Abstract

Objective: This article evaluates the potential of smartphone audio data to monitor individuals recovering from mood disorders.

Methods: A comprehensive literature review was conducted based on searches in nine bibliographic databases.

Results: Seven articles were identified that used smartphone audio data to monitor participants with bipolar disorder from four to 14 weeks. The studies captured audio data in various contexts (e.g., in-person daily conversations, phone calls) and used common audio features (e.g., pitch and volume) to ascertain clinically relevant outcomes, including mood and social rhythm. Findings suggest that the utility of audio data in clinical and research contexts remains relatively unexplored and presents some challenges. For example, information on adherence and engagement among individuals recovering from bipolar disorder were often insufficient to gauge the generalizability of findings.

Conclusions and Implications for Practice: Despite growing interest, additional research is required to confirm clinical utility of smartphone audio data for mood disorders.

Mood disorders create a great burden, yet they are undertreated and under-recognized (Whiteford, 2013). Traditionally, diagnosis of mood disorders and monitoring of recovery has relied heavily on individual self-report. The key drawback of such self–reports is recall bias, which is prevalent and exists even for recall periods as short as one day (Shiffman, Stone, & Hufford, 2008). In practice, clinicians have used more objective data, such as speech recordings of therapy sessions and vocal exercises to supplement individuals' self-reports and chart therapeutic progress (Cummins et al., 2015). People with a bipolar disorder tend to experience changes in speech patterns as a result of (hypo)manic symptoms, such as unusual talkativeness, pressurized speech, and racing thoughts, whereas people with depression tend to experience a change in speech patterns as a result of psychomotor agitation or retardation (which could affect the larynx), depressed mood, fatigue, anhedonia, or inability to concentrate (Cummins, 2015). Empirically, greater loudness, higher pitch, and faster speech have been observed in people experiencing (hypo)mania, whereas reduced pitch, slower speaking rate, and articulation errors have been observed in people in a depressed state (Cummins et al., 2015). This established relationship between voice and mood disorders using recordings from vocal exercises and therapy sessions lead to the question whether smartphone audio data could provide individuals recovering from mood disorders important clinical insights into their symptoms. In the past few years, a plethora of smartphone applications for health have emerged, yet little evidence has been published in peer-reviewed journals regarding their scientific rigor. In contrast, the popular media has focused more on the promises of smartphone audio data to diagnose various diseases. Our aim is to evaluate such claims by providing a systematic review of smartphone applications utilizing audio data in the context of mood disorders to help inform practitioners and researchers on the state of this technology for clinical care.

Methods

Using nine bibliographic databases (PubMed, Embase, Web of Science, PsycINFO, CINAHL, EconLit, PAIS, ABI, and INSPEC), we conducted a comprehensive literature review of scientific papers published by April 29, 2016, to evaluate the feasibility of using smartphone audio data to monitor mood disorders. Working with a librarian, we identified controlled vocabularies (e.g., MESH terms) for the corresponding bibliographic database and text words to capture papers that feature mobile technology and mood disorders. The controlled vocabularies included “mobile applications,” “cell(ular) phones,” “mobile device,” and “software applications”; text words such as iPhone and Android phone; and all synonyms of smartphone applications to capture literature on mobile technology. We identified literature on mood disorders using controlled vocabulary and text words such as “affective disorder,” “dysthymic disorder,” “bipolar disorder,” “major depression,” and “mania.” We did not impose any limits on the date of publication. The inclusion criteria were use of smartphone audio data for the monitoring of mood disorders in at least one human subject who had received a clinical diagnosis of a mood disorder, whereas the exclusion criteria were papers written in languages other than English, no mention of smartphone applications, non-empirical studies (i.e., comments, opinions, narratives), empirical studies that involved non-human subjects, studies that used “app” in a different meaning (e.g., as an abbreviation for a protein), or dealt with a subject matter other than mood disorders. Other exclusion criteria included use of ecological momentary assessment (EMA) because a review on EMA for mood disorders has been recently published (Faurholt-Jepsen, Munkholm, Frost, Bardram, & Kessing, 2016). To ensure specificity and relevance of this rapidly progressing area of research, we also excluded computer-based, web-based, or text-based studies, or effects of smartphone use (e.g., addiction), as well as applications designed for healthy individuals or treatment purposes. The first author screened the titles and abstracts of 1845 non-duplicate articles for relevance, and excluded 1722 papers based on the stated exclusion criteria. The full texts of the remaining 123 records were assessed based on the inclusion and exclusion criteria, seven of which were included in this review.

