

Empathetic Deep Learning to the Rescue: Speech Emotion Recognition from Adults to Children

Elina Lesyk

Master Thesis Proposal

University of Erlangen-Nuremberg, Erlangen, Germany



Emotions as Indicators of Human States

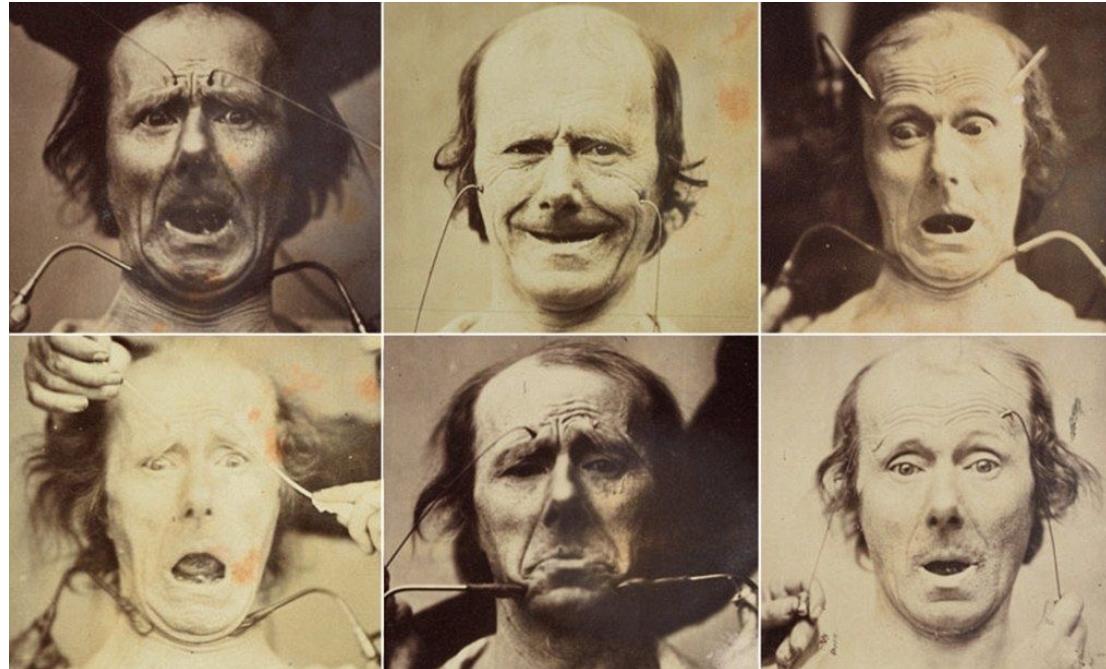


Fig. 1 The Expression of the Emotions in Man and Animals (Darwin 1872)
Taken from Darwin Correspondence Project, University of Cambridge

Emotions as Indicators of Human States

Impact on

- reasoning (Damasio 1994)
- decision making (Lerner et al 2015), stronger than rationality (Häusel 2014; Pispers and Dabrowski 2013)

Measured using



Self-reports



Facial
expressions



Sensor data



Observer
ratings



Vocal tract /
„the mirror to our soul“
(Sundberg 1998)

Emotional States in Theory

- “States elicited by rewards and punishers”, or “by instrumental reinforcing stimuli” (Rolls 2014, p.25)
- In such dimensions as arousal, valence, intensity, duration, frequency and time dimension (Feidakis 2011)

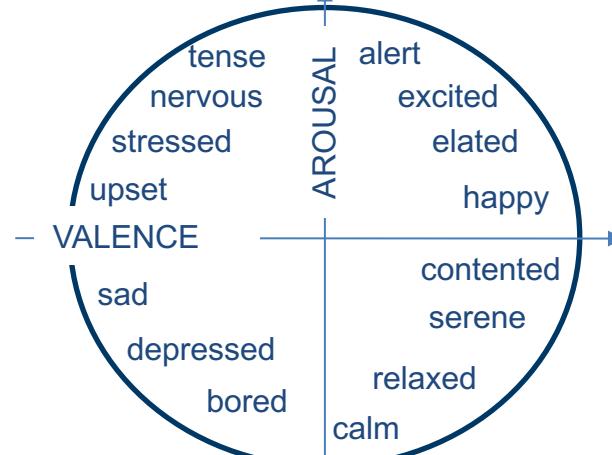


Fig. 2 A Circumplex Model of Affect (Russell 1980)

