Impact of Adaptive Multimodal Empathic Behavior on the Quality of User Interaction in Therapeutic Applications

Paper 785

ABSTRACT

Empathic behavior between humans has often a positive effect, particularly in healthcare, since it facilitates relationship, improves engagement, and reduces stress and anxiety. Despite this importance, the effects of empathic behavior of embodied virtual agents that interact with patients in a multimodal and adaptive way have not been widely explored. In this article, we propose an empathic model which endows a therapeutic embodied virtual agent with multimodal adaptive empathic behavior during interaction with a user. This model relies on user-agent interaction relationship and focuses on (1) the interpretation of user's behavior using multimodal input, and (2) the generation of multimodal empathic behavior during interaction. An experimental study in the context of empathic interaction with students during COVID-19 pandemic is presented to evaluate the effect of adaptive empathic behavior of agent on the quality of the user interaction. The results show that using real-time adaptive multimodal empathic agent is perceived more empathic and improves engagement of user during interaction.

KEYWORDS

Human-Agent Interaction, Embodied Virtual Agent, Therapeutic Communication, Empathy, Multimodal Interaction

ACM Reference Format:

Paper 785. 2022. Impact of Adaptive Multimodal Empathic Behavior on the Quality of User Interaction in Therapeutic Applications . In *Proc. of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2022), Auckland, New Zealand, May 9–13, 2022*, IFAAMAS, 10 pages.

1 INTRODUCTION

During interaction, humans use both verbal (speech, prosody) and nonverbal (facial expressions, gaze, gestures, postures, etc.) modalities to communicate. Effective communication requires the ability to understand and react on others' behavior. Empathy is henceforth a key socio-emotional capability for interaction [36]. Using empathic behavior during human-human interaction has substantial advantages since it has a positive effect on relationship [5, 32], engagement [39], acceptance and trust [40], and reduces stress and anxiety [17]. Therefore, endowing interactive systems with empathic ability participates in improving the understanding of the relationship between affective and cognitive processes, as well as enhancing human-agent interaction.

Embodied conversational agents have been broadly used in several applications for human-machine interaction, such as education, artificial companions, training, game and healthcare... Studies have shown that these ECAs should take into account both verbal and nonverbal behaviors of their interlocutors, generate backchannels, display empathy and construct a trustworthy relationship with the user [40] during interaction. However, the effect of multimodal adaptive behavior and the role of different interaction modalities used by ECAsthat interact with patients have not been fully explored in the literature.

This article focuses on the empathic behavior of an agent interacting with a user. Multimodal user-ECA dialogues are exploited to improve the engagement in interaction. We propose an original empathic model that enables an embodied virtual agent to exhibit multimodal adaptive behavior during interaction with the user. This model relies on user-agent interaction relationship and focuses on (a) the interpretation of the user's behavior using multimodal input, and (b) generating multimodal empathic behavior during interaction. In particular, we focus on how adaptation processes can be exploited to produce this behavior. The study of other aspects of ECAs such as the automatic generation of non-verbal behavior is out of the scope of this article.

To validate the contribution of this model, an experimental study in the context of a human-agent interaction in healthcare is carried out. The objective of this study is to gain insight into the role and impact of empathic agent behaviors on the user in the context of a patient-agent interaction. We aim at answering the following research question: *"Do adaptive multimodal empathic behavior and different modalities of Embodied virtual agent have an impact on the quality of patient-agent interaction?"*, using interaction engagement, relationship of the agent with the user and acceptability indicators.

Section 2 describes the current state of art regarding the application of ECAs in patient interaction and theoretical and computational models of empathy. Section 3 contains the proposed model of adaptive empathic behavior for ECA. Section 4 elaborates on the components of the proposed architecture. Section 5 describes the experimental study and discusses its results in order to validate our purposed model. The last section summarizes the contribution and proposes recommendations for future work.

2 RELATED WORK

2.1 Empathic Embodied Conversational Agents in Healthcare

The use of ECAs with the ability to express empathy toward the user isgrowing, specifically in healthcare applications ([15, 16, 24, 33, 40, 42, 43]). For example, empathic agents have been used in applications such as a virtual reality exposure to evoke clinical symptoms e.g. fear in public speaking, and to guide patients in their emotional responses [3]. Therapeutic virtual agents can identify these symptoms by interviewing patients about their mental health [17, 35, 40]. There is also evidence that people tend to disclose more personal information when they interact with a virtual human rather than with a real human [15]. Moreover, empathic responses emitted by agents results in improved user engagement [45], leads to more positive interaction [34], and helps building trust [40].

Proc. of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2022), P. Faliszewski, V. Mascardi, C. Pelachaud, M.E. Taylor (eds.), May 9–13, 2022, Auckland, New Zealand.

These characteristics of empathic agents are helpful for patient engagement during treatment in order to foster strong patient-doctor relationship [9, 10, 12, 13, 20, 40].

Empathy is a complex phenomenon associated with the ability of perceiving, understanding and experiencing another person's emotions. According to Hoffman [23], natural empathy is defined as *"A psychological process that makes a person have feelings that are more congruent with another's situation than with his own situation."*. Empathy can involve cognitive or affective attributes. Cognitive attributes of empathy include the reasoning to understand the feelings of the user and to communicate that understanding. Emotional or affective attributes of empathy involve physiological arousal and spontaneous expression reacting to someone else's display of emotions, such as mimicking user's perceived expressions of joy [38, 53]. However, in a healthcare context, therapeutic empathy is defined as *when the therapist is sensing the feelings and personal meanings which the patient is experiencing at each moment, when he can perceive these from 'inside', as they seem to the patient, and when he can successfully communicate something of that understanding to his patient* [44]. This particular type of empathic behaviour is clinically relevant and includes the mechanism to produce useful or adequate emotions in a therapeutic process.

During an interaction, an empathic ECA is expected to provide proper backchannels to the user as well as an emotional feedback [47]. Backchanneling can have an emotional effect on the perception of the user [20]. Behaviors such as mimicry and affective matching result from the innate capability of resonating with others during social interaction. Mimicry is achieved by imitation of the perceived facial expressions of others based on their emotional signals.In contrast to mimicry, Affective matching [15] allows the ECA to present and regulate emotions that are better perceived by the users. During empathic interaction, not only the empathic mechanism but also the content of the verbal response can affect the perception of the empathic response [53]. Thus, the ability to generate empathic responses during interactions also requires cognitive processing such as perspective-taking, that is the ability of an observer to imagine how the other feels. This mechanism implies to combine the appraisal of the situation given the context and the understanding of what caused the other's emotional state, in order to communicate the understanding of the user's action [23, 38].

