A Virtual Negative Symptom Schizophrenic Patient for Psychiatrist Training

Ajay Vishwanath, Shihao Xu, Lee Kar Chuen, Tan Jing Yi, Lee Jun Jie, Clifton Choo, Chua Yi Han Victoria, Justin Dauwels and Nadia Thalmann

Abstract Negative symptoms of schizophrenia include symptoms such as poverty of speech, blunted affect and prolonged response time. We develop a system which can simulate these symptoms in the form of a virtual character. Being autonomous, our character interacts with a person, say a psychiatrist, and delivers appropriate responses. These responses are triggered based on a perception-decision-action loop which processes the psychiatrist's dialog, body language, emotion, and audio signal. Additionally, these signals are used to objectively analyze the interaction. This analysis, coupled with a simulation of negative symptoms, can be used as an educational platform to train a psychiatrist. In this paper, we propose such a system and detail our approach to make apparent these symptoms in our virtual character.

1 Introduction

According to the latest report [1] by World Health Organization (WHO), there are only 9 mental health workers per 100,000 population globally. This number is even lower [1] in developing countries. WHO, along with several mental health institutes, currently recommends [2] to train non-mental health workers in identifying and treating mental health issues. This brings about a need for more training programmes and, additionally, patients for the trainee to practice upon. In graduate and undergraduate medical education, use of Simulated Patient [6] is widely adopted in psychiatry schools. Despite being a step above using real patients, it comes with evaluation obstacles [5] and lack of empathy [6] between the actor and trainee.

Ajay Vishwanath
Institute for Media Innovation, Nanyang Technological University, Singapore e-mail: avishwanath@ntu.edu.sg

Shihao Xu
School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore e-mail: xush0019@ntu.edu.sg

We move a step further by introducing a method to coach a psychiatrist to give a more objective assessment on schizophrenia patients. This system consists of a semi-structured interview to simulate the negative symptoms of schizophrenia. The avatar dialog system is used to simulate the real diagnostic interview, which aims to give a psychiatrist trainee better understanding of the patient's symptoms. In addition, our system could also help them by giving more accurate measurement and assessment of how they deal with the patient. We include the different severities of schizophrenic symptoms to the avatar by manipulating its dialog, conversational cue, vocal features, and affective system. In this paper, we develop two types of avatars, *healthy* and *patient* avatars, which refer to our avatar with default setting and a patient diagnosed with negative symptoms of schizophrenia, respectively. In our platform [7], we can alter the chatbot, affective system, speech and gaze of the avatar. In addition, the evaluation of the psychiatrist can be performed by the storage of audio, skeletal, video and language features. Thus, the result of objective evaluation from these features is will act as an aid to the subjective evaluation of an examiner.

Virtual character Justina [4] offers encouragement to pursue our work by using a virtual patient with Post Traumatic Stress Disorder (PTSD) and providing a framework to evaluate the psychiatrist trainee. However, it is not an end-to-end modular solution that can be replaced by a new chatbot corpus or manipulate the severity of symptoms. On the other hand, there are works such as SimSensei Kiosk [3] which develops a virtual human psychiatrist, designed to create an engaging face-to-face interaction with patients having anxiety, depression, and PTSD. Similar to SimSensei, our virtual character is fully autonomous and expressive, but calming, empathetic speaker and attentive listener [3]. Except that our virtual character is the patient with negative symptom schizophrenia. In the next section, we outline the architecture and various modules of our platform.

2 System Architecture

Our platform[7] is designed based on a perception-decision-action loop, which is shown in Fig. 1. The perception module comprises devices to capture audio and visual signals from the environment. The decision module is a policy-based system which uses the audio-visual features. Using these policies, the action module renders the appropriate response, which could be in the form of speech, animation or facial expression. We use this architecture for both, *healthy* and *patient* avatars.

