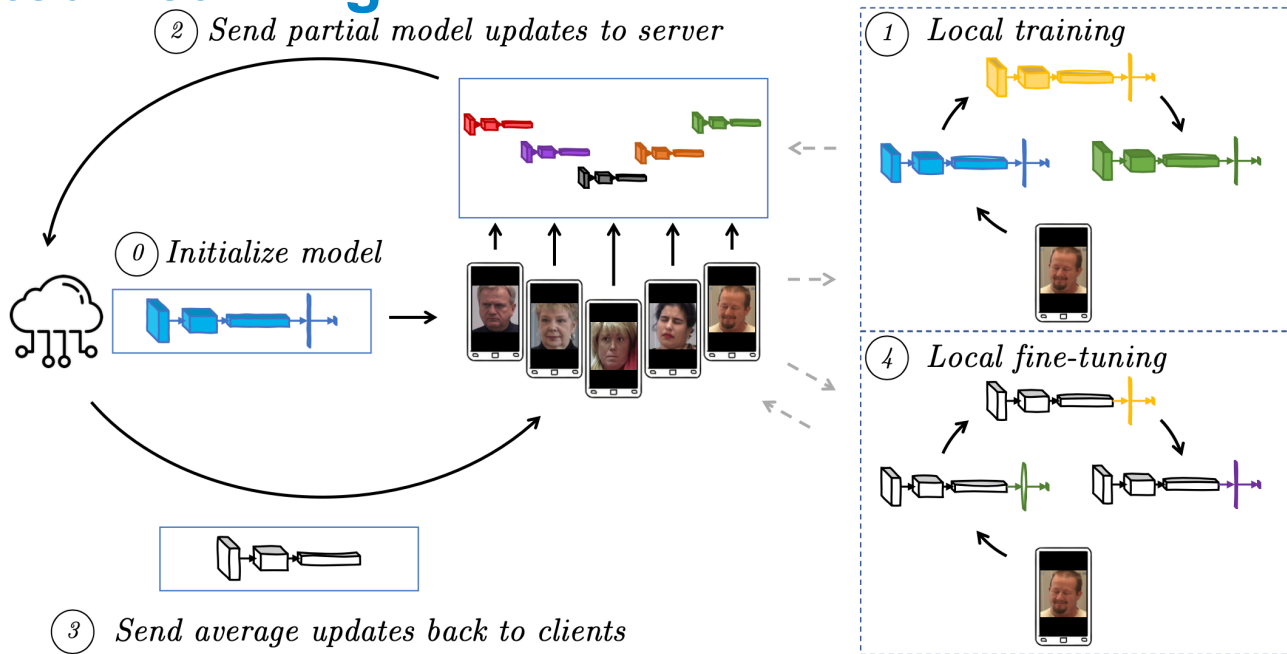
# Reinforced.

%UA	Base	RL
Face	48.6	77.5
Body	60.1	80.5
Physiology	30.6	78.9
Audio	49.0	76.4
Fusion	56.6	82.3

