Mental Phenotyping with 3D Cameras

Daphna Weinshall

Hebrew University

Research Question

- **Background**: Non-verbal cues such as facial expressions, voice and body movement are essential for the diagnosis of mental/neurological disorders
- **Our goal:**
  - **Quantitatively** characterize **non-verbal** behavior in patients (Parkinson & Schizophrenia)
  - Develop **automatic** tools for describing and analyzing clinically relevant measures of these conditions

ICRI-CI
Intel Collaborative Research Institute Computational Intelligence

Non-verbal cues

- *Speech prosody*: Schizophrenic patients may have “indifferent tone”, depressed patients may show small variability
- *Facial expressions*: Schizophrenic patients may show “flat affect”, with often inappropriate facial expressions; Parkinson patients may show “mask face”
- *Body gestures*: various aspects of Parkinson disease are judged based on the patient’s movement

Technological Solution

- Use 3D cameras to track body and facial movement
- Start from technology developed primarily for gaming (e.g., MS kinect camera)
- Advance the technology to deal with abnormal behavior and the characterization of facial expressions

Non-verbal cues

- *Speech prosody*: Schizophrenic patients may have “indifferent tone”, depressed patients may show small variability
- *Facial expressions*: Schizophrenic patients may show “flat affect”, with often inappropriate facial expressions; Parkinson patients may show “mask face”
- *Body gestures*: various aspects of Parkinson disease are judged based on the patient’s movement

