COMPONENTS FOR HUMAN-MACHINE INTERACTION

FENECH KRISTIAN, NIPG
ELTE IK, MESTERSÉGES INTELLIGENCIA TANSZÉK

Outline

- Motivations and Use cases
- Hand pose and gesture estimation
- Eyegaze
- Personality Estimation

Motivation and use cases

- Medical therapy
- Human-Machine Interaction
- Personality Estimation
- Persuasive computing

Medical therapy

- AI for Autism (Argus and Rush Medical College), PTSD
  - AI systems developed to aid the clinician in the diagnosis procedure
  - Similar systems could be developed to assess improvement of condition after therapy

- Guided physical rehabilitation (HumanE-AI)
  - Accurate body pose information can be used to assess exercise performance
  - Dialogue systems to engage and instruct the user

Example: Gesture recognition for autism diagnosis

Frame Number

Get/Grab/Take   Put   Brush
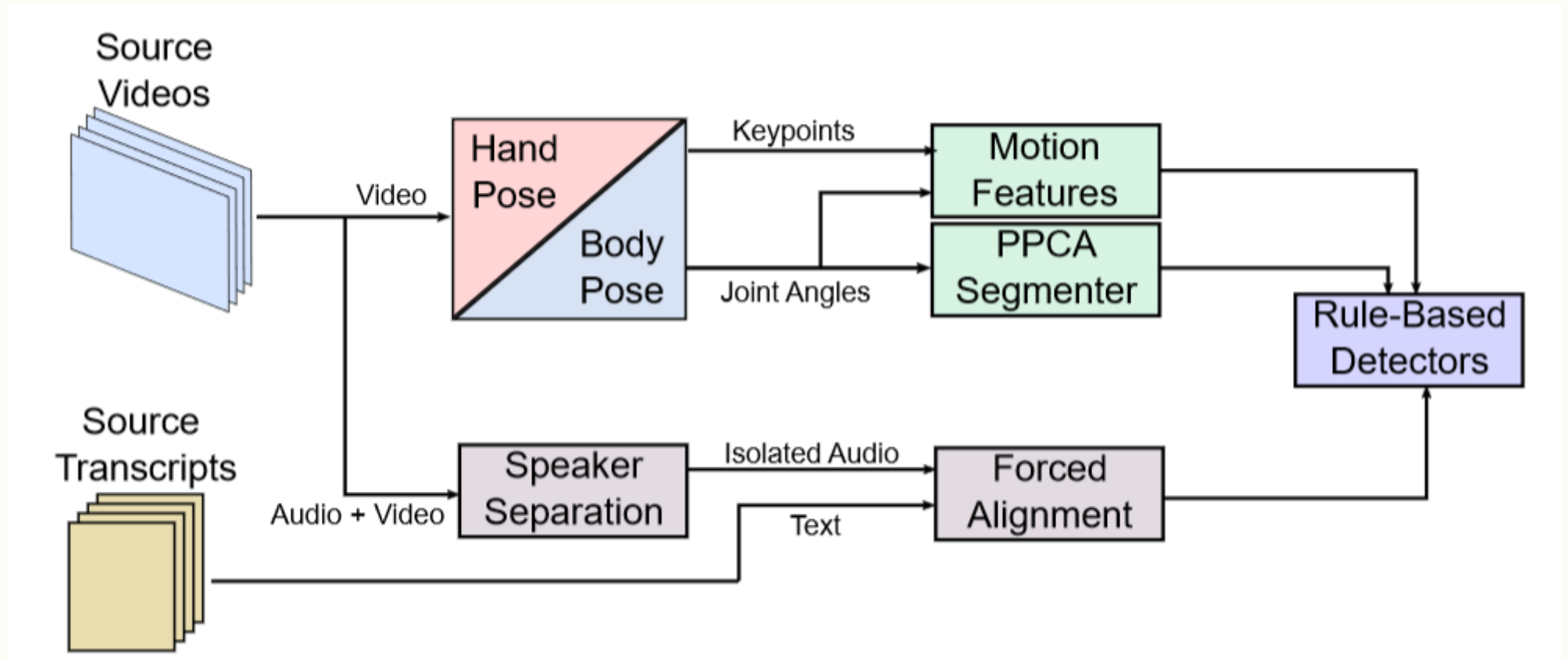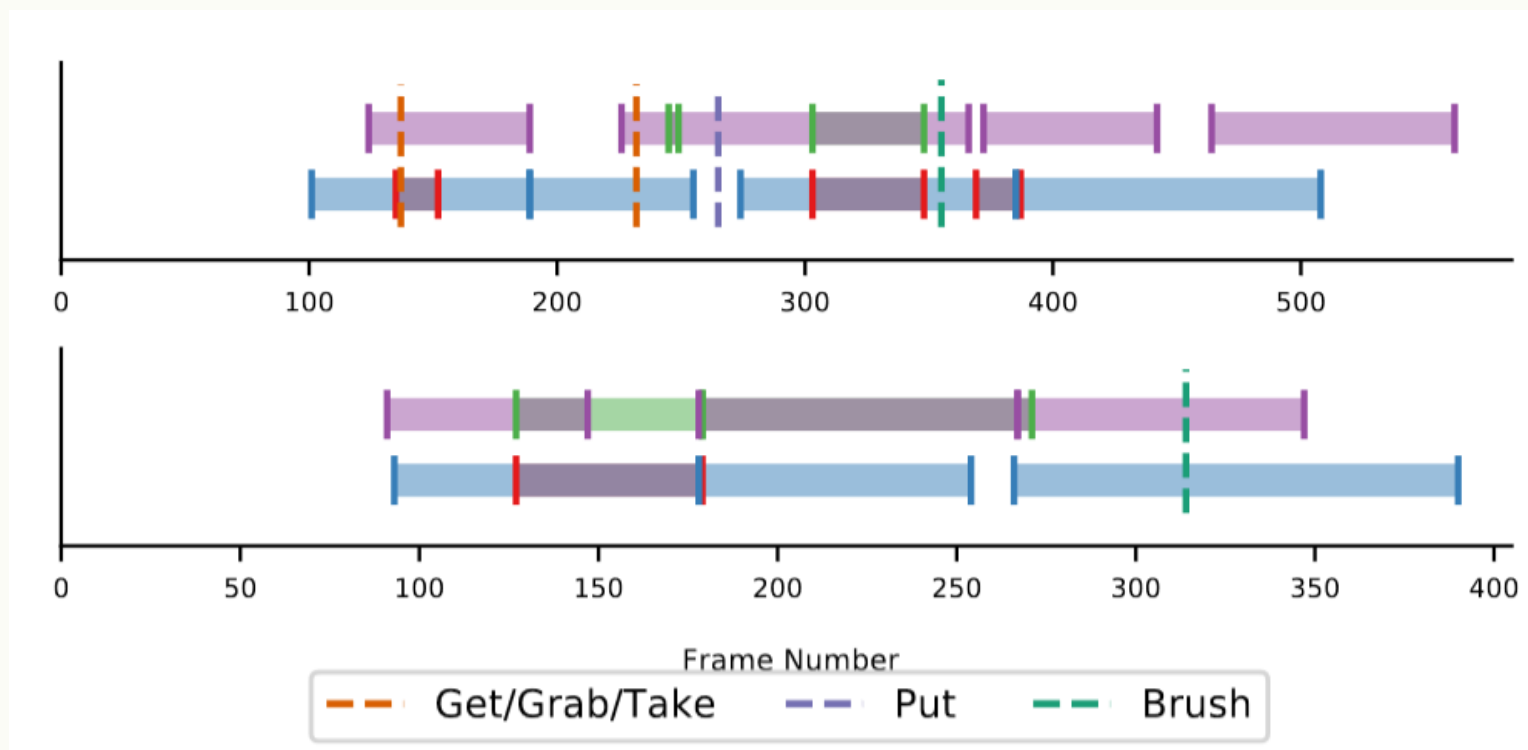
Results

