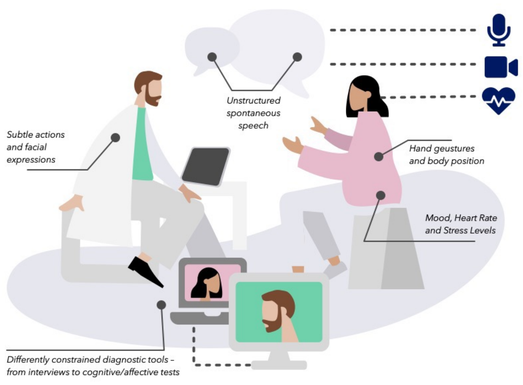
# Objective measures of psychiatric symptoms derived from speech and visual features of clinical interaction

Alexandra König[1,2], Hali Lindsay[3], Philippe Mueller[3], Danilo Postin[4], Eric Ettore[5], Michal Balazia[6], Amandine Lecomte[7], Benedikt Wirth[3]

1 Ki:elements
2 Cobtek (Cognition-Behaviour-Technology) Lab, FRIS-University Côte d'azur, Nice, France
3 German Research Center for Artificial Intelligence (DFKI), Saarbrücken, Germany
4 Department of Psychiatry, Carl von Ossietzky University of Oldenburg
5 Department of Psychiatry, Hopital Pasteur, CHU de Nice, France
6 Institut national de recherche en informatique et en automatique, Stars Team, France
7 INRIA, Semagramme Team, Nancy, France

## BACKGROUND

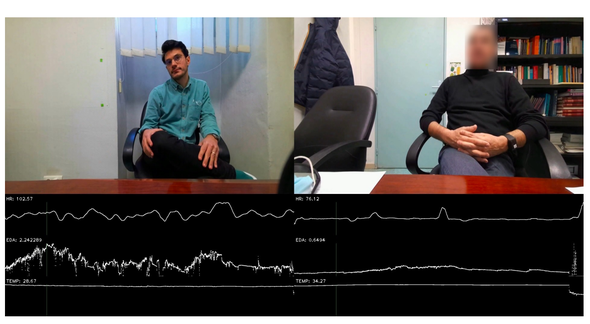
Compared to the progress seen in other branches of medicine, psychiatry lacks reliable and objective biomarkers to individually tailor treatments. Today, clinical states are measured by using question-based scales related to specific symptomatic domains which may be subject to biases. Thus, identifying objective markers of psychiatric disease states, including behavioral-based phenotypes, to support transdiagnostic dimensional approaches, is necessary for disease classification.




Since psychiatric disorders are disorders of social interaction emphasizing the importance of studying behavioral dynamics in real-life social interaction at the dyadic level rather than solely individual behavior. **Interaction-based phenotyping** to quantify the level of social behavior exchange or interpersonal synchrony of behavior could deliver additional data to generate an observer-independent picture of a patient's mental state ultimately leading to earlier and more sensitive identification of abnormalities. An **Interaction-based sociometric approach** combined with new automated monitoring technologies delivers novel state-associated biomarkers for diagnostic and therapeutic guidance. It leverages advancements in machine learning**, computer vision, and natural language processing** to extract meaningful information from patients' speech and visual cues.

## AIMS

We aim to identify and formalize a set of **novel multimodal digital biomarkers** derived from audio and visual interaction data and to develop predictive models within the scope of depression and schizophrenia. On this basis, we aim to develop models aiding in differential diagnosis, forecasting the patient's status (e.g., relapse prediction), and predicting therapeutic alliance. Important outcomes include technical and organizational methods for the management of medical data, demonstration scenarios covering patients' journeys including early detection, diagnosis support, relapse prediction, therapy support, an annotated corpus.



## METHODS

A longitudinal multicenter observational study is performed across three clinical sites in Germany and France for which clinical interviews with psychiatric patients are audio and video recorded after they give consent. The following types of data is being analyzed: clinical scores (questionnaires and scales), speech, video and physiological measures.

The audio analysis from dyadic interactions, the data will be pre-processed and speaker diarization and speaker labelling will be performed. In addition, the speech will be either manually annotated or automatically transcribed. Both paralinguistic and linguistic features will be extracted on the utterance level. Examples of speech features include pitch, speech rate, spectral and prosodic characteristics as well as word usage and content.

