

Feasibility and Acceptability of a Mobile-Based Emotion Recognition Approach for Bipolar Disorder

H. Daus^{1,2,3}, M. Backenstrass^{2,4} *

¹ Psychotherapy Practice, Dipl.-Psych. Henning Daus, Ellhofen (Germany)

² Institute of Clinical Psychology, Centre for Mental Health, Klinikum Stuttgart (Germany)

³ Faculty of Science, Eberhard Karls University Tübingen (Germany)

⁴ Department of Clinical Psychology and Psychotherapy, Institute of Psychology, Ruprecht-Karls-University Heidelberg (Germany)

Received 26 October 2020 | Accepted 9 June 2021 | Published 30 August 2021



ABSTRACT

Over the past years, the mobile Health approach has motivated research projects to develop mood monitoring systems for bipolar disorder. Whereas mobile-based approaches have examined self-assessment or sensor data, so far, potentially important emotional aspects of this disease have been neglected. Thus, we developed an emotion-sensitive system that analyzes the verbal and facial expressions of bipolar patients in regard to their emotional cues. In this article, preliminary findings of a pilot study with five bipolar patients with respect to the acceptability and feasibility of the new approach are presented and discussed. There were individual differences in the usage frequency of the participants, and improvements regarding its handling were suggested. From the technical point of view, the video analysis was less dependable than the audio analysis and recognized almost exclusively the facial expressions of happiness. However, the system was feasible and well-accepted. The results indicate that further developments could facilitate the long-term analysis of expressed emotions in bipolar or other disorders without invading the privacy of patients.

KEYWORDS

Assistance Systems,
Bipolar Disorder,
Emotion Recognition,
Intelligent Systems,
Mobile Health.

DOI: 10.9781/ijimai.2021.08.015

I. INTRODUCTION

EMOTIONAL expressions are an important indication of how we are feeling. Thus, sudden or severe changes in those expressions can be signs for psychopathological syndromes or diseases. One of the most common mental disorders worldwide is bipolar disorder [1], which is defined by its depressive or (hypo-) manic episodes [2]. These episodes are characterized by typical changes in the emotional experiences of patients: For example, patterns of elevated happiness and anger during mania or sadness and disgust during depression [3]. Beyond that, patients with bipolar disorder often show emotion-related difficulties like deficits in emotion regulation [4]–[8] and recognition [9]–[12] or in the afore-mentioned expression of emotions [13], [14]. Therefore, in bipolar disorder, the processing of emotional information and the verbal or facial expressions following emotional stimuli might be influenced by mood-dependent biases or regulation strategies like emotional avoidance [15]–[17]. Since all of these areas are important to the everyday lives of patients, the deficits affect their overall functioning [18]–[20].

Depending on the severity and the course of this disease, its personal and economic consequences can be enormous [21], [22]. Thus, within patient care, strategies that complement the medical and psychological

treatment are of great interest [23], [24]. Over the past years, the increasing technical possibilities have strengthened the *mobile Health* (mHealth) approach in general and, just as much, in bipolar disorder [25], [26]. Especially the advantages regarding the availability, accessibility and cost efficiency of technical systems show their potential to benefit the ambulatory care system. Many mobile-based approaches for bipolar disorder include self-assessment or sensor data in mood monitoring or recognition approaches [27]–[33]. Yet, none of the existing systems have focused on the emotional aspects of this disease [34]. The expressed emotions of patients, however, could be an important indication of their emotional reactivity [14] and possibly of their affective state or recurring episodes [34].

Within the research project *Emotion-sensitive Assistance systems for the reactive psychological Interaction with people* (EmAsIn), an emotion-sensitive mHealth approach for bipolar disorder has been developed. The system, which has already been described in detail [34], combines the common self-assessment (SA) or sensor approaches with other, innovative features. It incorporates an emotion-sensitive module that retrieves audio and video data from short and actively user-triggered recordings of patients. The auditive and visual information is analyzed in respect to emotional cues or content [34]. Beyond that, the system includes optional, external assessments or third-party assessments (TPA), which are realized by involving close contacts into the ambulatory approach.

In this article, we examine the feasibility and acceptability of several system features. Therefore, data from a pilot study with patients with

* Corresponding author.

E-mail addresses: mail@praxis-daus.de (H. Daus), m.backenstrass@klinikum-stuttgart.de (M. Backenstrass).

bipolar disorder is presented and discussed. As indications regarding the feasibility of the emotion-sensitive module, the SA or the TPA approach, some usage-related aspects (like the completion of the participation periods or the frequency of feature usage) were assessed. Concerning the emotion-sensitive module, the accuracy of the word and face recognition, the detected emotional expressions and the convergence between the emotion recognition within both sources (auditive and visual) were seen as further indications of its feasibility. In order to examine the acceptability of the features, interviews with the participants were conducted. All features were expected to be feasible and, in accordance with research suggesting a higher technical affinity of younger patients [36], [37], a higher feasibility with younger participants was expected (A). In more detail, a good usability (A.1) and convergent results of the recognition approach (A.2) were expected. Since participants of an earlier interview study of the authors had been open towards the innovative features [35], all features were expected to be well-accepted (B).

II. METHODS

A. Participants

From April to December 2018 five ambulatory patients with remitted bipolar disorder participated in the pilot study for an individual duration of several weeks (see results section). The study sample included three men and two women. Four participants, according to DSM-V, could be diagnostically categorized into bipolar I disorder, the fifth person into bipolar II disorder. The participants (between 24 and 51 years old) had an average age of 39.40 years ($SD = 9.94$).

B. System Features

The most important study instruments were smartphones with a pre-installed system application and a wearable device that participants wore on their wrists. The overall system acquires data from multiple sources in order to recognize mood state changes of patients. Next to the features described below, the system analyzes physiological or sleep data (e.g., heart rate or sleep duration) and behavioral patterns (e.g., recognized activities or smartphone usage). Since the overall concept of the system has been presented on another occasion [34], the following sections were reduced to essential technical aspects of the emotion recognition approach and the other two examined features.

