Contents lists available at ScienceDirect

Personalized Medicine in Psychiatry
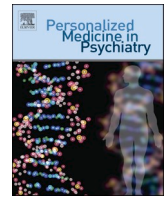
Multimodal phenotyping of psychiatric disorders from social interaction: Protocol of a clinical multicenter prospective study

Alexandra König [a,b,*], Philipp Müller [c], Johannes Tröger [c], Hali Lindsay [c], Jan Alexandersson [c], Jonas Hinze [d], Matthias Riemenschneider [d], Danilo Postin [e], Eric Ettore [b,f], Amandine Lecomte [g], Michel Musiol [g], Maxime Amblard [g], François Bremond [a,b], Michal Balazia [a], Rene Hurlemann [e]

[a] Institut national de recherche en informatique et en automatique (INRIA), Stars Team, Sophia Antipolis, Valbonne, France
[b] CoBTeK (Cognition-Behaviour-Technology) Lab, FRIS-University Côte d'Azur, Nice, France
[c] German Research Center for Artificial Intelligence, DFKI GmbH, Saarbrücken, Germany
[d] Department of Psychiatry and Psychotherapy, Saarland University Medical Center, Homburg, Germany
[e] Department of Psychiatry, School of Medicine and Health Sciences, Carl von Ossietzky University of Oldenburg, Bad Zwischenahn 26160, Germany
[f] Department of Psychiatry, Hopital Pasteur, CHU de Nice, 06000 Nice, France
[g] Institut national de recherche en informatique et en automatique (INRIA). Sémagramme Team. Nancy. France

ABSTRACT

Identifying objective and reliable markers to tailor diagnosis and treatment of psychiatric patients remains a challenge, as conditions like major depression, bipolar disorder, or schizophrenia are qualified by complex behavior observations or subjective self-reports instead of easily measurable somatic features. Recent progress in computer vision, speech processing and machine learning has enabled detailed and objective characterization of human behavior in social interactions. However, the application of these technologies to personalized psychiatry is limited due to the lack of sufficiently large corpora that combine multi-modal measurements with longitudinal assessments of patients covering more than a single disorder. To close this gap, we introduce Mephesto, a multi-centre, multi-disorder longitudinal corpus creation effort designed to develop and validate novel multi-modal markers for psychiatric conditions. Mephesto will consist of multi-modal audio-, video-, and physiological recordings as well as clinical assessments of psychiatric patients covering a six-week main study period as well as several follow-up recordings spread across twelve months. We outline the rationale and study protocol and introduce four cardinal use cases that will build the foundation of a new state of the art in personalized treatment strategies for psychiatric disorders.

Introduction

Psychiatric disorders represent a major challenge to global healthcare systems and society. Major depressive disorder (MDD) is one of the most common disorders with a lifetime prevalence of 16% [1] and schizophrenia is estimated to affect 1–2% of the European population, with roughly 20 million cases being reported globally in 2017. According to a report by the World Health Organization, MDD is the leading cause of disability worldwide [2,3]. The current COVID-19 pandemic is just increasing the prevalence showing a clear rise in depressive and anxiety disorders [3].

Left undiagnosed or untreated, psychiatric disorders increase the risk of hospitalizations, reoccurrences, co-morbidities, and suicide. Conditions like major depression, bipolar disorder, and schizophrenia have a complex network of overlapping behavioral symptoms which can be difficult to distinguish based on classical assessments during clinic consultation as clinicians lack objective markers. This makes disease management difficult and possibly ineffective with serious side effects due to potential misdiagnosis.

Current challenges of classical psychiatric assessment

Compared to the progress seen in other branches of medicine, psychiatry lacks sensitive indicators to tailor treatments to individuals.

Today, clinical states are measured by using question-based scales related to specific symptomatic domains which may be subject to biases [4]. One challenge is the heterogeneity of symptoms and variability between patients which can ultimately lead to missed diagnoses or incorrect treatments [5]. Another challenge is the discovery of pathognomonic symptoms for these different pathologies.

In the absence of objective markers of symptoms and their level of severity, it is a challenge to create effective and efficient methods for defining flexible medication treatment strategies. For instance, a mainstay treatment for long-term relapse prevention are mood stabilizers such as lithium, anticonvulsants or antipsychotics, which can cause side effects leading to a significant reduction of patient quality of life. Hence, methods for early identification of relapse, as well as respective medication and psychotherapy adjustments, and continuous monitoring of patient status for assessing impact of treatment are acutely needed.

The lack of objective markers is entailed by the absence of sufficient quantities of longitudinal, observational data. Recently, the National Institute of Mental health (NIMH) established a new diagnostic approach for researchers that states to combine biological, behavioural, and social factors to create *precision medicine for psychiatry*, highlighting the need for diverse data for better diagnosis of mental disorders [6]. Thus, identifying objective markers of psychiatric disease states, including *trans*-diagnostic, behavioural-based phenotypes, is necessary for improved disease classification and treatment [7]. With the current rise of the use of Artificial Intelligence (AI) in healthcare, personalized management of mental disorders is moving forward. Hence, technology-based behavioral sensing may prove to be effective in measuring subjective communicative functioning, making inferences about symptoms, and guiding treatment management [8].

Thus, two main challenges persist in modern psychiatry; the absence of clear quantitative markers of disease progress and lack of sufficiently large corpora that combine multi-modal measurements with longitudinal assessments [9].

Social interaction as a new study target

Many valuable diagnostic relevant information is extracted from the interaction between clinician and patient. This clinical interaction (e.g. conversation between patient and clinician including verbal as non-verbal behavior) is traditionally a clinician's most important source of information about patients' social skills, mood and motivation levels. However, a comprehensive clinical interview requires sufficient consultation time as well as strong clinical competencies and expertise to be able to detect early subtle signs of changes in communication. Moreover, for detecting these changes during a clinical conversation no standardized objective measures exist leaving a lot of room for speculations and subjective biases. Introducing ways to assess in a quantitative manner behavioral dynamics during real-life social interaction could help indicate for instance level of reciprocity and therapeutic alliance, which until now is merely left to clinical intuition.

Need for precise and sensitive digital markers

To develop and test new measures of mental illness, a movement from traditional markers and phenotyping to *digital markers* and *digital phenotyping* is needed. 'Digital phenotyping' refers to the moment-to-moment quantification of human behaviour in everyday life using data from digital devices [6]. Digital phenotyping suggests to collect patient data allowing for non-intrusive and continuous monitoring of behavioural and mental states, ultimately revealing clinically relevant information. Similarly, 'digital markers' are digitally-obtained disease indicators that can be used to define a digital phenotype [9].

Interaction-based phenotyping could provide various additional data to generate an observer-independent assessment of behaviour during a social interaction which reflects as a mirror the current symptomology of a patient. Additionally, interaction-based measures such as social

synchrony may have predictive value for treatment outcome.

Recent progress in computer vision, speech processing and machine learning has enabled detailed and objective characterization of human interaction behavior [10–12]. Applying these advanced methods of artificial intelligence (AI) provides new opportunities to identify digital markers of patient behaviour. Such markers have the potential to provide objective and continuous assessments of symptomatology in the context of patients' daily lives [13], thereby allowing to precisely tailor treatment to the concrete patient trajectory.

Many so far developed techniques are based solely on verbal information during interviews; however interpersonal communication often occurs nonverbally. Thus, merging computer vision-based measurement in a multi-modal approach would enhance the quality of analysis by allowing to detect changes in the quality of communication as alterations in the dyadic interaction patterns.

Digital markers and methods

In recent years, behavior recognition methods based on artificial intelligence (i.e., machine or deep leaning) have become increasingly effective in a variety of tasks, including action classification [14], body language and gestures [15]; gaze estimation [16], eye contact detection [10]; facial action units [17]; as well as affect extracted from single- or multiple modalities [18,19].

A growing number of approaches make use of this progress in human behaviour sensing to analyze clinical interaction data (e.g. therapy sessions); linguistic and paralinguistic characteristics from speech [20,21]. As psychiatric disorders (MDD, bipolar, schizophrenia) impact the quality of social interactions there is emphasis on studying these quantifiable behavioral dynamics in real-life social interaction at the dyadic level rather than solely individual behavior [22,23].

Automatic speech analysis shows promising results for the detection of affective states in patients with depression [24], bipolar disorder [25], schizophrenia [26] or dementia [27]. Certain mobile technologies have demonstrated feasibility for tracking depression that could inform models for predicting relapse [28]. In schizophrenia, passive smartphone data has been used to examine user behavior and predict clinical relapse events [29]. The CrossCheck Study found strong associations and predictive power between self-reported symptoms and both active and passive data [30].