From the video streams, we will extract a comprehensive set of visual behaviour descriptors from both patient and clinician. These include gaze, head pose, facial expressions, posture, and body movements. We will make use of state-of-the-art approaches in computer vision that enable automatic extraction of such behaviours.

## MAIN USE CASES

1. **Supporting differential diagnosis in depression by detecting specific aetiologies**
   - Depressive episodes can be due to unipolar depression, bipolar depression, or a traumatic event
   - Identifying early markers for specific aetiologies could help clinicians improve diagnosis and treatment plan

2. **Quantifying therapeutic alliance by means of social synchrony**
   - Develop digital markers for measuring social synchrony (e.g., back channelling, shared language)
   - Improve patient outcome and by quantifying markers for social synchrony

3. **Relapse prediction from negative symptoms in schizophrenia**
   - Negative symptoms – e.g., reductions in emotion, speech, social activity – are thought to precede a schizophrenic episode.
   - Determine earliest markers for diminished emotion/affect (speech/video) and social activity (interviews/wearables)

4. **Robust and Objective Measurement of Formal Thought Disorder (FTD) in Schizophrenia**
   - Cognitive and linguistic aspects of FTD are heavily studied to better understand psychosis in Schizophrenia
   - Quantifying syntactic complexity, discourse cohesion, eye movements could lead to better disease understanding

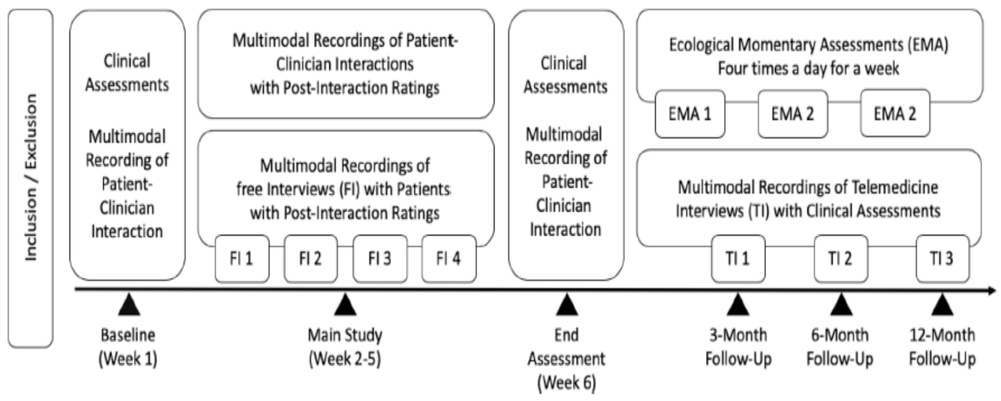
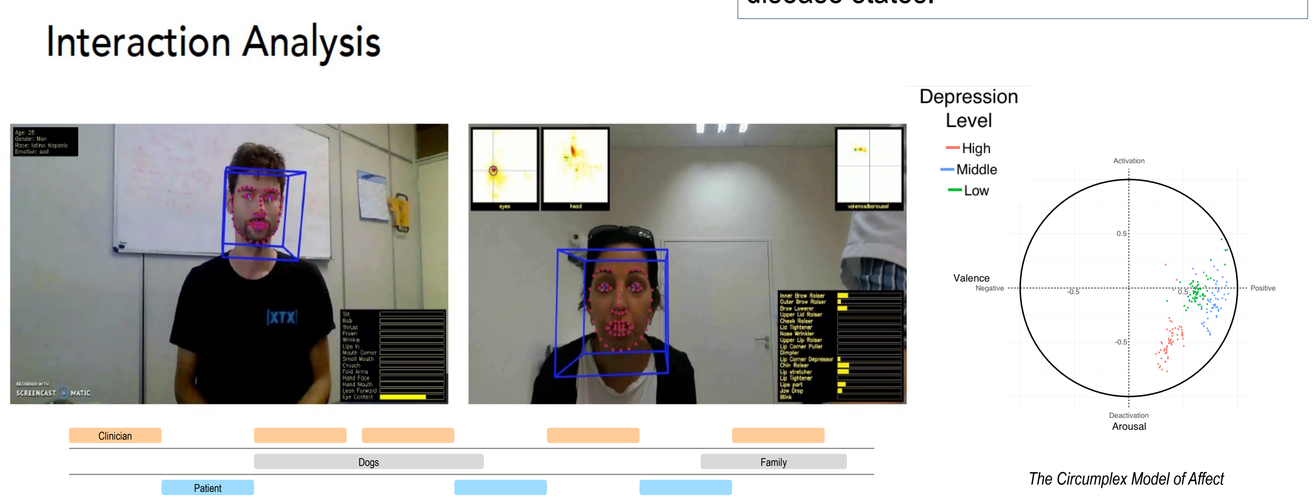



Fig. 1. The overall study design.