2.1 Perception

We use a standard microphone to input speech into the platform and a Microsoft Kinect 2.0[1] to capture the RGB, skeletal and depth videos. When the interviewer starts to interact with the avatar, the signals are processed to obtain cues such as

[1] https://developer.microsoft.com/en-us/windows/kinect/

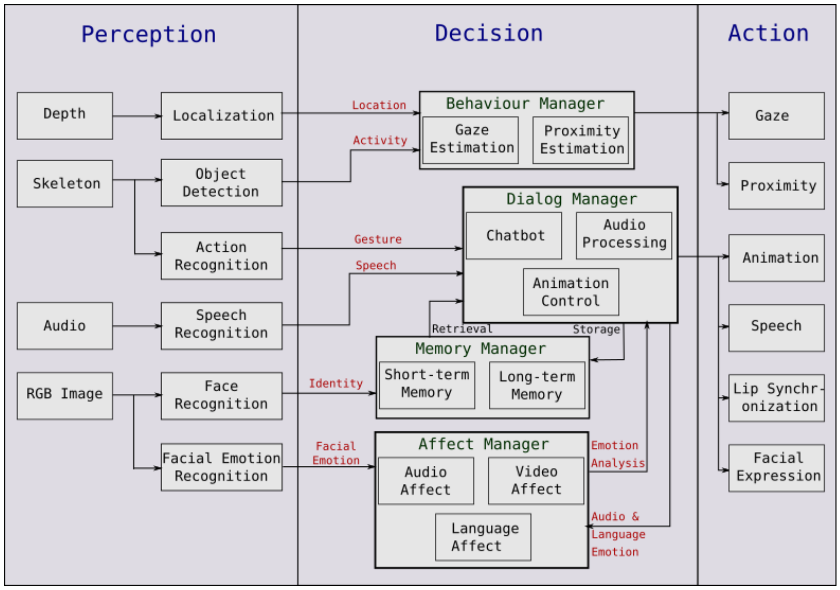

Fig. 1 System architecture diagram which summarizes the components of the perception-decision-action loop.

location, **speech**, **gesture** and **activity**. Also, we use OpenSmile[2] for audio feature extraction, Affectiva[3] for facial emotion recognition and OpenFace[4] for face recognition.

2.2 Decision

Decisions on the processed signals are based on a pre-defined policy. This starts with the memory manager in Figure 1, which interacts with the dialog manager for storage and retrieval of information. A psychiatrist can register onto our system using face authentication and the OpenFace eigenvector representation is saved in the long-term memory. To aid real-time retrieval, the short-term memory loads information about the psychiatrist, such as audio features, emotions, gestures and conversation logs from its long-term counterpart. In addition, it adds context to generic chatbot responses such as inserting the name of the psychiatrist, inserting adjectives referencing past conversation etc.

The affect manager determines the emotion of the patient using facial, verbal and audio cues. This is crucial because we need to manipulate the facial expression and tone of the avatar. Using the Personality Characterised Mood Dynamics model [8], we calculate the values of Pleasure, Arousal and Dominance to determine the degree

[2] https://github.com/naxingyu/opensmile

[3] https://www.affectiva.com/product/emotion-sdk/

[4] https://cmusatyalab.github.io/openface/

of effect of the avatar. For the *healthy* and *patient* avatars, we present the personality and initial mood to match the behaviour of an average person and a schizophrenic patient respectively. This shall be further explored in the next section.

The location of the psychiatrist is determined using Kinect skeletal data. Based on this data, the behaviour manager sets the head gaze and proximity of the avatar. Finally, the dialog manager, which is a policy-based decision center of our platform controls the speech, animation, facial expression and lip synchronization of the avatar based on inputs from the effect and memory manager. The avatar response is obtained from a chatbot developed using Chatterbot[5]. In addition, the animations and facial expressions of the avatar are triggered, based on user's speech, pose and past conversations. We vary the animations, chatbot, facial expression and speech to distinguish between the *healthy* and *patient* avatars.

2.3 Action

The avatar is rendered using Unity3D[6] to which we communicate using Thrift[7]. Using Maya, we create appropriate animation for the *healthy* and *patient* avatars. Based on the policies in the Decision module, the avatar animation, facial expression, gaze, lip synchronization and speech are determined. Currently, we use 6 facial expressions, 10 animations and 4 voice modulations.

3 Development

The *healthy* and *patient* avatars are designed based on the above perception-decision-action loop, to coach the psychiatrists. The typical negative symptoms [10] (*alogia*, social withdrawal, apathy, lack of motivation and drive, emotional blunting, and attention impairment) are embedded into the *patient* avatar by manipulating the social behaviour cues. This makes these symptoms observable and eventually assessed by the interviewer. The following subsections summarize specific behavioural cues of both *healthy* and *patient* avatars and how they are used to differentiate a negative symptom schizophrenic patient and healthy person.