2.2 Computational Models of Empathy

Several models of empathy generation have been proposed in the literature [7, 27, 34, 38, 54]. Some of the work focus on the binary classification of empathy of an agent as empathic or non empathic, and positive or negative empathy [8]. Ochs et al. [34] provides a theoretical model based on Scherer's appraisal theory [46] while concentrating mostly on the cognitive evaluation (or appraisal) of emotions by the agent. The computation model proposed by Boukricha et al.'s work [7] is founded on three processes including empathy mechanism (the process by which an empathic emotion arises), empathy modulation (the process by which both an empathic emotion is modulated and a degree of empathy is determined), and expression of empathy (process by which an empathic emotion is communicated and actions are taken). Recently, Yalcin et al. [52, 54] have proposed a hierarchical framework to model empathy for ECAs based on the "Russian Doll" model of empathy [14] which allows different levels of empathic behaviors. Furthermore, [27] have also proposed a decision tree based module for generating empathic response to the user. This model emulates both cognitive and affective empathy, by capturing and processing user's affective states, and combining it with affect-related information elicited from utterances to decide empathic responses for the agent.

Empirical approaches, relying on corpora collected in studies that examine peoples' social and emotional interactions, have also been proposed. For example, an adaptive engine proposed in [42] enables an agent to adapt its empathic behavior depending on the user model, and the adaptive rules for the user model are generated using a decision tree. The user model includes complex elements such as the user's verbal responses, personality, preferences, and emotional state. Moreover, empirical evidences also supports the clear relationship between different empathic features such as dialogue act, emotions and conversation management [26, 50, 56, 57]. For instance, when one adopts the *encouraging* dialogue act, she usually expresses the *EM caring* instead of *approval*. If one expresses empathy with *exploration*, she mostly adopts the *questioning* dialogue act and shows *surprise*.

These aforementioned computational models do not directly provide a real-time adaption of empathic behavior during interaction. Moreover, the adaptive engine proposed in [42] relies on user profile and computes the behavior of the agent before the interaction with the user. These models also focus on the generation of "natural" empathy, however to the best of our knowledge there exists no model that focuses on "therapeutic" empathic behavior of the ECA.

2.3 Interaction Models in Healthcare

Several studies have focussed on the use of empathic ECAs in the field of healthcare. These models of interactions [11, 13, 19] between a medical person (doctor/nurse) and a patientgenerate improved interaction and better health outcome. For example, the attachment theory of Cassedy et al. [11] provides different attachment styles as the measure of interaction that directly affect health outcomes. These categories illustrate the levels of physiological markers of stress, risky behavior *e.g.* smoking and drinking, symptom reporting, and adherence to treatment. The Interaction Model of Client Health Behavior (IMCHB) [13] is designed to explain patient-doctor interaction that affects health outcome and relies on three main concepts. Firstly, the patient singularity includes static variables such as demographic characteristics as well as dynamic variables such as the patient's emotion knowledge or belief about her health (cognitive appraisal) and her reaction to her health situation (affect). Secondly, the patient-doctor interaction in which the doctor attends and listens to the patient, includes affective support, health information, and decision control. The third component is the health outcome which includes health care status and satisfaction. This model provides the bidirectional relationship between patient and docto,r and unidirectional relation with health outcomes.

The Transactional Model of Stress and Coping (TMSC) [19] describes the relationship between the transactional process (e.g. stress) and the coping output. Two mediating processes occur in response to a transaction event or stressor. Initially, primary appraisal process perceives the impact of the event, and then appraises how well the event can be controlled or emotionally managed. After this process, coping response depends on coping effort, meaning-based coping, and social support. The coping effort is broken down into problem management (seeking more information about the issue and actively engaging in problem resolution) and emotional regulation (strategies that affect coping through the feeling about the situation). Meaning-based coping refers to strategies that focus on how to effectively handle events such as positive reframing of the situation. Social support affects not only the perception of the events but also the effectiveness of coping effortsoutcomes. The IMCHB model provides the patient-doctor relationship but lacks emphasis on behavioral mechanisms that influence emotional outcomes, whereas the TMCC model uses appraisal to compute an appropriate behavior that can help in achieving health goal and user acceptance, but does not enlighten user relationships.

2.4 Discussion

Therapeutic empathic agents exhibit different characteristics: (a) therapeutic empathy requires understanding of the user emotions as well as cognition [38], (b) positive emotions are indicative of active continuation [32], (c) they can improve user satisfaction [45], (d) sources of events affect one's emotions (appraisal theory), and (e) empathic feedback is the key to building relationships. However, several studies ([41] [42] [40]) conclude that these characteristics are not always justifiable. For example, [42] stated that although their agent is adaptive, these agents failed to produce appropriate empathic response to the user, and sometimes provided empathic utterances inappropriately. Furthermore, the use of empathic agents does not always improve outcome and relationship with the agent [41]. Pierre et al. [40] also demonstrated that only adding an empathic agent does not provide better results. Indeed, the observed acceptability is attributed to the information provided by the agent rather than the agent itself. There is a need for an integrated model of empathy behavior generation that takes the advantages of healthcare theories and that is also compliant with the results of empirical findings. Our approach towards a computational model of therapeutic empathic agent relies on a multimodal approach for the recognition of user's emotions in order to understand her socio-emotional state, focussing on bridging the gap between the interpretation and the generation of empathic behavior of the agent. In this context, we propose an empathic agent model that unifies the theories guided by the user-agent relationship, and generates comprehensive therapeutic empathic behavior.

3 ADAPTIVE EMPATHIC BEHAVIOR MODEL

In the context of human-agent interaction in healthcare, the multimodal Adaptive empathic behavior model that we propose is inspired from two fundamental models : the IMCHB model [13] and the TMSC model [19]. Cox's IMCHB model emphasizes the relationship between the user and the health provider, however behavioral mechanisms that influence the socio-emotional outcome of the model are not taken into account, whereas the TMSC model focuses on personal coping abilities to achieve health goals and improve treatment acceptance. Thus, combining TMSC with IMCHB model enables to provide a comprehensive empathic behavior model in the context of human-agent interaction. The integrated model for multimodal adaptive empathic behavior of an Embodied Medical Agent (EMA), summarized in Figure 1, includes the following components.

The **stressors** are stressful events or causes that occur during the interaction and affect the agent's perception of the user.

The **static variables** include demographic characteristics e.g., age, gender etc., and social influences such as environmental resources. The **dynamic variables** include the user's emotional reaction during the interaction (cognitive appraisal from IMCHB). It is responsible for the interpretation of the current state of the user. This integrated model also includes the primary and secondary appraisals from TMSC model as the sub-module of cognitive. These last two appraisals occur as the response to stressor events. The primary appraisal is the perceived impact of the event, e.g., the severity of the event whereas the secondary appraisal is how well the event can be emotionally managed. The static and dynamic variables influence the agent's perception of the user.

Dynamic variables are affected by the **user-patient interaction** which includes providing emotional support and communicating health information to the user.

