"Hear me out": Smart Speaker Based Conversational Agent to Monitor Symptoms in Mental Health

Raju Maharjan, Per Bækgaard, Jakob E. Bardram
Copenhagen Center for Health Technology
Technical University of Denmark
Lyngby, Denmark
{rajm,pgba,jakba}@dtu.dk

ABSTRACT

Difference in features of voice such as tone, volume, intonation, and rate of speech have been suggested as sensitive and valid measures of mental illness. Researchers have used analysis of voice recordings during phone calls, response to the IVR systems and smartphone based conversational agents as a marker in continuous monitoring of symptoms and effect of treatment in patients with mental illness. While these methods of recording the patient's voice have been considered efficient, they come with a number of issues in terms of adoption, privacy, security, data storage etc. To address these issues we propose a smart speaker based conversational agent – "Hear me out". In this paper, we describe the proposed system, rationale behind using smart speakers, and the challenges we are facing in the design of the system.

CCS CONCEPTS

• **Human-centered computing → Sound-based input / output**.

KEYWORDS

Smart Speakers; Mental Health; Voice User Interface; Voice Feature Analysis; Conversational Agent; Patient Self Reports

ACM Reference Format:
Raju Maharjan, Per Bækgaard, Jakob E. Bardram. 2019. "Hear me out": Smart Speaker Based Conversational Agent to Monitor Symptoms in Mental Health. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2019 International Symposium on Wearable Computers (UbiComp/ISWC '19 Adjunct), September 9–13, 2019, London, United Kingdom.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3341162.3346270

UbiComp/ISWC '19 Adjunct, September 9–13, 2019, London, United Kingdom

ACM ISBN 978-1-4503-6869-8/19/09.
https://doi.org/10.1145/3341162.3346270

1 INTRODUCTION

Effective monitoring of symptoms can help mentally ill patients recognize so-called 'early warning signs' resulting in timely clinical interventions [1]. The most common methods of monitoring and assessing mental illness are based on clinical evaluation including interviews, questionnaires (e.g., the 17-item Hamilton Depression Rating Scale (HAMD) [7] and the Young Mania Rating Scale (YMRS) [22]) and self-reports reflecting the patient's mental state (e.g., mood, stress or other symptoms). Studies have shown that these methods are subject to individual observer bias, lack of ecological validity due to recall bias, and incomplete assessment of behavior [2, 8].

To address these issues, researchers have leveraged the advancement in smartphone technology enabling them to collect Ecological Momentary Assessments (EMA) from the patients in real-time, in real-world settings, over time and across contexts [18]. Smartphone based self monitoring systems like MONARCA [1], PSYCHE [21], MORIBUS [17] and other similar systems allow continuous collection of automatically generated objective data on behavioral activities (e.g., mobility, log of text messages and phone calls per day) as well as self-reported subjective data (e.g., mood, sleep length, activity level, stress, and medicine compliance). Given the fact that there is a lack of biomarkers for mental illness [9], analysis of these data seems to be a highly promising approach to help mentally ill patients manage and cope with their diseases [1].

A recent review of the literature suggests that self-reports in these types of systems are often collected from the patients in graphical and textual forms [5] which do not capture emotional nuances such as tone, volume, intonation, and rate of speech present in the human voice. These features of voice reflect the patient's emotional state which is a significant indicator of mental health conditions [15, 19]. Further, variations of speech have been suggested as one of the sensitive and valid measures of mental illness such as bipolar disorder and depression [13].

Recently, researchers have analyzed voice features as an objective state marker of the patient's mental health by sampling from the patient's phone calls (e.g. [6]), interactive

voice responses (IVRs) (e.g. [13]), and smartphone based conversational agents (e.g. [12]). While these methods of collecting voice samples have been considered efficient, there exists a number of issues in terms of adherence, privacy, data storage and security [6, 20]. For example, Faurholt-Jepsen et al. [6] reported that the participants did not carry their smartphones with them at all times, calling from other devices and thereby not providing voice features during all their phone calls. It can be speculated that this method is not fully acceptable to the participants, perhaps due to the intrusiveness and privacy concerns. Further, the large number of voice features collected in the study proved to be a challenge in terms of computational costs and storage space. Likewise, Torous et al. called for a need to design and employ these technologies focusing on the user's perspective to increase adherence and acceptance [20].