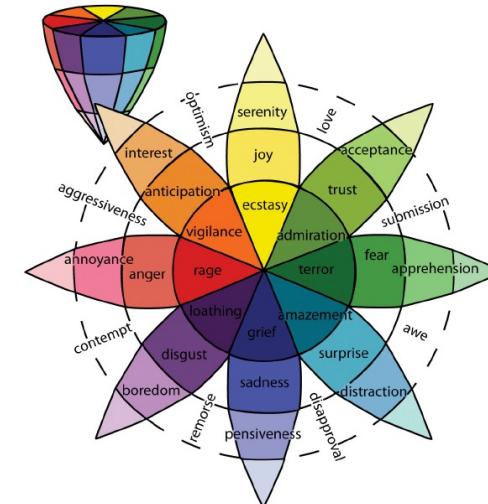
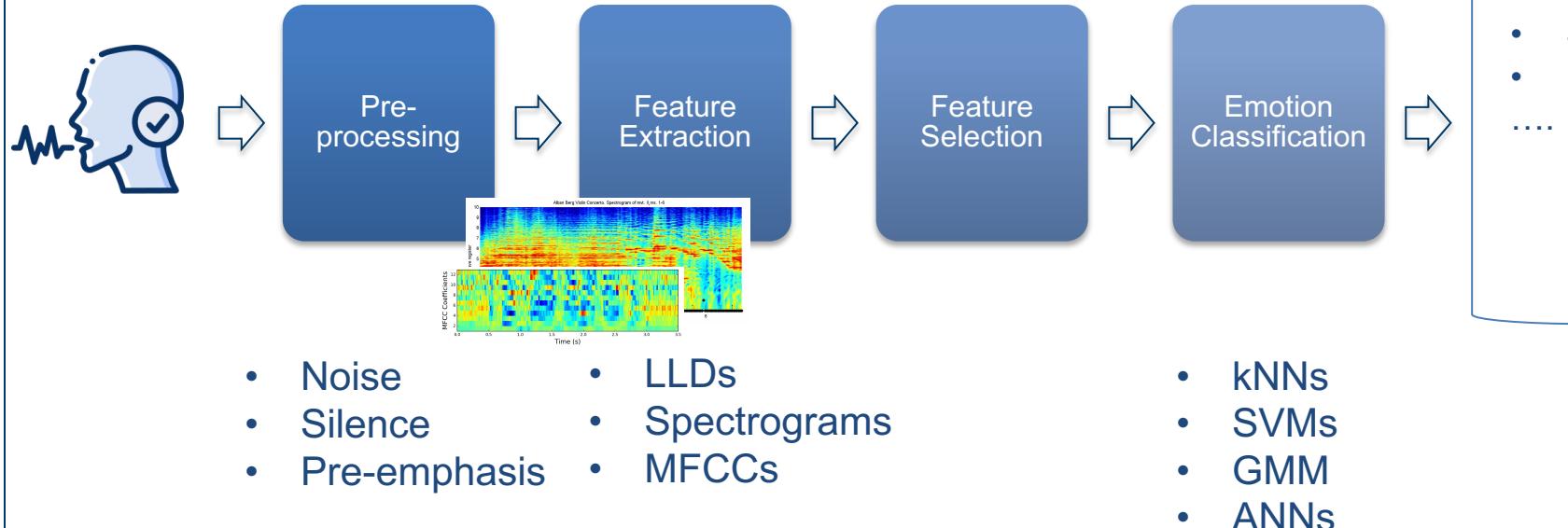


Fig. 3 Wheel of Emotions (Plutchik 1980)