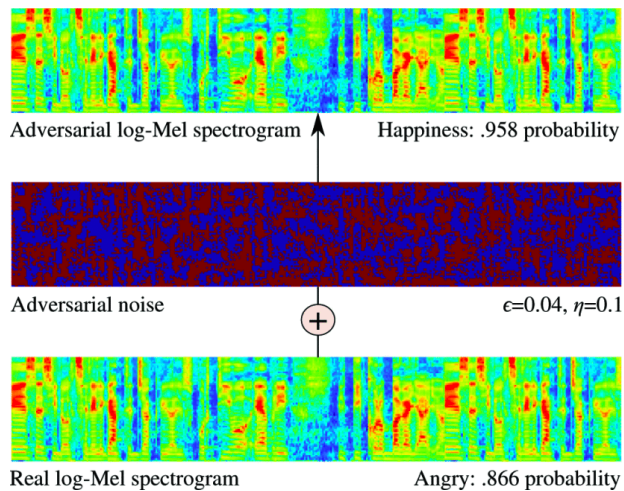
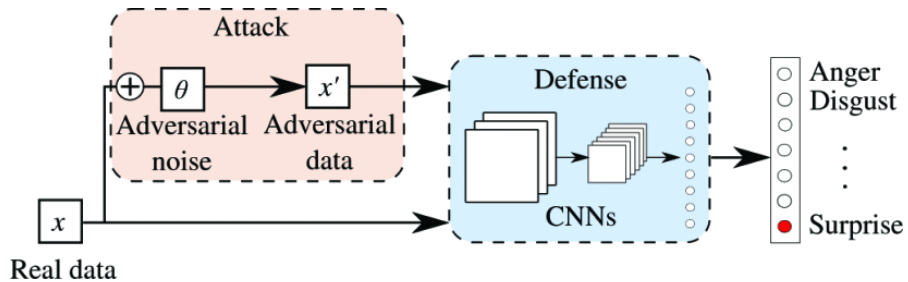