## RESULTS

The study is currently still ongoing with over 200 hours of collected recordings of more than 100 psychiatric patients.

We will calculate the relationship between the extracted audio and video measures and several clinical scales, thus symptom severity. The goal is to develop algorithms that can accurately classify and predict disease states.

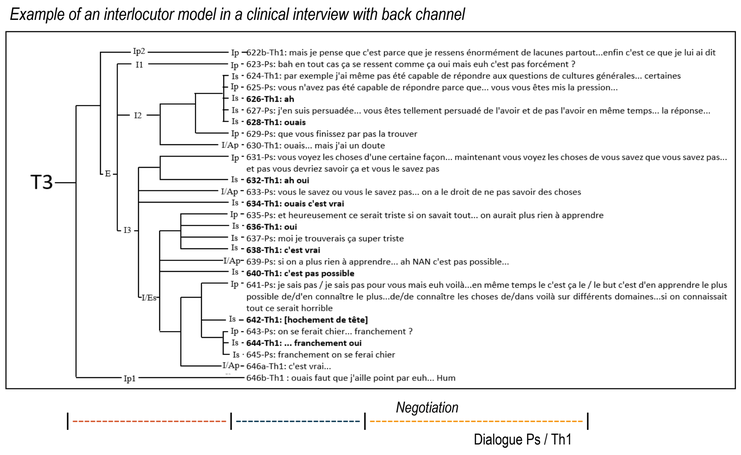


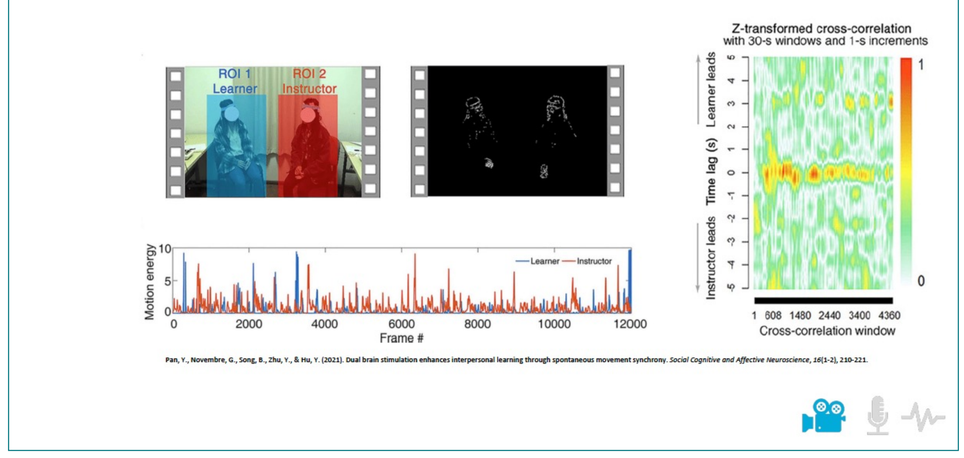

*The Circumplex Model of Affect*

*Descriptives features of Schizophrenia patient*

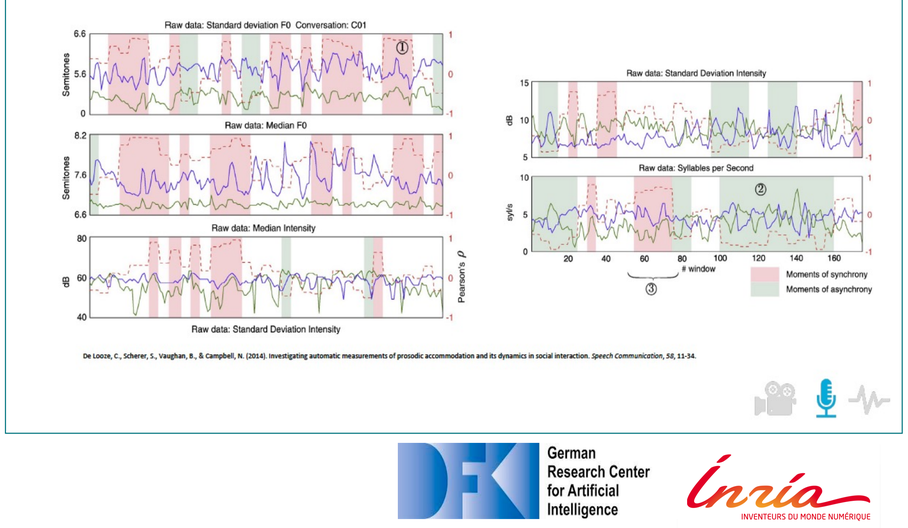
| Patient | Average of Eye Contact | Average of Brow Lowerer | VALENCE Min. of Valence | VALENCE Max. of Valence | VALENCE Average of Valence | VALENCE Ecart Valence | AROUSAL Min. of Arousal | AROUSAL Max. of Arousal | AROUSAL Average of Arousal | AROUSAL Ecart Arousal | PANSS N1 Blunted affect | PANSS N2 Emotional withdrawal | PANSS N3 Poor rapport | PANSS N4 Social withdrawal | SPT neg total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Patient 1 | 0.55 | 0.02 | -0.89 | 0.49 | -0.03 | 1.38 | -0.53 | 0.45 | -0.14 | 0.98 | 2.00 | 1.00 | 1.00 | 3.00 | 11 |
| Patient 2 | 0.56 | 0.18 | -0.92 | 0.31 | -0.12 | 1.23 | -0.87 | 0.59 | -0.26 | 1.46 | 1.00 | 2.00 | 1.00 | 2.00 | 12 |
| Patient 3 | 0.57 | 0.04 | -0.65 | 0.46 | -0.04 | 1.11 | -0.36 | 0.31 | -0.07 | 0.67 | 2.00 | 2.00 | 1.00 | 1.00 | 13 |