While these initial results are promising, this research needs to be accelerated by further development of digital phenotyping technology focusing on scalability and equity, by establishing shared longitudinal data repositories and by fostering multidisciplinary collaborations between clinical stakeholders, including patients, computer scientists, and researchers [31].

Study Overview

Design

The overall study design is presented in Fig. 1. and consists of two phases. During the *main study phase*, interactions between the patients and clinician will be recorded multimodally, i.e. with video, audio, and physiological sensors. In the succeeding *follow-up phase* videoconference-based recordings and ecological momentary assessments will be audio and video recorded with a videoconferencing system during a period of 12 months.

At the beginning of the main study phase (week 1), each included patient will undergo a set of clinical assessments that serve as baseline measures. Interactions between patient and clinician during these assessments will be recorded and represent a baseline observation of interaction behaviour. Subsequently, each patient will be recorded in four planned interview sessions spread over four weeks (weeks 2–5). In addition, we will record as many additional interactions between clinicians and patients as feasible during this time period. At the end of the
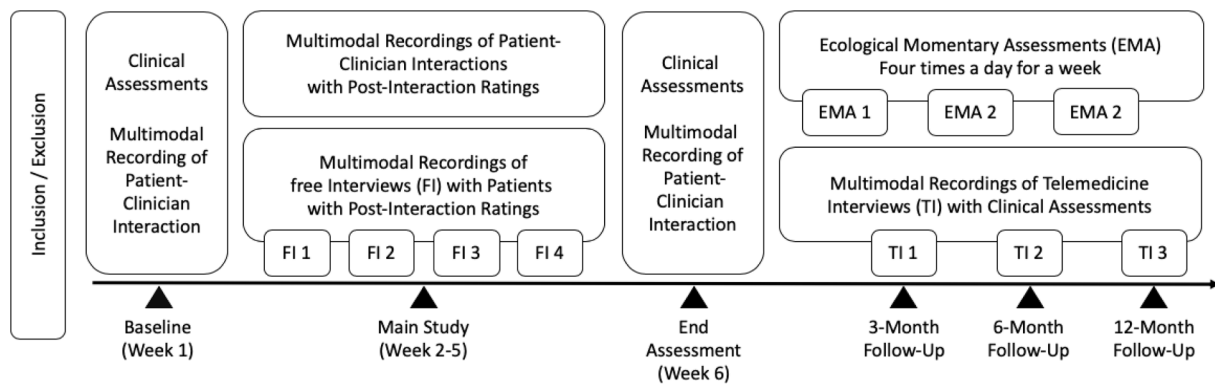
Fig. 1. The overall study design.

main study phase (week 6) a second set of clinical assessments and multimodal interaction recordings will be captured.

During the follow-up phase, three telemedicine interviews including clinical assessments of the patients will be recorded via videoconferencing system. These interviews will take place three months, six months and 12 months after the end of the main study phase. Between the first and second, as well as between the second and third follow-up, a one-week long block of ecological momentary assessments (EMA) will be recorded.

A detailed description of the planned assessments as well as the technical sensor setups is provided in chapters 2.5 and 2.6. In the following we will introduce the four cardinal use cases in MePheSTO and discuss how they will be addressed using the general study design.

Use cases

MePheSTO has a solid foundation of clinically motivated scenarios and use-cases synthesized jointly by the clinical partners of the project. Each use-case has been introduced to support clinicians at several points during the patient's journey, from the initial diagnosis and treatment planning to long-term psychiatric care (e.g. relapse prevention).

Each use case encompasses different research questions all centred around the scientifically sound validation of digital phenotypes for psychiatric disorders. In order to address these research questions, the MePheSTO project aims to provide objective marker of these phenotypes based on multimodal input, including speech, discourse and dialogical coherence, video, and bio-signals from clinical social interactions. Complementary, ecological momentary assessments (EMA) are collected to extend the data acquisition on a daily life basis. In the following, the four cardinal use cases of the MePheSTO projected are presented in detail:

Use case A

Supporting differential diagnosis for major depressive episode.

Major depressive episode (MDE) as defined by the current criteria (American Psychiatric Association (APA)) describes a large heterogeneous clinical syndrome comprising more than 1490 combinations of symptoms [32]. The various possibilities of fulfilling MDE criteria, including opposite symptoms such as insomnia and hypersomnia, decrease and increase in appetite or agitation and psychomotor retardation, highlight this heterogeneity.

According to international classifications (APA), MDE symptoms are the same whether in Major Depressive Disorder (MDD) or in Bipolar Disorder (BD). The differential diagnostics of MDD and BD during a MDE constitutes a major challenge in clinical practice. Indeed, about 20% of people suffering from MDE would be a misdiagnosed with BD [33]. And, the delay for a correct BD diagnosis can vary from almost seven to ten years after first mood symptoms.

In the same way, international classifications do not distinguish

whether MDE occurred in a context of traumatic exposure [34]. In fact, posttraumatic stress disorder (PTSD) is not the only psychiatric condition that may develop in the aftermath of trauma as evidenced by high comorbidity between MDE and PTSD [35]. MDE can also be the expression of a former traumatic exposure or chronic in this case [36]. Hence, the early exploration of the pre- or absence of psychological trauma in patients with MDE constitutes an important diagnostic step entailing a major impact on further therapeutic care.

Research questions:

- Are these different pathogenetic profiles of MDE characterized by different digital phenotypes?
- Can interaction-based, digital markers provide valid indicators for a better differentiation between the different clinical profiles of MDE?

Scientific approach:

Measures distinguishing between these clinical profiles of MDE will be determined based on observations of patient behaviour and physiology during patient-clinician interactions as well as on measures extracted during the patient's daily life:

- **Differential diagnosis of MDD and trauma triggered MDD comorbidity:** To extract markers indicative of trauma, the reaction to potentially trauma-associated topics will be analysed while patient-clinician interactions. Changes in verbal (e.g. speech), nonverbal (e.g. eye-contact) and physiological measures (e.g. skin conductance) are considered to be promising candidates [37].
- **Differential diagnosis of MDD and BD:** To assist the diagnosis of BD, both, data recorded during clinical interactions will be used. During interactions, speech represents an interesting modality for extracting digital markers indicating BD [38]. Furthermore, physiological measures extracted from wearable sensors can potentially indicate a manic phase in the course of the disease.

Use case B

Quantifying therapeutic alliance by means of social synchrony.

Therapeutic alliance, i.e. how a patient and a therapist connect, behave, and engage with each other, was shown to be connected to therapy outcome measures robustly across different disorders and therapeutic approaches [39]. However, the psychotherapeutic process is both dynamic and complex. It constitutes a most complex bio-psycho-social system in which language, cognition, and emotions are intertwined and influenced through the interactional dynamics between therapist and patient [40,41]. The therapeutic alliance can be associated with interpersonal coordination during human interaction in behaviour, physiology, emotional and cognitive states. Numerous studies have made connections to therapeutic processes and outcomes: vocal coordination [42], body movements [43], and physiology [44].

These markers for therapeutic alliance have the potential to support

clinicians during treatment and to assist in increasing the fit between patient and clinician. An important goal of MePheSTO is to develop such digital markers for therapeutic alliance and validate these markers in a multi-lingual, multi-disorder setting.

Research questions:

- How can the different aspects of social synchrony in therapeutic interactions be measured automatically by interaction-based, digital markers?
- Can digital markers for therapeutic alliance, extracted from clinical interactions, predict the subjectively perceived therapeutic alliance and clinical outcomes?
- Do digital markers for therapeutic alliance have the potential to support clinicians during treatment and to assist in increasing the fit between patient and clinician?

Scientific Approach:

In MePheSTO we will develop digital markers based on the in-Sync mode by Koole and Tschacher [39]. To measure different aspects of social synchrony, video, audio and physiological recordings taken from clinician and patient will be analysed. To validate the measures, clinician- and patient ratings of therapeutic alliance will be utilized alongside therapy outcome measures.

- **Movement synchrony**: To quantify movement synchrony we will integrate different modalities automatically extracted from video recordings (e.g. head pose, gesticulation, and eye gaze), e.g. via motion energy analysis.
- **Common language:** We will develop automatic measurements of linguistic behaviour matching based on audio recordings during therapy interactions, e.g. via Language Style Matching [45].
- **I-sharing**: We will detect sharing of subjective experiences between patient and therapist by language analysis on the recorded audio.
- **Affective co-regulation**: Detect via synchronization of breathing patterns extracted from video, as well as synchronized skin conductance levels measured by wearable sensors. Furthermore, we will detect complementary therapist behaviour with the goal to return to a homeostatic balance, e.g. in response to an upset patient.