Pain	%UA
Initial	44
Original Data	73
Federated	76

# Federated Learning.

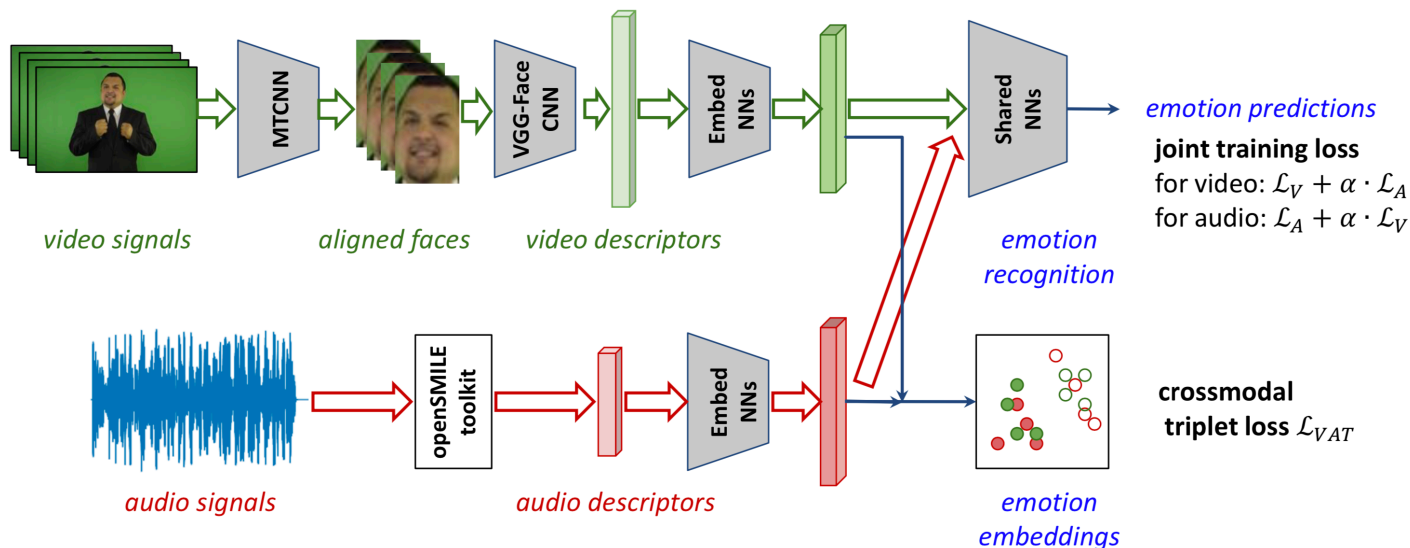


# Attacks.



# Training Support Modalities.

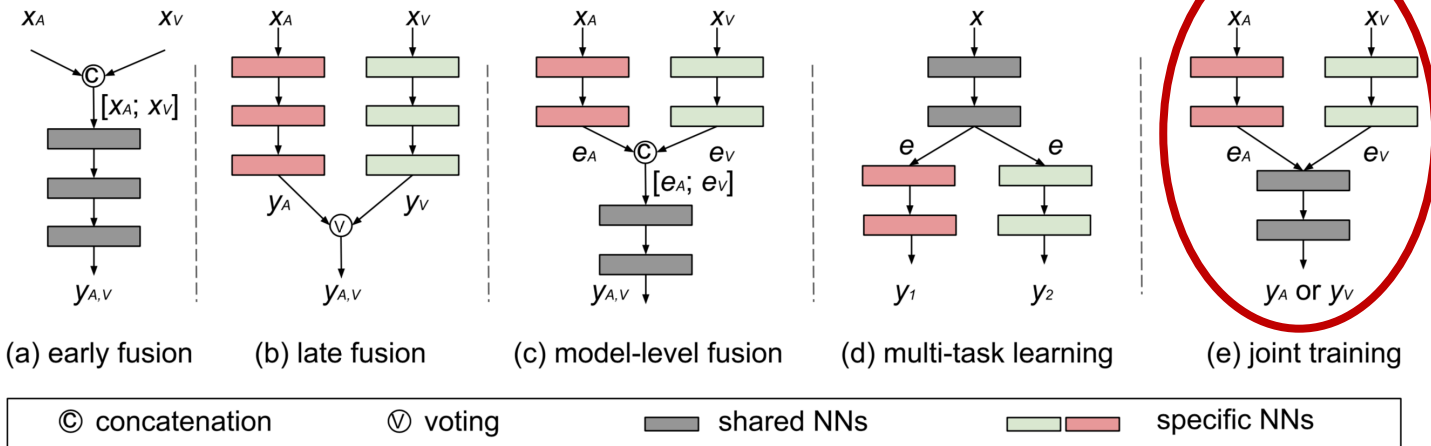
	Audio	Video
CCC RECOLA	-Va	-Ar
Monomodal	.381	.411
EmoBed	.439	.475