| Patient 4 | 0.38 | 0.21 | -0.62 | 0.54 | 0.05 | 1.16 | -0.38 | 0.30 | -0.10 | 0.68 | 2.00 | 2.00 | 1.00 | 3.00 | 15 |
| Patient 5 | 0.30 | 0.14 | -0.93 | 1.00 | 0.04 | 1.93 | -0.51 | 0.74 | -0.07 | 1.25 | 2.00 | 3.00 | 1.00 | 4.00 | 17.00 |
| Patient 6 | 0.71 | 0.24 | -0.82 | 0.44 | -0.01 | 1.26 | -0.21 | 0.45 | -0.01 | 0.66 | 2.00 | 1.00 | 1.00 | 3.00 | 18 |
| Patient 7 | 0.44 | 0.12 | -0.82 | 0.45 | -0.49 | 1.27 | -0.31 | 0.47 | 0.11 | 0.78 | 3.00 | 3.00 | 1.00 | 2.00 | 18 |
| Patient 8 | 0.37 | 0.40 | -0.71 | 0.54 | -0.10 | 1.25 | -0.40 | 0.38 | -0.16 | 0.79 | 3.00 | 5.00 | 2.00 | 4.00 | 21 |
| Patient 9 | 0.58 | 0.03 | -0.99 | 0.31 | -0.32 | 1.30 | -0.52 | 0.41 | -0.08 | 0.93 | 5.00 | 3.00 | 4.00 | 3.00 | 23.00 |
| Patient 10 | 0.28 | 0.04 | -0.22 | 0.30 | 0.07 | 0.52 | -0.58 | -0.03 | -0.32 | 0.55 | 5.00 | 5.00 | 4.00 | 4.00 | 26 |

*Example of an interlocutor model in a clinical interview with back channel*




Pan, Y., Novembre, G., Song, B., Zhu, Y., & Hu, Y. (2021). Dual brain stimulation enhances interpersonal learning through spontaneous movement synchrony. *Social Cognitive and Affective Neuroscience, 16*(1-2), 210-221.




De Looze, C., Scherer, S., Vaughan, B., & Campbell, N. (2014). Investigating automatic measurements of prosodic accommodation and its dynamics in social interaction. *Speech Communication, 58*, 11-34.

German Research Center for Artificial Intelligence

INVENTEURS DU MONDE NUMÉRIQUE