With an aim to address these issues, we propose a smart speaker based conversational agent – "Hear me out" – that will collect voice features of the patient's self-reports on *sleep, mood and activity level* in the home settings. We hypothesize that actively asking patients to self-report at his/her comfort while at home is less intrusive than other methods. Therefore, the proposed system is designed to record voice features exclusively when the patient self-reports, which will increase adherence and help reduce the burden of data storage and computational cost.

In this paper, we describe (1) the proposed system, (2) rationale behind using smart speakers, and (3) the challenges we are facing in the design process. During the workshop, we would like to get valuable feedback on the design of the system and suggestions to overcome our technical and methodological challenges.

2 "HEAR ME OUT"

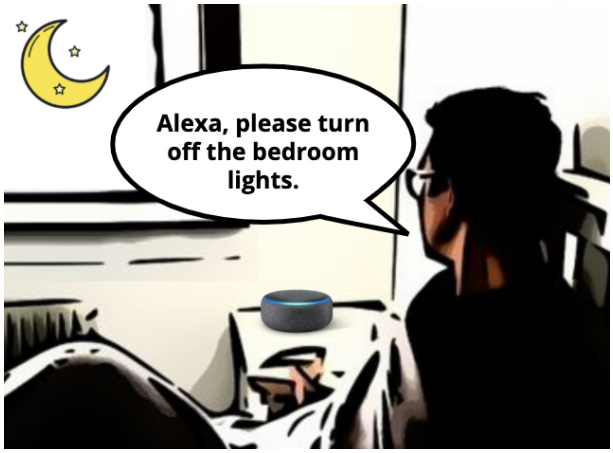
We took a scenario based approach [4] to design the proposed system that records self-reported data on sleep, mood and activity level from the patients. We interviewed two psychiatrists, and conducted a focus group which included general practitioners, psychologists, and an anthropologist. Based on the interviews and the focus group, we developed a patient persona (John Doe, a 40 years old patient recently diagnosed with bipolar disorder, unemployed, lives by himself, spends most of his time at home and does not have many friends to talk to) and scenarios 1, 2 and 3.

In scenario 1, Alexa will ask the patient about the length and quality of the sleep in the morning when he wakes up to the alarm; in scenario 2, it will ask him about the mood and the level of activity in the evening when he asks Alexa to turn off the light to go to bed; and in scenario 3, he will be prompted about the missed notification(s) asking him to record the subjective data retrospectively (will be allowed up to 2 days).

Table 1: Scenario 1– 7:00 AM. John has set an alarm in Alexa to wake him up at 7:00 AM. The alarm goes off ...

Alexa:	Good Morning John, How did you sleep?
John:	I slept well.
Alexa:	And how long did you sleep?
John:	About 8 hours.
Alexa:	Sounds good. Have a great day.

Table 2: Scenario 2– 10:00 PM. The patient is in the bedroom where the lighting is controlled by Alexa.



Alexa:	Hi John, How was your day today?
John:	It was alright [with a sigh...]
Alexa:	How was your mood like?
John:	It was good — most of the time.
Alexa:	And, how active were you?
John:	Wasn't very active today.
Alexa:	Alright then, have a good night!

Table 3: Scenario 3– 7:00 PM. The patient sees the orange light in Alexa indicating missed notification(s).

Alexa:	You did not report your mood and activity yesterday. Do you want to report them now?
John:	Sure.
Alexa:	Okay. Tell me how was mood yesterday?
John:	I was cheerful most of the time.
Alexa:	And, how active were you?
John:	I was very active. I went on a hike.
Alexa:	Okay, thank you for answering the questions.