Use case C: Treatment outcome and relapse prediction from negative symptoms in schizophrenia

Despite the progress in psychotherapeutic and pharmacological treatment in recent decades, remission rates in patients with schizophrenia stagnate around 20% [46]. One of the main goals when treating Schizophrenia is remission via relapse prevention. To achieve this, early detection of a relapse is crucial [47]. Relapse in schizophrenia is defined as a return of disease symptoms after a partial recovery, resulting in a negative impact on an afflicted person's life, their daily activities, and often leading to hospitalization [48]. However, the earliest indicators (negative symptoms) of relapse can be difficult to detect. Furthermore, patients quite often show difficulties in several domains of social functioning, namely deficits in social skills [49,50], social cognition disorders [51], loneliness [52], reduced social network size [53], impaired social motivation [54], and elevated social anhedonia [55]. These circumstances may additionally impact their medication maintenance or the development of clinical signs of relapse. Due to the life-long nature of this disease, current research focuses on a longitudinal understanding of schizophrenia symptoms and how to first predict and then prevent relapses. Assessing digital, interaction-based markers for the negative symptomatology along with changes in the daily life of people with schizophrenia would make it possible to identify the appearance of signs predictive of relapse [56].

Research questions:

- Can negative symptoms be assessed by digital, interaction-based markers?

- Can digital markers provide a reliable prediction of disease progression?
- Can a relapsing episode in schizophrenia be predicted via multi-modal digital markers prior to full onset?

Scientific Approach:

Digital phenotypes of negative symptoms will be determined based on observations of patient behaviour during patient-clinician interactions as well as on measures extracted during the patient's daily life:

- **Alogia and thought poverty (via speech data):** Defined via the amount of speech, average pause length, lack of articulation, average length of response, and conversational implicatures [57].
- **Anhedonia and affective flattening (via speech and video data):** Detected by facial and body movements, gaze fixation and through emotions and expressions from facial recognition.
- **Avolition and Social withdrawal (via EMA):** Assessed by the use of short prompted questionnaires assessing the participants' daily life [58].

Use case D: Uncovering formal thought disorders in schizophrenia

"Psychiatrists rely on language and speech behaviour as one of the main clues in psychiatric diagnosis" [59]. As language 'impairment' is one of the factors of psychosis and a major symptom for a schizophrenia diagnosis, understanding formal thought disorder in schizophrenia is crucial. The connection between language and Schizophrenia is well documented; phonetics, morphology, syntax, semantics, pragmatics and discourse structure have been studied [59,60]. Patients show difficulties in coordinating thoughts and actions according to their goals. This thinking is partly reflected in the subject's language at the lexical-semantic and syntactic levels, with either a productive side (e.g. vague, tangential, abstract, uninformative speech), or a deficient side (impoverished speech, lack of words, stereotypy). Furthermore, the pragmatic and dialogical approach has made it possible to highlight other elements that would be pathognomonic of schizophrenia, particularly discursive discontinuities [61]. Discursive discontinuities describe conversational breaks during verbal interactions between a schizophrenic patient and a clinician. Via modelling these conversational breaks [61], we aim to identify neurocognitive impairments (e.g. associated with deficits in attention, memory, language and executive function) and to improve the understanding of the mechanisms of social cognition, and specifically the processes underlying deficiencies in interpersonal relationships.

Research Questions:

- To what extent is a multimodal-type discourse analysis methodology (e.g. assessed via head movements, gaze movements, and facial expressions) likely to facilitate the identification of verbal interaction structures underpinning formal thought disorders?
- What congruent relationships can be established between "evolving formal thought disorders" and evolving intensity of schizophrenic disorders in general?

Scientific Approach:

From Speech and video recordings of clinical sessions between the clinician and patient, phonetics, morphology, syntax, semantics and pragmatics can be evaluated. Many measures have been reported for the understanding of formal thought disorder. We have listed a few:

- **Syntactic complexity and grammaticality** in schizophrenia have been reported as being "more grammatically deviant" and "less syntactically complex" compared to controls. However, this specific

finding was felt to be linked with earlier onset of illness, longer duration of illness, and negative symptoms [62,63].

- Using both the clinician and patient's **speech discourse cohesion** and the **communication disturbance index** can be used to evaluate the overall discourse level including disorganization of thoughts and speech [22,23,64–66].
- Measuring ability of the patient to encode or infer communicative intentions to build **common ground** with the interlocutor during the course of dialogue.
- Specific **eye-movement patterns**, i.e. lack of fixation in eye area or increased percentage of number of switches from one area to another one will be analyzed as well [67–69].

Methods

Sample population

Sample size

The target sample size of 450 participants, 150 per site ideally equally split among MDE (75) and schizophrenia patients (75) is justified by considering similar data sets with psychiatric patients. For each participant, the goal is to record a minimum of four clinical interactions which would result in a total of 1800 audio-visual recordings (600 per site) combined with physiological measures. This is a significant increase beyond existing corpora. For example, the DAIC-WO (Distress Analysis Interview Corpus - Wizard-of-Oz) database consists of audio and video recordings of 189 sessions of clinical interviews designed to support diagnosis of psychological distress conditions such as depression or post-traumatic stress [70]. Gavrilescu [71] analyzed facial expressions of 128 patients using the and the Facial Action Coding System (FACS) for predicting depression and anxiety scores. Cohn [72] collected sensor data of 23 unipolar & bipolar depressed and 32 healthy controls. For schizophrenia, studies show often similar or even smaller sample sizes. Bishay [73], automatically analysed the facial behaviour of 91 outpatients during clinician-patient interviews showing that certain expressions significantly correlated to symptoms of schizophrenia. Elvevåg [74] presents automated methods to perform discourse analysis for predicting psychosis with N = 83; 53 patients and 30 controls. Based on our review on the use of these digital markers and namely speech and video features in psychiatric patients we assume that chosen sample number is adequate to investigate the objectives of the study.

Inclusion and exclusion criteria

Participants will be recruited through the three clinics involved in the study. The following inclusion and exclusion criteria are presented in Table 1. and will be adhered to:

Co-enrolment in all study types will be allowed.

Setting

The MePheSTO study is built on a joint work program between INRIA, the French national research institute for the digital sciences, and DFKI, the German Research Centre for Artificial Intelligence, two of the world's largest non-profit research institutes active in the field of AI. The two institutes have decided to team up to address the challenges relating to the foundations and the applications of modern AI, including machine learning, speech and language processing, computer vision and healthcare. The presented study will be performed within this framework.

We aim to perform a prospective longitudinal multi-center observational study across three clinical sites, two psychiatric clinics in Germany (Homburg, Oldenburg) one in France (Nice). Each clinical site will apply separately for ethical approval at their local Ethics Committee. In France approval has been granted by the Ethics Committee ('Comité de protection des personnes', CPP Ouest II in Angers, France. In Germany, the protocol has been submitted for review at the Ethics Committee at the Saarland University Medical Center, Homburg, and at the University Clinic Karl Jaspers, Bad Zwischenahn. The study will be conducted according to the Declaration of Helsinki. Written informed consent will be gathered from all subjects prior to the study. Trained clinicians, psychiatrists and psychologists will be mainly involved in the recruitment and data collection of patients.

Protocol

Participants will be recruited in the participating clinics (inpatient and outpatient clinics) via their treating psychiatrists or a member of the research team. The participant will receive a detailed information sheet about the study, its process, potential risks and benefits. Prior to starting recording any missing demographic data, medical history or medication information will be collected from all participants. Sex, age, education (years), living status and current medications will be collected. This is to be able to control for any medications that may interfere with the social behavior to be analyzed during recordings. In regards to the current COVID-19 pandemic, all clinical interactions may be performed and recorded remotely via a videoconference system.

The different steps of the study are presented in Fig. 1. & Table 2. and should be completed in the following order:

1. Baseline assessment: research team member will start with a general screening interview (to last approximately 1 h30 to 2 h) with the participant using the SCID-CV tool. The screening interview will be video and audio recorded.

In order to determinate a subsample of prodromal patients, we will use the "Comprehensive Assessment of At-Risk Mental States" tool (CAARMS).