Gesture	With Keyword			Without Keyword		
	Precision	Recall	F_1	Precision	Recall	F_1
Brush	0.286	0.364	0.32	0.25	1.0	0.4
Put	0.40	0.266	0.32	0.083	0.5	0.142
Grab/Get/Take	0.125	0.111	0.118	0.1	0.2	0.133

Next steps on Autism

- Care needs to be taken when determining which video segments can be analysed. Even with state of the art methods for body and hand pose, video resolution, motion blur and occlusion are significant bottlenecks

- Current analysis only performed on children with Autism. Unclear if low F1 score is due to poor detector function, or poor representation of gestures by children with autism

- Analysis and comparison on Typically Developing children ongoing with Konya Leon

Gaze Estimation

- Gaze estimation provides useful information on social interactions [1]
  - Focus of attention, engagement, communication, indication of turn-taking
- Two forms of gaze estimation needed depending on the scenario
  - Screen-based (2D) point of gaze
  - 3D gaze estimation

[1] Cañigueral Roser, Hamilton Antonia F. de C. *The Role of Eye Gaze During Natural Social Interactions in Typical and Autistic People,* Frontiers in Psychology Vol 10 2019

- 2D point of gaze estimation evaluated with commercial software

- Software was used in the MemoryGame HumanE-AI project with TU Delft.

- Given a video of a recorded zoom session, determine which participant a user is looking at

- No knowledge of user parameters (Screen Size, Camera pose, Camera Intrinsics)

- Pixel level accuracy not needed

- 3D gaze estimation with current state of the art method ETH-X Gaze [2]

- Tested on the ELEA dataset, with known camera intrinsics and approximate camera pose provided.

[2] Xucong Zhang and Seonwook Park and Thabo Beeler and Derek Bradley and Siyu Tang and Otmar Hilliges. *ETH-XGaze: A Large Scale Dataset for Gaze Estimation under Extreme Head Pose and Gaze Variation,* European Conference on Computer Vision (ECCV) 2020

Group personality estimation

- Groups have been shown to coalesce in terms of affect [3]
- We have begun to explore if a similar group personality emerges
- First experiments with the ELEA corpus
  - Collaborative exercise
  - 3 or 4 member groups
 - Participants discuss which items to bring in a plane crash survival scenario