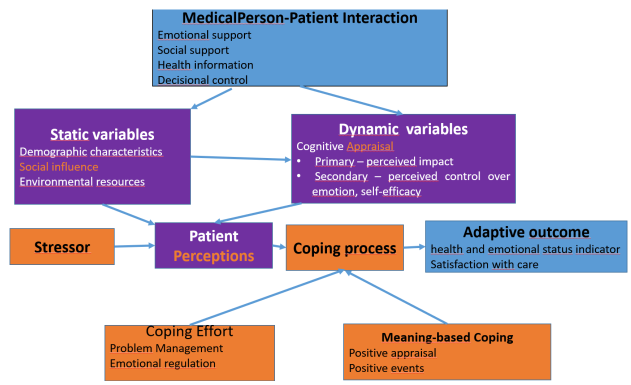

Figure 1: Integrated Adaptive empathic Behavior model for an EMA

The agent responds to the events by considering the variables and interaction context through the **coping process**. It is influenced by coping effort and meaning-based coping. Coping efforts play the role of a mediator between the effect of the event and the appraisal of emotional outcome. Coping efforts are further divided into problem management and emotional regulation. The problem management can include seeking more information from the user or continue moving towards the goal. Emotional regulation intends to create a change of belief or feeling about the situation [21]. Meaning-based coping refers to strategies to effectively handle stressors such as positive re-framing of the situation and beliefs to find a positive meaning for the patient.

The coping process leads to the **adaptive outcome**, which includes the health and emotional status indicator, and the satisfaction with the system.

4 ADAPTIVE MULTIMODAL EMPATHIC AGENT ARCHITECTURE

This section describes the different modules implemented to understand the socio-emotional state of the user and to generate the multimodal empathic responses during the interaction.

4.1 Reflexive Listening and Affect Matching

This module intends to firstly understand the user's feelings and secondly convey the idea to the speaker that her feelings have been correctly understood. It also includes displaying appropriate backchannel behavior and mirroring the socio-emotional state of the user, reflecting an emotional state with words and nonverbal behavior while the user is speaking. Following the *Russian-Doll* model of empathy [14], the agent mimics the user's emotions using affect matching behavior. This allows fast reaction to the behavior of the user, relying on a perception-action mechanism while the user talks. The process uses dynamic properties to regulate the emotional state of the agent, requiring first the emotion recognition and then the representation of the matching emotional state.The candidate emotion is determined using the weighted sum of the emotions [55] from video and audio input signals from the user when she is speaking. The intensity, duration and speed of the agents' expressions depend directly on the values from the perceived emotions. Then, once the user has finished its turn, the agent evaluates the current socio-emotional state of the user and uses the cognitive reasoning in order to choose an appropriate empathic response.

4.2 Multimodal Emotion Recognition in Conversation

The Multimodal Emotion Recognition in Conversation (MERC) model predicts the emotional attitude associated with an utterance by modeling the conversation history as well as notions of common sense, while distinguishing the speaker from other participants (listeners). It leverages the use of pre-trained architectures to create feature vectors used as inputs for the main architecture. The MERC model takes as input the following modalities: (1) the transcript of the utterance and (2) the video clip associated with this utterance. We employ deep-learning based pre-trained models to extract relevant feature vectors: (a) a language model [28] to represent textual information from the current utterance and its historical context; (b) a commonsense model [6] to extract the speaker's intent, the effect of the speaker's utterance on herself and the listener's reaction to the utterance; (c) the OpenCV toolkit to preprocess the video associated to the utterance. Following [18, 29] we model memory banks via deep neural recurrent networks [4]), representing different variables: the speakers' state and the influence of the utterance on the listeners. The importance of stored memory slots is weighted via a multi-headed attention mechanism [25]. The final representation, used for classification, is the combination of speaker representation, influence on listeners and context-enriched utterance.

In the context of this experiment, the proposed MERC model was trained using the IEMOCAP dataset, which provides social interactions similar to our scenario. Given the full transcript and the video clip associated to an utterance, the final output of our component is the likelihood distribution of each label provided in the training data (neutral, happiness, sadness, anger, frustration, excitation).

4.3 Emotion and Mood Regulation

In this process, the agent evaluates the socio-emotional state of the user when the user speech is over.

Based on the current emotion of the user determined by the MERC model, the agent first updates the current emotional state and regulates its dynamic variables *e.g.,* mood. The intensity of the agent's emotions are calculated in function of the intensity of the desirability of the events (user's verbal and nonverbal attitudes), the current mood of the agent and the influence of the mood on the emotion of the user. As the agent reacts emotionally and appraises the event, the OCC cognitive structure of emotion [37] is used to generate agent's emotions. The intensity of the emotion variation is calculated as:

$$intensity_e += intensity_{desirability} * scale * (mood_{value} * Factor_{MoodInfluenceOnEmotion})$$

The events also affect the mood of the agent taking into account the desirability of this event and the influence of emotion on the mood. The mood of the agent is updated as:

$$mood_{value} += scale * (Intensity_e * Factor_{EmotionInfluenceOnMood})$$

The mood of the agent changes the dynamic properties of the emotions. It allows the emotions to be sustained and decay, enabling consistent behavior over time.

4.4 Empathic Dialogue Adaptation

The therapeutic empathy Module is one of the main component of our architecture. It enables the virtual agent to communicate empathically with the user during intervention. During the adaptive empathic behavior generation, the agent uses both the socio-emotional state of the user, the semantic or contextual information from user's input utterance and its own socio-emotional state in order to choose the empathic response. The module is in charge of determining what, when and how to adapt the therapeutic empathic behavior during interaction with the user.

In order to provide affective support to the user during interaction, we focus on empathic dialogue cues during the interaction following the NURSE model of empathic cues [2, 22], that includes naming, understanding, respecting, supporting and exploring cues.
Naming means giving the emotion of the user as a way of showing that agent is attuned to what she is experiencing.
Understanding is important to determine the feeling or the meaning of the user's utterance. It may require some exploration, and the active listening can be an effective way to validate user emotions.
Respecting and acknowledging a patient's emotions is an important step towards showing empathy. It indicates that not only the user's emotions is appropriate but also important during the interaction.
Supporting the user can be done by the agent in various ways such as acknowledging the understanding of the user's current situation, and that the agent is available for the help.
Exploring refers to the asking question or expressing interest.

The model describes the techniques for expressing an empathic behavior to the patient. This expression can be verbal or nonverbal such as emotional gestures and sighing. In particular, this model focuses on the semantics of the verbal response aligned with the appropriate empathic emotion chosen by the agent in order to convey empathy to user. However, the NURSE model [2] does not take into account the social dimension, which is an important aspect of the effective conversation. Therefore, we extend the existing NURSE model by adding social dialogue cuesin ordre to build rapport with the user [5] during interaction. The characteristics of the resulting *NURSES* (NURSE + Social cues) model is also in line with empirical findings [57] that, in a particular situation, a specific category of empathic phrases can be used for interaction.

The decision making for the empathic dialogue adaptation relies on medical psychological description based on the analysis of patient-nurse interaction data proposed in [22], also substantiated by [32, 57]. We use a rule based approach that constructs the decision tree based on several parameters : the perceived user emotion, the semantic information of the user's input (user's DA and speech sentiment), the performance index PI_d, the agent's previous DA, user's interest, and the response delay. The sentiment of the user's utterance is also calculated using the deep learning based model which uses the french language model CamenBERT [30] and transformers [51] for sentiment prediction. The perceived emotion intensity is calculated with respect to the performance index of the user as described in Youssef et al. [55]. The Figure 2 illustrates the decision making mechanism in order to generate empathic dialogue cue in conjunction with the current emotional state of the agent.