The application analyzes intentionally initiated recordings of its users by activating the smartphone camera and microphone, as soon as this function is selected. Users then are instructed to tell their *Story of the Day* (SotD), a narrative of an important event of their day. In the study setting, smartphone holders and external microphones were included to ensure a good quality of the recordings. Only if the user approves, the system saves the recording locally and analyzes the given information. In that case, automatic transcriptions of the verbal information are analyzed in regard to the word count of different emotional categories by following the *Linguistic Inquiry and Word Count* (LIWC) approach [38]. For example, the specific categories of sadness, anxiety or anger include words like “to regret”, “nervous” or “outraged”. The more comprehensive categories of positive or negative emotions contain words of all the specific, emotional word categories with positive or negative connotations. The visual information, on the other hand, is automatically separated into multiple video frames per second. All captured facial expressions then are analyzed regarding the presence of one of four basic emotions (happiness, sadness, anger and anxiety) by following the *Facial Action Coding System* (FACS) [39].

The app-based SA approach has also been developed within the research project and has not been clinically evaluated yet. It includes six 7-point items regarding the most important symptoms of bipolar

disorder. All items can be assessed by choosing values from -3 to 3 with negative values or positive values representing depressive or (hypo) manic symptoms. The related and newly developed TPA of partners or related parties are realized using separate web applications (and are, therefore, not necessarily assessed via smartphone). They also include six items that are almost identical with the SA items, and are answered using the same scale [34].

The participants of the pilot study were asked to keep their daily routines as they were and to integrate the assistance system as far as possible. Whereas some system components were designed to be optional [34], the three presented features (or, two features in case of missing third parties), should have been used regularly (e.g., on a daily basis). The SotD was supposed to take about two minutes per day and to contain narrations of recent events that had affected the emotional experiences of the narrator. Furthermore, the stories were supposed to be recorded in a well exposed room without the presence of other persons and with a frontally placed camera (smartphone). During the pilot study, the assistance system was not yet fully developed and did not automatically react to the assessed data or recordings.

C. Semi-structured Interview

Potential technical issues and the individual usage experiences were examined at the end of the study participation during semi-structured interviews. The interview guide included questions about all system features. The section about the emotion recognition approach was especially detailed and concerned aspects like the perceived emotional intensity or potential confounding factors while telling the SotD. For example, the participants were asked, “Do you think that your perceived emotions or your emotional expressions were influenced by the missing dialog partner?” All questions could be answered on 5-point scales from 1 “negative” or “not at all” to 5 “positive” or “exactly”. Beyond that, every issue allowed for explanations or comments. The nine interview questions that are relevant to this article (seven of them SotD-related) took about ten minutes.

D. Indicators of Feasibility and Acceptability

The measurement of the feasibility was operationalized through several usage-related aspects: The duration of the study participation of all patients (in reference to the targeted duration, see section E) was seen as an indicator regarding the feasibility of the overall system. Furthermore, for all features, the usage frequency (in reference to the study duration) was assessed and seen as a feasibility measure. As for the SotD module, the duration of the recordings was seen as an additional measure of their usability. Moreover, the technical functionality of this module was assessed in order to infer its (technical) feasibility. Thus, the count of recognized words was assessed, and the congruence of automatic transcriptions with a sample of manual transcriptions was analyzed (see section E). The count of the recognized faces was assessed and compared to the count of video frames in order to evaluate the recognition. The amount of the recognized emotional expressions of several emotional categories (in reference to the count of words and frames with faces) was assessed and seen as a further (technical) feasibility measure of the module. Finally, by analyzing the congruence of the emotion recognition results of both sources, the measure of their convergent validity was included as a feasibility indicator, because it could be seen as a potential proof of concept. The acceptability of the features was operationalized and assessed through questions of the interviews. Therefore, the selection of the participants regarding the possible answer choices or additional comments delivered the acceptability measures.

E. Design and Analysis

As for most of the research questions, the five participants were considered single cases and their data was analyzed separately.

However, in spite of the small number of participants, in regard to some specific questions (e.g., the age effect or the acceptability), correlation analyses or other methods were applied within a statistical group design. The long-term assessment of the feasibility data (see section D) was realized within the ambulatory environment of the patients (ambulatory assessment). An individual assessment period of twelve weeks (about 84 days) per participant was aimed for. With respect to the potential age effect on the usage behavior, the correlations between the age of the participants at the beginning of the study and the frequency of their SA or the SotD assessments were calculated. The congruence between the automatic SotD transcriptions and a sample of ten manual transcriptions (as a feasibility measure of the emotion-sensitive approach) was assessed by comparing all recognized LIWC [38] word categories (e.g. pronouns, numbers or specific themes like “money”) within both sources. The percentage of the recognized emotional expressions within all recognized words and faces of each SotD assessment was analyzed following the LIWC [38] and FACS [39] approaches. The congruence of both recognition sources (convergent validity) was analyzed by calculating the correlations between corresponding or contradicting LIWC and FACS categories. The interviews, which were developed to assess acceptability aspects, were conducted in German with a single measurement time point at the end of the study participation. The answers as well as the comments were recorded, translated into English and delivered quantitative and occasional qualitative data. All descriptive or correlative analyses were conducted using *Microsoft Excel V. 16.40* and *IBM SPSS Statistics V. 26*.

III. RESULTS

A. Feasibility

The individual study duration of the five participants was at least 57 and up to 134 days. On average, the patients of the pilot study participated for 87.40 days ($SD = 32.14$). As illustrated in Table I, the individual study duration of some participants considerably exceeded or fell below the targeted assessment period.

TABLE I. STUDY DURATION, ASSESSMENT NUMBERS AND SOTD^a DURATION FOR EACH PARTICIPANT

Code	Study duration in days	SA ^b		TPA ^c		SotD	
		N	N	N	N	Duration in s (M, SD)	
P1	134	110	1	14	94.36	38.73	
P2	97	100	0	43	92.81	36.42	
P3	92	46	17	15	104.47	31.91	
P4	57	64	0	6	96.17	47.91	
P5	57	54	0	48	148.02	43.91	

^aSOTD = Story of the Day; ^bSA = self-assessments; ^cTPA = third-party assessments.

