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 Al.	NG Knowledg	ge tation	Planning
& Probi Solving	A.I		Learning
Perception	Socioemo Intellige	tional nce	NLP
creativ.	General Intelligence	Motion Manip	ulation

Emotions? Survive Motivate Make Decisions Inform yourself Communicate



AEI.



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

> The tech.

End-to-End.

CCC	Arousal	Valence
Audiovisual	.770	.612



"End2You -- The Imperial Toolkit for Multimodal Profiling by End-to-End Learning", arXiv, 2018.



"Evolving Learning for Analysing Mood-Related Infant Vocalisation", Interspeech, 2018.

Imperial College					# Cla	sses	%UA/*AUC/+CC
LUNUUN			Dec	eption		2	/2.1
			Sind	erity		[0,1]	65.4+
Challenge.			Soc	ial Signals		2x2	92.7*
.......			Con	flict		2	85.9
	# Classes	%UA/*AUC/	+CC Emo	otion		12	46.1
Interest	[-1,1]	4	2.8+				
Emotion	5		44.0				
Negativity	2		71.2				
			# Classe	s %UA/*Al	UC/+CC		
	Affect	Atypical		4	45.0		
	Affect	Self-Ass.		3	68.4		
INTER SPEECH	Crying	J		3	78.6		
	Heart	Beats		3	56.2		
	Elderh	y A/V	<u>2x</u>	3	63.8		
	Escala	ation		3	(59.8)		



"Attention-augmented End-to-End Multi-task Learning for Emotion Predicition from Speech", ICASSP, 2019.



"Hierarchical Component-attention Based Speaker Turn Embedding for Emotion Recognition", IJCNN, 2020.

Attention.



w/o

Att.

81.4

86.5

87.8

w/Att.

85.2

88.5

91.0

"Context Modelling using Hierarchical Attention Networks for Sentiment and Self-Assessed Emotion Detection in Spoken Narratives", submitted.

Uncertainty.

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





"Modelling Sample Informativeness for Deep Affective Computing", ICASSP, 2019.





"Dynamic Difficulty Awareness Training for Continuous Emotion Prediction ",

IEEE Transactions on Multimedia, 2019.



"Affect Recognition by Bridging the Gap between End-2-End Deep Learning and Conventional Features", ICASSP, 2018.



"Attention-based Atrous Convolutional Neural Networks: Visualisation and Understanding Perspectives of Acoustic Scenes", ICASSP, 2019. "Attention-augmented End-to-End Multi-task Learning for Emotion Predicition from Speech", ICASSP, 2019.

Confidence.



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



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

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



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



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

"From Hard to Soft: Towards more Human-like Emotion Recognition by Modelling the Perception Uncertainty", ACM Multimedia, 2017.



"Single-Channel Speech Separation with Auxiliary Speaker Embeddings", arXiv, 2019.

Package Loss?





CCC	wlo	CNet
Arousal	W /O	CINEL
Original	.769	-
35% drop	.681	.769
50% drop	.583	.754
85% drop	.564	.676
CCC		
CCC Valence	w/o	CNet
CCC Valence Original	w/o .432	CNet
CCC Valence Original 35% drop	w/o .432 .267	CNet - .420
CCC Valence Original 35% drop 50% drop	w/o .432 .267 .140	CNet - .420 .399



"ConcealNet: An End-to-end Neural Network for Packet Loss Concealment in Deep Speech Emotion Recognition", arXiv, 2020.

Data.



%UA

Freezing Intoxication

Screaming

3 hrs

70.2

72.6 97.0

"CAST a Database: Rapid Targeted Large-Scale Audio-Visual Data Acquisition via Small-World Modelling of Social Media Platforms", ACII, 2017.

Timing.

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



Timing.



Timing.

Methods	IEM	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	73.1	66.3	41.1



[&]quot;Towards Conditional Adversarial Networks for Predicting Emotions from Speech", ICASSP, 2018. (nominated best paper)

Imperial College

Conversion.

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



"StarGAN for Emotional Speech Conversion: Validated by Data Augmentation of End-to-End Emotion Recognition", ICASSP, 2020.

Embedded.

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





"Squeeze for Sneeze: Compact Neural Networks for Cold and Flu Recognition", Interspeech, 2020.

Holism.





Adaptive.







"Synchronization in Interpersonal Speech", Frontiers in Robotics and AI, 2019.

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





"Personalized Federated Deep Learning for Pain Estimation From Face Images", arXiv, 2021.



"Speech analysis for health: Current state-of-the-art and the increasing impact of deep learning", Methods, 2018.

Training Support Modalities.

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



"EmoBed: Strengthening Monomodal Emotion Recognition via Training with Crossmodal Emotion Embeddings", IEEE Transactions on Affective Computing, 2019.

Trainng Support Modalities.

• Mono = Multimodal.

-IS Emotion Challenge task - 2 classes





"EmoBed: Strengthening Monomodal Emotion Recognition via Training with Crossmodal Emotion Embeddings", IEEE Transactions on Affective Computing, 2019.

Lifelong Learning









"Internet of Emotional People: Towards Continual Affective Computing cross Cultures via Audiovisual Signals", Future Generation Computing Systems, 2020.



"Reading the Author and Speaker: Towards a Holistic and Deep Approach on Automatic Assessment of What is in One's Words", CICLing, 2017.

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

"Reading the Author and Speaker: Towards a Holistic and Deep Approach on Automatic Assessment of What is in One's Words", CICLing, 2017.

Ethical, Moral.





"Ethics and Good Practice in Computational Paralinguistics", IEEE Transactions on Affective Computing, 2020.

> 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



"Emotion-Augmented Machine Learning: Overview of an Emerging Domain", ACII, 2017.

Emotional Learning.

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

Optimisation.







> Making it work.

Self-Learning Repsentation AccompLISHED





Learn w/ Few

Big/ger Data

IIIc) Self-Supervised





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



Reading the Tea Leaves...



Thank You!

AEI may soon know you better than....

your spouse, family, psychologist...

Let's be technically prepared and

responsibly design it for You and I