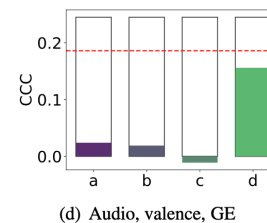
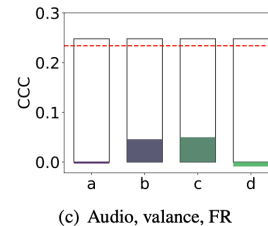
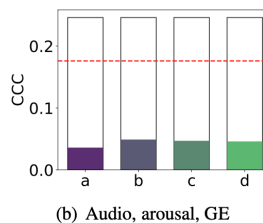
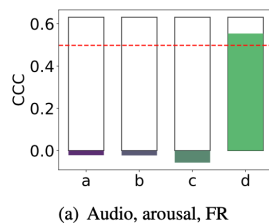
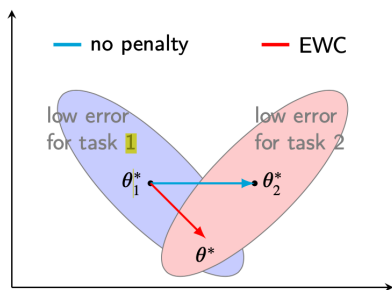
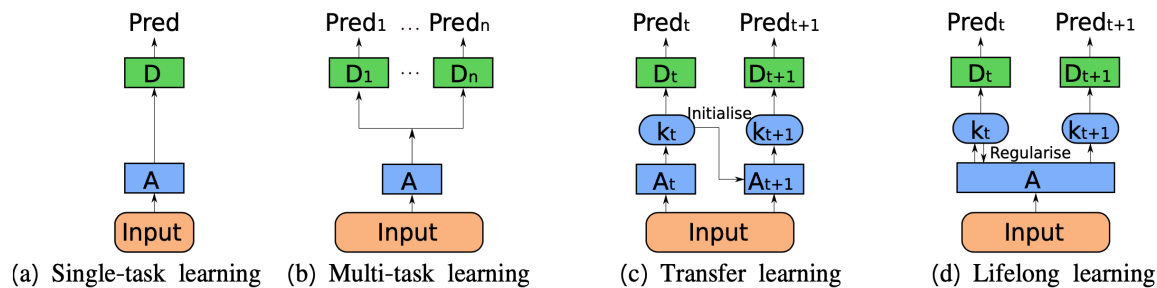
## Training Support Modalities.

