Empathy Framework for Embodied Conversational Agents

Özge Nilay Yalçın

School of Interactive Arts and Technology
Simon Fraser University,
250 -13450 102 Ave. Surrey
BC, CANADA

Abstract

Empathy is a complex socio-emotional behavior that results from the interaction between affective and cognitive mechanisms. Equipping embodied conversational agents (ECAs) with empathic capacity can benefit from the integration and evaluation of these low and high level capabilities in a hierarchical manner. Following the theoretical background on empathic behavior in humans, this paper presents a framework to equip ECAs with real time multi-modal empathic interaction capabilities. We present the implementation of this framework, which includes basic dialogue capabilities as well as three levels of empathic behavior in a conversational scenario. Our approach is an inclusive stand on modeling levels of empathy and provides a baseline behavior for empathic interaction.

Keywords: Empathy, Affective Computing, Social Interaction, Virtual Agents, Embodied Conversational Agents, Human-Computer Interaction

1. Introduction

The ability to understand and react towards the emotions of others, empathy, is a crucial socio-emotional behavior for smooth interpersonal interactions. There has been an increasing amount of contributions to model empathy in agents with various types of behavioral capabilities[1] as an attempt to enhance the interaction between humans and artificial agents and understanding the empathy mechanism better. Embodied conversational agents (ECAs) shows promise as an application area for artificial empathy due to their focus on the importance of natural multi-modal interaction.

However, modeling empathic capacity on interactive agents is a challenging task. Empathy can be attributed to a range of behavior from mirroring, affective matching to empathic concern, altruistic helping and perspective taking [2, 3]. These low and high level behaviors result from the complex interaction of hierarchical components which can be clustered as emotional communication mechanisms, emotion regulation mechanisms and cognitive mechanisms [4]. It is beneficial to understand how much each of these components contributes to the perception of empathy in interactive agents.

In this paper, we present a framework for creating empathy-enabled embodied conversational agents, a critical problem in ECAs. The framework leverages three hierarchical levels of capabilities to model empathy for ECAs, as suggested by the theoretical models [4, 5]. This framework can be used as a building block for integrating additional empathic and behavioral capabilities while having an efficient and responsive conversational behavior. Our system incorporates these baseline empathic behaviors by equipping emotion communication capabilities, which are perceiving and expressing emotions. We present an empathic dialogue scenario to evaluate the implementation of our framework in isolation, focusing on the changing states of the dialogue during the conversation.

In the following sections, we examine the related work on empathy and embodied conversational agents. Then, we provide the interaction scenario and the proposed framework for our 3D virtual avatar system. We further explain how the information processing flows between different components can lead to the levels of empathic behavior.

2. Related Work

Computational Empathy is a relatively new research area that focuses on implementing empathy in virtual agents. Embodied conversational agent (ECA) research as an active and successful field has been attempting to integrate empathy to the existing ECA frameworks. However, while applying the know-how of this mature area to the novel space of computational empathy, it is important to pay attention to a possible overlap between the existing components of these frameworks with the requirements of empathy. A close examination of the theories of empathy would allow us to develop a framework that can account for this possibility.

Current research on empathy in embodied conversational agents (ECAs) follows a variety of different theoretical approaches to define and model empathic behavior [1]. Most of the studies focus on a binary classification of empathy in artificial agents [6, 7]. Ochs and colleagues [8] provide a formal model of emotions based on appraisals while concentrating mostly on the cognitive evaluation of emotions by the agent. Another approach by Rodrigues and colleagues [9] incorporates emotion regulation components such as similarity, affective link, mood and personality that allows different intensities of emotion to be presented. Boukricha and colleagues [10] addressed the modulation of empathy according to the mood and familiarity of the agent and the perceived emotion. A third-person evaluation of empathy included three levels of empathy in a scale from feeling cold-towards and feeling-with in different familiarity and liking conditions.