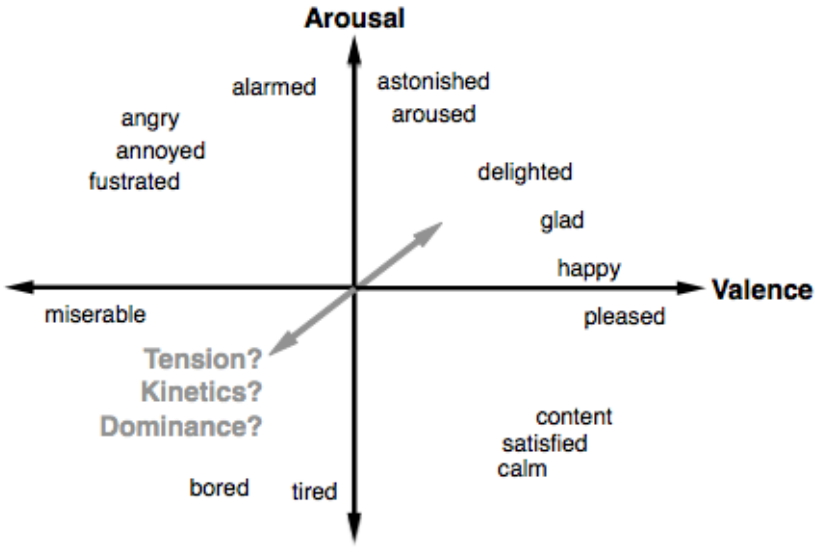
Facial Expressions - Theory

- Prototype Emotions

Categorial Vs. Dimensional approach

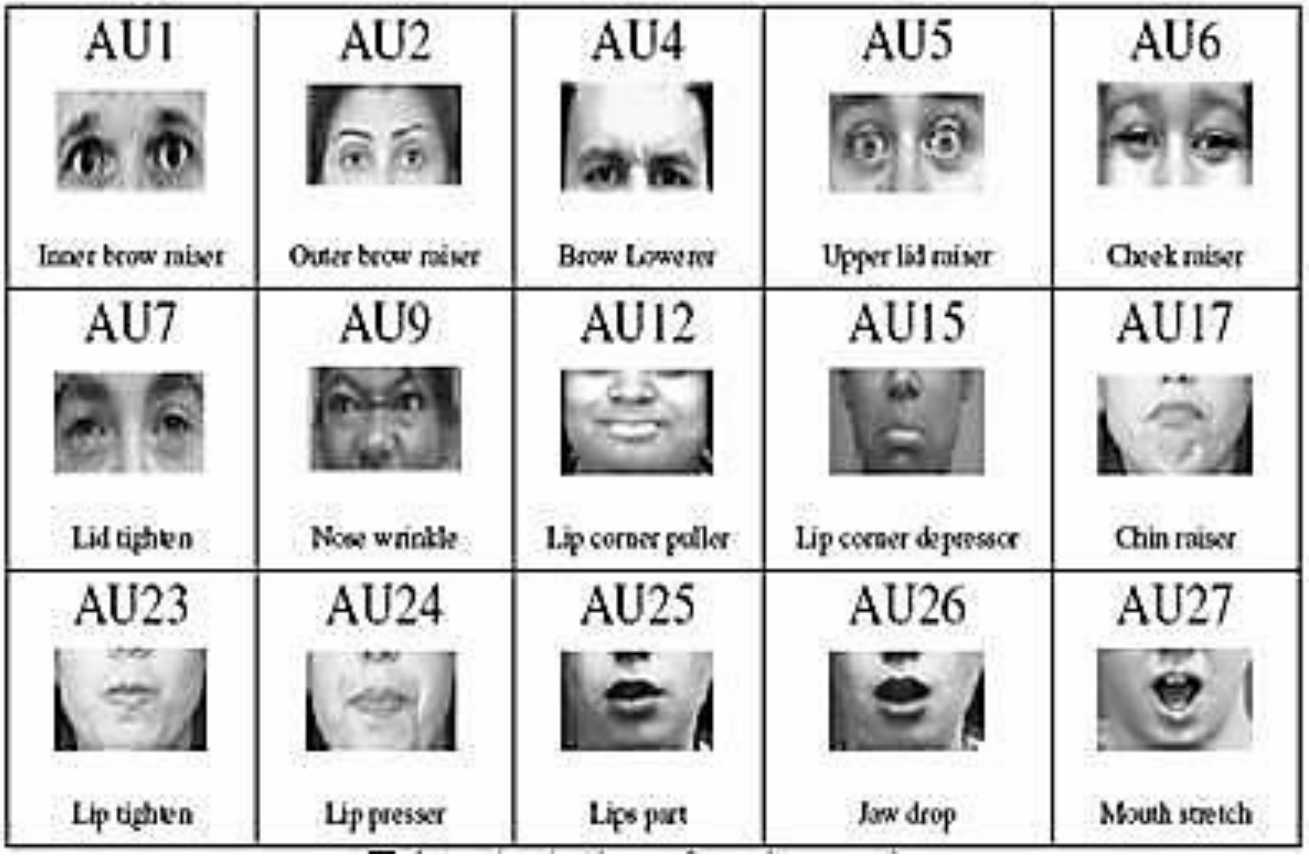
Ekman (1971)

- Dimensional



Russel (1980)

Facial Action Coding System (FACS)

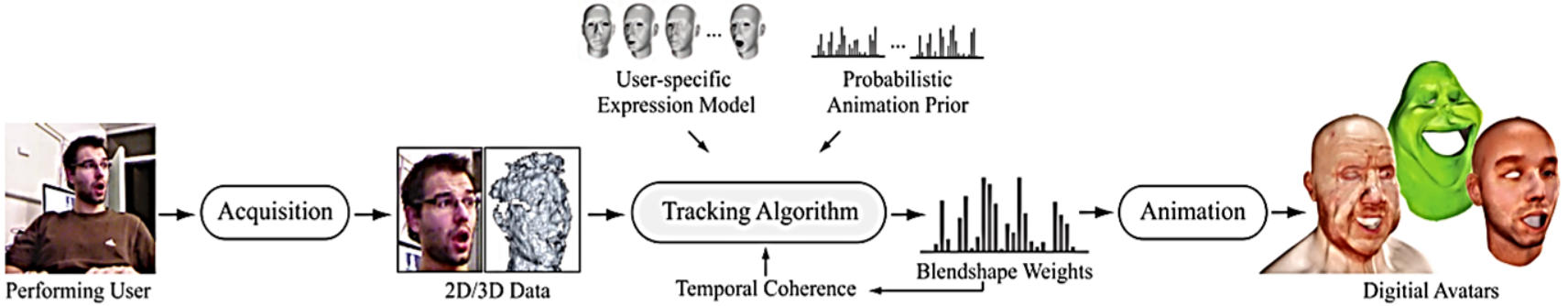


Paul Ekman and Erika L Rosenberg. What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS). Oxford University Press, 1997.

