

An Affectively Aware Virtual Therapist for Depression Counseling

Lazlo Ring
Northeastern University
Boston, MA
lring@ccs.neu.edu

Timothy Bickmore
Northeastern University
Boston, MA
bickmore@ccs.neu.edu

Paola Pedrelli
Massachusetts General Hospital
Boston, MA
ppedrelli@mgh.harvard.edu

ABSTRACT

Depression is one of the leading causes of disability in the world effecting more than 350 million people. Unlike other treatable epidemics that effect our society however, less than half of those suffering receive treatment. To address this issue, researchers have developed computer based cognitive behavioral therapy systems to let patients receive therapeutic care from the comfort of their own home.

While effective, these computer-based interventions are still inferior to traditional face-to-face therapy sessions with human counselors. This discrepancy is due, in part, to the inability of these systems to detect and react to a user's emotional state during therapy sessions.

We describe an affectively-aware virtual therapist for depression counseling, whose design is based on theories of emotions in psychotherapy, along with the results of a pilot study exploring the efficacy of this approach.

Author Keywords

Affective Computing; Embodied Conversational Agents; Computer Based Cognitive Behavioral Therapy

ACM Classification Keywords

H.5.2. Information interfaces and presentation: User Interfaces

INTRODUCTION

Depression is the leading cause of disability in the world with over 350 million people affected worldwide [17]. Unlike other treatable epidemics that affect our society, however, less than half of those suffering from depression receive any form of treatment [17]. The reason behind this is the cost, availability, and social stigma related to the treatment of depression [14]. To alleviate this issue, several computer-based cognitive behavioral therapy (CBT) systems have now been developed that allow patients to receive therapeutic care from the comfort of their own

homes. While they are superior to no treatment, these systems are inferior to traditional face-to-face therapy with human counselors. For example Dowrick et al. found that retention in computerized CBT interventions is significantly lower than that for face-to-face interventions [3]. We believe these shortcomings are due to the lack of affective responsiveness in these systems. In traditional face-to-face psychotherapy, proper empathy feedback plays a key role improving the recovery rate of patients with depression [1]. To create systems that provide proper empathic feedback to patients, however, we must first understand the role of emotions in psychotherapy.

In this paper, we review literature on the role of patient emotions during psychotherapy sessions, and discuss how this knowledge could be used to develop an affectively-aware automated counseling system. We also present the design and implementation of a system for depression counseling and the results of a pilot study exploring the efficacy of our system.

THEORETICAL BASIS

We have identified several principles from the psychotherapy literature that can be leveraged in an affect-aware automated counseling system:

Emotions require proper cognitions to promote growth

Researchers such as Greenberg [6] have demonstrated that being in an emotional state has no lasting benefits without understanding what brought you there. This is especially true in cognitive behavioral therapy, which focuses on helping patients understand the connection between their thoughts and feelings. To increase the effectiveness of counseling systems, they should both identify the emotional state of a user and the cause behind it to permit reflection to occur.

Positive emotions are indicative of activity continuation

As shown by Williams et al. [16], the display of positive emotions is one way we express our interest in continuing to do an activity. By identifying a strong correlation between a user's positive emotions and the tasks we are discussing, we can gain an understanding of the user's likes and dislikes. This can be used, for example, in guiding the counseling system's therapeutic agenda-setting strategy.

Paste the appropriate copyright/license statement here. ACM now supports three different publication options:

- ACM copyright: ACM holds the copyright on the work. This is the historical approach.
- License: The author(s) retain copyright, but ACM receives an exclusive publication license.
- Open Access: The author(s) wish to pay for the work to be open access. The additional fee must be paid to ACM.

This text field is large enough to hold the appropriate release statement assuming it is single-spaced in Times New Roman 8-point font. Please do not change or modify the size of this text box.

Each submission will be assigned a DOI string to be included here.

The perceived source of an event effects one's emotional response to it

Weiner demonstrated that people experienced different emotional responses to an event based on their perception of who they felt was responsible for it [15]. By utilizing a user's perceptions of an event, one could predict their emotional responses as a way to increase the accuracy of an emotion detection system.