An inclusive model of empathy would require the development of three hierarchical levels of empathic capabilities that can be categorized as communication competence, affect regulation and cognitive mechanisms [4]. It is crucial to develop and evaluate the system components separately while being mindful about the resulting behavior that each component is responsible for, due to the complexity of the interaction of these components. Research on empathy from various disciplines suggests the underlying mechanism for any empathic behavior to be the perception, recognition and expression of affective stimuli [2, 1]. Figure 1 shows a model of empathy by Yalcin and DiPaola [4], which suggests a three-level hierarchy of empathic behavior inspired by the work of de Waal [2]. In this model, communication competence allows for mimicry and

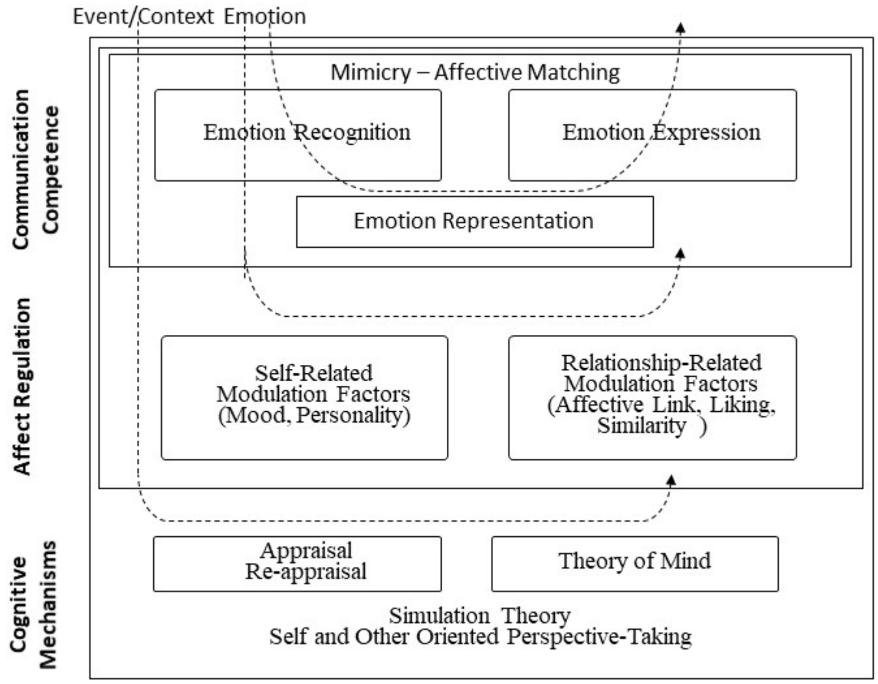

Figure 1: A model of empathy that involves hierarchical levels that are responsible of various types of empathic behavior.

affect matching behaviors while forming a foundation to higher levels of empathy. Emotional mimicry is considered as the congruent affective response to the observed individual's emotions, either by matching the motor expressions or the emotional representation of it in the observer [11]. This capability is essential for any type of empathic behavior [5] as well as emotional intelligence in general [12].

Moreover, for an ECA to be perceived as a social agent that is capable of natural interaction, requires it to follow some basic principles of human conversation [13] as well as showing emotions [14]. As a socio-emotional behavior, it is expected that the perception of empathy also affected by this. Natural human interaction consists of non-verbal and verbal behavior [15] that includes multi-modal synchronous behavior speech, intonation, gaze, head movements and gestures make meaning together using different channels at different timescales [16]. Listening behavior in humans consists of a combination of head nods, vocalizations and facial feedback that show agreement and acknowledgment, which is called backchannels [17]. Turn-taking and backchanneling acts [18, 19], as well as the context of the dialogue, are what determines which of these will be used by the conversation partners while they are in a speaker or a listener role. However, the integration of these behaviors in an empathic ECA should be carefully examined as the output gestures and synchronization might intersect with the expression of empathic behaviors during an interaction.

The ECA literature provides a detailed examination of several types of backchannel behavior during the listening act, but offer little insight on empathy behavior. Sensitive Artificial Listener (SAL) [13] is a multimodal dialogue system that is capable of nonverbal interaction based on speech, head movement and facial expressions of the user. This important work takes into account the user's emotion during listener feedback and can provide mimicry based on the detected facial action units (AUs) of the user. The system is also tested on different characters that had various personality factors that effect the selection and expression of emotional feedback. It was found that the emotionally congruent listening feedback results in better rapport and perception of social presence. However, empathy perception was not a part of this study and there was no distinction between the different types of affective feedback during evaluation.