Technology

- Structured light 3D camera (Carmine 1.09)

- AUs extraction out of 3D video (*Faceshift*™)

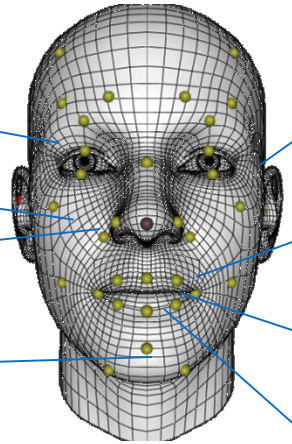


faceshift

Basic features

Faceshift™ returns 4 output types:

1. Intensity level of 51 Action Units over time, with sampling rate of 19 Hz



2. Eye gaze and position
3. 3D head coordinates
4. 3D position of facial markers

Facial Expressions & Parkinson

Collaborators:

- Nomi Vinokurov
- Prof. Hagai Bergman, Dr. David Arkadir

Hypomimia (mask face)

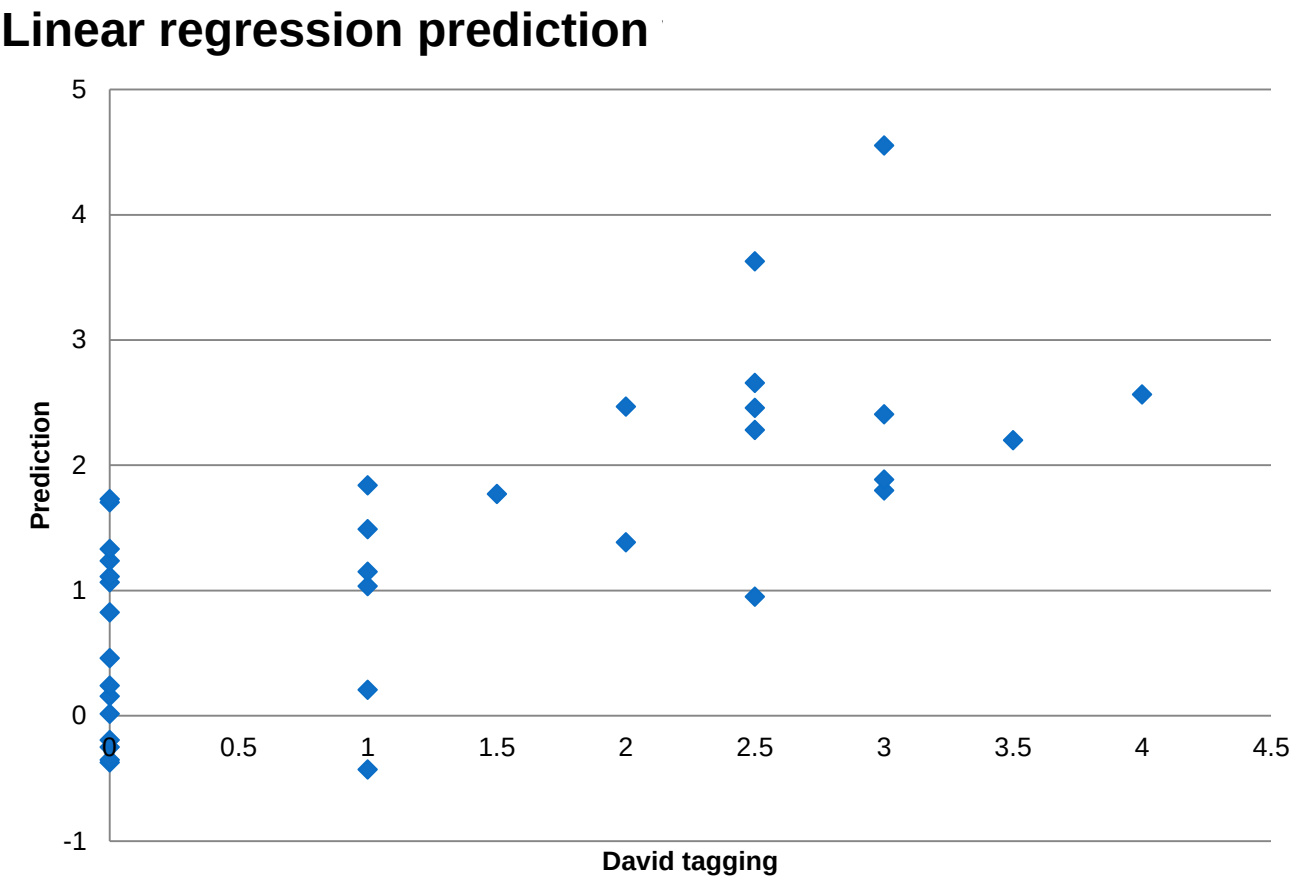
- Reduced degree of facial expression, reduced facial mobility
- Often early sign of Parkinson