2. The participant will be debriefed in the usage of the online survey as well as the wearable sensors
3. After responding sufficiently to the diagnosis criteria of depression or/and schizophrenia/ patients UHR or HR, further psychiatric assessments will be performed using classical standard questionnaires or scales (clinician administered and self-report):
- For all participants with **Major Depressive Episode:**
  - Beck Depression Inventory/ BDI (self-report)
  - Montgomery Asberg Depression Rating Scale/ MADRS (clinician administered)
  - Young Mania Rating Scale (YMRS) (clinician administered)
  - Childhood Trauma Questionnaire (CTQ)
- For all participants with **Schizophrenia:**

Table 1
Inclusion and Exclusion criteria.

Inclusion criteria	Exclusion criteria
• According to the Structured Clinical Interview for the DSM-V (SCID-5-CV) [75] OR the Mini International Neuropsychiatric Interview (MINI) [76] diagnosed with either Major Depressive Episode *OR* Schizophrenia According to the Comprehensive Assessment of At-Risk Mental States (CAARMS) [77]: High Risk (HR) and Ultra High Risk (UHR) individuals likely to develop one or the other of these pathologies • Aged between 16 and 55 • Have capacity to consent or if participants are under 18, the consent form will also be signed by at least one of the parents	o Do not have capacity to consent o Acute suicidal tendencies o Not native speakers or had any major hearing or language problems o Disabling or non-stabilized somatic pathology. o Patients with language, cognitive or neurological deficiencies (neurological lesions, dysphasia, aphasia, mutism...) evaluated in particular during the neurocognitive testing phase. o History likely to modify the cerebral anatomy. o Patients requiring a drug or therapeutic solution that is incompatible with the quality of evaluation in the research. o Current diagnosis of substance abuse or dependency

Table 2
Overview over the patient journey in MePheSTO.

POPULATION	MEASUREMENTS		Baseline Screening (Week 1)	Week 2	Week 3	Week 4	Week 5	Discharge (Week 6)	EMA	3 Months	EMA	6 Months	EMA	12 Months
Depression	Informations	Demographic	x											
		Medication	x					x		x		x		x
	Tests	SCID-5-CV	x											
		MADRS	x					x		x		x		x
		Semantic Fluency	x											
		Phonetic Fluency	x											
		Digit Span F + B	x											
		TMT A & B	x											
	Self-report	WHOQOL	x					x		x		x		x
		CTQ	x											
		BDI	x					x		x		x		x
		WAI Patient		x	x	x	x	x						
		WAI Clinician		x	x	x	x	x						
	Interview*	F2F OR Video	x	x	x	x	x	x		x		x		x
	Physiological	Empatica	x	x	x	x	x	x		x		x		x
Schizophrenia	Informations	Demographic	x											
		Medication	x					x		x		x		x
	Test*	SCID-5-CV	x											
		PANSS	x					x		x		x		x
		Semantic Fluency	x											
		Phonetic Fluency	x											
		Digit Span F + B	x											
		TMT A & B	x											
	Self-report	WHOQOL	x					x		x		x		x
		SNS	x					x		x		x		x
		WAI Patient		x	x	x	x	x						
		WAI Clinician		x	x	x	x	x						
	Interview*	F2F OR Video	x	x	x	x	x	x		x		x		x
	Physiological	Empatica	x	x	x	x	x	x		x		x		x

* = audio and video recording with a dictaphone and Kinect.

- Brief Negative Symptoms Scale/BNSS or Self-report Negative Symptoms/SNS (self-report)
- Positive and Negative Syndrome Scale/PANSS (clinician administered)

Every included participant will reply to a questionnaire on subjective quality of life, the "World Health Organization Quality-of-Life Scale" (WHOQOL).

4. Every participant will undergo a short cognitive test battery consisting of:
- Semantic Verbal fluency, Phonetic Verbal fluency (1 min)
- Digit Span forward + backward
- Trail Making Test A & B
5. Afterwards the participants will undergo their standard clinical pathway with its regular consultations. This includes medical and therapeutic consultations which will be each time recorded for the length of the study (which is four weeks for inpatient clinics and 6 months for outpatient clinics). During the consultation clinician and participant will be equipped with a wearable sensor to collect physiological measures (heartrate variability, electrodermal activity, accelerometer, temperature).

Before each recording session the patient will be asked on his or her current medication status (if there were any modifications since the last recording).

6. Short ratings will be completed after each recorded session by patient and clinician (via smartphone) on their perceived quality of the clinical interaction.

We will use specific items of the Working Alliance Inventory-short revised (WAI-SR).

- Item 1: After this session, I feel better
- Item 2: I feel the things we discussed today will help me to accomplish the changes that I want.
- Item 3: I feel __ cared about me during this session.
- Items 4: I felt that my interaction partner understood what I want to change in this session
- Item 5: Item 5: In this interaction, I felt the clinician and I were on the same page

- Item 6 (only for patient): In this interaction, I felt it easy to share personal experiences with the clinician
- Option to annotate particular observations
7. End of study participation at discharge (or after 4 free interview sessions)
   A minimum psychometric assessment will be performed (for symptom improvement, therapy success) at all clinical sites; see C) consisting of
1. Semi-structured Interview and a self-report
   1. MDE: BDI, MADRS, YMRS
   2. Schizophrenia: BNSS, SNS, PANSS
   3. All: WHOQOL
2. Assessment of the overall perceived quality of care: Working Alliance Inventory-short revised /full version (WAI-SR)
8. After discharge, daily life measures will be collected in form of short regular surveys and EMA. A wearable device will be provided to a subsample of participants to record additional information on sleep quality, physical activity, etc.
9. Follow up / re-evaluation at M3, M6, M12 after the end of the study participation (via phone, videoconference or face to face consultation):
1. Minimum psychometric assessments (for symptom improvement, therapy success) at all clinical sites; see C)
2. Screening for medication adherence, any clinical care in between (hospitalisation, outpatient counselling, relapse, etc.)

Technical setup

The recording setup will be as minimal as possible to be not invasive and ensure natural behaviour during the interactions. The room in which the recordings take place will be a regular consultation room in the clinics. They require to be relatively quiet and bright in light in order to capture audio and video of sufficient quality for further analyses.

The patient will sit in front of the clinician at least 2 m apart, eventually with a table in between them. On the table small discrete cameras will be placed in between.

In accordance with current COVID-19 protection ordinances, required protective measures (e.g. safety distance, plexiglass, masks) will be respected throughout the entire duration of the study.

We will use the following methods and technology for data recordings and data acquisition:
Audio recording

The speech of the participants and their interactions with the clinician will be recorded as audio files via a microphone. We will directly record the speech of the patient and the clinician with the device placed in between them. From the audio files, we will use automatic speech recognition (ASR) and manual transcriptions to obtain textual transcripts of the recordings. A subset or all of the data will be manually transcribed for comparison between automated and manual transcriptions. We will use either the internal microphone of the device (PC or tablet) or an external microphone for better recording quality. The recorded data are automatically stored on the secured server.

Video recording

We intend to support the proposed speech analysis by a complementary computer-vision based analysis. Such analysis exploits advanced methods related to automated face analysis, tracking, detection and recognition, as well as human behavior analysis.

Firstly, we plan to record 3D & 2D video-data and depth data (RGBD & RGB) from all participants for the computer vision-based analysis. Then we intend to study this data with focus to find facial/gestural behaviors that are representative for psychiatric symptoms during a social interaction.

We intend to acquire the video data using an external camera. The recorded data are automatically stored on the secured server. For a subsample of patients, follow up evaluations will be made via a videoconference system which will allow to gather additional audio and video data. The video data collected will allow us on a subsample to perform eye tracking analysis.

Physiological measures

We would like to explore the use of additional objective markers of stress levels within this study. For this we will extract physiological data during the recorded clinical interactions via a bracelet device worn around wrist. The following measurements will be collected:

- Electro-Dermal Activity (EDA): measures sympathetic nervous system activity manifested through the skin, by measuring the constantly fluctuating changes in certain electrical properties of the skin;
- Heart Rate Variability (HRV): derived from measuring Blood Volume Pulse (BVP);

Online questionnaires

Participants will receive text messages on their mobile phone or email address with a link to go on a specific website to complete remotely clinical questionnaires on their current symptoms as well as questions on daily life activities, so called Ecological Momentary Assessments (EMA). The information will be stored on a secured server.

Videoconference system

The system is a web-based platform fully dedicated to diagnosing, screening and monitoring mental disorders. This tool is developed by INRIA using the latest advances in information and communication technologies to provide remote care through a web platform using any internet browser. This web platform allows an easy and direct connection between a clinician and his patient. Both connect to the web platform using their respective identifiers and passwords.