Similarly, Skowron [20] focuses on a specific type of conversational agents, which is called Affect Listeners that are capable of detecting affective responses, reasoning with them and responding. However, the system and the evaluation is based on task-specific scenarios that are aimed at increasing user satisfaction and usability only. The Rapport Agent of Gratch and colleagues [21] found that random feedback is worse than contingent feedback, where the frequency of feedback was constant. They found the mimicking of head nods according to the prosodic features were perceived as an increased emotional rapport compared to random feedback. This suggests that even the specific backchannel behavior which seemingly does not have an emotional value attached to it when observed in isolation can have an emotional effect on the perception of the user. This highlights the importance of equipping the agent with backchannel feedback while providing a comparison based on the perception of empathy. Previous research repeatedly shown that affect sensitive feedback improves the interaction [22, 14, 23]. However, the literature does not give us insight into how the perception of empathy might be affected by this change.

In this paper, we propose a framework for embodied conversational agents (ECAs) that would allow us to implement levels of empathic capacity along with basic conversational behaviors. We aim to provide a hierarchical implementation of each empathy level along with controlled integration of conversational capabilities, to be able to test and compare each level and component with each other. This paper provides the framework with the current implementation for an empathic listening agent and explains how levels of empathic behavior can arise from different information processing cycles. In the following sections, we will present an empathy framework for our ECA starting with the description of the interaction scenario.

3. Agent and the Context

Our framework is intended for an embodied agent capable of maintaining conversational interaction with its human counterpart. This requires certain behavioral capabilities, such as being able to receive conversational input, process it and act in a social and emotionally suitable manner. In this context, our framework is implemented in a human-like embodied conversational agent. Our socially-situated 3D virtual character system can perform a set of verbal and non-verbal behavior that allows for a realistic conversation with the interaction partner. These behaviors include facial expressions, gaze, head and body gestures, as well as verbal behaviors.

The interaction scenario includes a face-to-face conversation between the agent and a human interaction partner, similar to a video-conference. According to the state of the dialogue, the behavior of the agent can change and adapt to the user. While the interaction partner is speaking, the agent enters the listening state. Listening mode will be activated via the speech and video input from the agent that was described in the perceptual module section. In this state, the agent is expected to provide proper backchanneling behavior as well as the emotional feedback. After the speech of the interaction partner is completed, the agent will enter the thinking state. In this state, the agent will be finished gathering information from the perceptual model and start processing the speech input for generating a response. This response generation process will make use of the context of the dialogue as well as the emotional content of the message. Lastly, the agent will enter the speaking state, where it executes the prepared response via its output channels including voice, facial expression and body gestures.

Note that, our approach includes discrete stages of the dialogue, where each stage sequentially follows each other. Nevertheless, thinking, speaking, and idle states of the agent can be interrupted by the user, which allows for more complex interaction. Moreover, the same framework can be used for agents equipped with different output channels, as long as it allows for a mapping between the emotion representations and the behavioral capabilities. Next, we will go over how the agent framework incorporates levels of empathic capacity within this conversational interaction scenario.

4. Empathy Framework

Our framework for an empathic agent is intended to implement an ECA that is capable of responding to an emotional conversation with the user using verbal and non-verbal behaviors. The implementation achieves levels of empathic behavior through the interaction between the components of perceptual and behavior generation modules with the central behavior controller. The perceptual module processes the visual and verbal input from the user to analyze the context and emotional value. This information is then sent to the behavior controller to be reasoned with according to the selected empathy mechanisms.

According to the level of empathic behavior and the context, an emotional response is selected and sent to the behavior generation module to prepare the response to be displayed with the behavior realizer. Figure 2 shows the overall framework of our system.