In this architecture, the Empathy Module provides both affective and cognitive empathy. The proposed system perceives the user's emotions using multimodal input and semantic information of the user's input and reacts emotionally (verbal and nonverbal) to convey affective empathy to the user.

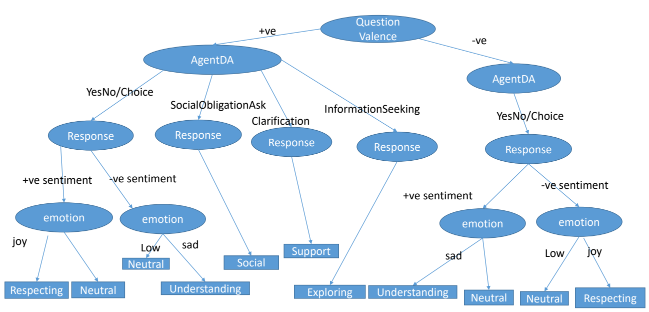

Figure 2: Empathic decision tree

The agent can expect whether her interaction will be pleasant or unpleasant for the user. As the user reacts emotionally and appraises the event, the OCC [37] cognitive structure of emotion is used to predict agent's emotions. The interpretation of the socio-emotional state of the user and the appraisal of the events produced by the user results in the generation of an empathic response. In our model, the desirability appraisal variable from the OCC model is taken into account. If the event produced by the user is appraised by the agent as desirable event (*e.g., if the user takes medicines on time or the user indicates self-reported improved health, and the goal of the agent is the well-being of user*) and the user's current potential emotion state is positive, the agent expresses joyful emotion in conjunction with the empathic dialogue and sustains its emotional state computed during affect regulation until the next decision (*aah! that sounds really good*). However, if the user's event is appraised as undesirable by the agent (*e.g., if the user's emotion is distress*), the agent starts with an empathic response matching with the socio-emotional state of the use (*I'm really sorry to hear that*).

Displaying positive empathic support from agent triggers the positive emotions in the user [32]. However expressing negative empathy by taking the perspective of the user may imply a sense of agreement with the user's negative expressions. Thus, in order to provide positive empathic response, the agent applies a meaning-based coping process [19] which includes strategies for emotion regulation. We have adapted the cognitive change strategy of emotion regulation described in [21] in order to evaluate the event with a different (more positive) perspective through the secondary appraisal (re-appraisal) process. This type of cognitive change involves *changing a situation's meaning in a way that alters its emotional impact* [21]. That is, if the primary appraisal process results in the generation of a negative empathic response, then this response is followed by the encouraging empathic response with joyful emotion in order to construct more positive meaning of the original event through interaction. The agent then sustains its emotional state computed during affect regulation until the next decision. Mimicking of positive emotion of a person interacting with a virtual agent helps to build rapport and liking [32] and also it may reduce the stress level of the user as the user may observe the supportive behavior of the agent.

4.5 Discussion

This architecture has been integrated into a virtual environment where the adaptive therapeutic agent is embodied as a virtual character. Figure 3 shows a screen-shot of the interaction scenario where the agent plays the role of a therapeutic medical person that can interact with the user using natural language interaction. The originality of the proposed model, compared to computational models of "natural" empathy (*e.g.* [7, 52]) lies firstly in the role of the MERC Model that computes dynamically the cumulative emotion of the user, which is then used for the affect regulation of the agent, and secondly in the Therapeutic nature of the adaptive empathic dialogue model, where the agent uses the *NURSES* model and adaptive dialogue model to produce therapeutic appropriate behavior. This behavior is based on the integrated model of IMCHB and TMSC interaction models in healthcare, that allows the generation of multimodal responses suitable to the healthcare context. Compared to the adaptive engine [42], the proposed empathic behavior exhibit both cognitive and affective empathic behavior and can adapt the empathic behavior during interaction.

5 EXPERIMENTAL EVALUATION

The goal of this experimental study is to gain insight into the effects of the therapeutic multimodal empathic behavior of the EMA.

In this experiment, the EMA collects information concerning the stress and fear during the Covid-19 period through a dialogue. To evaluate the effects of adaptive multimodal empathic behavior of ECA on user's engagement, following two experimental conditions are considered.

Baseline Adaptable ECA (BA-ECA): One of the state of art approach to exhibit empathic behavior is by using an adaptive engine to produce tailored empathic responses as proposed in [42]. The agent uses the user profile to decide the preference of the user for certain types of empathic responses. The user profile includes the information about user's demographics, personality, emotional state and attitude. However, the agent only displays the lip-sync while speaking and does not have any non-verbal behaviors except a smile. The agent does not demonstrate any backchannel or low level empathic behavior. It uses the user's verbal response (utterance) along with the user profile to select empathic responses.

Adaptive Therapeutic Multimodal EMA (ATM-EMA): The proposed therapeutic multimodal empathetic behavior of the agent uses both verbal and nonverbal modalities to understand the socio-emotional state of the user, and to generate appropriate empathic response. The agent uses both verbal (speech) and nonverbal behavior (backchannel, emotion expression, lip-sync and gestures) to express empathy. In this model, while the user speaks, the agent provides low level empathic response using affect matching and backchannel. Once the user speech is over, the agent decides which empathic behavior should be selected by taking into account the perceived socio-emotional state of the user and the semantic information of the utterances. The nonverbal gestural behavior is automatically generated by mapping the semantics of the utterance to gestures.

We aim at answering the following research question: *Do adaptive empathic behavior, and different modalities of embodied conversational agent have any impact on patient-agent interaction?* This question is studied through two aspects of evaluation which include (1) user perception of the empathic engagement of the ECA, and (2) the engagement of the user during the interaction with the ECA. In order to evaluate these aspects, we have defined the following hypotheses:

- H[1]: By exhibiting both low level affect matching mechanism and high level affect regulation and cognitive mechanism, the adaptive therapeutic EMA is perceived as more empathic during the interaction with the user compared to the baseline adaptive Agent.
- H[2]: Multi-modal adaptive empathic behavior of therapeutic EMA increases user engagement during interaction.
  - H[2.1]: with multimodal Empathic behavior, the user is more involved in interaction.
  - H[2.2]: with multimodal Empathic behavior, the user experiences more enjoyment.

5.1 Method

5.1.1 Participants. For this experiment, 30 students from an engineering school in France were recruited through a call for participation. Because we focus on face-to-face multimodal verbal interaction with an ECA, and to ensure the consistency of the study panel, we imposed a controlled condition that the participants must be fluent French speakers. There were 13 men and 17 women between the ages of 19 and 23 years (mean 20.41 years, SD = 1.37) and we randomly associated them with one of the evaluation conditions.

5.1.2 Data Collection. In order to evaluate the impact of the behavior of the ECA, several subjective and behavioral measures were used. Subjective measures are carried out using questionnaires in three parts. The first part of the questionnaire includes the assessment of demographic, psychological and mental state of the subject. The second part of the questionnaires focuses on the user's perception of empathic behavior of the ECA adapted from consultation and relational empathy (CARE) measure [31] (Table 1). As in both conditions, the agent uses the same scenario (only the empathic behavior were different), there should be no significant difference for questions q2, q7 and q8 in both conditions. The third part of the questionnaires focuses on the evaluation of user's engagement with the ECA (Table 2).

q1	The agent's manner made me feel completely at ease.
q2	The agent let me tell my story.
q3	The agent listened to everything I had to say with her full attention.
q4	The agent seemed genuinely interested in me as a person.
q5	The agent was very sympathetic about my problems.
q6	The agent seemed to understand exactly the way I have been feeling.
q7	The agent explained things in a way I could fully understand.
q8	The agent had a positive attitude.