Study outcomes and hypotheses

The overall project goal is to acquire a multimodal dataset of patient-clinician interactions, which will be annotated and clinically labelled for the scientifically sound validation of digital phenotypes for psychiatric disorders. To this end, we aim to identify and formalize a set of novel multimodal digital biomarkers derived from the 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. Based on the previously outlined research, we hypothesise the following:

1. The clinical profiles of major depressive disorder, bipolar disorder, and posttraumatic stress disorder show significantly different digital phenotypes in patient-clinician interactions.
2. Different level of therapeutic alliance can be distinguished by automatically extracted, interaction-based digital markers.
3. Furthermore, digital markers of therapeutic alliance constitute a significant predictor for disease progression in clinical outcomes of schizophrenia and depression.
4. Negative symptoms in schizophrenia (alogia, thought poverty, anhedonia, affective flattening, avolition and social withdrawal) can be assessed by digital, interaction-based markers.
5. Furthermore, negative symptomatology assessed by digital markers provides a significant predictor of disease progression and relapse occurrence in schizophrenia.
6. Formal thought disorder in schizophrenia can be captured via automatically extracted digital markers of verbal, e.g. "discourse discontinuities" and non-verbal interaction structures, e.g. "gaze fixations".

Data management and analysis

Data management

Collection of data will be made via the different recording devices. Digital data (audio -or speech, video, physiological measures, recorded scores, answers to questionnaires) as well as paper data (written records) will be collected.

Concerning the paper data, they will be stored at a designated place with limited access to clinicians participating in the study (a key is needed for access). Each involved clinical partner will store these papers in his clinic. The data will be digitized through a web-platform in order to conduct the research work. Concerning the digital data, demographic and all clinical data will be stored in the secured and certified Health Data Hosting infrastructures of each clinical partner.

All Investigators involved with this study must comply with the requirements of the appropriate data protection legislation (including the General Data Protection Regulation and Data Protection Act) with regard to the collection, storage, processing and disclosure of personal information. Computers used to store the data will have limited access measures via user names and passwords. Published results will not contain any personal data and be of a form where individuals are not identified and re-identification is not likely to take place.

An end-to-end encryption methodology will be employed for transfer of data, using asymmetric encryption, as detailed in the guidelines of the German Federal Office for Information Security (BSI - Bundesamt für Sicherheit in der Informationstechnik). Data will be encrypted at the clinical sites on the recording computer that is not connected to the internet. After copying this encrypted data to a computer that is connected to the internet, the data will be transferred via the internet to the technical partners. The data will be kept encrypted at the technical sites and only be de-crypted right before de-crypted data is needed (e.g. to re-structure, annotate, extract features, or train machine learning models with the data).

Data analyses

In this study, we will mainly work on the analysis of the following types of data: clinical scores (questionnaires and scales either obtained during face-to-face visits or remotely), speech, video and physiological measures.

To achieve the primary objective of the study comparison analysis will be performed between the new digital markers and the standard clinical measures. We will perform a multi-modal analysis using combined data modalities as well.

Speech analysis

From the collected synchronised data, the audio will be extracted and pre-processed. For recordings from dyadic interactions, speaker diarization and speaker labelling will be performed. In addition, the speech will be either manually annotated or automatically transcribed depending on the use case. Both paralinguistic and linguistic features will be extracted on the utterance level. Paralinguistic, sometimes referred to as acoustic features, will include those have been shown to be correlated with psychiatric conditions (); Harmonic-to-noise ratio, jitter, shimmer and pitch. Linguistic analysis will include gauging complexity of speech through syntactic complexity measures [78], semantic processing [79], and topic coherence [80]. To gauge levels of synchrony between clinician and patient methods for backchannel identification and quantification will be employed (TODO) as well as language stye matching [45] and reciprocal language style matching [45]. The mentioned methods will be used on multiple scales; multiple measures over a single conversation as well as longitudinally to monitor between consultations.

Video analysis

From the synchronized 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 recent state-of-the-art approaches in computer vision that enable automatic extraction of such behaviours. For facial landmark detection, head pose-, and facial action unit estimation we will employ OpenFace 2.0 [14]. Furthermore, we will extract body posture using the OpenPose framework [8]. To achieve the maximal possible accuracy in gaze estimation, we will combine a recent state-of-the-art approach to gaze estimation [13] with additional calibration information recorded during the interactions. To quantify movement energy, we will employ the well-established motion energy analysis framework of Ramseyer [81]. Based on these raw features, we will compute higher-level representations that encode clinically relevant behaviours [82]. In cases where automatic extraction of behaviours achieves only insufficient accuracy, we will employ human annotators, potentially in a human-in-the-loop annotation scheme [83].

Physiological measures

Physiological measurements will be synchronized with audio and video recordings to allow for unimodal as well as multimodal analysis. The raw physiological measurements of skin conductance will be fed through standard pre-processing pipeline [84]. First of all, the skin conductance signal will be visually inspected for periods with poor contact or sharp square wave spiking. If artifacts are found, the data will be down-sampled [85]. After that a moving-average procedure or a notch-filter (50 Hz) depending on the quality of data will be applied [86]. For further analysis we will decompose the data into its phasic and tonic components. The heart rate will be automatically calculated with a proprietary algorithm based on the blood volume pulse recorded from a wristband [87].

Discussion

Significance of the collected corpus

As just stated recently in Lancet Psychiatry, clinicians desperately need new science and technology to deal with rising public mental health problems [88]. However, even though methods to analyse digital measures of mental health are growing exponentially, only small-scale data sets are available of monitoring patients for a short amount of time [89] (and current research efforts to collect more data seem rather exploratory [31]. For robust identification of markers regarding mental illness onset, treatment response or relapse, larger and longitudinal corpora are needed for effective analysis. Most studies only investigated the use of one digital measure, not necessarily combining multiples

sensor outputs. In addition, apart from a few ethological analyses, most studies analyzed the patient alone, without being in interaction. However, social interaction especially during clinical interviews seem to be today the most important source of information for clinical assessment.

To address this problem, a multi-center, multi-disorder longitudinal corpus creation effort was designed to develop and validate novel multimodal markers for psychiatric conditions. Additionally, we will collect objective information of social dynamics between patient and clinician rather than focusing on the patients' behaviour alone. This will allow to enrich the understanding of behavioral mechanisms of transdiagnostic social impairments in psychiatric disorders.

The major scientific achievements that will directly result from the corpus will be the development of novel approaches to automated behaviour analysis in unconstrained interactions. Our results fostered by multidisciplinary collaborations between clinical stakeholders (including patients), computer scientists, and researchers, will fill the gap of insufficient available multi-modal corpora and support the future use of digital markers for precision psychiatry. Several applications with high clinical relevance have been proposed. For instance, identifying a distinct digital phenotype for MDD can contribute to earlier etiological diagnosis and improved tailored therapeutic care. Objective measures of social synchrony could help clinicians with timely behavioral adaptation in order to increase therapeutic alliance. Predicting treatment response and detecting potential signs of relapse can lead to prevention strategies among risk patients before potential re-hospitalisation. Discourse analysis may facilitate the identification of verbal interaction structures underpinning formal thought disorders and other cognitive deficits in psychiatric patients.

Nevertheless, psychiatric disorders are complex and trajectories are heterogeneous; it will remain challenging to capture the nuances of using these specific technologies and even more challenging to attribute a clinical meaningfulness to the data acquired. Therefore, it is reasonable to foresee digital and interactional phenotyping as support tools to be used additionally to classical methods and professional assessments to better detect the nuances of intra-individual variabilities over time [5].

Limitations

While the MePheSTO corpus is an important improvement over the state of the art in multi-modal psychiatric corpora, several limitations remain that need to be addressed in the future. Due to the complex interdependencies of psychiatric phenotypes with e.g. the presence of somatic pathologies, major cognitive or neurological deficiencies, or substance abuse, we exclude such patients from our study. While this is in line with common practice in psychiatric phenotyping research [90], it also limits the applicability of findings to the excluded population. Future research needs to find ways to incorporate- and model all patient groups to realise inclusive digital phenotyping.

Another important influence on behaviour in interactions can be a patients' cultural background [91]. While our study covers cultural differences between German and French patients, dedicated corpus creation efforts are needed to understand the influence of culture on the expression of psychiatric phenotypes on a global scale. Lastly, the study is taking place during the COVID19 pandemic which may will complicate recruitment as well as the data acquisition process, but a video-conferencing modality was therefore implemented in the protocol.