Speech as Facilitator for Detection



Deep Learning as Tool for Emotion AI

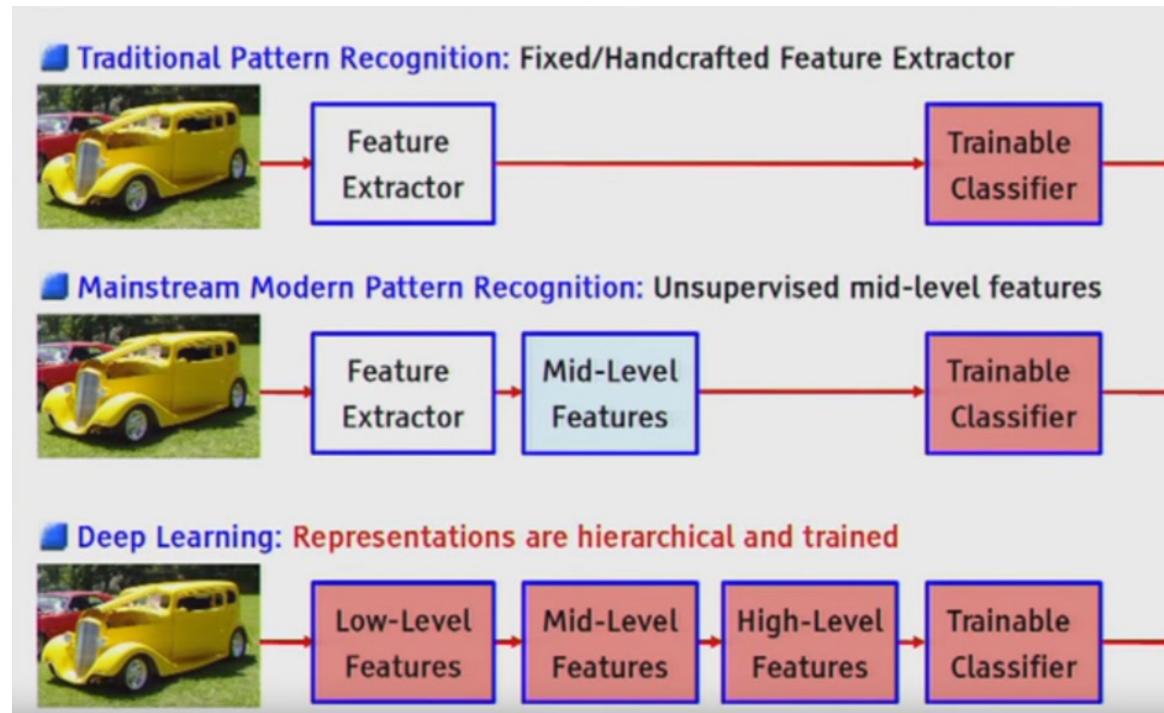


Fig. 4 Deep Learning Benefits (LeCun 2016)

Deep Learning Methods for SER

Reference	Audio Parameters	Techniques	Accuracy
Lee and Tashev 2015	Segment-level features + DNN	ELM	52.13% WA
Xia and Liu 2017	1582 features from IS10 DBN with MTL	SVM	60.9% WA
Tzinis and Potamianos 2017	Statistical features over 3 segments	LSTM	64.16% WA
Satt et al. 2017	STFT spectrograms	CNN–BiLSTM	68.8%WA
Luo et al. 2018	6373 HSFs features, Log-mel spec.	DNN/CRNN	60.35% WA
Ma et al. 2018	Variable length spectrograms	CNN–BiGRU	71.45% WA
Yenigalla et al. 2018	Phoneme embedding and spectrogram	2 CNN channels	73.9%WA
Wu et al. 2019	Spectrograms	CNN–GRU–SeqCap	72.73% WA
Xi et al. 2019	Magnitude spectrograms	Residual Adapter on VoxCeleb2	72.73% WA
Mustaqeem and Kwon 2019	Noise reduction Spectrograms	DSCNN	84% WA

Table 1. Methods used for IEMOCAP dataset (Lieskovska 2021)

Emotion AI State-of-the-Art

<i>Keywords</i>	Emotion AI	Emotion Deep Learning	Speech emotion recognition	Emotion Detection	Emotion Classification	Children speech emotion recognition
2017	49	220	635	658	902	26
2018	96	428	819	839	1208	31
2019	149	679	988	953	1429	40
2020	224	883	994	1085	1566	34
2021	180	692	686	770	1091	25

Table 2. Scopus Results for Keywords

Children in Speech Emotion Recognition



- Children are particularly sensitive to certain emotional wordings (Vesker et al. 2018)
- Children responded differently than adults did to happy and angry expressions (Kestenbaum et al 1992)
- How can speech emotion recognition be applied to different age groups? Are emotional states described differently in children?