Table 1: Perception of empathic behavior, adapted from [31]

q9	How engaging was the interaction?
q10	How relaxing or exciting was the experience?
q11	How completely were your senses engaged?
q12	I enjoy the agent talking to me.
q13	I enjoy participating in this session with the the agent.
q14	I find the agent enjoyable.
q15	I find the agent fascinating.

Table 2: User engagement during interaction

5.1.3 Procedure. The experimental process involves three steps. In the first step before the experiment, participants were informed on the general context of the activity to interact with virtual agent and about the general course of interactions. The proposed EMA-user interaction scenario includes the COVID-19 Student Stress Questionnaire [58] and The Fear of COVID-19 Scale questionnaires [1]. In this interaction scenario, the aim of the agent, playing the role of the doctor, is to obtain the information from students in order to perform their psychological evaluation. Figure 3 shows a screenshot of the evaluation scenario where the user interacts with the ECA. The scenario does not contain any empathic responses of the agent: those are only generated by the empathic models.

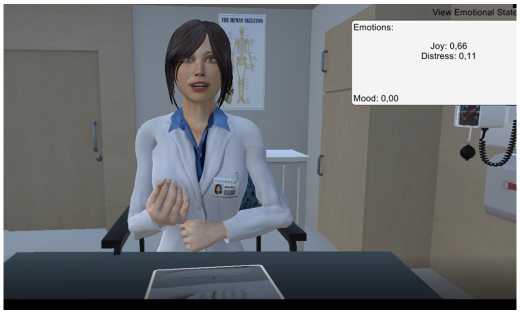

Figure 3: Empathic Agent showing joyful empathic response to user

In the second step, the subject is invited to perform the experiment in one of the experimental conditions. The subject is left alone in the room with experimental setup to avoid any effect of the presence of the observer. After the experiment, the subject is asked to fill the questionnaires regarding her experience. The questions assessment is made on a Likert scale of 5 categories (1- strongly disagree, 2- disagree, 3- neutral, 4- agree and 5- strongly agree).

5.1.4 Design Analysis. In this study, since we have 28 participants (therefore 14 participants in each condition), in order to obtain cumulative results, we decided to split the Likert scale in two groups (the first group with values of strongly disagree + disagree, and the second group with the last three scales). Given the multiple dependent measures, we use one-way analysis of variance (ANOVA) to determine whether the mean of a dependent variable is the same in two or more unrelated, independent groups. The significant level for all of the analysis was set to 0.05.

5.2 Results

This sections describes the subjective evaluation of various factors such as empathic engagement and user engagement with the agent, based on the post-interaction questionnaires filled by the users.

5.2.1 Empathic Engagement of the User with an Agent. We want to assess the user perception of empathic engagement with the ECA. We have first run a univariate ANOVA on the user's view about ECA's efforts for making her feel at ease in function of the condition type of empathic agents (question q1). There is a significant effect of the ECAs empathic condition ($F(1, 27) = 8.49$; $p = .007$). The user has a more positive view about ATM-EMA's efforts towards introducing herself to user, explaining the context of the intervention, being warm towards user and treating user with respect (Mean score=3.14, SD= 1.1), as compare to the BA-ECA condition (mean score=2.14, SD= 0.54). Moreover, there is a significant effect of the agent conditions on the user's perception of the agent letting her tell her story ($F(1, 27) = 18.56$; $p = .000$) (question q2). The user perceived that the ATM-EMA gave more time to user in order to fully describe her condition in her own words; not diverting the user while the user is telling her story (mean score=3.71, SD=1.2) as compare to the BA-ECA (mean score=2.0, SD=0.8). The results of the univariate ANOVA also indicate significant difference on the user's perception of agents attentive listening behavior ($F(1, 27) = 15.81$; $p = .000$) (question q3). The user perceived that the adaptive multimodel ECA paid close attention to what she was saying (mean score=3.5, SD=1.28) as compare to the BA-ECA (mean score=1.93, SD=0.73). The reason is that unlike baseline adaptive ECA, the ATM-EMA shows low level empathic behavior (backchannel) while the user is speaking.

There is a significant difference in user's perception of whether the agent is genuinely interested in her as a whole person during the interaction ($F(1, 27) = 5.08$; $p = 0.034$) (question q4). Users perceived that the adaptive ATM-EMA was interested in knowing relevant details about her situation (mean score=4.0, SD=0.55) as compare to the BA-ECA (mean score=3.21, SD=1.1). The ATM-EMA uses the *exploring* empathic cues when the user responds to an information seeking question, and the *support* empathic cues to convey acknowledgement and understating in response to the clarification information provided by the user. There is significant difference in the user's perception about the ECA showing care and compassion during the interaction ($F(1, 27) = 7.15$; $p = 0.013$) (question q5). The ATM-EMA is seemed as more genuinely concerned, connecting with user on a human level and not being detached from the user during the interaction (mean score=3.86, SD=1.03) as compared to the BA-ECA (mean score=2.86, SD=0.95). The reason is that when the user is speaking, the ATM-EMA provides low level empathic response (backchannel and affect generation), and generates positive empathic behavior using meaning-based coping in case when the user is distressed, that builds the sense of rapport with the ECA.

There is a significant difference on the user's view about the agent's understanding of her feeling ($F(1, 27) = 5.41$; $p = 0.028$) (question q6). The users considered that the ATM-EMA had communicated that she had accurately understood user's concerns and anxieties (mean score=2.5, SD=1.2) better than BA-ECA (mean score=1.64, SD=0.63). The reason is that the proposed ATM-EMA does not only display an empathic feedback while the user is speaking, but also uses the overall emotion during the conversation recognised by MERC model along with semantic information of user's utterance to determine the appropriate empathic response.

As the ECA in both conditions use the same interaction scenario, the user perceived no significant difference in agent's ability of explaining things clearly and giving her adequate information ($F(1, 27) = 2.93$; $p = 0.099$) (question q7). However, due to the presence of rhythm, stress or intonation on the adaptive multimodal agent's voice, the ATM-EMA scored higher (mean score=3.76, SD=1.12) as compared to the BA-ECA (mean score=3.0, SD=1.30) having speech with no prosody. Moreover, the univariate ANOVA reveals no significant difference in user's view on both ECAs being positive ($F(1, 27) = 3.2$; $p = .085$) as both ECAs show joyful emotion expression (question q8). However, ATM-EMA scored higher (mean score=4.71, SD=0.47) for having a positive attitude and being not negative about user's problems as compare to BA-ECA. The reason is that the ATM-EMA exhibits therapeutic empathic behavior, whereas the BA-ECA shows natural empathy.

