Audio-visual Data Fusion for Behavioral Coding

Sneha Das[1](sned@dtu.dk), Flavia D. Frumosu[1], Nicole N. Lønfeldt[2], A. Katrine Pagsberg[2,3,4] and Line K. H. Clemmensen[1]

[1]Department of Applied Mathematics and Computer Science, Technical University of Denmark, Kongens Lyngby, Denmark
[2]Child and Adolescent Mental Health Center, Copenhagen University Hospital – Mental Health Services CPH, Hellerup, Denmark
[3]Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
[4]Department of Clinical Biochemistry, Hospital Glostrup - University Hospital, Glostrup, Denmark

Introduction

Observation serves as a fundamental instrument for understanding and researching human behavior and mental conditions.

Coding human behavior is a resource-intensive task, with reliability and bias as potential issues.

Machine learning methods can enhance coding reliability, reduce costs, and scale behavioral coding in clinical and research applications [1,2].

Here, we present how to use open-source computational tools to generate codes from commonly used behavioral coding manuals. The focus is on a modular structure that is easy to follow even for a non-technical audience.

Research objectives

- Help mental health professionals to automatize coding human behavior.
- Can the existing pre-trained ML models be used for coding human behavior?
- How can we combine visual and speech signals coming from videos for tracking human behavior?

Challenges

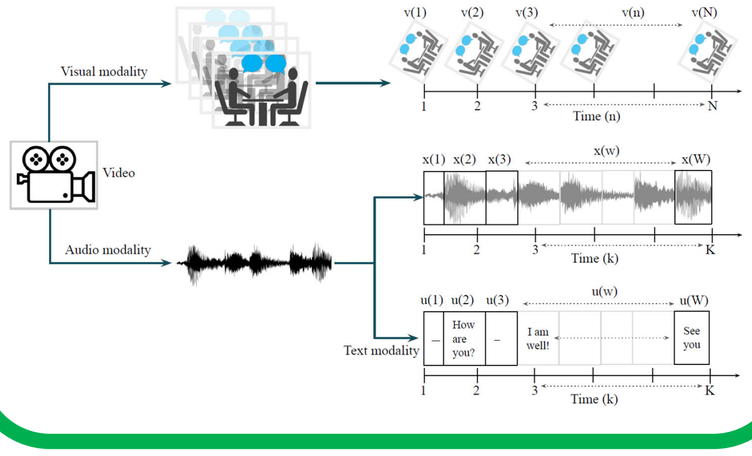
- Data resource constraints.
- Generated outputs from different models are not usually aligned to psychology applications.
- Non-transparency/black-box qualities of deep learning models
- Multi-modal data fusion for non-synchronous sampling.

Video data pipeline

During therapy, the behavior of the child, parent, and therapist is widely relevant within clinical psychology.

Video can be recorded from different angles, but the most important is that the faces and interaction are visible.

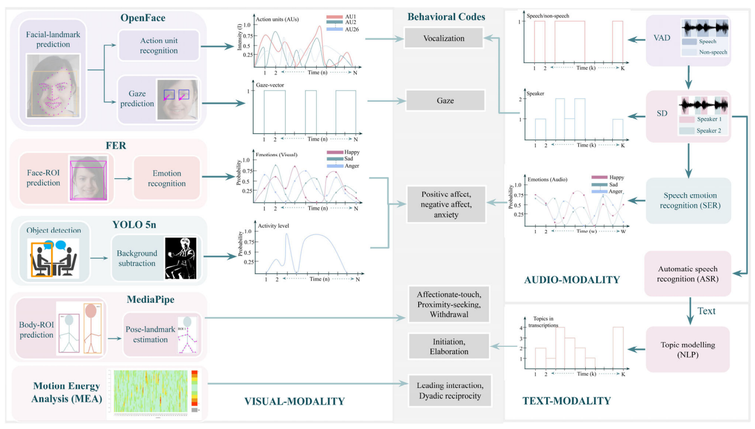
Representation for the visual and signal modalities.