Data Collection

- 17 Parkinson patients
- 18 healthy controls
- Hypomimia independently evaluated by 2 neurologists
- Protocol includes: short interview, photo and video presentations, reading session

Evaluation Algorithm

- Derived features: 4 moments of AUs, bilateral differences (left and right), measures of AU's change and speed
- Discriminative feature selection down to 4 features
- Fit linear regression classifier to training sample
- Predict hypomimia level for LOO subject



- Pearson coefficient of predictor: 0.73
- Pearson coefficient between neurologists: 0.8

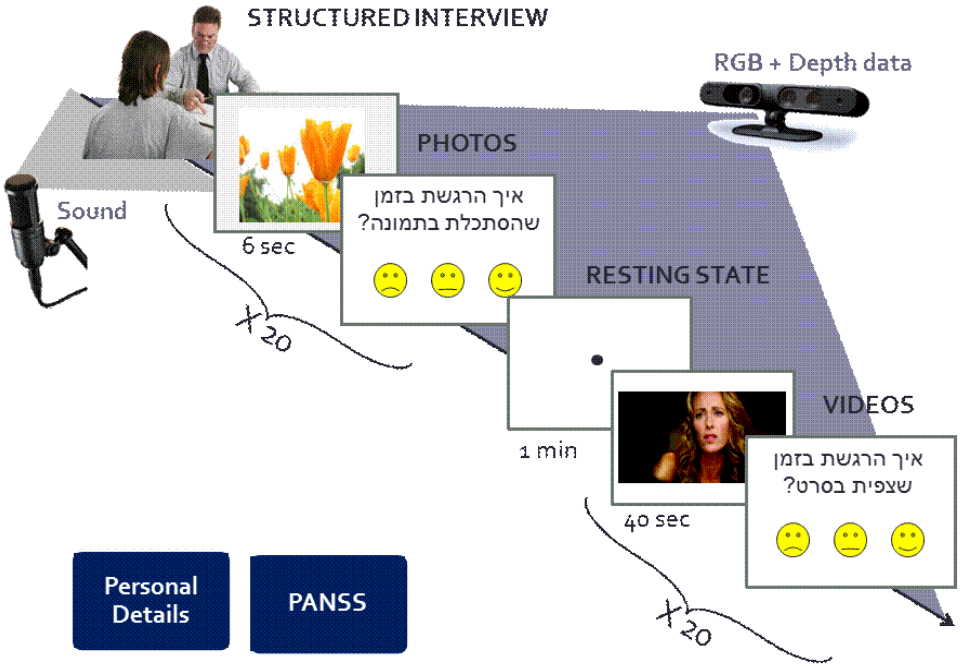
Facial Expressions & Schizophrenia

Collaborators:

- Talia Tron
- Prof. Avraham Peled (Shaar Menashe mental hospital)

Data Collection

- 34 Schizophrenic patients
- 34 healthy controls
- Evaluated by 2 psychiatrists



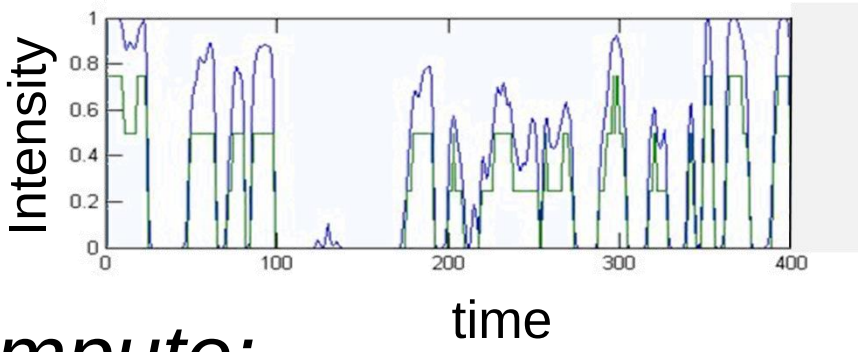
Prediction of PANSS

Goal: Learn to predict the Positive And Negative Syndrome Scale of Schizophrenia

- *Positive symptoms* - an excess or distortion of normal functions (e.g., hallucination and delusions)
- *Negative* symptoms - a diminution or loss of normal functions (e.g., blunted affect and emotional withdrawal)
- *General symptoms* - depression, motor retardation, anxiety, etc