The system will consist of a smart speaker, Alexa[1] installed with a custom skill[2] and an Android smartphone with a voice recording app. In this set up, smartphone will record the conversation between the patient and Alexa. In order to capture a wider spectrum of the voice features, the system will allow the patient to self-report the subjective data in unconstrained natural language. The conversation can be initiated either by the patient or Alexa. In Figure 1, (A) represents the conversation initiated by Alexa as shown in scenario 1, (B) represents the conversation initiated by the patient as shown in scenario 2, and scenario 3, and (C) represents the recording of the conversations on the smartphone.

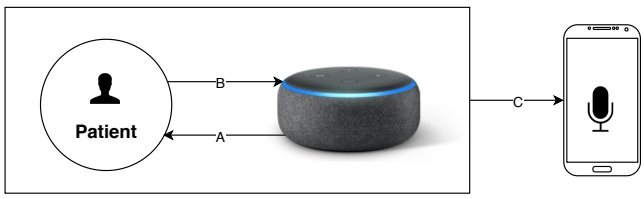


Figure 1: "Hear me out" system overview

[1]https://developer.amazon.com/alexa

[2]https://developer.amazon.com/en-US/alexa/alexa-skills-kit

3 NOW WE ARE TALKING

The ability of psychologists and clinicians to directly observe and intervene mentally ill patients in their homes has been limited by methodological and logistical barriers [14]. Smart speakers may, however, offer an opportunity for observing and intervening in homes, where they are most used [11], thereby providing new insights into patients' mental health and opportunities for home-based care.

Furthermore, the rapid growth in adoption of smart speakers due to high user acceptance and purchase rate has made these devices a dominant tool in everyday life [10] such as for reminders and scheduling, playing music, phone calls, and online shopping [11]. A recent survey reflects a 40% rise in smart speaker ownership over the previous year and reports 133 million smart speakers in use as of 2019 [10]. It is projected to top 175 million by 2020 in the U.S alone [16]. This trend suggests that these devices will soon become ubiquitous, as have smartphones and smartwatches which have fostered research and interventions in mental health [1, 17, 21].

Another reason for using smart speakers is that they are usually connected to a continuous power supply, which reduces the risk of failing to collect data due to battery-life limitations of the smartphone-based solutions applied currently. Also, this takes the burden off the patient's shoulder of frequently charging the battery drained by continuous sensing and recording of the data from his/her smartphone.

4 ENOUGH TALKING

Our formative assessment of the proposed system with clinicians (N=5), including psychologists, general practitioners and an anthropologist, indicated that the proposed system has potential in extending the current personal monitoring systems. However, we have identified a set of technical and methodological challenges that we would like to discuss in this workshop.

Alexa records all the conversations when the recording is not turned off. These recordings can be accessed to be listened to and deleted. However, the recordings cannot be downloaded, which prevents any custom / offline analysis of the recordings both in terms of voice feature analysis, speech recognition, and content-based analysis. We did not find any other smart speaker that allows downloading of the voice recordings. We would like to know if there are any other smart speakers with similar features like Alexa that additionally allows downloading of the voice recordings. Our system design couples Alexa with an Android smartphone to overcome the limitation of downloading the voice samples. In an attempt to couple Alexa with the smartphone, we leveraged Alexa's built-in "routines" feature. As shown in Figure 2, (C) represents the trigger to the IFTTT [3] applet, initiated by (A)

[3]https://ifttt.com

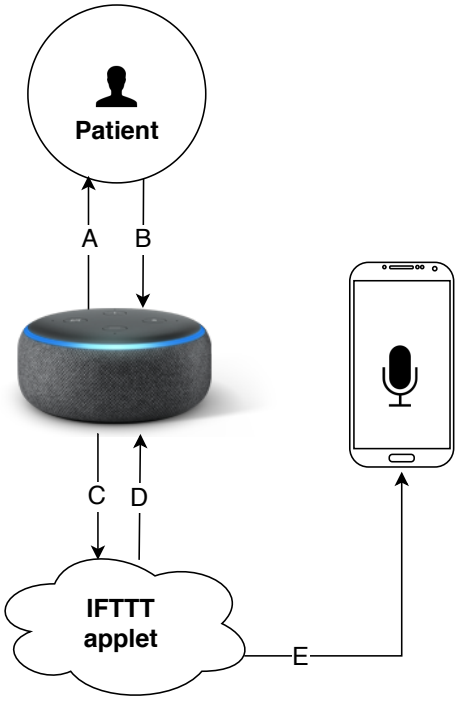


Figure 2: Our attempt to couple Alexa with a smartphone via an IFTTT applet.