[3] Group Affect, Sigal G. Barsade and Andrew P. Knight, Annual Review of Organizational Psychology and Organizational Behavior 2015 2:1, 21-46
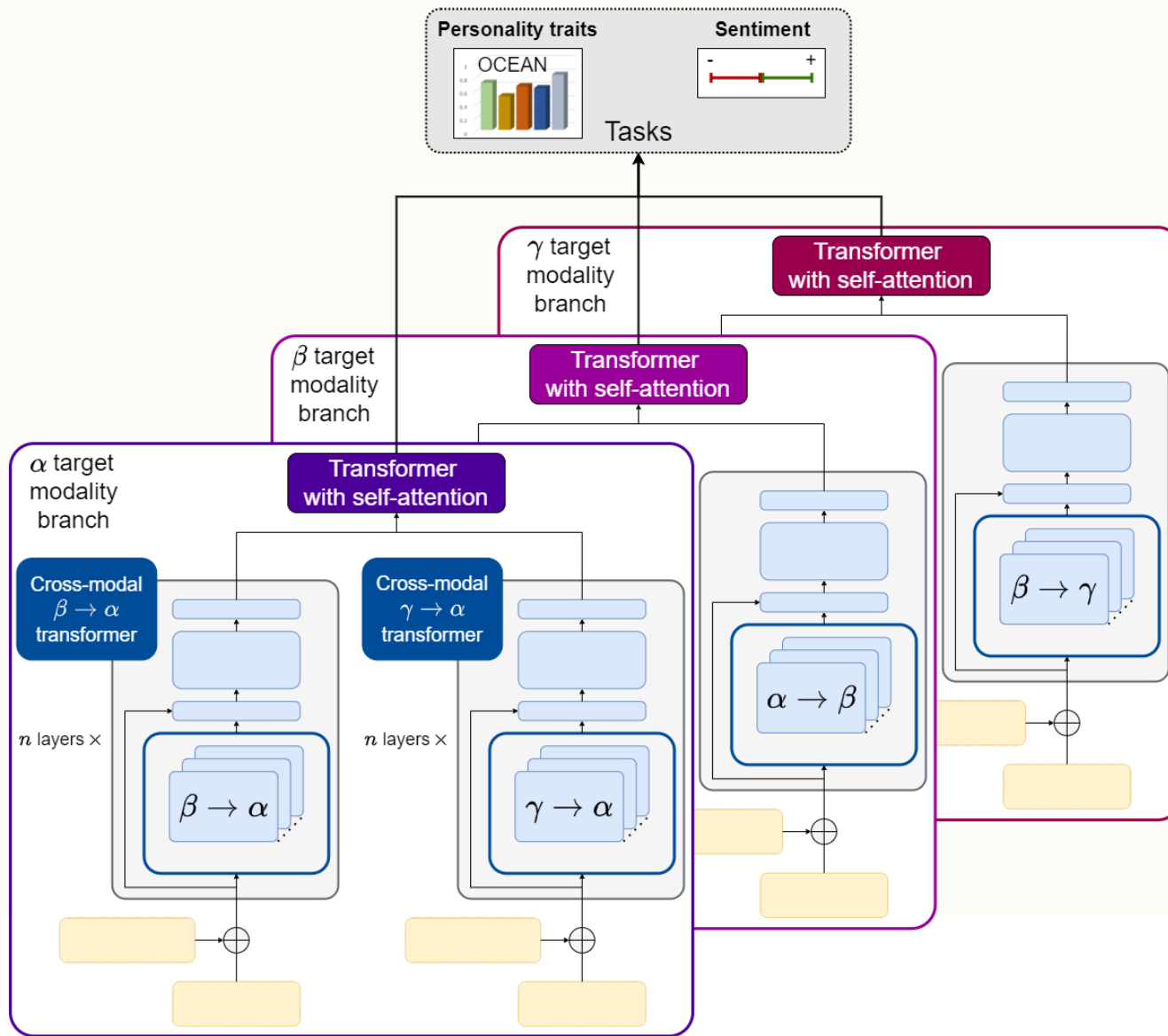
- Individual personality can be represented by 5 traits
  - **O**penness, **C**onscientiousness, **E**xtraversion, **A**greeableness, **N**euroticism (OCEAN)
- Traits predicted from video with a multi-modal transformer (Video, Audio, Text)
- Model trained on ChaLearn First Impressions V2 dataset [4]

[4] Escalante, H. J.; Kaya, H.; Salah, A. A.; Escalera, S.; Gucluturk, Y.; Guclu, U.; Baró, X.; Guyon, I.; Jacques Junior, J. C. S.; Madadi, M.; Ayache, S.; Viegas, E.; Gurpinar, F.; Wicaksana, A.S.; Liem, C.C.S.; van Gerven, M. A. J.; van Lier, R. "Modeling, Recognizing, and Explaining Apparent Personality from Videos," IEEE Transactions on Affective Computing (TAC), 2020.