Analysis, pre-processing

- Selection of discriminative AUs, followed by quantization



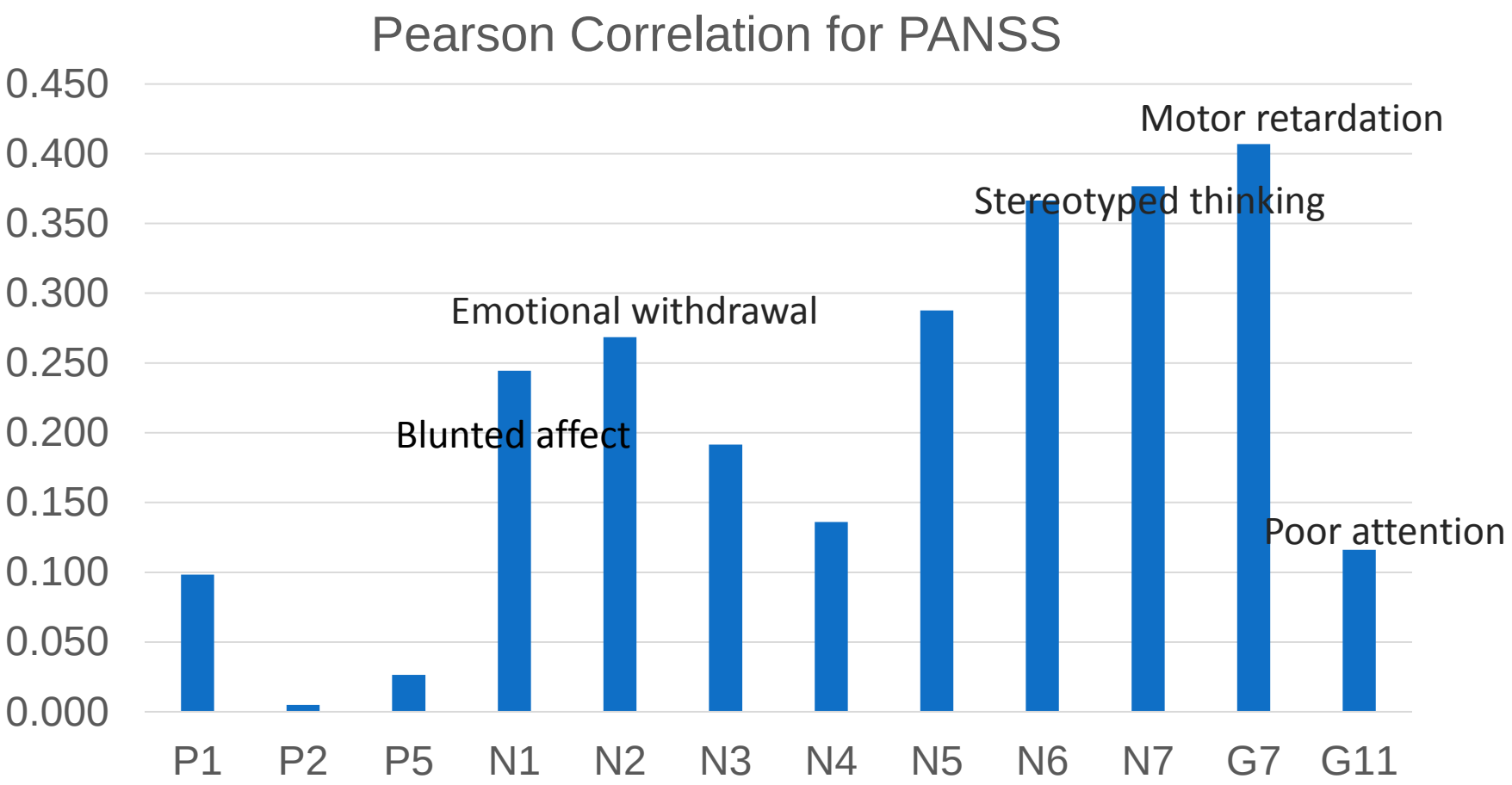
- For each *AU compute:*
  1) Expression Ratio (activation percentage)
  2) Expression Level
  3) Expression Length
  4) Change Ratio (how often the activation changes)
  5) Fast Change Ratio