In order to evaluate the overall empathic engagement through the perception of empathic behavior of ECA by the user, we computed the sum of the scores of these 8 questions. The univariate ANOVA reveals that there is a significant difference in the user's perception of overall empathic engagement of the ECA in function of the condition type of empathic agents ($F(1, 27) = 8.49$; $p = .007$).

The ATM-EMA is perceived more empathic (mean score=29.21, SD=4.31) as compare to the BA-ECA (mean score= 20.92, SD=5.06). Thus, according to the CARE measure of empathy [31], the ATM-EMA shows medium empathic level whereas the BA-ECA shows low level empathy aspects for the user.

We can conclude that these results support hypothesis *H[1]* which states that, by exhibiting both low level affect matching mechanism and high level affect regulation and cognitive mechanism, the adaptive therapeutic EMA (ATM-EMA) is perceived as more empathic during the interaction with the user as compare to the baseline adaptive Agent (BA-ECA).

5.2.2 User Engagement during the interaction with an Agent. In order to evaluate the engagement of user, several questions were directed at eliciting the user' feelings on how involved they were with the EMA (questions q9-q11) and how much they enjoyed the interaction with the EMA (q12-q15).

Impact of empathic behavior of EMA on user's involvement during interaction. In order to evaluate the involvement of the user during the interaction with ECA, we run the univatiate ANOVA on the user's opinion about her engagement in function of the type of the agent. There is a significant effect of the type of ECA on the user's engagement during interaction ($F(1, 27) = 10.06$; $p = .004$). The users feel more engaged with ATM-EMA (mean score=4.07, SD=0.61) as compared to the BA-ECA (mean score=3.21, SD=0.8) . Moreover, users find the interaction with ATM-EMA more relaxing and exciting (mean score=3.92, SD=0.83) as compared to the interaction with the BA-ECA (mean score=2.43, SD=0.94) and the effect of the type of ECA is significant ($F(1, 27) = 20.12$; $p = .000$). There is also a significant difference ($F(1, 27) = 4.75$; $p = .039$) in the sense of engagement for user in function of the ECA type. Taken together, there is significant difference in the score of the user's involvement with agent in function to the condition types of ECA ($F(1, 27) = 20.75$; $p = .000$). That is, the user is more involved with ATM-EMA (mean score= 8.0 , SD=1.36) as compared to BA-ECA (mean score= 5.64 , SD=1.39). One of the reasons for this can be that in the BA-ECA condition, the ECA does not show any backchannel or nonverbal behavior while user speaks. Moreover, it computes the list of empathic cues suitable for the user based on user profile before the beginning of the interaction using a machine learning based decision-tree algorithm whose accuracy is below 75%. On the contrary, the ATM-EMA not only listens actively (backchannel and affect maching empathic behavior) to the user when the user speaks, but also adapts its multimodal empathic behavior during interaction thanks to the empathic dialogue adaptation process and the *NURSES* model. These results support the hypotheis *H[2.1]* that with multimodal Empathic behavior, the user is more involved in the interaction.

Impact of empathic behavior of EMA on user's perceived enjoyment with the Agent. The perceived enjoyment is defined as *the extent to which the activity of using computers is perceived to be enjoyable in its own right* [49]. It also referred to as an intrinsic motivation variables such as the doing of an activity for satisfaction rather than for spcific outcomes. There is no significant difference ($F(1, 27) = 3.54$; $p = .07$) in the opinion about the perceived enjoyment of the agent talking to the user. However, as the prosody and mimicry play an important role during interaction [48], users enjoyed more the ATM-EMA agent talking to them (mean score= 4.0, SD=0.78) as compared to the BA-ECA talking to them (mean score=3.35, SD=1.0). Moreover, the users enjoyed more participating in the narrative interaction with the ATM-EMA (mean=4.0, SD=0.88) as compare to an interactive session with BA-ECA (mean score=3.0, SD=1.07), the difference begin significative ($F(1, 27) = 6.295$; $p = .019$). The reason is that the ATM-EMA communicates with user using both verbal and nonveral modalities (speech with prosody, facial emotion expressions and gestures), whereas the BA-ECA only uses monotonic voice and only smiles and lip-sync to communicate. Furthermore, users find the ATM-EMA more enjoyable (mean score= 4.7, SD=0.73) than that of BA-ECA (mean score= 3.28, SD=1.68) with significant difference ($F(1, 27) = 5.157$; $p = .03$) in their scores. However, there is no significant difference in the score about how fascinating the agent is ($F(1, 27) = 0.317$; $p = .578$). Taken together, univariate ANOVA reveals that there is significant difference ($F(1, 27) = 4.997$; $p = .034$) in aggregated scores of user's perception of enjoyment with the EVA in function to the condition types of agent. These results validate the hypothesis *H[2.2]* that, with multimodal Empathic behavior of EMA, the user perceives more enjoyment during the interaction with the EMA.

In order to evaluate the overall impact of the empathic behavior of ECA on user engagement, we compute the aggregated score of these questions concerning the involvement and enjoyment of the user during interaction. The ANOVA reveals that there is a siginifant effect of the empathic agent conditions on the overall scores for user engagement during the interaction with an Agent ($F(1, 27) = 14.734$; $p = .0.01$). The user feels more engaged with ATM-EMA (mean score=27.00, SD=3.94) as compare to BA-ECB (mean score=21.42, SD=3.756). during interaction. We can conclude that the hypothesis *H[2]* stating that the multi-modal adaptive empathic behavior of therapeutic EMA increases user engagement during interaction, is supported by the results.

6 CONCLUSION

An empathic model is proposed in this article which endows a therapeutic embodied virtual agent with multimodal adaptive empathic behavior during interaction with a user. We then presented an experimental study to evaluate the impact of the proposed empathic behavior model on the quality of user interaction in therapeutic application. The adaptive agent has the means to communicate empathic responses to the user. The results indicate that the users feel involved in interaction with the proposed adaptive multimodel empathic therapeutic agent and enjoy the interaction with the agent, compared to an adaptive baseline. Results confirm the initial anticipation that the real-time adaptive multimodal empathic agent is perceived as more empathic and improves engagement of user during interaction.

Future lines of work include a thorough exploitation of objective measures collected during the experiment in conjunction to questionnaire results. Furthermore, the literature [42] shows that some users prefer more neutral behavior from virtual medical persons, so coupling profile information and dynamic adaptation might yet improve the quality of interaction.

REFERENCES

[1] Daniel Kwasi Ahorsu, Chung-Ying Lin, Vida Imani, Mohsen Saffari, Mark D Griffiths, and Amir H Pakpour. 2020. The fear of COVID-19 scale: development and initial validation. *International journal of mental health and addiction* (2020), 1–9.

[2] Anthony L Back, Robert M Arnold, Walter F Baile, James A Tulsky, and Kelly Fryer-Edwards. 2005. Approaching difficult communication tasks in oncology 1. *CA: a cancer journal for clinicians* 55, 3 (2005), 164–177.