Empathic feedback is key to building relationships

Researchers like Feller and Cotton have shown that there is a strong connection between therapeutic displays of empathy by a therapist and the quality of the relationship between a patient and the therapist [5]. This suggests that providing real time empathic feedback to users could result in better therapeutic outcomes since there is a strong connection between therapeutic relationships and outcome in psychotherapy [7].

AN AFFECTIVELY AWARE AGENT FOR DEPRESSION COUNSELING

Based on the rules we identified, we developed a prototype of an affectively-aware depression counseling system. Since the goal of our system was to recreate one-on-one face-to-face therapeutic interactions with a human counselor, we decided to use an embodied conversational agent, an animated computer character that emulates face-to-face conversations through both verbal and non-verbal behavior, as our interface (Figure 1).



Figure 1. Affectively-Aware Agent.

The overall system architecture is shown in Figure 2. The counseling dialogue is based on a manualized cognitive behavioral therapy intervention for depression, developed by our clinical collaborator, Dr. Pedrelli. The dialogue is authored in an XML-based scripting language that allows dialogue authors to easily control the agent's utterances and conversational flow. The underlying dialogue system is designed so that additional utterances, such as reflections and empathic messages, can be automatically inserted into the counseling dialogue without affecting the overall conversation. The virtual counselor speaks using a text-to-speech engine, and its conversational nonverbal behavior, including hand gestures, posture shifts, and facial displays, are automatically generated using the BEAT engine [2].

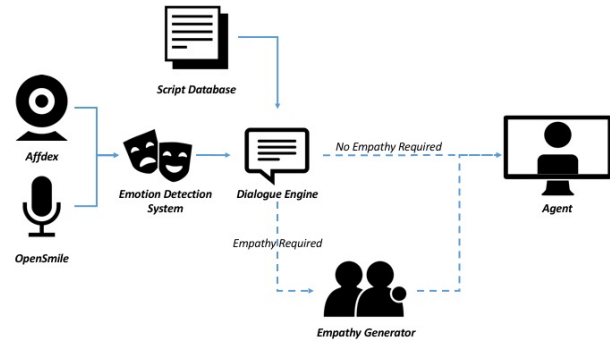


Figure 2. System Architecture

Speech and Affect Detection:

Users interact with the virtual counselor using constrained speech or touch screen input. At each point in the dialogue at which they can contribute to the conversation, a list of possible responses is displayed (right side of Figure 1) and the user must respond with one of these by speaking the prompt verbatim or touching it on the touch screen display. This interaction format allows for a natural-feeling conversation and speech-based affect detection, while avoiding the dangers and ethical issues associated with unconstrained natural language input.

User speech input is passed to two sub-systems responsible for speech recognition and affect detection. The speech recognition system uses pocket sphinx [8] to create a grammar-based speech recognizer using US-English acoustic model and dictionary. Audio recordings of user utterances are also passed to the affect detection system, which classifies the utterance into the three valence categories of happy, neutral, or sad. The classifier was trained using the OpenSmile system with libsvm [4] on the emotional prosody and transcripts database [18] for these three categories of emotion using approximately 160 samples per category. Due to the short duration of the recorded utterances, we combined the audio from multiple utterances within a dialogue segment for more accurate results.

User affect was also classified based on their facial display using the Affdex SDK [11]. This system ran in parallel to the speech-based affect detection system, but was only used when the user choose to not interact with the system via speech.

Emotional Dialogue Generation:

Empathic responses by the agent to negative user affect were automatically generated by the dialogue system. This sub-system uses the user's valence and arousal as input (based on the circumflex model of emotion [13]). This sub-system determined if an empathic response should be generated based on predefined variables that represented the accuracy thresholds associated with the connected sensors. The empathic responses generated by the system were selected from a list written by our clinical collaborator and presented to the user, along with a prompt that allowed them to pause interaction in order to calm and compose

themselves (Figure 3). Arousal based responses were handled in a similar manner, in which the system would present the user with a response to re-engage them with the interaction and would offer them the opportunity to pause if needed.

Figure 3. Example Empathic Statement:

After detecting negative affect during a script.

Agent: You sounded pretty sad, do you need a minute before we go on?

Button Prompts:

- Yah.
- I'm fine, let's continue.