1. Usage-related Aspects

The individual assessment numbers for the SA, TPA, and SotD for each participant are also shown in Table I. On several occasions, the same features were used several times during the same days. That is why, for example, one participant completed 64 SA within 57 days (see Table I). According to the study instruction of daily usage, Fig. 1 illustrates solely the percentage of participation days for each patient, on which they (or their related parties) used the three concerned features at least once. In reference to the SA, the usage frequency for all participants ranged from 40.22 % to 96.49 % of the participation days. The TPA were only used by related parties of two of the participants, resulting in a minimum of 0 % usage and a maximum of 17.39 % of

the participation days. For the SotD assessments, the usage frequency ranged from 8.21 % to 82.46 %. Together, the five participants of the pilot study recorded 126 SotDs. As one can see in Table I, the individual mean values for the audio file duration ranged from 92.81 s to 148.02 s.

There were strong, negative correlations between the age of the participants and the percentage of days, on which they used the SA [$r(3) = -.60, p = .285$] and the SotD [$r(3) = -.70, p = .188$].

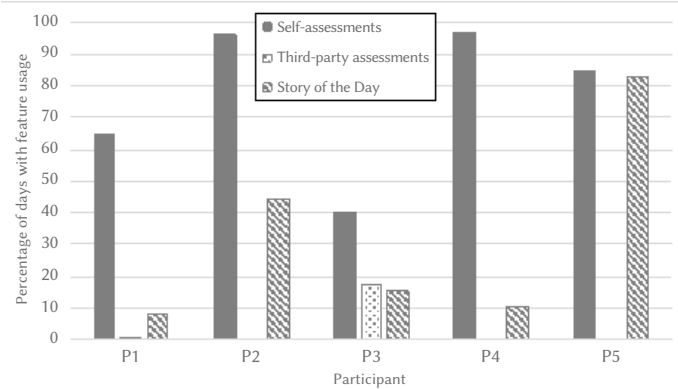


Fig. 1. Usage frequency for specific system features and each participant in reference to the individual study duration (in days).

2. Functionality of the Emotion Recognition Approach

Within the automatic transcriptions of the SotD audio files of each participant, on average, the LIWC program [38] recognized between 106.29 and 240.77 words (see Table II). A random sample of ten audio files was also transcribed manually and then analyzed with the LIWC software. All recognized LIWC word categories (not only emotional categories) within the manual and the corresponding automatic transcriptions matched from 72.00 % to 93.00 % ($M = 87.30, SD = 6.70$). The LIWC analysis of the audio data of all participants found, on average, between 3.93 and 11.15 single words corresponding to the category of positive emotions. This means that, for example, all audio files of one participant contained on average 5.47 single words belonging to that specific category. The corresponding mean values for each participant are illustrated in Table II. For the category of negative emotions, the individual mean values ranged from 1.33 to 4.52 words per file (see Table II). Regarding the more specific emotional category of sadness, the LIWC analysis resulted in mean values between 0.43 and 1.04 words per file (Table II). For the other two afore-mentioned, specific emotional LIWC categories (anxiety and anger), all corresponding mean values resulting from the analysis were ≤ 0.67 .

TABLE II. LIWC^a ANALYSIS OF THE SOTD^b RECORDINGS FOR EACH PARTICIPANT

Code	Number of words		Positive Emotions		Negative Emotions		Sadness	
	M	SD	M	SD	M	SD	M	SD
P1	106.29	81.40	3.93	3.17	2.21	1.67	0.43	0.51
P2	140.02	59.40	5.93	2.87	1.91	1.82	0.53	0.74
P3	173.33	58.17	5.47	3.20	3.00	2.56	0.93	1.16
P4	149.67	89.91	6.00	3.85	1.33	1.21	0.83	0.98
P5	240.77	92.79	11.15	4.87	4.52	2.42	1.04	1.01

^aLIWC = Linguistic Inquiry and Word Count; ^bSotD = Story of the Day; P1 ($N = 14$), P2 ($N = 43$), P3 ($N = 15$), P4 ($N = 6$), P5 ($N = 48$).

In percent (compared to the total wordcount of the audio files and for each participant separately), there were average amounts from 3.00 % to 5.89 % of words belonging to the category of positive emotions in the recordings (see Fig. 2). In regard to the category of negative emotions, the individual mean values ranged from 0.93 % to 2.01 % of words per file. The more specific category of sadness accounted for 0.34 % to 0.58 % of the wordcounts (see Fig. 2).

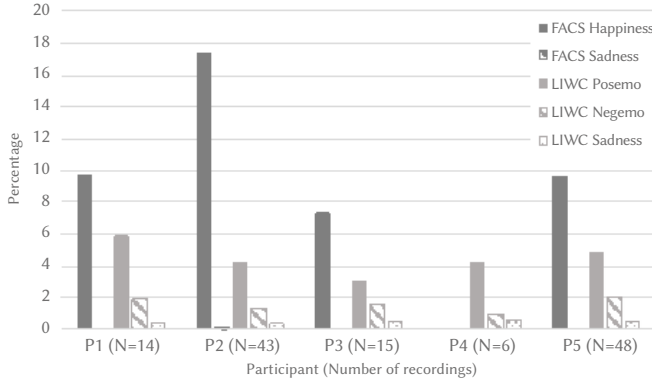


Fig. 2. Percentage of recognized emotions of Facial Action Coding System (FACS) categories and Linguistic Inquiry and Word Count (LIWC) categories in the Story of the Day recordings in reference to the individual number of frames with recognized faces (FACS) or the individual word count (LIWC).

The corresponding SotD video files were separated by the assistance system into, on average and for each participant separately, 2787.42 to 4443.42 video frames. For two assessments, the smartphone application apparently did not work accurately, and no video data was recorded. Both assessments belonged to P1 and had to be excluded from the calculations concerning the FACS [39] analysis. With four participants, within more than 94 % of the video frames there were recognized faces. As for P4, only 0.31 % of the video frames contained recognized faces. Thus, the individual mean values of frames with recognized faces that could be analyzed by the FACS-based software ranged from 7.00 to 4438.25 (see Table III). When the 6 video files of P4 without face recognition are not taken into account, for the remaining four participants, the FACS analysis found individual mean values of 209.93 to 538.51 frames with recognized happiness in the facial expressions (see Table III). However, the automatic FACS analysis detected almost no sadness in the facial expressions, resulting in individual mean values of all participants of ≤ 0.60 frames with recognized sadness (see also Table III). Beyond that, the FACS analysis detected none of the other two examined emotional FACS categories (anxiety or anger). Regarding the participants with successful face recognition, the individual mean percentage of frames with happiness out of all relevant frames (with faces) ranged from 7.38 % to 17.44 % (see Fig. 2). The mean percentage of sadness for all participants was ≤ 0.02 % (see Fig. 2).