Future work

Even though there is a wide range of available technologies and methods presented in the literature, only a few have been adopted by clinicians and none have been approved by governmental agencies for clinical psychiatric use [92]. It has been suggested that one of the reasons for the challenges in implementing these technologies might be a lack of psychometrics to effectively evaluate and understand the

extracted measurements. Insufficient validity and reliability have been of major concern and barrier to adoption. However, psychiatric phenotypes are not static across time and space and when measuring profoundly certain constructs such as affect and cognition values can vary considerably. An interesting approach may be to take temporal and spatial features underlying psychiatric phenotypes into account as sort of dynamic data which can potentially by scaled over time providing opportunities for understanding the progress of disorders and for personalizing pharmacological und non pharmacological treatments. In future work, we aim to build upon this validation approach of dynamic 'resolution' data with the MePheSTO corpus which includes various multiple continuous data streams. The next step towards fully realizing the vision of precision psychiatry will be to incorporate- and evaluate our novel digital markers into clinical care pathways and design well-adapted interfaces. In addition to the enormous potential digital phenotyping holds to improve treatment in the clinic, digital markers could also be highly valuable to prevent admission to a clinic by allowing to take immediate action once a clinical phenomena is detected. This can be in form of triggered alarms, personalized treatments or even digital tailored interventions that give advice and guidance how to handle symptoms and experiences [93,94]. This offers potential avenues to reduce important treatment costs [95].

Conclusion

This paper presented the MePheSTO (Digital Phenotyping for Psychiatric Disorders from Social Interaction) study, a joint initiative of two large European research institutes in the field of Artificial Intelligence: INRIA and DFKI. We described the rationale and methodology consisting of a prospective longitudinal multi-center observational study with the major aim of creating a multimodal corpus of patient-clinician interactions within the context of major depressive episodes and schizophrenia. The study is built around four distinct use cases form which research hypotheses are developed: A: Supporting differential diagnosis for major depressive disorder etiology; B: Quantifying therapeutic alliance by means of social synchrony; C: Relapse prediction from negative symptoms in schizophrenia; D: Uncovering formal thought disorders in schizophrenia. The collected corpus will serve the creation of novel multi-modal digital markers designed to improve diagnosis and treatment of psychiatric disorders and. As such, it represents a significant contribution towards the vision of precision psychiatry realized by in-depth AI-supported analysis of patient behavior.

Data availability

A minimal dataset generated by this study will be made available upon reasonable request.

CRediT authorship contribution statement

Alexandra König: Writing – original draft. **Philipp Müller:** Writing – original draft. **Johannes Tröger:** Conceptualization. **Hali Lindsay:** . **Jan Alexandersson:** Conceptualization. **Jonas Hinze:** Writing – original draft. **Matthias Riemenschneider:** Visualization, Supervision. **Danilo Postin:** Writing – original draft. **Eric Ettore:** Writing – original draft. **Amandine Lecomte:** Writing – original draft. **Michel Musiol:** Conceptualization. **Maxime Amblard:** Conceptualization. **François Bremond:** Conceptualization. **Michal Balazia:** . **Rene Hurlemann:** Visualization, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Kessler RC, Berglund P, Demler O, Jin R, Koretz D, Merikangas KR, et al. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). JAMA 2003;289(23):3095.

[2] Friedrich MJ. Depression is the leading cause of disability around the world. JAMA 2017;317(15):1517. https://doi.org/10.1001/jama.2017.3826.

[3] Santomauro DF, Mantilla Herrera AM, Shadid J, Zheng P, Ashbaugh C, Pigott DM, et al. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. The Lancet 2021; 398(10312):1700–12.

[4] Arevian AC, Bone D, Malandrakis N, Martinez VR, Wells KB, Miklowitz DJ, et al. Clinical state tracking in serious mental illness through computational analysis of speech. PLoS ONE 2020;15(1):e0225695.

[5] Brietzke E, Hawken ER, Idzikowski M, Pong J, Kennedy SH, Soares CN. Integrating digital phenotyping in clinical characterization of individuals with mood disorders. Neurosci Biobehav Rev 2019;104:223–30. https://doi.org/10.1016/j.neubiorev.2019.07.009.

[6] Insel TR. Digital phenotyping: technology for a new science of behavior. JAMA 2017;318(13):1215–6. https://doi.org/10.1001/jama.2017.11295.

[7] Schilbach L. Using interaction-based phenotyping to assess the behavioral and neural mechanisms of transdiagnostic social impairments in psychiatry. Eur Arch Psychiatry Clin Neurosci 2019;269(3):273–4. https://doi.org/10.1007/s00406-019-00998-y.

[8] Deif R, Salama M. Depression from a precision mental health perspective: utilizing personalized conceptualizations to guide personalized treatments. Front Psychiatry 2021;12:650318. https://doi.org/10.3389/fpsyt.2021.650318.

[9] Andrea A, Agulia A, Serafini G, Amore M. Digital biomarkers and digital phenotyping in mental health care and prevention. European Journal of Public Health 2020;30(Supplement_5). https://doi.org/10.1093/eurpub/ckaa165.1080.

[10] Müller P, Huang MX, Zhang X, Bulling A. "Robust eye contact detection in natural multi-person interactions using gaze and speaking behaviour," presented at the Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications, Warsaw, Poland, 2018. [Online]. Available: https://doi.org/10.1145/3204493.3204549 .

[11] Cao Z, Hidalgo G, Simon T, Wei SE, Sheikh Y. OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. IEEE Trans Pattern Anal Mach Intell 2021;43(1):172–86. https://doi.org/10.1109/TPAMI.2019.2929257.

[12] Tzirakis P, Zhang J, Schuller B. End-to-End Speech Emotion Recognition Using Deep Neural Networks. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2018. p. 5089–93.

[13] Ebner-Priemer U, Santangelo P. Digital phenotyping: hype or hope? Lancet Psychiatry 2020;7(4):297–9. https://doi.org/10.1016/S2215-0366(19)30380-3.

[14] Das S, Thonnat M, Bremond F. "Looking deeper into Time for Activities of Daily Living Recognition," 2020.

[15] Liu X, Shi H, Chen H, Yu Z, Li X, Zhao G. *iMiGUE: An Identity-free Video Dataset for Micro-Gesture Understanding and Emotion Analysis.* 2021, pp. 10626-10637.

[16] Sinha N, Balazia M. *FLAME: Facial Landmark Heatmap Activated Multimodal Gaze Estimation.* 2021.

[17] Baltrusaitis T, Zadeh A, Lim YC, Morency L. "OpenFace 2.0: Facial Behavior Analysis Toolkit," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, 2018 2018, pp. 59-66, doi: 10.1109/FG.2018.00019. [Online]. Available: https://ieeexplore.ieee.org/document/8373812/.

[18] Müller PM, Amin S, Verma P, Andriluka M, Bulling A. "Emotion recognition from embedded bodily expressions and speech during dyadic interactions. International Conference on Affective Computing and Intelligent Interaction (ACII) 2015;2015: 663–9. https://doi.org/10.1109/ACII.2015.7344640. [Online]. Available: https://ieeexplore.ieee.org/document/7344640/.

[19] Sharma G, Dhall A. "A survey on automatic multimodal emotion recognition in the wild," in *Advances in Data Science*. In: Phillips-Wren G, Esposito A, Jain LC, editors. (Intelligent Systems Reference Library. Cham Switzerland: Springer; 2021. p. 35–64.

[20] Hinzen W, Rossello J. The linguistics of schizophrenia: thought disturbance as language pathology across positive symptoms. Front Psychol 2015;6:971. https://doi.org/10.3389/fpsyg.2015.00971.

[21] Birnbaum ML, Ernala SK, Rizvi AF, Arenare E, R. Van Meter A, De Choudhury M, et al. Detecting relapse in youth with psychotic disorders utilizing patient-generated and patient-contributed digital data from Facebook. NPJ Schizophr 2019;5(1). https://doi.org/10.1038/s41537-019-0085-9.

[22] Kuperberg GR. Language in schizophrenia Part 1: an Introduction. Lang Linguist Compass 2010;4(8):576–89. https://doi.org/10.1111/j.1749-818X.2010.00216.x.

[23] Kuperberg GR. Language in schizophrenia Part 2: What can psycholinguistics bring to the study of schizophrenia…and vice versa? Lang Linguist Compass 2010;4(8): 590–604. https://doi.org/10.1111/j.1749-818X.2010.00217.x.

[24] Kiss G, Vicsi K. Mono- and multi-lingual depression prediction based on speech processing. Int J Speech Technol 2017;20(4):919–35. https://doi.org/10.1007/s10772-017-9455-8.