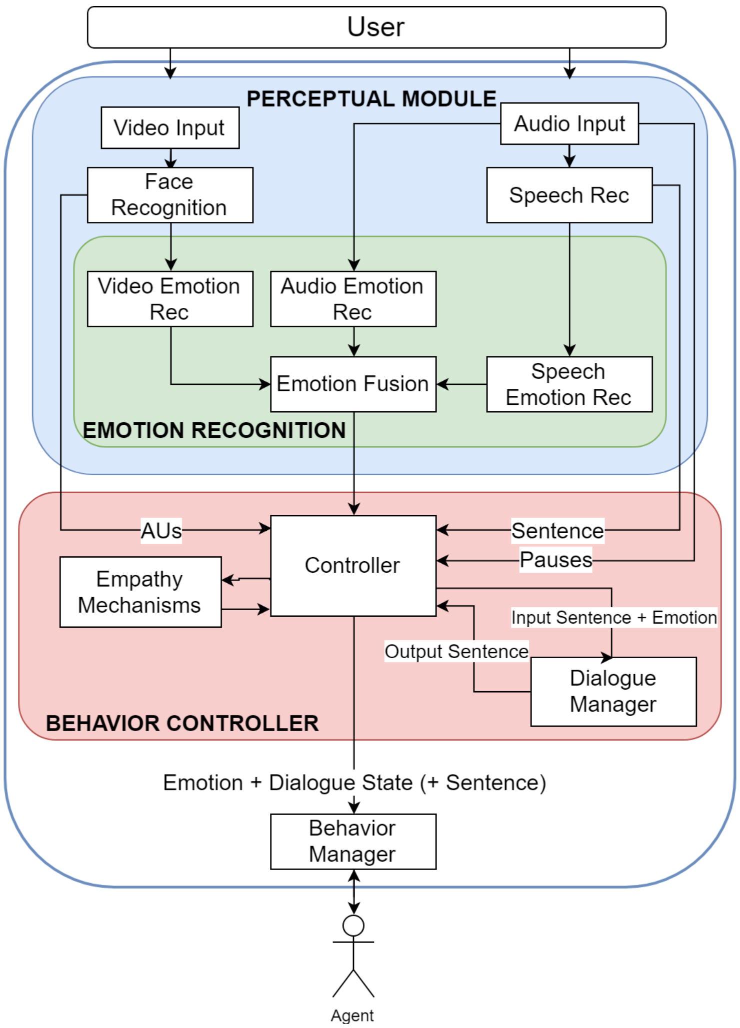

Figure 2: Our framework that includes perceptual, behavior controller and behavior manager modules. Perceptual module includes a multi-modal emotion recognition sub-module. Empathy mechanisms are selected via the controller and impact the behavior of the virtual agent accordingly.

Levels of empathic behavior are achieved by following various processing pathways within this framework. Low-level empathic behavior such as mimicry and affective matching goes through minimal processing in the behavior controller, which allows for a fast response. Mid-level empathic behavior such as affect regulation requires handling the dynamic properties of emotions that can change according to temporal parameters such as mood and personality. Lastly, the higher-level empathic behavior includes reasoning with information about the context. Due to the use of inputs from the lower processing levels, the high-level reasoning results in a slower response.

The following sections will provide detailed information about the modules of the framework and its implementation. The framework is implemented using Python 3.5 programming language, and the output of the framework is realized using Smartbody behavior realizer in Windows platform [1]. Inputs are gathered using a standard web-cam and a microphone.

4.1. Perceptual Module

The perceptual module manages the visual and audio inputs received from the user and generates representations of these inputs to be used by the behavior controller. These representations can be low-level such as the changes in the energy levels from the user's voice based audio signal. Higher-level representations such as emotion categories from the voice and facial emotion recognition, as well as the words and sentence-level sentiment are also gathered in this module. Figure 3 shows a detailed graph of the processing streams within this module.

Audio input includes verbal signals from the user, where the initiation, pauses and termination of the speech signal are gathered to inform the behavior controller about the state of the conversation. This information is gathered using the energy levels from the audio signal with PyAudioAnalysis library [24], and an active speech channel is recognized via a Push-to-Talk functionality that we developed. Speech is recognized by using the real-time Text-to-Speech (TTS)

[1] The implementation code and documentation is released at https://github.com/onyalcin