Methodology

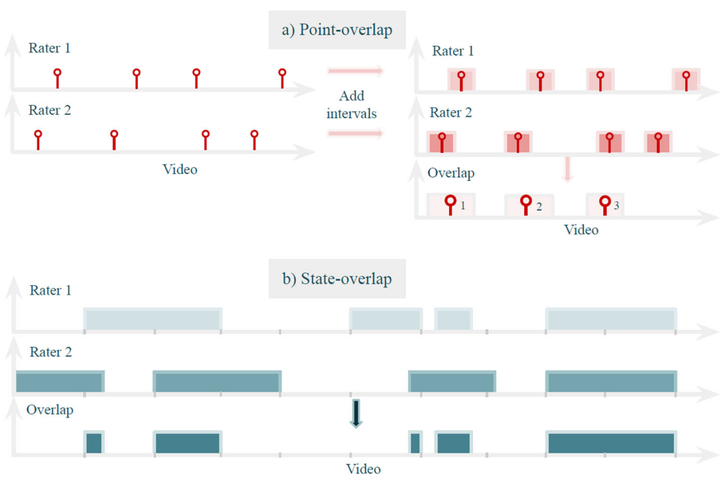
Coding manuals within developmental and clinical psychology provide behavioral codes. The aim is to use pre-trained computer vision and audio models to generate behavioral codes.

A modular approach is employed. Specific behavior codes are mapped via features extracted using pre-trained models.



Metric between raters

Types: **state**, **point,** and **global**.
State events are measuring overlap time duration. **Point events** indicate occurrence number. **Global variables** are measured over the duration of one interaction. The assigned scores are from 1-5, where higher scores are an indication of higher duration, frequency, and behavior intensity.



Results

Behavioral code	Modality			Type
	Visual	Audio	Text	
Child, parent, therapist (Individual codes)				
Positive affect[a]	FER	SER		State
Vocalization[b]	OpenFace	SD		State
Initiation [a]	–	ASR	TM	Point
Elaborating [a]	–	ASR, SD	TM	Point
Child + parent (Individual codes)				
Gaze[b]	OpenFace	–	–	State
Sadness[a]	FER	SER	SA	State
Anger[a]	FER	SER	SA	State
Anxiety[a]	Background subtraction, FER	SD, SER	SA	State
Affectionate touch[a]	MediaPipe, FER	–	–	State
Proximity-seeking[c]	MediaPipe	–	–	State
Withdrawal[c]	MediaPipe	–	–	State
Child (Individual code)				
Avoidance[c]	OpenFace, MediaPipe	–	IM	State
Approach[c]	OpenFace, MediaPipe	–	IM	State
Parents + therapist codes (Individual codes)				
Praising	–	–	SA, IM, TM	State
Validating	–	–	SA, IM, TM	State
Accommodating[c]	MediaPipe	–	SA, IM	State
Dyadic relation (Interaction codes)				
Leading interaction[b]	MEA	SD	TM	Global
Dyadic reciprocity[a]	MEA	SD	TM	Global

FER: Facial Expression Recognition. **SER**: Speech Emotion Recognition. **MEA**: Motion Energy Analysis. **SD**: Speech Diarization. **ASR**: Automatic Speech Recognition. **TM**: Topic Modelling. **SA**: Sentiment Analysis. **IM**: Intent Modelling. Gray is optional inclusion. [a]Codes from[3].[b]Codes defined by authors. [c]Codes from [4].

Conclusions

In this work, we propose different computer vision and audio pre-trained models for behavior coding.

For validation with human raters, we propose a new metric system.

Future work includes validation on clinical data sets.

Acknowledgements

The work is funded by the Novo Nordisk Foundation (grant number: NNF19OC0056795) via the Wrist Angel Project.

Find out more

Bibliography

1. Lønfeldt, N. N., Frumosu, F. D., Mora-Jensen, A. R. C., Lund, N. L., Das, S., Pagsberg, A. K., & Clemmensen, L. K. (2022). Computational behavior recognition in child and adolescent psychiatry: A statistical and machine learning analysis plan. arXiv preprint arXiv:2205.05737.

2. Lønfeldt, N. N., Das, S., Frumosu, F.D., Mora-Jensen, A. R. C., Pagsberg, A. K., & Clemmensen, L. (2023). Scaling-up Behavioral Observation with Computational Behavior Recognition.

3. R. Feldman, Coding interactive behavior (CIB) manual, Unpublished manuscript. Bar-Ilan University (1998).

4. K. G. Benito, C. Conelea, A. M. Garcia, J. B. Freeman, CBT specific process in exposure-based treatments: Initial examination in a pediatric OCD sample, Journal of Obsessive-Compulsive and Related Disorders 1 (2) (2012) 77–84.