- **Mono = Multimodal.**  
–IS Emotion Challenge task – 2 classes

%F1 OMG	Audio	Video
Others	33.0	37.0
Monomodal	36.5	37.9
Joint A/V	40.2	42.1
X-Modal Triplet	40.7	41.0
EmoBed	41.7	43.9

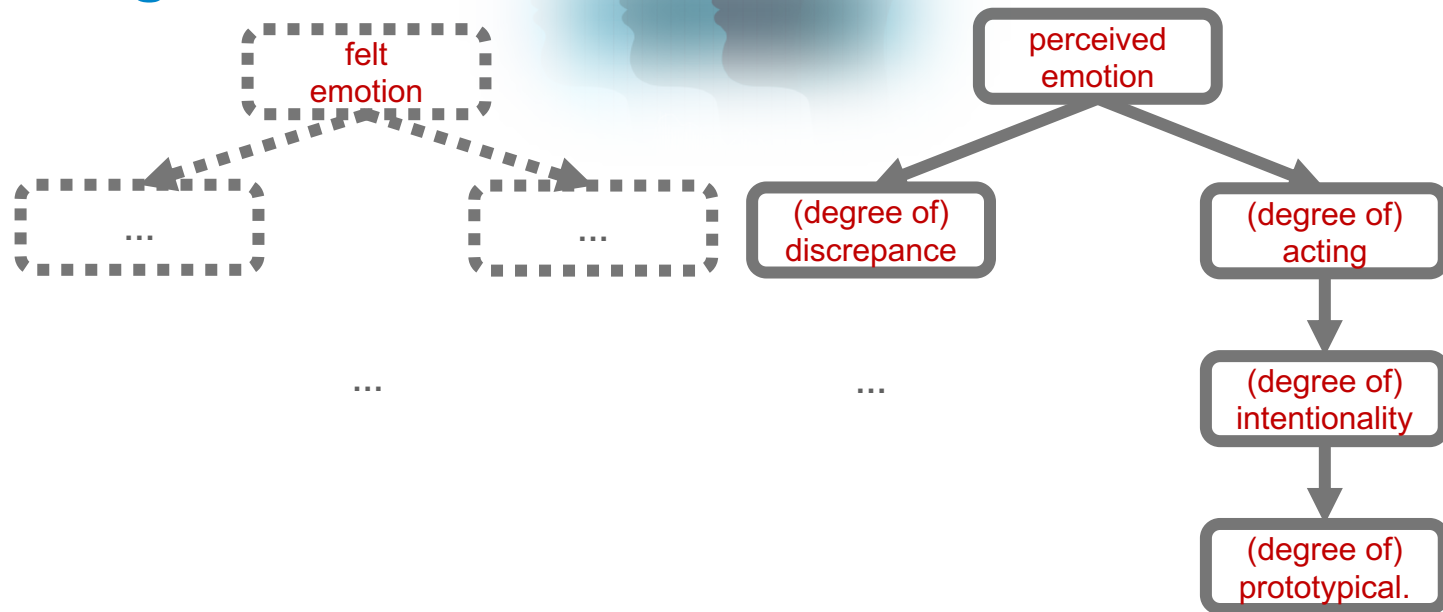


# Lifelong Learning





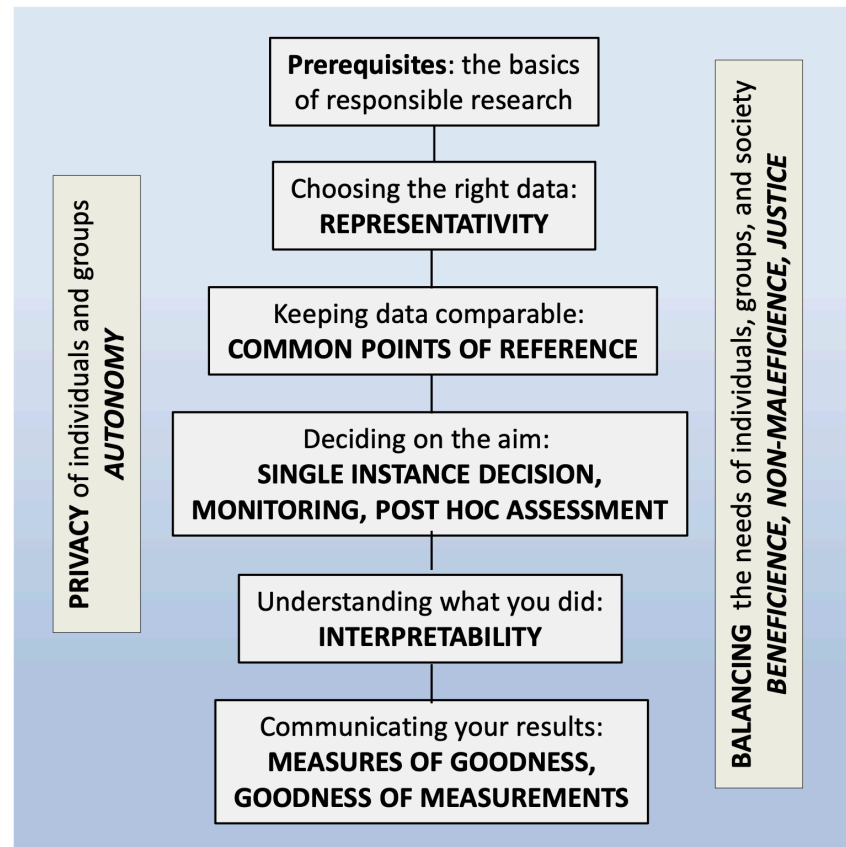
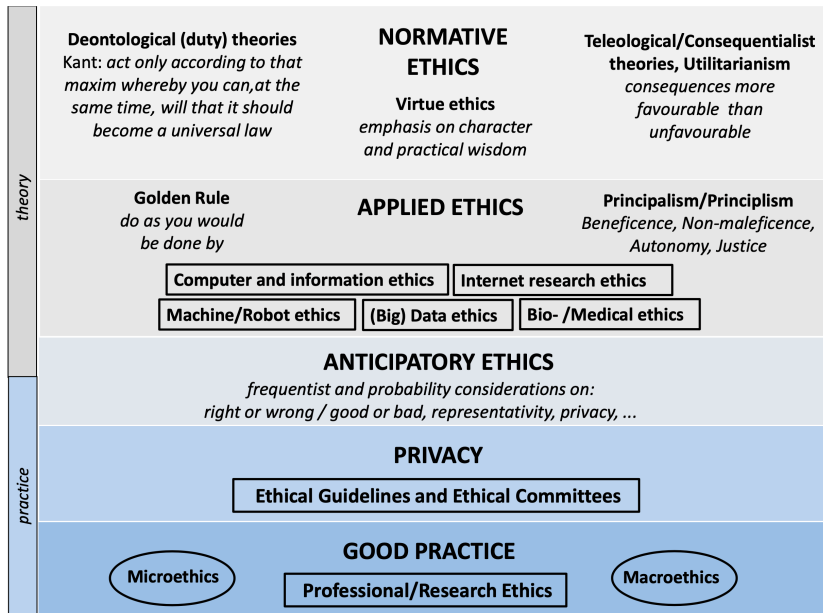
## Modelling.