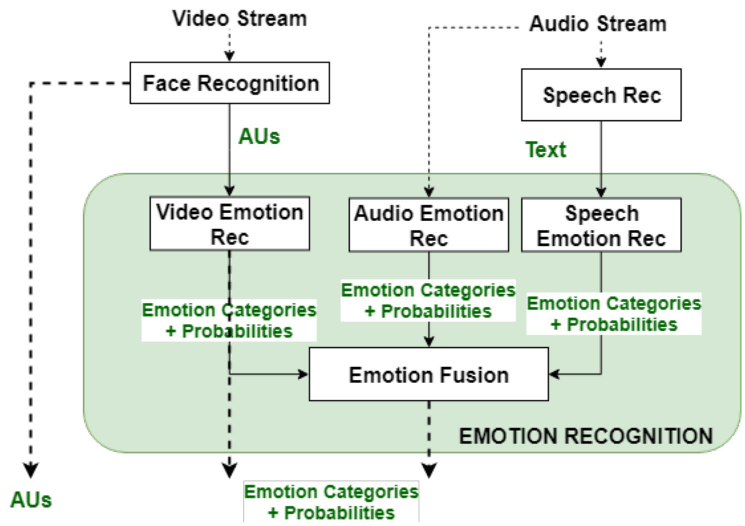

Figure 3: Perceptual module processes video and audio inputs received from the user on separate streams. The outputs from this module are sent to the Behavior Controller according to the state of the agent and the level of selected empathic behavior.

system by using Python's SpeechRecognition [25] library with Google Cloud Speech API for Python [26]. The streaming speech signal and the pauses within the speech signal are sent directly to the Controller during listening to be used to create backchannel behavior, such as head nods and tilts, in real-time. The audio and the text output are also sent to the emotion recognition sub-module for further processing.

Video input is used to extract facial landmark points to be used in the emotion recognition system. We currently use Affdex real-time expression recognition toolkit [27] to extract facial expressions for emotion recognition. During the interaction, the perceptual module collects information about the facial action units (AUs) that include AU1 (inner brow raiser), AU2 (outer brow raiser), AU4 (brow loweverer), AU5 (upper lid raiser), AU6 (cheek raiser), AU7 (lid tightener), AU9 (nose wrinkler), AU10 (upper lip raiser), AU12 (lip corner puller), AU14 (dimpler), AU15 (lip corner depressor), AU17 (chin raiser),AU18 (lip puckerer), AU20 (lip stretcher), AU24 (lip pressor), AU26 (jaw drop), AU28 (lip suck), and AU43 (eyes closed). These are used both directly during the mimicry process and indirectly during affective matching while first passed through the emotion recognition sub-module before being sent to the Controller.

4.1.1. Emotion Recognition

The emotion recognition module is a component of the perceptual module that manages the recognition and fusion processes using the information gathered from the user inputs. Our implementation includes three types of modalities for the purposes of emotion analysis: facial emotion recognition from video, tone analysis from low-level audio information, speech emotion recognition from linguistic information. These sub-components are activated according to the selected empathy mechanism as well as the state of the dialogue. During listening, emotion recognition is not activated if the selected empathy mechanism is low-level mimicry behavior. For affective matching behavior during the listening state, the agent uses the immediate information from the video and audio emotion recognition sub-components. After the user is finished speaking, the agent enters the thinking state, where the overall emotion of the interaction partner is calculated by including the emotion recognition from the speech signal and sent to the controller for a proper response.

Video emotion recognition component uses OpenCV library [28] for face detection and the CK Dataset [29] for training. The face images are categorized in basic emotion categories (Anger, Disgust, Fear, Joy, Sadness, Surprise and Contempt) as well as Valence value in a weighted scoring system [2]. Emotions are recognized based on frames in 60 frames per second. Audio emotion recognition component is intended to use the audio signal from the audio input stream to detect the emotions from the speech signal. We used RAVDESS audio dataset [30] to train our CNN-based emotion recognition model[3]. The speech emotion

[2]The facial emotion recognition model can be found in https://github.com/onyalcin/face_emotion_recognition

[3]The audio emotion recognition model can be found in https://github.com/onyalcin/audio_emotion_recognition

recognition component is activated after the speech recognition component fully processes the speech signal, which explained in the previous section. The recognized speech is further processed in this component using the SO-CAL sentiment analyzer [31] and NRC-Canada System [32].

The outputs from these components are passed to the emotion fusion component along with the output from audio emotion recognition. We aim to use late-fusion classifier for combining the outputs of the different modalities to detect the emotion of the user during the user's speech. This process is done by providing a weighted scoring system for each basic emotion category: anger, disgust, fear, joy, sadness, surprise and contempt.

4.2. Behavior Controller

The central component of this framework is the behavior controller, which acts as a decision-maker that provides a link between perceptions and actions. Depending on the state of the conversation and the empathy mechanisms to be used, the behavior controller selects the input channel to be used, processed and evaluated for preparing a response. The decision to react to the emotional input from the user emphatically is made in this central module, where the emotion of the agent itself is regulated based on the percepts. During a conversation, the agent should be able to decide the conversational state depending on the interaction: listening, thinking, speaking and idle. Furthermore, at each one of these states, the behavior of the agent should change according to the emotional value (valence, arousal or the emotion category) as well as the empathic behavior the agent is assigned to.