TABLE III. FACS^a ANALYSIS OF THE SOTD^b RECORDINGS FOR EACH PARTICIPANT

Code	Frames with faces		Happiness		Sadness	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
P1	2919.17	1051.87	243.42	205.14	0.00	0.00
P2	2749.60	1122.88	538.51	638.62	0.60	3.81
P3	2961.33	963.49	209.93	219.27	0.00	0.00
P4	7.00	12.25	0.17	0.41	0.00	0.00
P5	4438.25	1315.08	435.88	384.13	0.02	0.14

^aFACS = Facial Action Coding System; ^bSOTD = Story of the Day; P1 (N = 12), P2 (N = 43), P3 (N = 15), P4 (N = 6), P5 (N = 48).

In order to examine, if the measures for the verbal and facial emotional expressions had delivered convergent results, the correlations between the most important LIWC and FACS categories were analyzed for each participant separately. For all participants with working face recognition, there were moderate or strong correlations between specific verbal and facial emotion recognition categories (see Table IV for all correlations). In regard to P1, the most prominent results were strong negative and significant correlations between the FACS category happiness and the LIWC category positive emotions (see Table IV). Thus, in this case, a higher amount of recognized expressions of happiness within the video frames coincided with less emotionally positive words within the LIWC analysis.

Concerning P2, the most prominent results were strong positive and significant correlations between FACS happiness and LIWC positive emotions or between FACS sadness and the LIWC sadness categories (see Table IV). Consequently, more video frames with expressed happiness coincided with a higher amount of positive words. And more recognized sadness within the FACS analysis was associated with more sadness-related words in the LIWC analysis. As for P3, there were no significant results, and there was only one moderate correlation indicating that more facially expressed happiness was associated with the use of more emotionally positive words (see Table IV). Regarding P5, there were several moderate and significant correlations: The FACS happiness categories were positively associated with LIWC positive emotions and negatively associated with LIWC negative emotions (%). Therefore, more recognized happiness within the facial expressions coincided with more emotionally positive and less emotionally negative words within the spoken language of this participant.

B. Acceptability

During the semi-structured interviews at the end of their individual study participation, the five participants showed positive attitudes towards the app-based SA ($M = 3.80$, $SD = 1.10$). One participant, nevertheless, emphasized the relevance of external assessments due to a potentially less-reliable self-perception (see Table V). Yet, only two participants were able to use the TPA with the help of related parties. These two participants perceived the assessments very positively ($M = 4.50$, $SD = 0.71$). Other participants still commented on this feature and expressed their interest in external assessments that could compensate for biased retrospective assessments regarding their mood (see Table V). With respect to the SotD module, the overall view of the five participants was relatively positive ($M = 3.40$, $SD = 1.14$) and comments mentioned positive “side-effects” of this feature (see Table V). The instruction of this module was easy to understand ($M = 4.60$, $SD = 0.55$), and most participants could imagine using this feature on a regular basis ($M = 4.00$, $SD = 0.71$). However, comments differed regarding the potential frequency of this usage or included recommendations for possible modifications (see Table V).

When asked, if their perceived emotions during the SotD recordings and during the described current events were of a similar intensity, the participants showed a moderate agreement ($M = 2.80$, $SD = 1.30$). They further agreed that the missing dialog partner, to some degree, had influenced their perceived and expressed emotions during the SotD ($M = 3.40$, $SD = 1.14$). Comments specified that some participants had experienced the strange feeling of being observed while talking into the camera. Yet, they further mentioned that this effect was not persistent (see Table V for comments). The “selfie mode” during the recordings was perceived as somewhat influencing, too, but less strong than the missing interaction ($M = 3.00$, $SD = 1.23$). However, some participants were somewhat irritated by seeing their own emotional expressions (see Table V). When asked, if they felt burdened by the SotD recordings, the participants did not agree ($M = 2.00$, $SD = 1.41$).

TABLE IV. PEARSON CORRELATIONS BETWEEN FACS^a AND LIWC^b CATEGORIES

Codec	FACS ^a	LIWC ^b					
		Positive Emotions	Positive Emotions (%)	Negative Emotions	Negative Emotions (%)	Sadness	Sadness (%)
P1	Happiness	-.73**	-.36	-.29	.06	-.03	.30
	Happiness (%)	-.72**	-.12	-.36	.02	-.09	.20
P2	Happiness	.49**	.14	.25	.02	.07	-.06
	Happiness (%)	.28	.13	.15	.04	.04	-.03
	Sadness	.00	-.03	.18	.14	.53**	.35*
	Sadness (%)	.00	-.03	.19	.14	.53**	.36*
P3	Happiness	.31	.18	.21	.05	.19	.10
	Happiness (%)	.14	.15	.00	-.09	-.07	-.12
P5	Happiness	.44**	.05	.05	-.31*	-.04	-.19
	Happiness (%)	.37**	.18	-.08	-.31*	-.08	-.16
	Sadness	.09	.17	-.21	-.20	-.15	-.15
	Sadness (%)	.09	.17	-.21	-.20	-.15	-.15

^aFACS = Facial Action Coding System; ^bLIWC = Linguistic Inquiry and Word Count; ^cP4 excluded due to failed face recognition, for P1 and P3 the FACS analysis did not detect any sadness; P1 (N = 12), P2 (N = 43), P3 (N = 15), P5 (N = 48); ** Correlation is significant at the 0.01 level (2-tailed); * Correlation is significant at the 0.05 level (2-tailed).