## Modelling.

state						trait
spontaneous						acted
complex						simple
measured						assessed
continuous						categorical
felt						perceived
intentional						instinctual
consistent						discrepant
private						social
prototypical						peripheral
universal						culture-specific
uni-modal						multi-modal

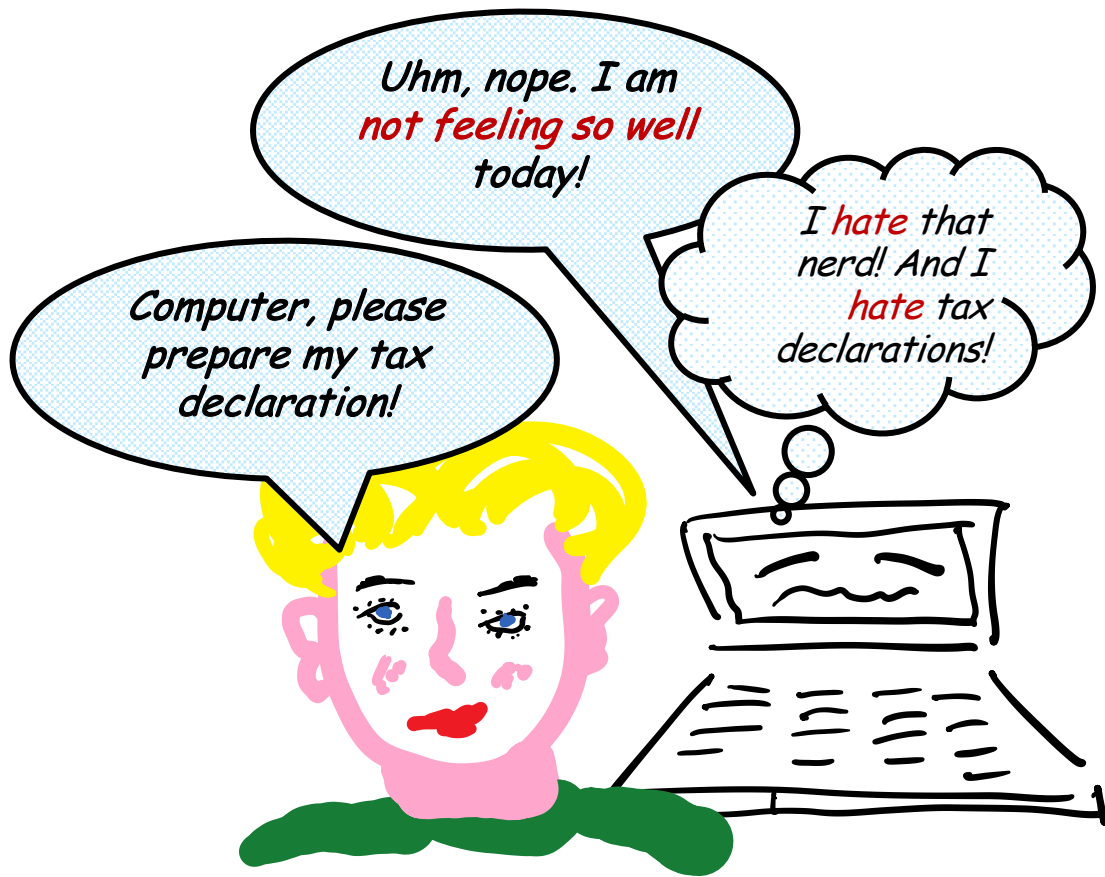
# Ethical, Moral.



> Having  
Emotions?