Behavior Controller includes three sub-modules: Central Controller, Empathy Mechanism, and the Dialogue Manager. These sub-modules process the information received from the perceptual module and generate a response using different strategies according to the state of the dialogue the agent is in. The Controller is responsible for the timing and sequence of processing for each sub-module. It determines which channels to receive information from, how to process them, in what order to process them, and how to generate the response.

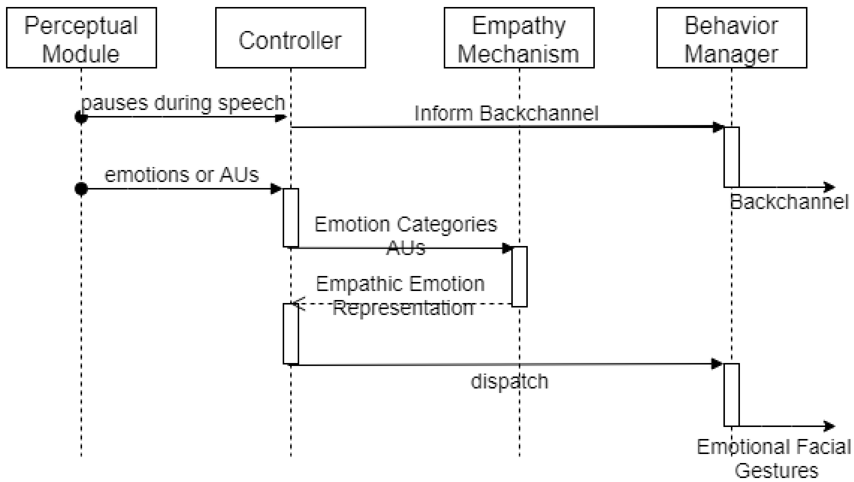

Figure 4: The information flow within the Behavior Controller during the listening state. It processes the pause signal from the speech and the emotion categories or AUs to give facial emotional feedback.

During the listening state, the agent uses the pauses during the user's speech to give backchanneling behavior and the facial emotion recognition results to provide a matching emotional gesture (see Figure 4). After the speech of the user is over, the agent will enter the thinking state while processing the message within the Behavior Controller according to the selected empathy mechanism (see Figure 5).

Language has an essential role in the perception and experience of emotions and has been suggested to be equally important for empathy[33]. Therefore, empathy mechanisms are highly linked with the dialogue manager in the behavior controller module. The Dialogue Manager is responsible for generating proper verbal responses according to the utterance of the interaction partner, as well as the emotional value of the desired response that is created by the Empathy Mechanisms sub-module. The Empathy Mechanisms sub-module embodies all the necessary emotion processing functions for generating levels of empathic behavior. The sequence of the processing steps changes according to the conversation state and the selected level within the empathy mechanism.

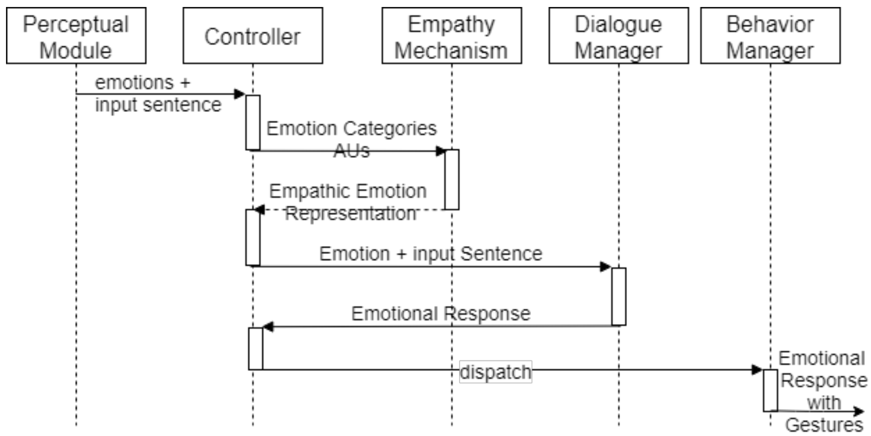

Figure 5: The information flow withing the Behavior Controller during the thinking and speaking cycles in low-level empathy behavior. The emotion of the user is first being processed to get a matching emotional behavior and sent with the input sentence to the dialogue manager to generate a response. This emotional response is then dispatched to the behavior manager to prepare synchronized behavior for the embodied agent.