[3] Ligia Batrinca, Giota Stratou, Ari Shapiro, Louis-Philippe Morency, and Stefan Scherer. 2013. Cicero - Towards a Multimodal Virtual Audience Platform for Public Speaking Training. In *Intelligent Virtual Agents*, Ruth Aylett, Brigitte Krenn, Catherine Pelachaud, and Hiroshi Shimodaira (Eds.). Springer Berlin Heidelberg, 116–128.

[4] Yoshua Bengio, Aaron Courville, and Pascal Vincent. 2013. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence* 35, 8 (2013), 1798–1828.

[5] Timothy Bickmore, Amanda Gruber, and Rosalind Picard. 2005. Establishing the computer–patient working alliance in automated health behavior change interventions. *Patient Education and Counseling* 59, 1 (2005), 21–30. https://doi.org/10.1016/j.pec.2004.09.008

[6] Antoine Bosselut, Hannah Rashkin, Maarten Sap, Chaitanya Malaviya, Asli Celikyilmaz, and Yejin Choi. 2019. COMET: Commonsense Transformers for Automatic Knowledge Graph Construction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 4762–4779.

[7] Hana Boukricha and Ipke Wachsmuth. 2011. Empathy-based emotional alignment for a virtual human: A three-step approach. *KI-Künstliche Intelligenz* 25, 3 (2011), 195–204.

[8] Scott Brave, Clifford Nass, and Kevin Hutchinson. 2005. Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent. *International Journal of Human-Computer Studies* 62, 2 (2005), 161–178. https://doi.org/10.1016/j.ijhcs.2004.11.002

[9] Adrián Bresó, Juan Martínez-Miranda, Cristina Botella, Rosa María Baños, and Juan Miguel García-Gómez. 2016. Usability and acceptability assessment of an empathic virtual agent to prevent major depression. *Expert Systems* 33, 4 (2016), 297–312.

[10] Lorainne Tudor Car, Dhakshenya Ardhithy Dhinagaran, Bhone Myint Kyaw, Tobias Kowatsch, Shafiq Joty, Yin-Leng Theng, and Rifat Atun. 2020. Conversational agents in health care: scoping review and conceptual analysis. *Journal of medical Internet research* 22, 8 (2020), e17158.

[11] Hannah F Cassedy, Richard A Enander, Richard C Robinson, H Monroe Evans Jr, Blake Frank, Christy Tucker, Paula D Miltenberger, Sandra Pitts, and C Allen Stringer. 2015. Attachment theory as a model of doctor–patient interaction. *Journal of Applied Biobehavioral Research* 20, 4 (2015), 151–178.

[12] Debaleena Chattopadhyay, Tengteng Ma, Hasti Sharifi, and Pamela Martyn-Nemeth. 2020. Computer-controlled virtual humans in patient-facing systems: systematic review and meta-analysis. *Journal of medical Internet research* 22, 7 (2020), e18839.

[13] Cheryl L Cox. 1982. An interaction model of client health behavior: Theoretical prescription for nursing. *Advances in nursing science* 5, 1 (1982), 41–56.

[14] Frans BM De Waal. 2007. The 'Russian doll'model of empathy and imitation. *On being moved: From mirror neurons to empathy* (2007), 35–48.

[15] D. DeVault, Ron Artstein, Benn G., Teresa Dey, Edward Fast, Alesia Gainer, Kallirroi Georgila, J. Gratch, Arno Hartholt, Margot Lhommet, Gale M. Lucas, Stacy Marsella, Fabrizio Morbini, Angela Nazarian, Stefan Scherer, Giota Stratou, Apar Suri, D. Traum, Rachel Wood, Yuyu Xu, A. Rizzo, and Louis-Philippe Morency. 2014. SimSensei kiosk: a virtual human interviewer for healthcare decision support. In *AAMAS*.

[16] Lucile Dupuy, Etienne de Sevin, Hélène Cassoudesalle, Orlane Ballot, Patrick Dehail, Bruno Aouizerate, Emmanuel Cuny, Jean-Arthur Micoulaud-Franchi, and Pierre Philip. 2021. Guidelines for the design of a virtual patient for psychiatric interview training. *Journal on Multimodal User Interfaces* 15, 2 (2021), 99–107.

[17] Lucile Dupuy, Jean-Arthur Micoulaud-Franchi, and Pierre Philip. 2021. Acceptance of virtual agents in a homecare context: Evaluation of excessive daytime sleepiness in apneic patients during interventions by continuous positive airway pressure (CPAP) providers. *Journal of Sleep Research* 30, 2 (2021), e13094. https://doi.org/10.1111/jsr.13094

[18] Deepanway Ghosal, Navonil Majumder, Alexander Gelbukh, Rada Mihalcea, and Soujanya Poria. 2020. COSMIC: COmmonSense knowledge for eMotion Identification in Conversations. *arXiv preprint arXiv:2010.02795* (2020).

[19] Karen Glanz and Marc D Schwartz. 2008. Stress, coping, and health behavior. (2008).

[20] Jonathan Gratch, Ning Wang, Jillian Gerten, Edward Fast, and Robin Duffy. 2007. Creating rapport with virtual agents. In *International workshop on intelligent virtual agents*. Springer, 125–138.

[21] J. J. Gross and R. A Thompson. 2007. *Handbook of emotionregulation.* The Guilford Press, Chapter Emotion Regulation: Conceptual Foundations, 3–24.

[22] M Jawad Hashim. 2017. Patient-centered communication: basic skills. *American family physician* 95, 1 (2017), 29–34.

[23] Martin L Hoffman. 2001. *Empathy and moral development: Implications for caring and justice.* Cambridge University Press.

[24] Shashank Jaiswal, Michel Valstar, Keerthy Kusumam, and Chris Greenhalgh. 2019. Virtual human questionnaire for analysis of depression, anxiety and personality. In *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*. 81–87.

[25] Wenxiang Jiao, Michael Lyu, and Irwin King. 2020. Real-time emotion recognition via attention gated hierarchical memory network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 8002–8009.

[26] Qintong Li, Hongshen Chen, Zhaochun Ren, Pengjie Ren, Zhaopeng Tu, and Zhumin Chen. 2020. EmpDG: Multi-resolution Interactive Empathetic Dialogue Generation. In *Proceedings of the 28th International Conference on Computational Linguistics*. 4454–4466.

[27] Christine Lisetti, Reza Amini, Ugan Yasavur, and Naphtali Rishe. 2013. I Can Help You Change! An Empathic Virtual Agent Delivers Behavior Change Health Interventions. *ACM Trans. Manage. Inf. Syst.* 4, 4, Article 19 (Dec. 2013), 28 pages. https://doi.org/10.1145/2544103

[28] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692* (2019).

[29] Navonil Majumder, Soujanya Poria, Devamanyu Hazarika, Rada Mihalcea, Alexander Gelbukh, and Erik Cambria. 2019. Dialoguernn: An attentive rnn for emotion detection in conversations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 6818–6825.

[30] Louis Martin, Benjamin Muller, Pedro Javier Ortiz Suárez, Yoann Dupont, Laurent Romary, Éric de la Clergerie, Djamé Seddah, and Benoît Sagot. 2020. CamemBERT: a Tasty French Language Model. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Online, 7203–7219. https://www.aclweb.org/anthology/2020.acl-main.645

[31] Stewart W Mercer, Margaret Maxwell, David Heaney, and Graham Watt. 2004. The consultation and relational empathy (CARE) measure: development and preliminary validation and reliability of an empathy-based consultation process measure. *Family practice* 21, 6 (2004), 699–705.