## Emotional AI?



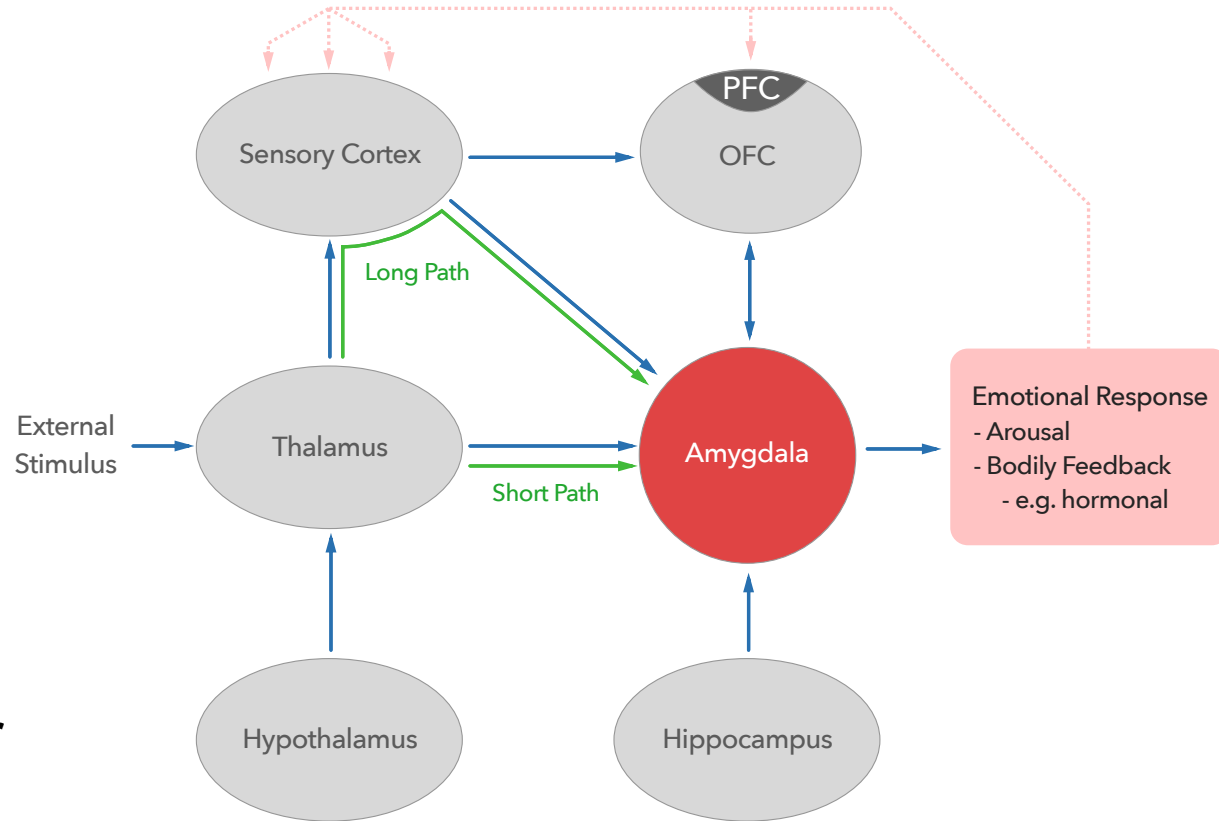
## Neuroscience.