3.1 Verbal cues

Most of the negative symptoms like apathy, social withdrawal, and lack of motivation and drive could be observed in the conversation dialogues. Based on the questions of negative symptom assessment interview [10] designed by psychiatrists in (Institute of Mental Health) IMH, we manually developed two Question and Answer (Q&A) scripts to simulate the *healthy* and *patient* avatars. More specifically, we constructed these questions based on five categories: emotion, daily activity, general interests, social interests and sense of purpose, which are aiming to assess

[5] https://github.com/gunthercox/ChatterBot

[6] https://unity3d.com/

[7] https://thrift.apache.org/

Fig. 2 Our avatar in a counselling room scene.

the negative symptoms rating skills of psychiatrists. There are total 100 question-answer pairs to address our purpose. Some significant examples of Q&A scripts are shown in Table 1.

Table 1 Examples of Question and Answer scripts.

Category	Question and Answer	
	Healthy Avatar	**Patient Avatar**
Emotion	Q1: How have you been feeling for the past week? A1: Quite OK. Been really busy at work and I'm looking forward to the weekend.	Q1: How have you been feeling for the past week? A1: I just want to feel good again. It has been tough and tiring for me
Daily Activity	Q2: What do you usually do during your free time? A2: Umm, hang out with my friend, or watch the movies.	Q2: What do you usually do during your free time? A2: Read books or newspapers.
General Interests	Q3: What are your hobbies? A3: like traveling, swimming, and I also like listening the music.	Q3: What are your hobbies? A3: Reading.
Social Interests	Q4: Do you have any friends? A4: Uh-huh, of course, my colleagues and classmates. I usually hang out with them.	Q4: Do you have any friends? A4: I live with my family.
Sense of purpose	Q5: What makes life meaningful for you? A5: I enjoy the happiness of life, and my friends make life more colorful.	Q5: What makes life meaningful for you? A5: I don't know.

3.2 Non-verbal cues

Alogia is most commonly found as a negative symptom of schizophrenia. It can be mild, where a person tends to reflect slowing or blocked thoughts, and often manifested as short, empty replies to questions. At times, the symptom can be so severe that a person with schizophrenia is impossible to understand. Thus, in the dialog manager, we have programmed a longer response time to *patient* avatar, where the distributions of response time of two avatars are determined by the real distributions in our early study [14] (Patient Avatar: Gamma distribution with parameter a=1.277 and b=1.080; Healthy Avatar: Gamma distribution with parameter a=1.233

and b=0.975). We also give *patient* avatar some random responses such as "I see a zombie" or "spoon full of sugar". There are a total of 10 random responses which are triggered based on the severity of negative symptoms. Higher severity implies a higher probability of random responses.

Many research [12] have found that affective flattening (or blunted affect) is common in the schizophrenia patients with negative symptoms. The voices of schizophrenia do not show the normal emotional range that most people display. Rather, they speak in a monotone [15] which means their voices don't rise and fall in pitch. Therefore, by changing the pitch parameters of Google text-to-speech, we set a lower pitched voice to *patient* avatar.

Attention impairment is also implemented in the *patient* avatar. In the behaviour manager, once we determine the location of the psychiatrist, we set the gaze of the *healthy* avatar in Unity3D accordingly. The animation of the avatar needs to be realistic and constantly gazing at the psychiatrist could make it fall in the uncanny valley [16]. Hence, we have a *lookAt* and *lookAway* duration in the avatar animation. The proportion of the duration [13] are varied while idle, talking and listening.

4 Conclusion and Future Work

The development of a psychiatrist training system has been proposed. We alter various modules such as the chatbot, speech, affect and gaze to differentiate a patient from a generic human. This virtual character can be deployed in mental health training institutes, universities and hospitals. In addition, our avatar can be used in the public domain to aid mental health awareness campaigns to influence societal stigma associated with mental health.

Our next step will focus on evaluating this system through qualified psychiatrists. They will interact with the *healthy* and *patient* avatars and provide feedback on whether the system accurately depicts the negative symptoms of schizophrenia. We can also modify our platform to potentially simulate other kinds of psychiatric disorders, such as PTSD, autism etc. In addition, with the advancement of AI algorithms, there is definitely potential for our system to achieve: higher accuracy in the objective evaluation of a trainee, a fully unstructured interview and higher quality affect animations of the avatar.