To emulate empathic listening, a key skill in improving therapeutic relationships [12], we rephrased participant's statements after eliciting emotionally sensitive content from them (Figure 4). In our prototype, we manually authored this dialogue, but we intend to automatically generate these responses in future versions.

Figure 4. Example Reflective Statement:

Agent: *What depressive symptoms are you experiencing?*

User selects: "Feelings of hopelessness, pessimism", "Feelings of guilt, worthlessness, helplessness", "Decreased energy, fatigue, feeling slowed down"

Agent: *So you have been feeling hopeless, helpless and exhausted recently?*

Pilot Study

We conducted a pilot study to evaluate the acceptance and feasibility of our automated counseling system among individuals with mild to moderate depression.

Study Design:

Using the affectively aware agent described above, we recruited participants to go through the first counseling session in a cognitive behavioral therapy intervention. In this interaction, participants discussed the following topics with the agent: How to interact with the agent; What is depression; A review of the user's depressive symptoms; What the user thought about therapy; A brief introduction into the concept of CBT.

Participants:

Participants were recruited via online postings and local flyers. Participants were eligible if they scored 5-14 on the PHQ-9 [9] (mild to moderate depression), and were not currently enrolled in therapy and not on anti-depressant medication.

Measures:

Participants received a depression screeners (PHQ-8 [10]) and a state anxiety questionnaire before and after interacting with the system. After interacting with the system, participants filled out 7-point scale measure questionnaires to evaluate the agent and participated in a semi-structured interview. All interactions with the agent

were video recorded for review and evaluation by our clinical collaborator.

Quantitative Results:

Ten participants, 5 male, 5 female, between the ages of 18-28 (Average = 22.4, SD = 2.4) with mild to moderate depression (PHQ9 score average: 6.6, SD = 2.7) were recruited to interact with the agent.

Pre-post testing conducted immediately before/following the agent interaction found no significance differences in depression or anxiety (Table 1), although both trended in a positive direction. Agent ratings were generally neutral across the board, with satisfaction (Mean = 4.5, SD = 1.35); desire to continue using (Mean = 4.2, SD = 1.6); trust (Mean = 3.9, SD = 1.66) and likeability (Mean = 4.4, SD = 1.7) scoring around the mid-point. Participants did however report that they did not feel very close with the agent (Mean = 2.1, SD = 1.29) and that they felt it was more like interacting with a stranger (Mean = 2.1, SD = 1.19). This was likely due to the short duration of the interaction and the minimum amount of social dialogue present in the system.

	Pre - Mean (SD)	Post - Mean (SD)	p
Depression (PHQ-8)	8.0 (4.62)	6.4 (3.13)	.38
Anxiety (State-Anxiety)	23.3 (6.46)	21.78 (7.61)	.65

Table 1. Outcome Measures from Pilot Study

Qualitative Results:

The transcripts of the semi-structured interviews were thematically analyzed for the evocation and understanding of emotion by the system. Our analysis found that half of the participants felt that the agent evoked emotional responses in them during the interactions, stating that they felt emotional because the agent was presenting them with "information they did not realize" (Patient 5). Two thirds of the participants expressed that they felt the agent understood their emotions, stating that they felt the agent "... could decipher some of my attitudes" (Patient 10) and "... understood my emotions because I felt that it gave me the right responses" (Patient 8).

With the aid of our clinical collaborator, we conducted a video review to validate the ability of the system to detect and respond to user emotion during the counseling sessions. Video segments from the beginning of a session to the first point of potential empathic feedback were extracted and reviewed for half of the participants. The expert was instructed to classify the valence expressed in each video segment on a scale from 0 to 100, along with a determination of whether empathy should have been expressed after the segment. Their ratings classified 80% (4/5) of the segments as requiring empathy and 20% (1/5) as not, while our emotion detection system rated all of the

segment's as requiring empathy. Valance ratings for the video segments were also higher for the expert (Mean = 55, SD = 5) when compared to the the ratings generated by our emotion detection system (Mean = 34.4, SD = 10.5).

This analysis confirmed that there were instances in which the agent successfully evoked and correctly responded to participant's emotional states during the conversation, primarily through offering pauses in the conversation during moments of emotional distress. However, these results do suggest that our emotion detection system may be bias towards detecting negative affect.