Following the theoretical background, our framework includes three levels.

4.2.1. Low-Level Empathy

Low-level empathic behavior includes mimicry and affect-matching behavior [4], which allows for fast response mechanisms while relying on the Perception-Action Cycle [2]. This level of behavior requires less processing from the higher levels of cognitive functions and is achieved by either matching the perceived AUs or the perceived emotions with agent's expressive repertoire. During the listening, the Controller gathers the AU or emotion representation input from the perceptual module and sends it to the Empathy Mechanisms sub-module for matching representations in real-time. This information can be used for generating a matching speech response in the Dialogue Manager module (see Figure5). After the response is ready, the Controller again sends the emotional response to the Behavior Generation module for generating an output.

First evaluations of these capabilities showed that the implementation of our framework was perceived significantly more empathetic than baseline backchanneling behavior during the listening, where the difference between mimicry and affect matching is significant during speaking state of the agent [34].

4.2.2. Affect Regulation

Affect regulation abilities form the middle layer of the empathic behavior, where the agent regulates the dynamic properties of the emotions according to self-related and relationship-related parameters [4, 35]. These can include mood, personality, liking and similarity parameters. As a part of the hierarchical process, regulation mechanisms would share the emotional representations generated by the lower-level functions (see Figure 6).

In the empathy mechanisms sub-module, mood and personality parameters changes the dynamic properties of the emotions in various time-scales. Mood allows the emotions to be sustained and decay, allowing for consistent behavior over time. The personality of the agent can change the frequency, amplitude and duration of the generated gestures. The mood of the agent can change during an interaction, while the personality parameters are not likely to change. Relationship-related parameters such as liking and similarity can also affect the strength of empathic behavior expressed by the agent. For example, predetermined liking and familiarity values can be used to compute the similarity of the agent is expressed emotion compared to the emotion of the user [10, 9].

4.2.3. Cognitive Mechanisms

At the highest level of the empathy hierarchy, we have goal-oriented behavior that takes the context information into account [4]. The Empathy Mechanism sub-module must include information about the goals of the agent, the event and the accompanying emotion in order to calculate the appraisal of the situation. The low-level information about the emotions will be received from the lower levels of the empathic mechanisms as well as emotion representations via the perceptual system. Information about relation from affect regulation level is also be used to compute the Information about context will be generated