TABLE V. COMMENTS OF THE PARTICIPANTS ON SELECTED ISSUES

Interview section and issue	Comment and participant number
Self-assessments	<i>"It's a good feature, but I didn't use it regularly. My self-perception isn't that good, thus, third-party assessments are very important."</i> P1
Third-party assessments	<i>"I would recommend using that feature and keeping it simple. Maybe one could also notify the partner to facilitate simultaneous comparisons."</i> P1
	<i>"If you would ask me how I felt last week, I wouldn't remember. Thus, third-party assessments would be helpful [in regard to future systems]."</i> P5
Overall view	<i>"It helps your self-perception and forces yourself to observe yourself."</i> P3
Story of the Day	<i>"If it was easier to use."</i> P1
	<i>"Two or three times a week, yes. Seven times would be difficult. It's mentally challenging and you don't have a story to tell each day."</i> P3
	<i>"In theory, yes. I don't like filming myself, but if there was a feedback and it would help my situation, yes."</i> P4
	<i>"If there was a feedback, yes. But not on a daily basis."</i> P5
Missing dialog partner	<i>"It irritates, yes. I feel strange when I'm being filmed, it would be different with a real counterpart."</i> P4 <i>"At the beginning it's strange to talk to the camera, because you feel observed. Later it's like keeping a diary."</i> P5
Selfie mode	<i>"I tried to focus on the camera, because it irritated me to see my own feelings on screen."</i> P5

IV. DISCUSSION

The mHealth approach offers many opportunities for the ambulatory care system and the disease management of patients with bipolar disorder. Within the EmAsIn project, to our knowledge, the first emotion-sensitive assistance system for bipolar disorder was developed and examined. The following sections of the current article first address the presented results regarding the (technical) feasibility and acceptability of the newly developed system. Subsequently, the implications of these findings with respect to clinical practice and further research as well as methodological issues will be discussed.

A. Feasibility

In order to evaluate the feasibility of the approach, several usage-related aspects (study participation, usage frequency, recording duration) and, concerning the SotD, rather functionality-related aspects (accuracy of word or face recognition, recognized emotions, congruence of verbal and facial emotion recognition) were assessed.

1. Usage-related Aspects

There were individual differences regarding the study participation and the usage frequency of the examined features. Three patients participated longer than proposed. Especially one participant was motivated to exceed the estimated study duration by seven weeks. Two of the participants were not able to participate for the originally intended duration of twelve weeks, because they were recruited only eight weeks before the project-internal assessment period ended. Consequently, no patient terminated the participation prematurely, which might indicate an overall feasible, ambulatory system. In reference to the specific features, the SA was the most feasible. Four out of five participants used it almost on a daily basis or even more frequently with multiple assessments on single days. Due to their well-established role in other monitoring approaches for bipolar disorder with comparable response rates [40], [41], this result was expected. The TPA were only used by related parties of two participants. Solely in one case there were multiple entries. Therefore, within the current study setting, this feature was not feasible. This result could

be explained within the context of the often-strained relationships of patients with bipolar disorders [21] or of the relationship status of the participants. Yet, another important issue might have been the web-based study approach that did not allow for push messages or reminders and might have been more effortful in its handling. In the future, this feature should be realized with associated smartphone applications. Thus modified TPA might be valuable and feasible with patients in reliable relationships and should be examined in larger studies, possibly conceptualized as family studies [42].

The SotD assessments were more effortful than the SA due to the installation of external study microphones and holders and because of the necessity to narrate a freely chosen story. Two of the participants still used the SotD on a regular basis (i.e., almost daily or almost every second day). The other three participants used the SotD much less frequently (i.e., on about ten to 16 percent of their participation days). Although these results show that the SotD (in its current version) might be feasible with some patients, modifications regarding its handling or the usage experience may be necessary to increase its general feasibility. The presented results further indicate that not only the SA but even more the SotD approach may be more feasible with younger patients. This confirms findings of earlier studies examining how the age of patients with bipolar disorder affects their new media usage [36], [37].

2. Functionality of the Emotion Recognition Approach

Following the insights of an earlier pre-study with healthy participants, the SotD study approach was extended by the external microphones and holders. This hardware change did improve the previously worse recording quality of the auditive information and, thereby, did increase the accuracy of the automatic transcriptions or the count of correctly recognized words to an acceptable level [43]. In concern to the video analysis, there were six assessments without working face recognition. The corresponding recordings were examined more closely and could be explained with unfavorable recording angles and a relatively low lighting. As a consequence, instructions should be even more specific. Furthermore, future systems should inform their users on missing face recognition or they should be trained to be less vulnerable for disturbances. By analyzing the expressions within both sources (auditive and visual), the recognition software found far more positive than negative emotions. Although several factors may have contributed to these results (see following sections), the recognized emotions might still reflect the expressed emotions of the participants: In order to explore further research questions, additional clinical data of the five study-patients was assessed. Most of them showed relatively stable mood throughout their study participation, including moderate (hypo) manic or mild depressive symptoms but no severe depressive episodes. Therefore, strong or persisting changes in their emotional experiences, that could have elevated the amount of expressed sadness during the SotD assessments, may not have been induced by psychopathological symptoms. Beyond that, most of the correlations regarding the emotional LIWC [38] or FACS [39] categories seem to support the consistency of the auditive and visual measures. At first glance, however, the strong negative correlations between the FACS parameters for happiness and the LIWC category of positive emotions with one participant seem conflicting.

Clinically speaking, these results could still be explained: The participant showed mild depressive symptoms and a limited emotional reactivity throughout the whole study but yet parathymic smiles during the bi-weekly clinical assessments. These deficits coincide with findings of earlier studies indicating difficulties of patients with bipolar disorders in the (facial) expression of negative emotions [13], [14]. Therefore, the emotion recognition results presented here might represent a further indication of disease-specific deficits

in the emotional expressions of bipolar patients. The verbally expressed information could have consciously been adapted to the assessment situation, while the missing correspondence within the facial expressions might have represented a more basic process of emotional avoidance. This interpretation might, in part, also explain the higher percentage of positive emotions within the FACS data of all patients (as compared to the LIWC results). Nevertheless, technical or conceptual issues may have further contributed to these results: Whereas the LIWC [38] and FACS [39] approaches delivered an empirically established framework for emotion recognition, the FACS-based video analysis might have been less sensitive to the recognition of negative emotions due to insufficient training data during the aforementioned pre-study.

Moreover, the results might suggest that short recordings of approximately two minutes or less could contain sparse emotional expressions altogether. In that case, ambient sound samples or random voice features, which have shown some potential regarding the prediction of mood state changes in bipolar disorder [27], [44]–[47], could increase the obtained information. However, as far as we know, none of the existing mobile-based approaches for bipolar disorder have analyzed the emotional content of verbal or facial expressions. It would already be difficult to realize the emotion recognition approach with random sound samples. The assessment of random video data with the emotion-sensitive approach would almost be impossible without harming the (perceived) privacy of patients. With this in mind, the SotD approach goes beyond the existing mHealth systems for bipolar disorder [34]. It allows for the analysis of even more sensitive and personal ambulatory data without harming privacy issues.