Results

The included studies aimed to demonstrate the potential use of smartphone audio data (Guidi et al., 2015) to ascertain mood states (Grunerbl et al., 2015; Karam et al., 2014; Muaremi, Gravenhorst, Grunerbl, Arnrich, & Troster, 2014; Osmani, 2015), mood and energy instances, and social rhythm (Abdullah et al., 2016) in people with a bipolar disorder. The gold standards used for these outcomes were participants' self-reports and clinical assessments using instruments such as the Young Mania Rating Scales (YMRS) and the Hamilton Rating Scale for Depression (HAMD). All studies provided participants with an Android study phone as a data collection tool, and no study made use of individuals' personal phones. Audio data were captured from subjects when they read a provided text or supplied a verbal description of pictures presented on the screen (Guidi et al., 2015), during in-person daily conversation (Abdullah et al., 2016) and during day-to-day phone calls with (Karam et al., 2014) or without (Grunerbl et al., 2015; Maxhuni, 2016; Muaremi et al., 2014; Osmani, 2015) calls with clinicians over a period of four to 14 weeks. None of these studies evaluated speech content, only statistical features of speech, and raw audio data were not stored in some studies to protect the privacy of the participants (Abdullah et al., 2016). Commonly examined statistical audio features included number of conversations, speaking length, pitch, and volume, all of which were used to ascertain clinically relevant outcomes, such as mood and social rhythm.

While it is feasible to use smartphone applications (e.g., MoodRhythm, MONARCA) to facilitate the collection of audio data in clinical populations, findings from our review suggest that the clinical utility of audio data is currently undetermined, where clinical utility is defined in terms of the usefulness, benefits and drawbacks of the technology to clinical practice, and how these different components should be weighed against one another (Smart, 2006). Most studies were pilots and accordingly employed small sample sizes (e.g., (Guidi et al., 2015)) and carried out mostly exploratory data analyses. The potential effects of various medications on the participants' voice were not considered in these studies. While the use of study phone might help streamline research protocol, we speculate that participants might have used study phones and personal phones differently, especially if the model of the study phone was different from their personal phone (Belisario, 2015). This possible discrepancy in phone usage patterns may give rise to selection bias. In one clinical study, a third of the participants with bipolar disorder refused to use study phones to make calls; some participants switched off sensors to save battery; and only data from approximately half of the participants were included in the statistical analyses across all papers (Grunerbl et al., 2015; Maxhuni, 2016; Muaremi et al., 2014; Osmani, 2015). In addition, sparse clinical assessments limited the availability of ground truth data relative to collected audio data (Grunerbl et al., 2015; Muaremi et al., 2014; Osmani, 2015).

Despite these limitations, audio features from daily calls allowed for classification of people with a bipolar disorder into one of seven possible mood states, ranging from severe depression to severe mania, with a classification accuracy of 70-82% (Grunerbl et al., 2015; Maxhuni, 2016; Muaremi et al., 2014; Osmani, 2015). These numbers are comparable to the 66-77% accuracy that was achieved with call metadata (non-audio data) alone, including call time, duration, and number of unique contacts (Grunerbl et al., 2015; Muaremi et al., 2014; Osmani, 2015). A combination of audio data and audio metadata did not necessarily improve classification accuracy (69-83%) (Grunerbl et al., 2015; Muaremi et al., 2014; Osmani, 2015); see Table 1. Since all of the reviewed studies used study phones, we speculate that use of personal smartphones could improve the information content of audio metadata. While smartphone applications enable behavioral monitoring beyond the clinic (Ben-Zeev, Scherer, Wang, Xie, & Campbell, 2015), audio data seem to differ in clinical utility by the context of the recording. Audio data collected from clinical interviews by phone were found to better differentiate depression and hypomania from euthymic state, whereas audio data collected from calls outside the clinic had difficulties differentiating depression from euthymia (Karam et al., 2014). Smartphone audio data collected during phone calls on the day of the clinical assessment did not improve mood-state classification by audio data from clinical assessment alone (Karam et al., 2014). Furthermore, contrary to the expectation that increased speech frequency was associated only with (hypo)mania, increased speech frequency in audio was also observed during the transition from hypomania to depression (Guidi et al., 2015). Finally, weak correlations were reported between audio features and mood (Abdullah et al., 2016; Guidi et al., 2015). The findings from the reviewed publications suggest that GPS, accelerometer, and other phone usage and sensor data could provide superior performance at classification and prediction of relevant clinical outcomes without requiring the collection of audio data from individuals recovering from a bipolar disorder (Grunerbl et al., 2015; Maxhuni, 2016; Muaremi et al., 2014; Osmani, 2015).