Analysis

- Features are calculated for each segment (~30 seconds) for each subject
- First training phase – learn prediction score for each segments
- Second training phase – learn prediction score based on moments of the predictor distribution over segments

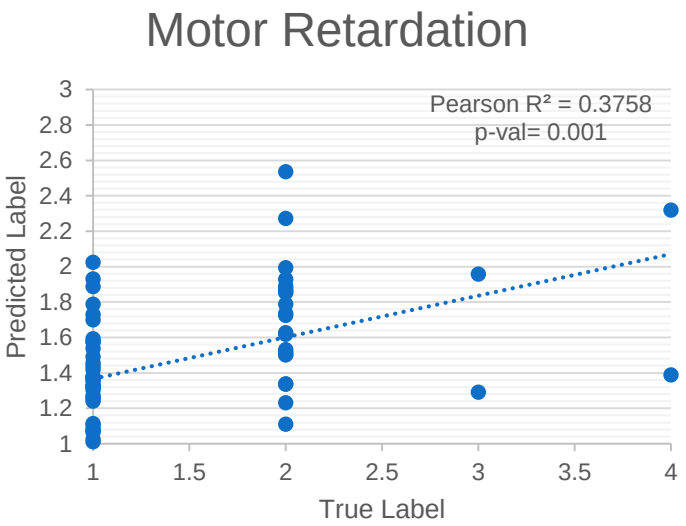
[Training using between subjects LOO cross validation]

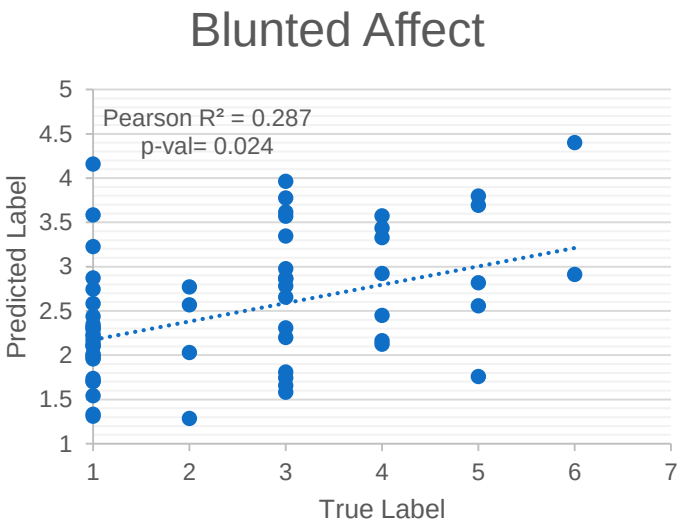
Results


Pearson Correlation for PANSS
0.450
0.400
0.350
0.300
0.250
0.200
0.150
0.100
0.050
0.000
Motor retardation
Stereotyped thinking
Emotional withdrawal
Blunted affect
Poor attention
P1
P2
P5
N1
N2
N3
N4
N5
N6
N7
G7
G11