B. Acceptability

In order to evaluate the acceptability of the approach, semi-structured interviews were conducted with each participant. In accordance with earlier studies that indicate the positive attitudes of bipolar patients towards technical assistance [35], [41], [48], the participants of the pilot study perceived all three examined system features positively. As for the TPA, although only two patients included other persons into the data acquisition, the participants emphasized the importance of this feature. This coincides with comparable opinions of patients with bipolar disorder during an earlier, project-related interview study [35]. Considering the discussed and not yet perfectly solved issues of the SotD module, the positive response of the participants on this feature is quite impressive. Although only two out of five study participants used the SotD regularly (about three to six times per week), all of them were open to a more regular usage. The mentioned conditions, like an easier usage experience (e.g., without microphone or holder) or automatic reactions towards the told stories, matched the original concept of the assistance system [34]. In accordance with the process of development, these aspects had not been implemented within the study setting. They should, however, be technically convertible in the future. Thus, from the acceptability point of view, the ambulatory long-term assessment of emotional expressions in bipolar disorder should be possible.

C. General Discussion

Overall, the results of the pilot study regarding the emotion recognition approach are promising (while the TPA were not feasible). The ambulatory study setting with participation times of several weeks increased the knowledge gain regarding the acceptability and feasibility of the approach by accounting for a long-term, realistic and natural environment [49], [50]. The results indicate that emotion-sensitive systems may be feasible and well-accepted, especially with younger patients. These findings coincide with the good feasibility or acceptability of mHealth systems using self-monitoring, sensor or wearable data with bipolar patients [25], [29], [30], [40] and with the

positive attitudes of this patient group towards innovative, technical strategies for disease management [35].

The explorative approach of the pilot study with five patients, who were mostly considered as single cases, allowed for detailed insights into an ambulatory application scenario and into personal usage experiences. Although the SotD module “solely” requires its regular usage to assess active and passive emotion-related data, it comes along with more effort than common SA. Consequently, younger patients, who show a higher technical affinity [36], [37], used the study version of the SotD more frequently than older participants. And all participants were very specific about the conditions of a regular (e.g., several times per week) long-term usage. In accordance with this, future developments should be more practical (e.g., without microphones or holders), less irritating (no selfie mode) and should realize a perceived system-interaction or feedback during the SotD assessments. Furthermore, in case of the informed consent of patients and after thorough consideration of all ethical and legal implications, future systems could include the attending physicians or therapists. Thus, individual and sudden changes in the emotional expressions of patients could be thoroughly reflected during subsequent sessions.

Of course, the current approach and the small sample size with individual differences within the participation times of the patients limit the generalizability of the findings. Beyond that, detailed cost-benefit analyses would be necessary before implementing emotion-sensitive modules into disease management approaches for bipolar disorder. Therefore, more research is needed and, aside from its acceptability and feasibility, the mobile-based emotion recognition approach would have to provide valid and reliable results regarding all relevant emotions of mood episodes in bipolar disorder [3], [51]. As a consequence, it could facilitate the monitoring and understanding of emotional aspects in this disease and enable following research to examine its potential clinical value or contribution towards mood state recognition. The strong relation between emotional deficits of bipolar patients and their global and social functioning [4], [6]–[9], [13], [52] might make the gained information helpful to therapeutic approaches [13] and relapse prevention. Last but not least, a well-functioning, mobile-based emotion recognition approach could help our understanding of emotional experiences or expressions and their ramifications in other disorders as well.

V. CONCLUSION

As far as we know, the examined assistance system incorporates the first, mobile-based emotion recognition approach for bipolar disorder. Whereas the openness of patient groups towards technical or mobile-based assistance in their disease management has been investigated on several occasions, the pilot study shows that even the ambulatory assessment of audio and video data may be well-accepted and feasible. Beyond that, the approach allows for the long-term analysis of verbally and facially expressed emotions without harming the perceived privacy of patients or data privacy. Thus, the emotion-sensitive mHealth approach could affect other research areas or fields of application as well. However, to that end, some methodological and technical issues have to be addressed by future developments, and further empirical studies with larger samples of patients are necessary to increase the generalizability of the results.

AUTHOR STATEMENT

The EmAsIn project was approved by the ethical committee of the Faculty of Behavioural and Cultural Studies, Ruprecht-Karls-University Heidelberg.

ACKNOWLEDGMENT

We would like to thank the participants of our study and the technical partners of our research project EmAsIn for their support. This work was supported by the German Federal Ministry of Education and Research [grant number 16SV7357].

CONFLICTS OF INTEREST

The authors herewith declare no potential conflict of interest in respect to research, authorship and/or publication of this article.