Alexa or (B) the patient to trigger (D) the custom skill in Alexa, and to start (E) voice recording on the smartphone. This method to couple Alexa with the smartphone was not successful due to the technical limitation that did not allow the "routines" and the IFTTT applet to trigger the custom skill in Alexa. Therefore, in this workshop, we would like to discuss the possible methods to integrate Alexa and the smartphone which will address the above issues.

Furthermore, the aforementioned scenarios are set in the patient's bedroom. Given the skepticism on privacy and the admittance by Amazon that they review the voice recording clips in an effort to improve user experience [3], bedroom may not be the most suitable room to place the proposed system. We would like to take this workshop as an opportunity to discuss other possible scenarios to collect self-reports using smart speakers.

Finally, we would like to discuss the privacy implication of the proposed system, and more broadly its applicability especially within mental health. During our design process, the design seemed to fit patients with affective disorders (unipolar and bipolar depression), but it was also discussed that this solution would not fit all types of mentally ill patients. For example, a patient suffering from schizophrenia might not be comfortable with a voice interface.

5 WHAT NEXT?

While conversational agents in smartphones have been considered efficient, smart speakers provide a unique user experience due to its static nature in the home environment. Therefore, it is worth exploring how it can be used along with current personal monitoring systems. With the knowledge we get from the workshop, we hope that we will be able to overcome the challenges we are facing to design the proposed system. We aim to implement a prototype of the system and investigate its feasibility in monitoring and diagnosis of mental health diseases by running a deployment study. We hope that the proposed system will call attention to smart speaker technology and provide an understanding of its usage in mental health.

ACKNOWLEDGMENTS

This project is financially supported by the Novo Nordisk Foundation, Denmark and the Copenhagen Center for Health Technology.

REFERENCES

[1] Jakob E. Bardram, Mads Frost, Károly Szántó, Maria Faurholt-Jepsen, Maj Vinberg, and Lars Vedel Kessing. 2013. Designing Mobile Health Technology for Bipolar Disorder: A Field Trial of the Monarca System. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 2627–2636. https://doi.org/10.1145/2470654.2481364

[2] Roy F Baumeister, Kathleen D Vohs, and David C Funder. 2007. Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science* 2, 4 (2007), 396–403.

[3] Natalia Drozdiak By Matt Day, Giles Turner. 2019. Amazon Workers Are Listening to What You Tell Alexa. Retrieved June 27, 2019 from https://www.bloomberg.com/news/articles/2019-04-10/is-anyone-listening-to-you-on-alexa-a-global-team-reviews-audio

[4] John M Carrol. 1999. Five reasons for scenario-based design. In *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences. 1999. HICSS-32. Abstracts and CD-ROM of Full Papers*. IEEE, 11–pp.

[5] Ezgi Dogan, Christian Sander, Xenija Wagner, Ulrich Hegerl, and Elisabeth Kohls. 2017. Smartphone-based monitoring of objective and subjective data in affective disorders: where are we and where are we going? Systematic review. *Journal of medical Internet research* 19, 7 (2017), e262.

[6] Maria Faurholt-Jepsen, Jonas Busk, Mads Frost, Maj Vinberg, Ellen M Christensen, Ole Winther, Jakob Eyvind Bardram, and Lars V Kessing. 2016. Voice analysis as an objective state marker in bipolar disorder. *Translational psychiatry* 6, 7 (2016), e856.