**Complex, intertwined  
Amygdala central to  
emotion**

**Multiple paths, some  
inhibitory e.g. OFC**

*Inhibition is not  
common in typical AI*

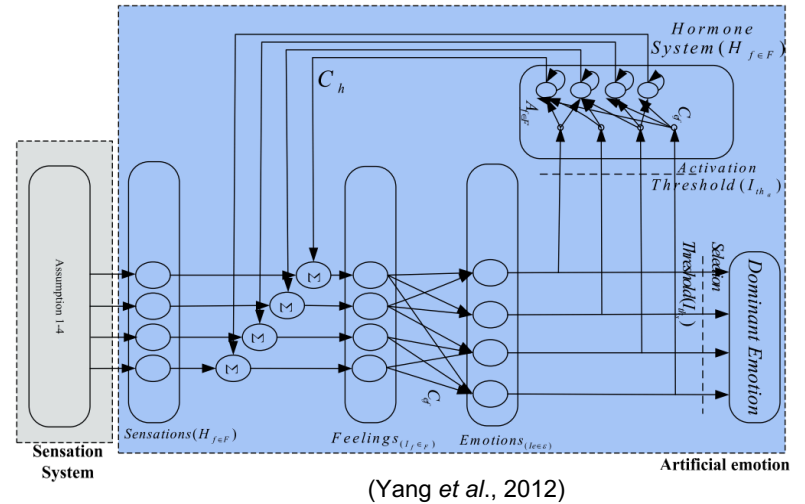
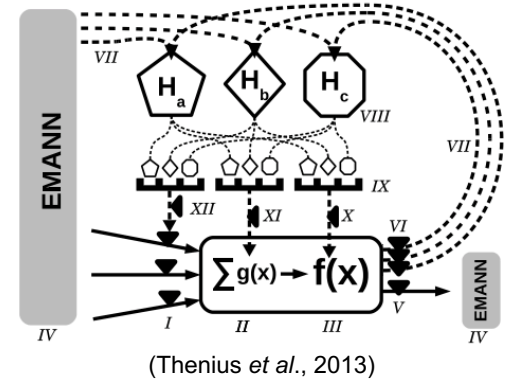
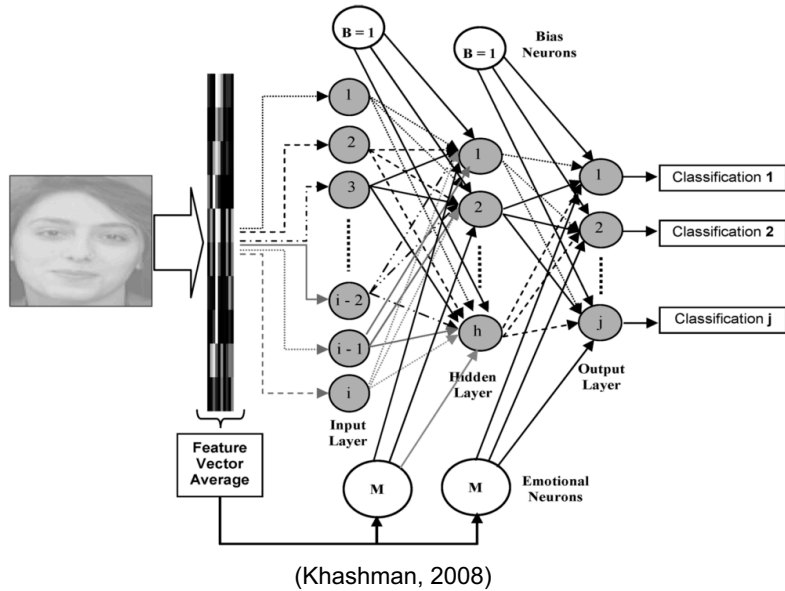
**Changes how you act  
and what you remember**



## Emotional Learning.

1. Optimisation methods
  2. Anatomical Models
  3. Reinforcement learning (many variants)
  4. Cognitive Architectures
-

# Optimisation.





> Making it work.



# I) Self-Learning Representation



# II) Couple Analysis + Synthesis

**CHALLENGE  
ACCEPTED**

III)

Reinforcement  
Learning

**WANTED!**

**IIIa)**

**Learn  
w/ Few**

**IIIb)**

**Big/ger  
Data**

IIIc)

**Self-Supervised**

Confidences Context Adversarial Robustness

Diarisation Separation ...  
Holism

# FULL PACKAGE!

Adaptation eXplanation  
Hiding  
Package Loss Concealment Green Processing

## Reading the Tea Leaves...


**AEI** may soon know you better than...

your spouse, family, psychologist...

Let's be technically prepared and  
responsibly design it **for You and I**





**GLAM + EIHW**  **Thanks!**

Imperial College  
London

## Reading the Tea Leaves...



**AEI** may soon know you better than...

your spouse, family, psychologist...

**Thank You!**

Let's be technically prepared and  
responsibly design it **for You and I**