REFERENCES

- [1] I. Grande, M. Berk, B. Birmaher, and E. Vieta, “Bipolar disorder,” *Lancet*, vol. 387, no. 10027, pp. 1561–1572, 2016.
- [2] J. V. Pinto *et al.*, “Remission and recurrence in bipolar disorder: The data from health outcomes and patient evaluations in bipolar disorder (HOPE-BD) study,” *Journal of Affective Disorders*, vol. 268, pp. 150–157, 2020.
- [3] L. A. Carolan and M. J. Power, “What Basic Emotions Are Experienced in Bipolar Disorder?,” *Clinical Psychology & Psychotherapy*, vol. 18, no. 5, pp. 366–378, 2011.
- [4] M. Paris, Y. Mahajan, J. Kim, and T. Meade, “Emotional speech processing deficits in bipolar disorder: The role of mismatch negativity and P3a,” *Journal of Affective Disorders*, vol. 234, pp. 261–269, 2018.
- [5] A. C. Bilderbeck *et al.*, “Associations between mood instability and emotional processing in a large cohort of bipolar patients,” *Psychological Medicine*, vol. 46, no. 15, pp. 3151–3160, 2016.
- [6] S. L. Johnson, C. S. Carver, and J. A. Tharp, “Suicidality in Bipolar Disorder: The Role of Emotion-Triggered Impulsivity,” *Suicide and Life-Threatening Behavior*, vol. 47, no. 2, pp. 177–192, 2017.
- [7] S. L. Johnson *et al.*, “Emotion in Bipolar I Disorder: Implications for Functional and Symptom Outcomes,” *Journal of Abnormal Psychology and Symptom Outcomes*, vol. 125, no. 1, pp. 40–52, 2015.
- [8] A. Aparicio, J. L. Santos, E. Jiménez-López, A. Bagny, R. Rodríguez-Jiménez, and E. M. Sánchez-Morla, “Emotion processing and psychosocial functioning in euthymic bipolar disorder,” *Acta Psychiatrica Scandinavica*, vol. 135, no. 4, pp. 339–350, 2017.
- [9] L. D. Branco, C. Cotrena, A. Ponsoni, R. Salvador-Silva, S. J. L. Vasconcellos, and R. P. Fonseca, “Identification and Perceived Intensity of Facial Expressions of Emotion in Bipolar Disorder and Major Depression,” *Archives of Clinical Neuropsychology*, vol. 33, no. 4, pp. 491–501, 2018.
- [10] J. Gray *et al.*, “Bipolar patients show mood-congruent biases in sensitivity to facial expressions of emotion when exhibiting depressed symptoms, but not when exhibiting manic symptoms,” *Cognitive Neuropsychiatry*, vol. 11, no. 6, pp. 505–520, 2006.
- [11] H. R. Venn *et al.*, “Perception of facial expressions of emotion in bipolar disorder,” *Bipolar Disorders*, vol. 6, no. 4, pp. 286–293, 2004.
- [12] C. M. Hoertnagl *et al.*, “Combined processing of facial and vocal emotion in remitted patients with bipolar i disorder,” *Journal of the International Neuropsychological Society*, vol. 25, no. 3, pp. 275–284, 2019.
- [13] G. Bersani *et al.*, “Facial expression in patients with bipolar disorder and schizophrenia in response to emotional stimuli: A partially shared cognitive and social deficit of the two disorders,” *Neuropsychiatric Disease and Treatment*, vol. 9, pp. 1137–1144, 2013.
- [14] I. Broch-Due, H. L. Kjaerstad, L. V. Kessing, and K. Miskowiak, “Subtle behavioural responses during negative emotion reactivity and down-regulation in bipolar disorder: A facial expression and eye-tracking study,” *Psychiatry Research*, vol. 266, pp. 152–159, 2018.
- [15] R. P. Bentall, P. Kinderman, and K. Manson, “Self-discrepancies in bipolar disorder: Comparison of manic, depressed, remitted and normal participants,” *British Journal of Clinical Psychology*, vol. 44, no. 4, pp. 457–473, 2005.
- [16] M. L. Inder, M. T. Crowe, P. R. Joyce, S. Moor, J. D. Carter, and S. E. Luty, “‘I really don’t know whether it is still there’: Ambivalent acceptance of a diagnosis of bipolar disorder,” *Psychiatric Quarterly*, vol. 81, no. 2, pp. 157–165, 2010.
- [17] L. M. Weinstock, T. Chou, C. Celis-deHoyos, I. W. Miller, and J. Gruber, “Reward and Punishment Sensitivity and Emotion Regulation Processes