- We examine the Big 5 traits in a reduced dimensionality the Big 2
  - Stability (Agreeableness + Conscientiousness + (-1*Neuroticism))
  - Plasticity (Extraversion + Openness)
- Initial evaluations on the ELEA corpus
- Traits calculated over a 3 second sliding window
- Personality traits show the average over the full session
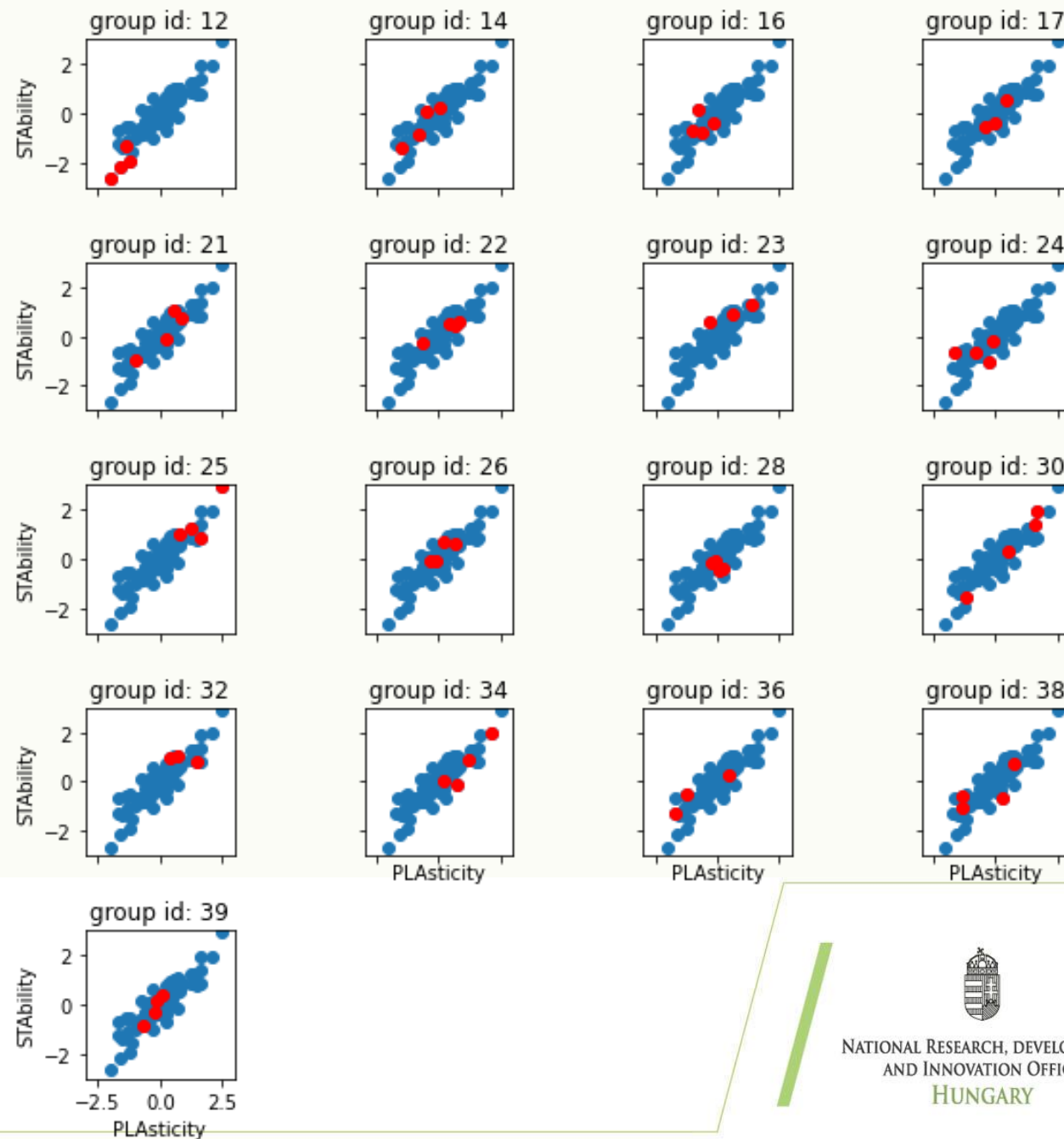
Group personality dynamics
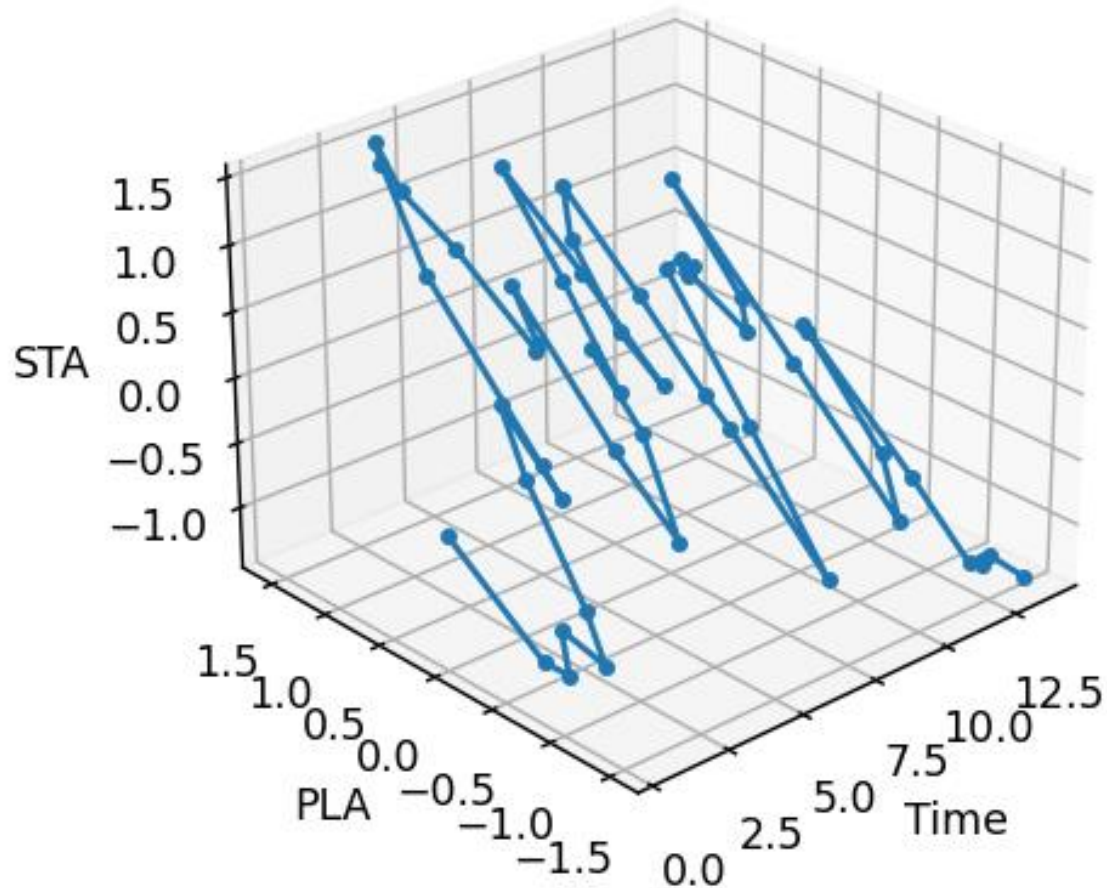
- Speaker interruptions may factor into the dynamics

- ELEA dataset is of limited size. Further analysis of the AMI dataset is considered to check if such dynamics can also be detected

- Initial analysis of the data finds that interruptions may not be relevant to all groups.

- Effect size may depend on the group personality

- Extension of interruptions to include 'back channels' is planned

- Robust Human-Machine Interaction needs understanding both verbal and non-verbal communication methods

- These technologies can be utilised in a wide variety of industries: therapy, rehabilitation, situational awareness, sales

- Hand gestures, body language and gaze provide important communicative signals

- Knowledge of personality and personality dynamics can lead to development of persuasive systems

https://www.inf.elte.hu/content/alkalmazasiterulet-specifikus-nagy-megbizhatosagu-informatikai-megoldasok.t.2299?m=437