[7] MAX Hamilton. 1967. Development of a rating scale for primary depressive illness. *British journal of social and clinical psychology* 6, 4 (1967), 278–296.

[8] Gabriella M Harari, Nicholas D Lane, Rui Wang, Benjamin S Crosier, Andrew T Campbell, and Samuel D Gosling. 2016. Using smartphones to collect behavioral data in psychological science: Opportunities, practical considerations, and challenges. *Perspectives on Psychological Science* 11, 6 (2016), 838–854.

[9] Nicholas C Jacobson, Hilary Weingarden, and Sabine Wilhelm. 2019. Digital biomarkers of mood disorders and symptom change. *npj Digital Medicine* 2, 1 (2019), 3.

[10] Bret Kinsella. 2019. U.S. Smart Speaker Ownership Rises 40% in 2018 to 66.4 Million and Amazon Echo Maintains Market Share Lead Says New Report from Voicebot. Retrieved June 27, 2019 from https://voicebot.ai/2019/03/07/u-s-smart-speaker-ownership-rises-40-in-2018-to-66-4-million-and-amazon-echo-maintains-market-share-lead-says-new-report-from-voicebot/

[11] Shanhong Liu. 2019. Smart speakers - Statistics Facts. Retrieved June 27, 2019 from https://www.statista.com/topics/4748/smart-speakers/

[12] Adam S Miner, Arnold Milstein, Stephen Schueller, Roshini Hegde, Christina Mangurian, and Eleni Linos. 2016. Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health. *JAMA internal medicine* 176, 5 (2016), 619–625.

[13] James C Mundt, Adam P Vogel, Douglas E Feltner, and William R Lenderking. 2012. Vocal acoustic biomarkers of depression severity and treatment response. *Biological psychiatry* 72, 7 (2012), 580–587.

[14] Benjamin W Nelson and Nicholas B Allen. 2018. Extending the passive-sensing toolbox: using smart-home technology in psychological science. *Perspectives on Psychological Science* 13, 6 (2018), 718–733.

[15] Stanley Newman and Vera G Mather. 1938. Analysis of spoken language of patients with affective disorders. *American journal of psychiatry* 94, 4 (1938), 913–942.

[16] Sarah Perez. 2019. Voice-enabled smart speakers to reach 55% of U.S. households by 2022, says report. Retrieved June 27, 2019 from https://techcrunch.com/2017/11/08/voice-enabled-smart-speakers-to-reach-55-of-u-s-households-by-2022-says-report

[17] Darius A Rohani, Nanna Tuxen, Andrea Quemada Lopategui, Maria Faurholt-Jepsen, Lars V Kessing, and Jakob E Bardram. 2019. Personalizing Mental Health: A Feasibility Study of a Mobile Behavioral Activation Tool for Depressed Patients. In *Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth'19)*. ACM, New York, NY, USA, 282–291. https://doi.org/10.1145/3329189.3329214

[18] Saul Shiffman, Arthur A Stone, and Michael R Hufford. 2008. Ecological momentary assessment. *Annu. Rev. Clin. Psychol.* 4 (2008), 1–32.

[19] Steven R Steinhubl and Eric J Topol. 2018. Now we're talking: bringing a voice to digital medicine. *The Lancet* 392, 10148 (2018), 627.

[20] John Torous, Paul Summergrad, and S Nassir Ghaemi. 2016. Bipolar disorder in the digital age: new tools for the same illness. *International journal of bipolar disorders* 4, 1 (2016), 25.

[21] Gaetano Valenza, Claudio Gentili, Antonio Lanatà, and Enzo Pasquale Scilingo. 2013. Mood recognition in bipolar patients through the PSYCHE platform: preliminary evaluations and perspectives. *Artificial intelligence in medicine* 57, 1 (2013), 49–58.

[22] Robert C Young, Jeffery T Biggs, Veronika E Ziegler, and Dolores A Meyer. 1978. A rating scale for mania: reliability, validity and sensitivity. *The British Journal of Psychiatry* 133, 5 (1978), 429–435.