Acknowledgements This study is in collaboration with Institute for Media Innovation at Nanyang Technological University (NTU) Singapore and Institute of Mental Health (IMH) Singapore, and funded by Design and Innovation Project (DIP) 2018/2019 in the School of Electrical and Electronic Engineering (EEE) at NTU. We are also grateful to the third year EEE students in NTU (Clifton Choo, Joel Lee, Lim Yu Qi, Muaaz bin Kamsari, Desmond Chua, Bronson Huang, and, Ng Yan Jun) for their contribution on this project.

References

1. World Health Organization, Mental Health Atlas 2017, https://www.who.int/mental_health/evidence/atlas/mental_health_atlas_2017/en/

2. Kakuma R, Minas H, van Ginneken N, Dal Poz MR, Desiraju K, Morris JE, Saxena S, Scheffler RM. Human resources for mental health care: current situation and strategies for action. The Lancet. 2011 Nov 5;378(9803):1654-63.
3. DeVault, D., Artstein, R., Benn, G., Dey, T., Fast, E., Gainer, A., ... Morency, L.-P. (2014). SimSensei Kiosk: A Virtual Human Interviewer for Healthcare Decision Support. In Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent Systems (pp. 10611068). Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems.
4. Kenny P, Parsons TD, Gratch J, Rizzo AA. Evaluation of Justina: a virtual patient with PTSD. InInternational Workshop on Intelligent Virtual Agents 2008 Sep 1 (pp. 394-408). Springer, Berlin, Heidelberg.
5. Brenner AM. Uses and limitations of simulated patients in psychiatric education. Academic Psychiatry. 2009 Mar 1;33(2):112-9.
6. Brown AM. Simulation in undergraduate mental health nursing education: A literature review. Clinical Simulation in Nursing. 2015 Oct 1;11(10):445-9.
7. Zhang J, Zheng J, Thalmann NM. MCAEM: mixed-correlation analysis-based episodic memory for companionuser interactions. The Visual Computer. 2018 Jun 1;34(6-8):1129-41.
8. Zhang J, Zheng J, MagnenatThalmann N. PCMD: personalitycharacterized mood dynamics model toward personalized virtual characters. Computer Animation and Virtual Worlds. 2015 May;26(3-4):237-45.
9. Edition F. Diagnostic and statistical manual of mental disorders. Arlington: American Psychiatric Publishing. 2013.
10. Andreasen NC. Negative symptoms in schizophrenia: definition and reliability. Archives of general psychiatry. 1982 Jul 1;39(7):784-8.
11. Chakraborty D, Yang Z, Tahir Y, Maszczyk T, Dauwels J, Thalmann N, Zheng J, Maniam Y, Amirah N, Tan BL, Lee J. Prediction of Negative Symptoms of Schizophrenia from Emotion Related Low-Level Speech Signals. In2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2018 Apr 15 (pp. 6024-6028). IEEE.
12. Carter JD, Bizzell J, Kim C, Bellion C, Carpenter KL, Dichter G, Belger A. Attention deficits in schizophreniapreliminary evidence of dissociable transient and sustained deficits. Schizophrenia research. 2010 Sep 1;122(1-3):104-12.
13. Ruhland K, Andrist S, Badler J, Peters C, Badler N, Gleicher M, Mutlu B, Mcdonnell R. Look me in the eyes: A survey of eye and gaze animation for virtual agents and artificial systems. InEurographics state-of-the-art report 2014 Apr 7 (pp. 69-91).
14. Tahir, Y., Chakraborty, D., Dauwels, J., Thalmann, N., Thalmann, D. and Lee, J., 2016, March. Non-verbal speech analysis of interviews with schizophrenic patients. In Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on (pp. 5810-5814). IEEE.
15. Xu, S., Yang, Z., Chakraborty, D., Tahir, Y., Maszczyk, T., Victoria, C.Y.H., Dauwels, J., Thalmann, D., Thalmann, N.M., Tan, B.L. and Keong, J.L.C., Automated Linguistic Analysis of Interviews with Schizophrenic Patients. International Conference on Digital Signal Processing (DSP) 2018
16. Mori M. The uncanny valley. Energy. 1970;7(4):33-5.