[25] Faurholt-Jepsen M, Busk J, Frost M, Vinberg M, Christensen EM, Winther O, et al. Voice analysis as an objective state marker in bipolar disorder. Transl Psychiatry 2016;6(7):e856.

[26] Tahir Y, Yang Z, Chakraborty D, Thalmann N, Thalmann D, Maniam Y, et al. Non-verbal speech cues as objective measures for negative symptoms in patients with schizophrenia. PLoS ONE 2019;14(4):e0214314.

[27] König A, Linz N, Zeghari R, Klinge X, Tröger J, Alexandersson J, et al. Detecting Apathy in Older Adults with Cognitive Disorders Using Automatic Speech Analysis. J Alzheimers Dis 2019;69(4):1183–93.

[28] Sequeira L, Perrotta S, LaGrassa J, Merikangas K, Kreindler D, Kundur D, et al. Mobile and wearable technology for monitoring depressive symptoms in children and adolescents: A scoping review. J Affect Disord 2020;265:314–24.

[29] Barnett I, Torous J, Staples P, Sandoval L, Keshavan M, Onnela J-P. Relapse prediction in schizophrenia through digital phenotyping: a pilot study. Neuropsychopharmacology 2018;43(8):1660–6. https://doi.org/10.1038/s41386-018-0030-z.

[30] Wang Y, Bilinski P, Bremond F, Dantcheva A. G3AN: disentangling appearance and motion for video generation. In: in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; 2020. p. 5264–73.

[31] Huckvale K, Venkatesh S, Christensen H. Toward clinical digital phenotyping: a timely opportunity to consider purpose, quality, and safety. npj Digital Med 2019;2 (1):1–11. https://doi.org/10.1038/s41746-019-0166-1.

[32] Ostergaard SD, Jensen SOW, Bech P. The heterogeneity of the depressive syndrome: when numbers get serious. Acta Psychiat Scand 2011;124(6):495–6. https://doi.org/10.1111/j.1600-0447.2011.01744.x.

[33] Angst J, Merikangas KR, Cui L, Van Meter A, Ajdacic-Gross V, Rossler W. Bipolar spectrum in major depressive disorders. Eur Arch Psychiatry Clin Neurosci 2018; 268(8):741–8. https://doi.org/10.1007/s00406-018-0927-x.

[34] Gootzeit J, Markon K. Factors of PTSD: Differential specificity and external correlates. Clin Psychol Rev 2011;31(6):993–1003. https://doi.org/10.1016/j.cpr.2011.06.005.

[35] Gros DF, Price M, Magruder KM, Frueh BC. Symptom overlap in posttraumatic stress disorder and major depression. Psychiatry Res 2012;196(2–3):267–70. https://doi.org/10.1016/j.psychres.2011.10.022.

[36] Kostaras P, Bergiannaki JD, Psarros C, Ploumbidis D, Papageorgiou C. Posttraumatic stress disorder in outpatients with depression: Still a missed diagnosis. J Trauma Dissociation 2017;18(2):233–47. https://doi.org/10.1080/15299732.2016.1237402.

[37] Huon P, Stutz N. "Linguistic markers of time and subjectivity in the narration of psychic trauma," (in French). Evol Psychiatr 2020;85(4):479–508. https://doi.org/10.1016/j.evopsy.2020.06.008.

[38] Wang B. *et al.*, "Learning to detect bipolar disorder and borderline personality disorder with language and speech in non-clinical interviews," *arXiv preprint arXiv: 2008.03408,* 2020.

[39] Koole SL, Tschacher W. "Synchrony in Psychotherapy: A Review and an Integrative Framework for the Therapeutic Alliance," *Frontiers in Psychology,* vol. 7, Jun 14 2016, doi: ARTN86210.3389/fpsyg.2016.00862.

[40] Gelo OC, Salvatore S. A dynamic systems approach to psychotherapy: A meta-theoretical framework for explaining psychotherapy change processes. J Couns Psychol 2016;63(4):379–95. https://doi.org/10.1037/cou0000150.

[41] Schiepek G, Fricke B, Kaimer P. "Synergetics of Psychotherapy," in *Self-Organization and Clinical Psychology: Empirical Approaches to Synergetics in Psychology*, Tschacher W, Schiepek G, Brunner Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1992, pp. 239-267.

[42] Reich CM, Berman JS, Dale R, Levitt HM. Vocal Synchrony in Psychotherapy. J Soc Clin Psychol 2014;33(5):481–94. https://doi.org/10.1521/jscp.2014.33.5.481.

[43] Ramseyer F, Tschacher W. Nonverbal Synchrony in Psychotherapy: Coordinated Body Movement Reflects Relationship Quality and Outcome. J Consult Clin Psychol 2011;79(3):284–95. https://doi.org/10.1037/a0023419.

[44] Marci CD, Ham J, Moran E, Orr SP. Physiologic correlates of perceived therapist empathy and social-emotional process during psychotherapy. J Nerv Ment Dis 2007;195(2):103–11. https://doi.org/10.1097/01.nmd.0000253731.71025.fc.

[45] Aafjes-van Doorn K, Porcerelli J, Muller-Frommeyer LC. Language Style Matching in Psychotherapy: An Implicit Aspect of Alliance. Journal of Counseling Psychology 2020;67(4):509–22. https://doi.org/10.1037/cou0000433.

[46] Jaaskelainen E, Juola P, Hirvonen N, McGrath JJ, Saha S, Isohanni M, et al. A systematic review and meta-analysis of recovery in schizophrenia. Schizophr Bull 2013;39(6):1296–306.

[47] Meisenzahl E, Walger P, Schmidt SJ, Koutsouleris N, Schultze-Lutter F. Early recognition and prevention of schizophrenia and other psychoses. Nervenarzt 2020;91(1):10–7. https://doi.org/10.1007/s00115-019-00836-5.

[48] Lader M. What is relapse in schizophrenia? Int Clin Psychopharmacol 1995;9:5–10.

[49] Bellack AS, Morrison RL, Wixted JT, Mueser KT. An Analysis of Social Competence in Schizophrenia. Br J Psychiatry 1990;156:809–18. https://doi.org/10.1192/bjp.156.6.809.

[50] Mueser KT, Bellack AS, Douglas MS, Morrison RL. Prevalence and Stability of Social Skill Deficits in Schizophrenia. Schizophr Res 1991;5(2):167–76. https://doi.org/10.1016/0920-9964(91)90044-R.

[51] Green MF, Horan WP, Lee J. Social cognition in schizophrenia. Nat Rev Neurosci 2015;16(10):620–31. https://doi.org/10.1038/nrn4005.

[52] Lim MH, Gleeson JFM, Alvarez-Jimenez M, Penn DL. Loneliness in psychosis: a systematic review. Soc Psychiatry Psychiatr Epidemiol 2018;53(3):221–38. https://doi.org/10.1007/s00127-018-1482-5.

[53] Gayer-Anderson C, Morgan C. Social networks, support and early psychosis: a systematic review. Epidemiol Psych Sci 2013;22(2):131–46. https://doi.org/10.1017/S2045796012000406.

[54] Fulford D, Campellone T, Gard DE. Social motivation in schizophrenia: How research on basic reward processes informs and limits our understanding. Clin Psychol Rev 2018;63:12–24. https://doi.org/10.1016/j.cpr.2018.05.007.

[55] Blanchard JJ, Horan WP, Brown SA. Diagnostic differences in social anhedonia: a longitudinal study of schizophrenia and major depressive disorder. J Abnorm Psychol 2001;110(3):363–71. https://doi.org/10.1037//0021-843x.110.3.363.

[56] Wunderink L, van Bebber J, Sytema S, Boonstra N, Meijer RR, Wigman JTW. Reprint of: Negative symptoms predict high relapse rates and both predict less favorable functional outcome in first episode psychosis, independent of treatment strategy. Schizophr Res 2020;225:69–76. https://doi.org/10.1016/j.schres.2020.11.046.

[57] Gazdar G. Pragmatics : implicature, presupposition and logical form. New York: Academic Press; 1979.

[58] Mote J, Fulford D. Ecological momentary assessment of everyday social experiences of people with schizophrenia: A systematic review. Schizophr Res 2020;216:56–68. https://doi.org/10.1016/j.schres.2019.10.021.

[59] Ratana R, Sharifzadeh H, Krishnan J, Pang S. A Comprehensive Review of Computational Methods for Automatic Prediction of Schizophrenia With Insight Into Indigenous Populations. Front Psychiatry 2019;10:659. https://doi.org/10.3389/fpsyt.2019.00659.