Conclusions

While use of smartphone audio data to monitor recovery for individuals living with a mental illness holds potential to facilitate personalized care, our review suggests that many fundamental scientific questions still need to be addressed before it can become a validated clinical tool. Few studies reported data on adherence and engagement, leaving questions open about whether those with mental illnesses are willing to have their audio data captured by their smartphone in the long term. Furthermore, properties of the collected data, particularly the degree of missingness, were often not described in sufficient detail to judge the robustness of the conclusions of many of the studies. One drawback of using audio data in general is that it requires active input from the participants, and their willingness to contribute such data is likely influenced by their mood state. Consequently, audio data might not be available when, for example, an individual recovering from a manic episode switches into depression, isolating himself from social contact and his phone. Further research is needed to examine how study phones versus personal phones capture fluctuations in speech patterns and other daily activities; how audio data captured from reading fixed passages of text versus those captured during phone calls differ in clinical utility; and finally how to model missing audio data and the potential biases that it presents. While the included studies found audio features of phone calls to outperform characteristics of call metadata, studies have yet to utilize call log metadata which are more often readily available and complete, without some of the legal and ethical complications of capturing sensitive and personal audio data. Despite growing interest in using audio data from smartphones in psychiatry, we conclude that additional research in clinical populations is required to confirm clinical utility of smartphone audio data for the diagnostic purpose of mood disorders. While smartphone-based voice technologies hold great potential for mental health, because of the many remaining questions, practitioners should assume caution if considering using them today as part of illness management.

References

Abdullah, S., Matthews, M., Frank, E., Doherty, G., Gay, G., & Choudhury, T. (2016). Automatic detection of social rhythms in bipolar disorder. *Journal of American Medical Informatics, 23*(3), 538-543. doi:10.1093/jamia/ocv200

Belisario, J. S. M. J. J., Huckvale K, O'Donoghue J, Morrison CP, Car J. (2015). Comparison of self-administered survey questionnaire responses collected using mobile apps versus other methods. *Cochrane Database Systematic Review, 27*(7).

Ben-Zeev, D., Scherer, E. A., Wang, R., Xie, H., & Campbell, A. T. (2015). Next-generation psychiatric assessment: Using smartphone sensors to monitor behavior and mental health. *Psychiatric Rehabilitation Journal, 38*(3), 218-226. doi:10.1037/prj0000130

Cummins, N., Scherer, S., Krajewski, J., Schnieder, S., Epps, J., & Quatieri, T. F. . (2015). A review of depression and suicide risk assessment using speech analysis. *Speech Communication, 71*, 10-49.

Faurholt-Jepsen, M., Munkholm, K., Frost, M., Bardram, J. E., & Kessing, L. V. (2016). Electronic self-monitoring of mood using IT platforms in adult patients with bipolar disorder: A systematic review of the validity and evidence. *BMC Psychiatry, 16*(1), 7.

Grunerbl, A., Muaremi, A., Osmani, V., Bahle, G., Ohler, S., Troster, G., . . . Lukowicz, P. (2015). Smartphone-based recognition of states and state changes in bipolar disorder patients. *IEEE Journal of Biomedical and Health Informatics, 19*(1), 140-148. doi:10.1109/jbhi.2014.2343154

Guidi, A., Salvi, S., Ottaviano, M., Gentili, C., Bertschy, G., de Rossi, D., . . . Vanello, N. (2015). Smartphone application for the analysis of prosodic features in running speech with a focus on bipolar disorders: System performance evaluation and case study. *Sensors (Basel), 15*(11), 28070-28087. doi:10.3390/s151128070

Karam, Z. N., Provost, E. M., Singh, S., Montgomery, J., Archer, C., Harrington, G., . . . Ieee. (2014). Ecologically valid long-term mood monitoring of individuals with bipolar disorder using speech *2014 Ieee International Conference on Acoustics, Speech and Signal Processing*. New York: Ieee.

Maxhuni, A., Muñoz-Meléndez, A., Osmani, V., Perez, H., Mayora, O., & Morales, E. F. . (2016). Classification of bipolar disorder episodes based on analysis of voice and motor activity of patients. *Pervasive and Mobile Computing*.

Muaremi, A., Gravenhorst, F., Grunerbl, A., Arnrich, B., & Troster, G. (2014). Assessing bipolar episodes using speech cues derived from phone calls. In P. Cipresso, A. Matic, & G. Lopez (Eds.), *Pervasive Computing Paradigms for Mental Health* (Vol. 100, pp. 103-114). Berlin: Springer-Verlag Berlin.

Osmani, V. (2015). Smartphones in mental health: detecting depressive and manic episodes. *Ieee Pervasive Computing, 14*(3), 10-13.

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology, 4*, 1-32.

Whiteford, H. A., Degenhardt, L., Rehm, J., Baxter, A. J., Ferrari, A. J., Erskine, H. E., ... & Burstein, R. . (2013). Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. *The Lancet, 382*(9904), 1575-1586.