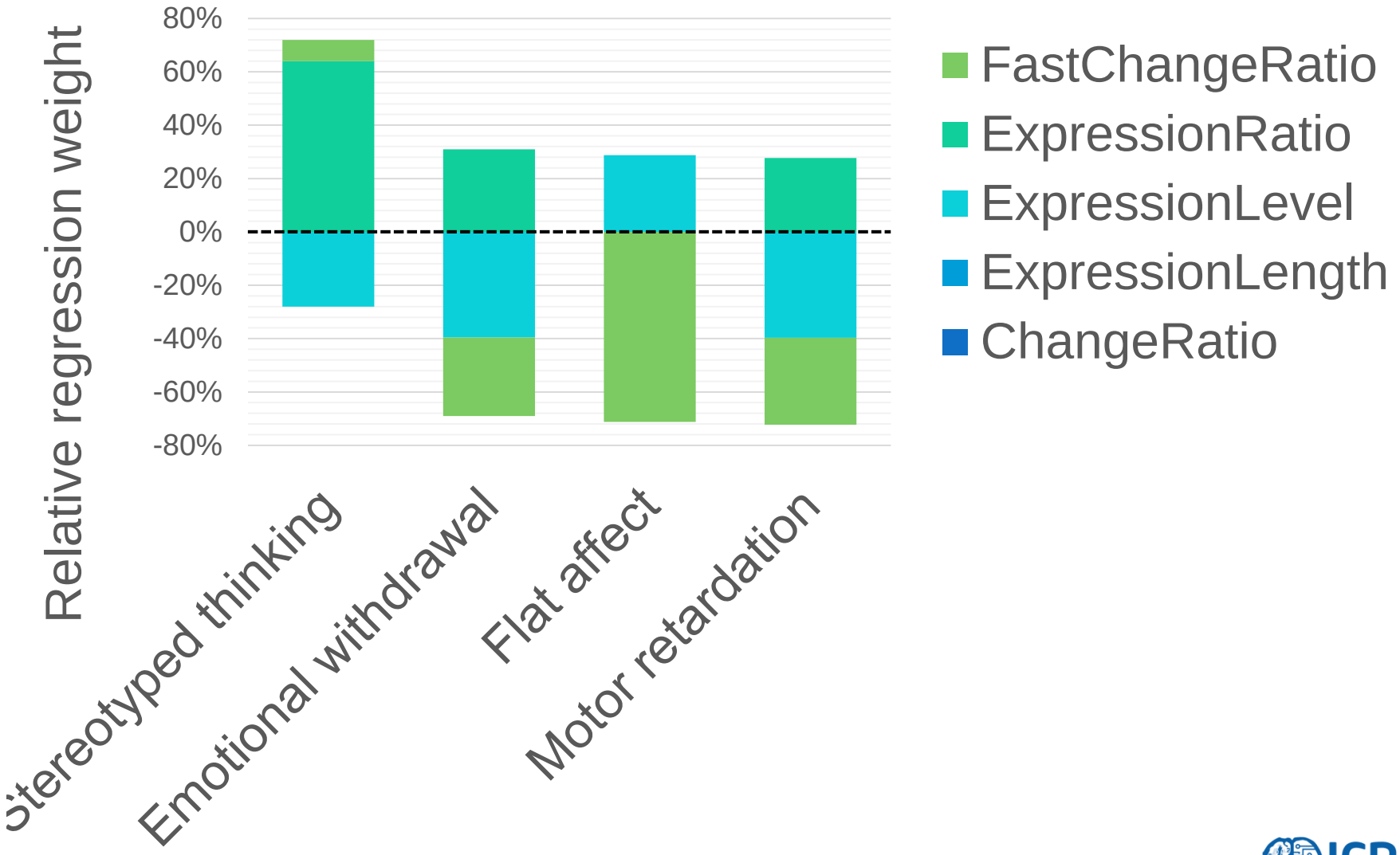
Results

	Motor retardation	Blunted Affect
test Correlation	R^2=0.380 (p<0.01)	R^2=0.276 (p<0.05)
train Correlation	R^2=0.609 (p<<0.001)	R^2=0.601 (p<<0.001)