- Differentiate Bipolar and Unipolar Depression,” *Cognitive Therapy and Research*, vol. 42, no. 6, pp. 794–802, 2018.
- [18] J. Pech, M. Akhøj, J. Forman, L. V. Kessing, and U. Knorr, “The impact of a new affective episode on psychosocial functioning, quality of life and perceived stress in newly diagnosed patients with bipolar disorder: A prospective one-year case-control study,” *Journal of Affective Disorders*, vol. 277, pp. 486–494, 2020.
- [19] F. Bennett *et al.*, “Predictors of psychosocial outcome of bipolar disorder: data from the Stanley Foundation Bipolar Network,” *International Journal of Bipolar Disorders*, vol. 7, no. 1, 2019.
- [20] A. López-Villarreal *et al.*, “Progression of the functional deficit in a group of patients with bipolar disorder: a cluster analysis based on longitudinal data,” *European Archives of Psychiatry and Clinical Neuroscience*, vol. 270, no. 8, pp. 947–957, 2020.
- [21] M. J. Gitlin and D. J. Miklowitz, “The difficult lives of individuals with bipolar disorder: A review of functional outcomes and their implications for treatment,” *Journal of Affective Disorders*, vol. 209, no. July 2016, pp. 147–154, 2017.
- [22] C. J. L. Murray *et al.*, “Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: A systematic analysis for the Global Burden of Disease Study 2010,” *Lancet*, vol. 380, no. 9859, pp. 2197–2223, 2012.
- [23] S. Mizziou *et al.*, “Psychosocial treatment and interventions for bipolar disorder: a systematic review,” *Annals of General Psychiatry*, vol. 14, p. 19, 2015.
- [24] E. Morton, E. E. Michalak, R. Hole, S. Buzwell, and G. Murray, “‘Taking back the reins’ – A qualitative study of the meaning and experience of self-management in bipolar disorder,” *Journal of Affective Disorders*, vol. 228, pp. 160–165, 2018.
- [25] D. Hidalgo-Mazzei, A. Mateu, M. Reinares, A. Matic, E. Vieta, and F. Colom, “Internet-based psychological interventions for bipolar disorder: Review of the present and insights into the future,” *Journal of Affective Disorders*, vol. 188, pp. 1–13, 2015.
- [26] E. Gliddon, S. J. Barnes, G. Murray, and E. E. Michalak, “Online and mobile technologies for self-management in bipolar disorder: A systematic review,” *Psychiatric Rehabilitation Journal*, vol. 40, no. 3, pp. 309–319, 2017.
- [27] M. Matthews *et al.*, “Development and Evaluation of a Smartphone-Based Measure of Social Rhythms for Bipolar Disorder,” *Assessment*, vol. 23, pp. 472–483, 2016.
- [28] C. A. Depp *et al.*, “Augmenting psychoeducation with a mobile intervention for bipolar disorder: A randomized controlled trial,” *Journal of Affective Disorders*, vol. 174, pp. 23–30, 2015.
- [29] D. Hidalgo-Mazzei *et al.*, “OpenSIMPLE: A real-world implementation feasibility study of a smartphone-based psychoeducation programme for bipolar disorder,” *Journal of Affective Disorders*, vol. 241, pp. 436–445, 2018.
- [30] M. Faurholt-Jepsen *et al.*, “Smartphone-based self-monitoring in bipolar disorder: evaluation of usability and feasibility of two systems,” *International Journal of Bipolar Disorders*, vol. 7, no. 1, pp. 1–11, 2019.
- [31] E. Mühlbauer *et al.*, “Effectiveness of smartphone-based ambulatory assessment (SBAA-BD) including a predicting system for upcoming episodes in the long-term treatment of patients with bipolar disorders: Study protocol for a randomized controlled single-blind trial 11 Medical a,” *BMC Psychiatry*, vol. 18, no. 1, 2018.
- [32] J. Zulueta *et al.*, “Predicting mood disturbance severity with mobile phone keystroke metadata: A biaffect digital phenotyping study,” *Journal of Medical Internet Research*, vol. 20, no. 7, pp. 1–10, 2018.
- [33] A. Cochran, L. Belman-Wells, and M. McInnis, “Engagement strategies for self-monitoring symptoms of bipolar disorder with mobile and wearable technology: Protocol for a randomized controlled trial,” *Journal of Medical Internet Research*, vol. 20, no. 5, 2018.
- [34] H. Daus, T. Bloecher, R. Egeler, R. De Klerk, W. Stork, and M. Backenstrass, “Development of an emotion-sensitive mobile Health approach for mood state recognition in bipolar disorder,” *JMIR Mental Health*, vol. 7, no. 7, pp. 1–10, 2020.
- [35] H. Daus, N. Kislicyn, S. Heuer, and M. Backenstrass, “Disease management apps and technical assistance systems for bipolar disorder_ Investigating the patients’ point of view,” *Journal of Affective Disorders*, vol. 229, pp. 351–357, 2018.
- [36] R. Bauer *et al.*, “Internet use by older adults with bipolar disorder,” *International Journal of Bipolar Disorders*, vol. 6, no. 1, p. 20, 2018.
- [37] R. Bauer *et al.*, “Internet use by patients with bipolar disorder: Results from an international multisite survey,” *Psychiatry Research*, vol. 242, 2016.
- [38] J. W. Pennebaker, M. E. Francis, and R. J. Booth, “Linguistic Inquiry and Word Count,” *Word Journal Of The International Linguistic Association*, pp. 1–21, 2001.
- [39] P. Ekman and W. V. Friesen, “Manual for the facial action coding system,” *Consulting Psychologists Press*, 1978.
- [40] S. Schwartz, S. Schultz, A. Reider, and E. F. H. Saunders, “Daily mood monitoring of symptoms using smartphones in bipolar disorder: A pilot study assessing the feasibility of ecological momentary assessment,” *Journal of Affective Disorders*, vol. 191, pp. 88–93, 2016.
- [41] D. Hidalgo-Mazzei *et al.*, “Psychoeducation in bipolar disorder with a SIMPLE smartphone application: Feasibility, acceptability and satisfaction,” *Journal of Affective Disorders*, vol. 200, pp. 58–66, 2016.
- [42] T. J. Rothausen, “‘Family’ in organizational research: A review and comparison of definitions and measures,” *Journal of Organizational Behavior*, vol. 20, no. 6, pp. 817–836, 1999.
- [43] M. M. Al-Aynati and K. A. Chorneyko, “Comparison of voice-automated transcription and human transcription in generating pathology reports,” *Archives of Pathology & Laboratory Medicine*, vol. 127, no. 6, pp. 721–725, 2003.
- [44] S. Abdullah, M. Matthews, E. Frank, G. Doherty, G. Gay, and T. Choudhury, “Automatic detection of social rhythms in bipolar disorder,” *Journal of the American Medical Informatics Association*, vol. 23, no. 3, pp. 538–543, 2016.
- [45] F. Or, J. Torous, and J.-P. Onnela, “High potential but limited evidence: Using voice data from smartphones to monitor and diagnose mood disorders,” *Psychiatric Rehabilitation Journal*, vol. 40, no. 3, pp. 320–324, 2017.
- [46] M. Faurholt-Jepsen *et al.*, “Voice analysis as an objective state marker in bipolar disorder,” *Translational Psychiatry*, vol. 6, p. e856, 2016.
- [47] Z. Pan, C. Gui, J. Zhang, J. Zhu, and D. Cui, “Detecting Manic State of Bipolar Disorder Based on Support Vector Machine and Gaussian Mixture Model Using Spontaneous Speech,” *Psychiatry Investigation*, vol. 15, no. 7, pp. 695–700, 2018.
- [48] K. E. A. Saunders, A. C. Bilderbeck, P. Panchal, L. Z. Atkinson, J. R. Geddes, and G. M. Goodwin, “Experiences of remote mood and activity monitoring in bipolar disorder: A qualitative study,” *European Psychiatry*, vol. 41, pp. 115–121, 2017.
- [49] T. J. Trull and U. Ebner-Priemer, “Ambulatory assessment,” *Annual Review of Clinical Psychology*, vol. 9, pp. 151–176, 2013.
- [50] U. W. Ebner-Priemer and T. J. Trull, “Ecological Momentary Assessment of Mood Disorders and Mood Dysregulation,” *Psychological Assessment*, vol. 21, no. 4, pp. 463–475, 2009.
- [51] American Psychiatric Association, “Diagnostic and statistical manual of mental disorders (5th ed.),” *American Psychiatric Publishing*, 2013.
- [52] I. M. M. Lima, A. D. Peckham, and S. L. Johnson, “Cognitive deficits in bipolar disorders: Implications for emotion,” *Clinical Psychology Review*, vol. 59, pp. 126–136, 2018.



Henning Daus

Henning Daus is psychologist and psychotherapist (CBT). He graduated from Eberhard Karls University Tübingen and worked as research associate at the Institute of Clinical Psychology at the Stuttgart Hospital. His research interests include affective disorders, psychotherapy research or mHealth developments in bipolar disorder.



Matthias Backenstrass

Matthias Backenstrass is currently head of the Institute of Clinical Psychology at the Stuttgart Hospital. He is also Professor at the Institute of Psychology, University of Heidelberg. His research interests include affective disorders, especially bipolar disorder and persistent depressive disorder, emotion regulation, and psychotherapy research. He is psychologist and psychotherapist (CBT).