Datasets in Use

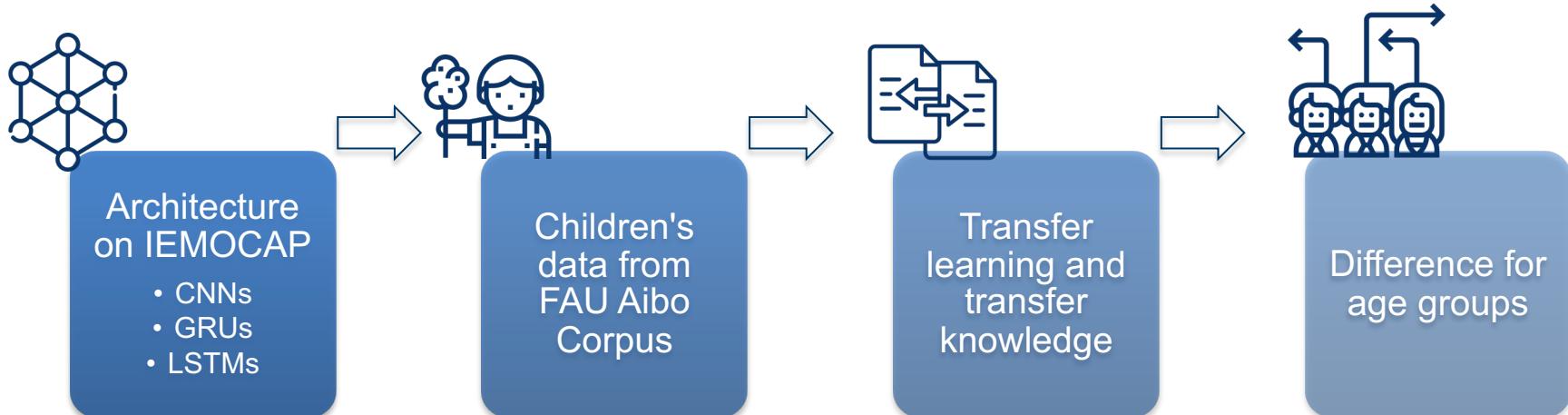
IEMOCAP

- 10039 recordings
- 10 actors (5 male and 5 female)
- Emotions: Anger, happiness, excitement, sadness, frustration, fear, surprise, other and neutral state

FAU Aibo Corpus

- 51 children (10-13 years) with spontaneous German speech; 30 with English
- Emotions: Anger, neutral, motherese, emphatic, joyful

Planned Thesis Outline



Potential Application Prospects



Remote monitoring



Teaching for autism



Health diagnostics

About Me



Thank you for attention!

Glad to answer your
questions.

Elina Lesyk, B. Sc.

- International Information Systems Master's Programme, University of Erlangen-Nuremberg
- elina.lesyk@fau.de

References

- University of Cambridge. (n.d.). *Emotion Experiment | Darwin Correspondence Project*. Retrieved September 14, 2021, from <https://www.darwinproject.ac.uk/commentary/human-nature/expression-emotions/emotion-experiment>
- Damasio, Antonio R. (1994): Descartes' error. Emotion, reason, and the human brain. London, New York: Penguin. Available online at <https://books.google.de/books?id=SLdYHlhqgKMC>.
- Lerner, Jennifer S.; Li, Ye; Valdesolo, Piercarlo; Kassam, Karim S. (2015): Emotion and decision making. In *Annual review of psychology* 66, pp. 799–823. DOI: 10.1146/annurev-psych-010213-115043.
- Häusel, Hans-Georg (Ed.) (2014): Neuromarketing. Erkenntnisse der Hirnforschung für Markenführung, Werbung und Verkauf. 3. Auflage. München: Haufe Lexware Verlag (Haufe Fachbuch)
- Pispers, Ralf; Dabrowski, Joanna (2013): Neuromarketing im Internet. Von der Website zum interaktiven Käuferlebnis. 2. Aufl. s.l.: Haufe Verlag (Haufe Fachbuch). Available online at <http://gbv.eblib.com/patron/FullRecord.aspx?p=1167377>.
- Sundberg, J. (2009). Expressivity in singing. A review of some recent investigations. <Http://Dx.Doi.Org/10.1080/140154398434130>, 23(3), 121–127. <https://doi.org/10.1080/140154398434130>
- Rolls, Edmund T. (2014): Emotion and decision-making explained. First Edition. Oxford: Oxford University Press.
- Feidakis, M., Daradoumis, T., Gr, D., & Caballé, S. (2011). *Emotion Measurement in Intelligent Tutoring Systems: What, When and How to Measure*. <https://doi.org/10.1109/INCoS.2011.82>
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161–1178. <https://doi.org/10.1037/H0077714>
- Plutchik, R. (1980). *Emotion Theory, Research, and Experience*. <https://www.google.de/books/edition/Emotion/-8NBPwAACAAJ?hl=en>
- LeCun, Y. (2016). *Accelerating Understanding: Deep Learning, Intelligent Applications, and GPUs - YouTube*. <https://www.youtube.com/watch?v=Qk4SqF9FT-M>
- Lieskovská, E., Jakubec, M., Jarina, R., & Chmulík, M. (2021). A Review on Speech Emotion Recognition Using Deep Learning and Attention Mechanism. *Electronics 2021, Vol. 10, Page 1163*, 10(10), 1163. <https://doi.org/10.3390/ELECTRONICS10101163>
- Vesker, M., Bahn, D., Kauschke, C., Tschense, M., Degé, F., & Schwarzer, G. (2018). Auditory Emotion Word Primes Influence Emotional Face Categorization in Children and Adults, but Not Vice Versa. *Frontiers in Psychology*, 0(MAY), 618. <https://doi.org/10.3389/FPSYG.2018.00618>
- Kestenbaum, R., & Nelson, C. A. (1992). Neural and behavioral correlates of emotion recognition in children and adults. *Journal of Experimental Child Psychology*, 54(1), 1–18. [https://doi.org/10.1016/0022-0965\(92\)90014-W](https://doi.org/10.1016/0022-0965(92)90014-W)

Empathetic Deep Learning to the Rescue: Speech Emotion Recognition from Adults to Children

Elina Lesyk

Master Thesis Proposal

University of Erlangen-Nuremberg, Erlangen, Germany