Behavioral Results:

The counseling dialogue allowed participants to question the relevance of the therapeutic content and, through a subsequent prompt, terminate the session early if desired. Only one of the ten participant's chose to click the first of these prompts, and none of the participants chose to terminate the session early.

Discussion:

This study demonstrates that it is possible to evoke and respond to a user's emotional state in real time during automated counseling sessions with an affectively aware agent. Although we did not power this study to produce significant changes in depressive symptoms, the majority of users expressed that they felt the agent understood their emotions and responded appropriately.

CONCLUSIONS AND FUTURE WORK

In this paper, we explore the use of an affectively aware agent for use in depression counseling. Based on a review of the use of emotion in psychotherapy, we have outlined general rules for how to make an affective-aware health counseling systems. Additionally, we demonstrate the potential efficacy of such a system through our pilot study in which depressed participant interacted with an affectively aware agent that guided them through a session of cognitive behavioral therapy.

We are currently expanding the automated counseling system to support a five-week longitudinal, home-based intervention for depression that will be evaluated in a much larger randomized trial. We are also exploring text classification techniques to automatically identify counseling topics that could trigger emotional responses in participants, and text generation techniques to automatically generate the agent's empathic responses. We plan to use these methods to provide more tailored empathic responses to participants.

REFERENCES

[1] Burns, D. and Nolen-Hoeksema, S. Therapeutic empathy and recovery from depression in cognitive-behavioral therapy: a structural equation model. *Journal of consulting and clinical ...*, (1992).

[2] Cassell, J., Vilhjálmsón, H., and Bickmore, T. Beat: the behavior expression animation toolkit.

Life-Like Characters, (2004).

- [3] Dowrick, C. Computerised self help for depression in primary care. *BMJ*, (2015).
- [4] Eyben, F., Wöllmer, M., and Schuller, B. Opensmile: the munich versatile and fast open-source audio feature extractor. *Proceedings of the international ...*, (2010).
- [5] Feller, C. and Cottone, R. The importance of empathy in the therapeutic alliance. *The Journal of Humanistic Counseling, ...*, (2003).
- [6] Greenberg, L. and Paivio, S. *Working with emotions in psychotherapy*. 2003.
- [7] Horvath, A. and Symonds, B. Relation between working alliance and outcome in psychotherapy: A meta-analysis. *Journal of counseling psychology*, (1991).
- [8] Huggins-Daines, D. and Kumar, M. Pocketsphinx: A free, real-time continuous speech recognition system for hand-held devices. ... , *Speech and Signal ...*, (2006).
- [9] Kroenke, K. and Spitzer, R. The PHQ-9: a new depression diagnostic and severity measure. *Psychiatr Ann*, (2002).
- [10] Kroenke, K., Strine, T., and Spitzer, R. The PHQ-8 as a measure of current depression in the general population. *Journal of affective ...*, (2009).
- [11] McDuff, D. and Kaliouby, R. El. Affectiva-MIT Facial Expression Dataset (AM-FED): Naturalistic and Spontaneous Facial Expressions Collected "In-the-Wild." *Computer Vision and ...*, (2013).
- [12] Myers, S. Empathic listening: Reports on the experience of being heard. *Journal of Humanistic Psychology*, (2000).
- [13] Russell, J., Weiss, A., and Mendelsohn, G. Affect grid: a single-item scale of pleasure and arousal. *Journal of personality and ...*, (1989).
- [14] Saxena, S., Thornicroft, G., Knapp, M., and Whiteford, H. Resources for mental health: scarcity, inequity, and inefficiency. *The lancet*, (2007).
- [15] Weiner, B. The role of affect in rational (attributional) approaches to human motivation. *Educational Researcher*, (1980).
- [16] Williams, D. and Dunsiger, S. Does affective valence during and immediately following a 10-min walk predict concurrent and future physical activity? *Annals of Behavioral ...*, (2012).
- [17] World Health Organization. Depression (Fact sheet No. 369). (2015).
- [18] Emotional Prosody Speech and Transcripts - Linguistic Data Consortium. <https://catalog.ldc.upenn.edu/LDC2002S28>.