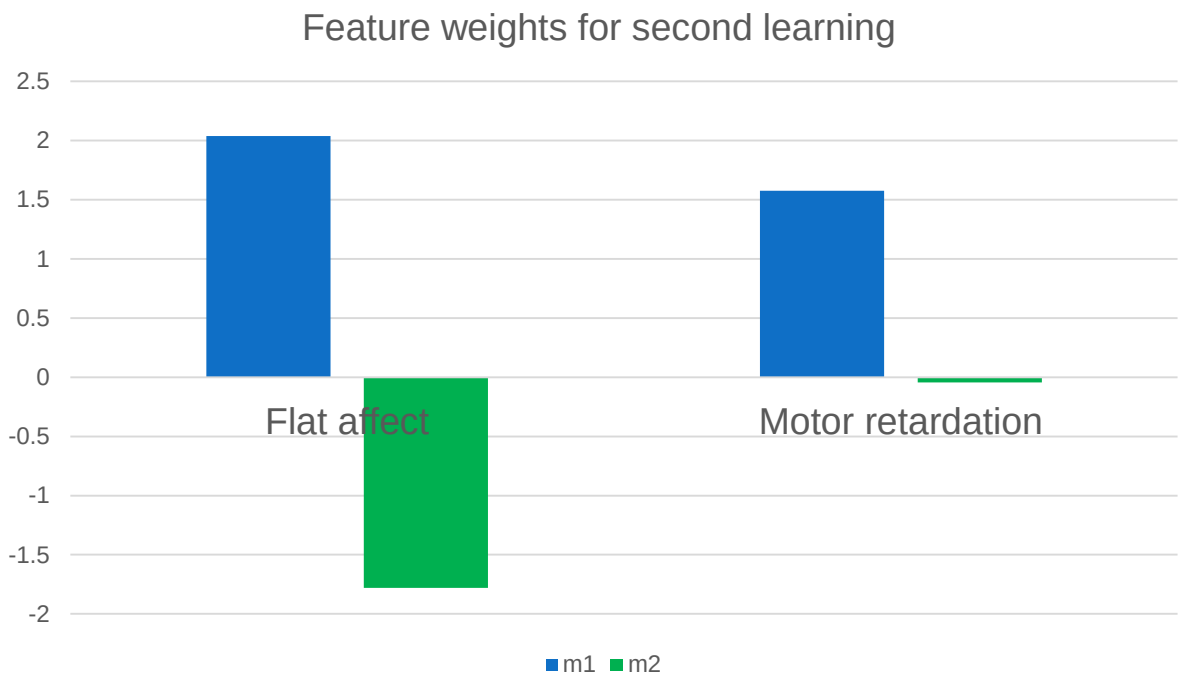


Feature Selection & observed behavior



Preliminary results

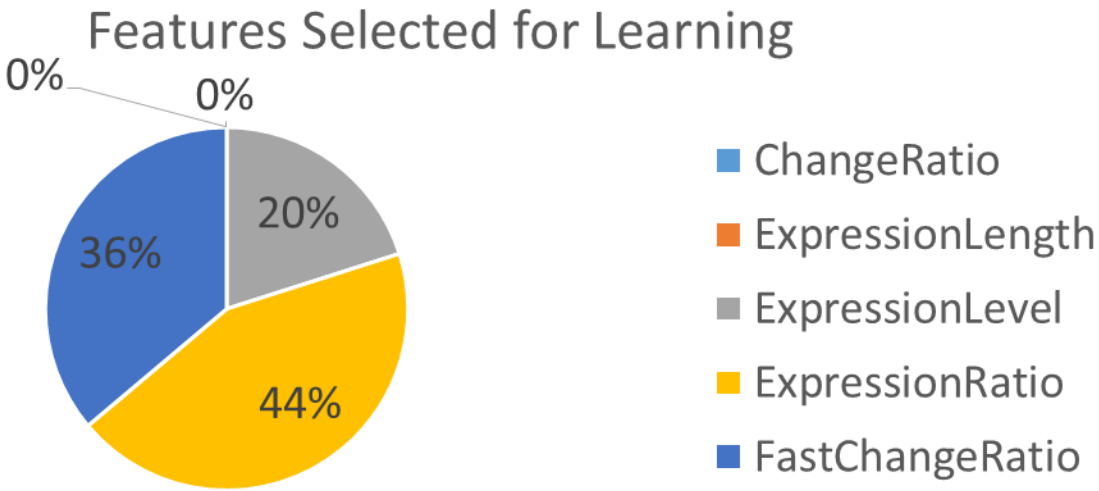
Second phase of learning: as expected, variance in time in the facial AUs is also useful for PANSS prediction



Preliminary Results

Patients Vs. Controls:

specificity	0.74
sensitivity	0.64
accuracy	0.69
ss_mean	0.69



ICRI-CI
Intel Collaborative Research Institute Computational Intelligence

Future work

- Integrate improved camera technology (better 3D resolution, better depth range), and improved software for facial 3D feature extraction
- Big data acquisition and analysis
- Move up to more semantic description of facial expressions (e.g., valence)
- Integrate additional cues, like voice prosody