[60] Covington MA, He C, Brown C, Naçi L, McClain JT, Fjordbak BS, et al. Schizophrenia and the structure of language: the linguist's view. Schizophr Res 2005;77(1):85–98.

[61] Musiol M, Verhaegen F. Investigating Discourse Specificities in Schizophrenic Disorders. In: Rebuschi M, Batt M, Heinzmann G, Lihoreau F, Musiol M, Trognon A, editors. Interdisciplinary Works in Logic, Epistemology, Psychology and Linguistics: Dialogue, Rationality, and Formalism. Cham: Springer International Publishing; 2014. p. 315–42.

[62] Amblard M, Fort K, Demily C, Franck N, Musiol M. "Analyse lexicale outillée de la parole transcrite de patients schizophrènes," (in French). Revue TAL 2015;55(3): 91–115.

[63] Hoffman RE, Sledge W. An analysis of grammatical deviance occurring in spontaneous schizophrenic speech. J Neurolinguistics 1988;3(1):89–101. https://doi.org/10.1016/0911-6044(88)90008-5.

[64] M. Constant and A. Dister, "Automatic detection of disfluencies in speech transcriptions," in *Spoken Communication*, vol. 1, M. Pettorino, A. Giannini, I. Chiari, and F. Dovetto Eds., no. 1): Cambridge Scholars Publishing, 2010, pp. 259-272.

[65] Maher BA, Manschreck TC, Linnet J, Candela S. Quantitative assessment of the frequency of normal associations in the utterances of schizophrenia patients and healthy controls. Schizophr Res 2005;78(2–3):219–24. https://doi.org/10.1016/j.schres.2005.05.017.

[66] Docherty NM, DeRosa M, Andreasen NC. Communication disturbances in schizophrenia and mania. Arch Gen Psychiatry 1996;53(4):358–64. https://doi.org/10.1001/archpsyc.1996.01830040094014.

[67] Langdon R, Corner T, McLaren J, Coltheart M, Ward PB. Attentional orienting triggered by gaze in schizophrenia. Neuropsychologia 2006;44(3):417–29. https://doi.org/10.1016/j.neuropsychologia.2005.05.020.

[68] D. Sun, R. Shao, Z. Wang, and T. M. C. Lee, "Perceived Gaze Direction Modulates Neural Processing of Prosocial Decision Making," *Front Hum Neurosci,* Original Research vol. 12, no. 52, 2018, doi: 10.3389/fnhum.2018.00052.

[69] Padroni S, Demily C, Franck N, Bocerean C, Hoffmann C, Musiol M. "Behavioral adjustment and saccadic eye movements in schizophrenia Ajustement comportemental et mouvements de saccades oculaires dans la schizophrénie," (in French). L'Évolution Psychiatrique 2016;81:365–79. https://doi.org/10.1016/j.evopsy.2016.01.008.

[70] J. Gratch *et al.*, "The Distress Analysis Interview Corpus of human and computer interviews," Reykjavik, Iceland, 2014: European Language Resources Association (ELRA), in Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), pp. 3123-3128. [Online]. Available: http://www.lrec-conf.org/proceedings/lrec2014/pdf/508_Paper.pdf.

[71] M. Gavrilescu and N. Vizireanu, "Predicting Depression, Anxiety, and Stress Levels from Videos Using the Facial Action Coding System," *Sensors-Basel,* vol. 19, no. 17, 2019, doi: ARTN369310.3390/s19173693.

[72] J. F. Cohn *et al.*, "Detecting depression from facial actions and vocal prosody," presented at the 2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, 2009. [Online]. Available: https://ieeexplore.ieee.org/document/5349358/.

[73] Bishay M, Palasek P, Priebe S, Patras I. SchiNet: Automatic Estimation of Symptoms of Schizophrenia from Facial Behaviour Analysis. IEEE Trans Affective Comput 2021;12(4):949–61. https://doi.org/10.1109/taffc.2019.2907628.

[74] Elvevåg B, Foltz PW, Rosenstein M, Delisi LE. An automated method to analyze language use in patients with schizophrenia and their first-degree relatives. J Neurolinguistics 2010;23(3):270–84. https://doi.org/10.1016/j.jneuroling.2009.05.002.

[75] Beesdo-Baum K, Zaudig M, Wittchen H-U. SCID-5-CV Strukturiertes Klinisches Interview für DSM-5-Störungen–Klinische Version: Deutsche Bearbeitung des Structured Clinical Interview for DSM-5 Disorders-Clinician Version von Michael B. First, Janet BW Williams: Rhonda S. Karg, Robert L. Spitzer. Hogrefe; 2019.

[76] Sheehan DV, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;vol. 59 Suppl 20:34–57.

[77] Yung AR, Yung AR, Pan Yuen H, Mcgorry PD, Phillips LJ, Kelly D, et al. Mapping the onset of psychosis: the Comprehensive Assessment of At-Risk Mental States. Aust N Z J Psychiatry 2005;39(11-12):964–71.

[78] H. Lindsay, J. Tröger, N. Linz, J. Alexandersson, and J. Prudlo, "Automatic detection of language impairment," *ExLing 2019,* vol. 25, p. 133, 2019.

[79] Wang K, Cheung EFC, Gong Q-Y, Chan RCK, Zhang XY. Semantic processing disturbance in patients with schizophrenia: a meta-analysis of the N400 component. PLoS ONE 2011;6(10):e25435.

[80] Salavera C, Puyuelo M, Antoñanzas JL, Teruel P. Semantics, pragmatics, and formal thought disorders in people with schizophrenia. Neuropsychiatr Dis Treat 2013;9:177.

[81] Ramseyer FT. Motion energy analysis (MEA): A primer on the assessment of motion from video. Journal of counseling psychology 2020;67(4):536.

[82] Troisi A. Ethological research in clinical psychiatry: the study of nonverbal behavior during interviews. Neurosci Biobehav Rev 1999;23(7):905–13.

[83] Monarch RM. Human-in-the-Loop Machine Learning: Active learning and annotation for human-centered AI. Simon and Schuster; 2021.

[84] Aqajari SAH, Naeini EK, Mehrabadi MA, Labbaf S, Dutt N, Rahmani AM. pyEDA: An Open-Source Python Toolkit for Pre-processing and Feature Extraction of Electrodermal Activity. Procedia Comput Sci 2021;184:99–106.

[85] Braithwaite JJ, Watson DG, Jones R, Rowe M. A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments. Psychophysiology 2013;49(1):1017–34.

[86] Boucsein W, editor. Electrodermal Activity. Boston, MA: Springer US; 2012.

[87] Milstein N, Gordon I. Validating measures of electrodermal activity and heart rate variability derived from the empatica E4 utilized in research settings that involve interactive dyadic states. Front Behav Neurosci 2020;14.

[88] The Lancet Psychiatry. Digital psychiatry: moving past potential. The Lancet Psychiatry 2021;8(4):259.

[89] Faurholt-Jepsen M, Bauer M, Kessing LV. Smartphone-based objective monitoring in bipolar disorder: status and considerations. International journal of bipolar disorders 2018;6(1):1–7.

[90] Jongs N, Jagesar R, van Haren NEM, Penninx BWJH, Reus L, Visser PJ, et al. A framework for assessing neuropsychiatric phenotypes by using smartphone-based location data. Transl Psychiatry 2020;10(1).

[91] Alarcón RD. Culture, cultural factors and psychiatric diagnosis: review and projections. World psychiatry 2009;8(3):131.

[92] Cohen AS, Cowan T, Le TP, Schwartz EK, Kirkpatrick B, Raugh IM, et al. Ambulatory digital phenotyping of blunted affect and alogia using objective facial and vocal analysis: Proof of concept. Schizophr Res 2020;220:141–6.

[93] Wahle F, Kowatsch T, Fleisch E, Rufer M, Weidt S. Mobile sensing and support for people with depression: a pilot trial in the wild. JMIR mHealth and uHealth 2016;4 (3):e5960.

[94] Weidt S, Wahle F, Rufer M, Hörni A, Kowatsch T. MOSS-Mobile Sensing and Support Detection of depressive moods with an app and help those affected. Therapeutische Umschau Revue therapeutique 2015;72(9):553–5.

[95] McCrone P, Knapp M, Proudfoot J, Ryden C, Cavanagh K, Shapiro DA, et al. Cost-effectiveness of computerised cognitive-behavioural therapy for anxiety and depression in primary care: randomised controlled trial. The British Journal of Psychiatry 2004;185(1):55–62.