[32] Takashi Numata, Hiroki Sato, Yasuhiro Asa, Takahiko Koike, Kohei Miyata, Eri Nakagawa, Motofumi Sumiya, and Norihiro Sadato. 2020. Achieving affective human–virtual agent communication by enabling virtual agents to imitate positive expressions. *Scientific reports* 10, 1 (2020), 1–11.

[33] Magalie Ochs, Daniel Mestre, Grégoire De Montcheuil, Jean-Marie Pergandi, Jorane Saubesty, Evelyne Lombardo, Daniel Francon, and Philippe Blache. 2019. Training doctors' social skills to break bad news: evaluation of the impact of virtual environment displays on the sense of presence. *Journal on Multimodal User Interfaces* 13, 1 (2019), 41–51.

[34] Magalie Ochs, David Sadek, and Catherine Pelachaud. 2012. A formal model of emotions for an empathic rational dialog agent. *Autonomous Agents and Multi-Agent Systems* 24, 3 (2012), 410–440.

[35] Stefan Olafsson, Teresa K O'Leary, and Timothy W Bickmore. 2020. Motivating Health Behavior Change with Humorous Virtual Agents. In *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*. 1–8.

[36] Becky Lynn Omdahl. 2014. *Cognitive appraisal, emotion, and empathy.* Psychology Press.

[37] Andrew Ortony, Gerald L Clore, and Allan Collins. 1990. *The cognitive structure of emotions.* Cambridge university press.

[38] Ana Paiva, Iolanda Leite, Hana Boukricha, and Ipke Wachsmuth. 2017. Empathy in virtual agents and robots: A survey. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 7, 3 (2017), 1–40.

[39] Catherine Pelachaud. 2014. Interacting with socio-emotional agents. *Procedia Computer Science* 39 (2014), 4–7.

[40] Pierre Philip, Lucile Dupuy, Marc Auriacombe, Fushia Serre, Etienne de Sevin, Alain Sauteraud, and Jean-Arthur Micoulaud-Franchi. 2020. Trust and acceptance of a virtual psychiatric interview between embodied conversational agents and outpatients. *NPJ digital medicine* 3, 1 (2020), 1–7.

[41] Hedieh Ranjbartabar, Deborah Richards, Ayse Bilgin, and Cat Kutay. 2019. First Impressions Count! The Role of the Human's Emotional State on Rapport Established with an Empathic versus Neutral Virtual Therapist. *IEEE Transactions on Affective Computing* (2019), 1–1. https://doi.org/10.1109/TAFFC.2019.2899305

[42] Hedieh Ranjbartabar, Deborah Richards, Ayse Aysin Bilgin, Cat Kutay, and Samuel Mascarenhas. 2020. Adapting a Virtual Advisor's Verbal Conversation Based on Predicted User Preferences: A Study of Neutral, Empathic and Tailored Dialogue. *Multimodal Technologies and Interaction* 4, 3 (2020). https://doi.org/10.3390/mti4030055

[43] Lazlo Ring and T. Bickmore. 2016. An Affectively Aware Virtual Therapist for Depression Counseling.

[44] Carl Ransom Rogers. 1995. *On becoming a person: A therapist's view of psychotherapy*. Houghton Mifflin Harcourt.

[45] T. Sandbank, Michal Shmueli-Scheuer, Jonathan Herzig, D. Konopnicki, and Rottem Shaul. 2017. EHCTool: Managing Emotional Hotspots for Conversational Agents. *Proceedings of the 22nd International Conference on Intelligent User Interfaces Companion* (2017).

[46] Klaus R Scherer. 2010. The component process model: Architecture for a comprehensive computational model of emergent emotion. *Blueprint for affective computing: A sourcebook* (2010), 47–70.

[47] Marc Schroder, Elisabetta Bevacqua, Roddy Cowie, Florian Eyben, Hatice Gunes, Dirk Heylen, Mark ter Maat, Gary McKeown, Sathish Pammi, Maja Pantic, Catherine Pelachaud, Bjorn Schuller, Etienne de Sevin, Michel Valstar, and Martin Wollmer. 2012. Building Autonomous Sensitive Artificial Listeners. *IEEE Transactions on Affective Computing* 3, 2 (2012), 165–183. https://doi.org/10.1109/T-AFFC.2011.34

[48] Ovidiu Serban, Mukesh Barange, Sahba Zojaji, Alexandre Pauchet, Adeline Richard, and Emilie Chanoni. 2017. Interactive narration with a child: impact of prosody and facial expressions. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction*. 23–31.

[49] Jieun Sung and Younghwa Yun. 2010. Toward a more robust usability concept with perceived enjoyment in the context of mobile multimedia service. *International Journal of Human Computer Interaction* 1, 2 (2010), 12–32.

[50] Anuradha Welivita and Pearl Pu. 2020. A taxonomy of empathetic response intents in human social conversations. *arXiv preprint arXiv:2012.04080* (2020).

[51] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. 2020. Transformers: State-of-the-Art Natural Language Processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. Association for Computational Linguistics, Online, 38–45. https://www.aclweb.org/anthology/2020.emnlp-demos.6

[52] Ozge Nilay Yalcin and Steve DiPaola. 2018. A computational model of empathy for interactive agents. *Biologically inspired cognitive architectures* 26 (2018), 20–25.

[53] Ozge Nilay Yalçın and Steve DiPaola. 2019. Evaluating levels of emotional contagion with an embodied conversational agent. In *Proceedings of the 41st annual conference of the cognitive science society*.

[54] Özge Nilay Yalçın and Steve DiPaola. 2019. Modeling empathy: building a link between affective and cognitive processes. *Artificial Intelligence Review* (2019), 1–24.

[55] Atef Ben Youssef, Mathieu Chollet, Hazaël Jones, Nicolas Sabouret, Catherine Pelachaud, and Magalie Ochs. 2015. Towards a socially adaptive virtual agent. In *International Conference on Intelligent Virtual Agents*. Springer, 3–16.

[56] Chengkun Zheng, Guanyi Chen, Chenghua Lin, Ruizhe Li, and Zhigang Chen. 2021. Affective Decoding for Empathetic Response Generation. *arXiv preprint arXiv:2108.08102* (2021).

[57] Chujie Zheng, Yong Liu, Wei Chen, Yongcai Leng, and Minlie Huang. 2021. CoMAE: A Multi-factor Hierarchical Framework for Empathetic Response Generation. In *Findings of the Association for Computational Linguistics: ACL 2021*.

[58] Maria Clelia Zurlo, Maria Francesca Cattaneo Della Volta, and Federica Vallone. 2020. COVID-19 Student Stress Questionnaire: Development and Validation of a Questionnaire to Evaluate Students' Stressors Related to the Coronavirus Pandemic Lockdown. *Frontiers in Psychology* 11 (2020), 2892. https://doi.org/10.3389/fpsyg.2020.576758