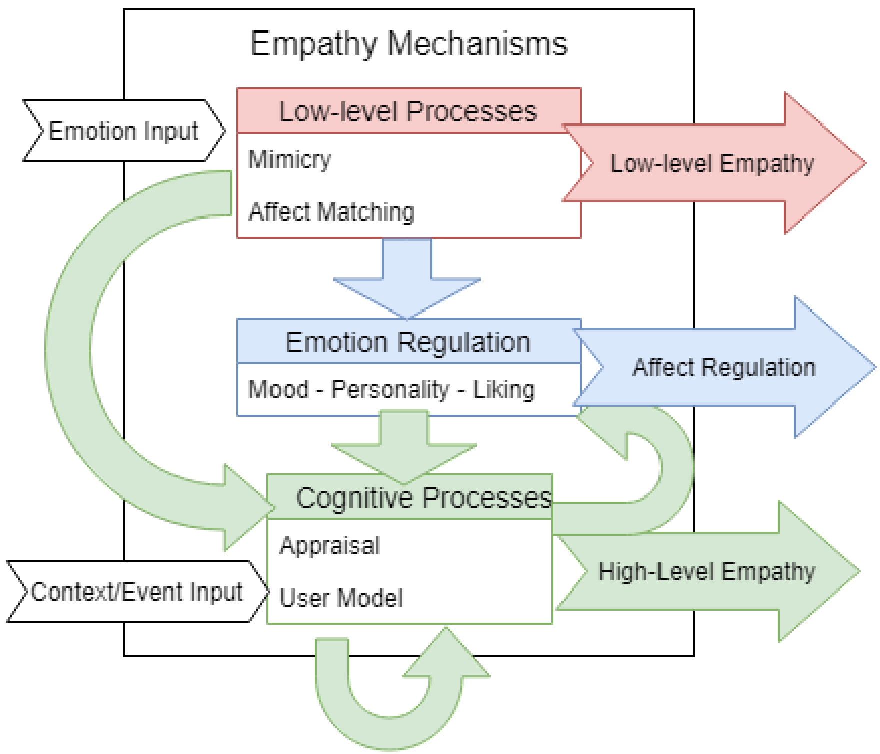

Figure 6: The process within Empathy Mechanisms sub-module allows for the generation of levels-of empathic behavior. The red colored arrows show low-level empathy flow, and blue arrows show the mid-level empathic behavior with affect regulation. Green arrows show multiple ways the higher-level empathy can be processed within the sub-module. The cyclic nature of information flow can be used for appraisal and re-appraisal of the situation with self and other-oriented perspectives.

from the conversation; hence, the Dialogue Manager. This suggests that, unlike the lower levels of empathic processing, cognitive mechanisms require the processing of contextual information before being processed by the empathy mechanisms sub-component. Moreover, this cycle can be re-processed multiple times to allow for appraisal, re-appraisal and higher level recursive reasoning. This repetitiveness in the processing would result in a longer thinking state during the conversational cycle (see Figure 6).

The implementation of the appraisal processes can follow different theoretical models [36, 37, 38, 39]. In their paper, Ochs and colleagues [8] provided a formal model of empathy according to the appraisal of the situation. Similarly, Broekens and colleagues [40] suggested a formal model for Scherer's appraisal theory [37] to use the representations of inputs from percepts to assign appraisal values based on some evaluation functions. However, an important distinction in cognitive processes related to empathic behavior is to model the other-focused appraisal of the situation, not the self-focused appraisal. A self-focused appraisal is the evaluation of the situation according to one's own goals and beliefs, where other-focused appraisal is the evaluation of the situation from the other's point of view. In that sense, a user model that relates to the Theory of Mind is a must. A good example of other-focused appraisal can be seen in the work of Lisetti and colleagues [41] that adopts user-modeling techniques for their goal-oriented embodied conversational agent.

4.3. Behavior Generation

The Behavior Generation module is responsible for generating the verbal and non-verbal response with proper timing and synchronization. It receives the emotion and proper output channels from the Behavior Controller module and prepares it for the agent interface, which is the Smartbody behavior realizer [42]. Smartbody allows for transforming messages received from the module into locomotion, object manipulation, lip syncing, gazing and nonverbal behaviors of the 3D agent in real time. We use the Stomp library [43] to provide a two-way communication between the framework and the behavior realizer with Behavior

Markup Language (BML) [44]. BML is a powerful standard in ECAs that allows for synchronization between multiple behaviors. Using BML standard, we can synchronize the exact timing of each gesture (body gestures, gaze, head movements, facial gestures) in relation to the speech of the agent.

The behavior generation component allows for synchronization of reflective behavior such as shifting body posture, breathing gaze or self-touching behavior as responses to the events happening in the environment and idle movement patterns during the lack of external input increases the quality of the interaction [45, 46, 47]. The consistency and coordination of these movements allow generating a sense of personality in affective agents [48].

5. Conclusion and Future Work

We have presented a framework and implementation of the levels of empathy behavior in embodied conversational agents following the theoretical background on empathy as well as the best practices in ECA research. We showed how each level in the hierarchical model of empathy could be realized via the interaction between different components of the framework in a conversational setting. This framework and implementation are intended to provide a blueprint for a wide range of applications with variations in input and output channels for emotion. Emotional representations linked to other bio-signals such as gaze behavior, breathing and heart rate, as well as non-anthropomorphic expressive behavior can be used following the layers of empathic processing.

6. Acknowledgements

This work was partially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) [RGPIN-2019-06767] and the Social Sciences and Humanities Research Council of Canada (SSHRC) [435-2017-0625].

References

[1] A. Paiva, I. Leite, H. Boukricha, I. Wachsmuth, Empathy in virtual agents and robots: a survey, ACM Transactions on Interactive Intelligent Systems (TiiS) 7 (3) (2017) 11.

[2] F. B. de Waal, S. D. Preston, Mammalian empathy: behavioural manifestations and neural basis, Nature Reviews Neuroscience 18 (8) (2017) 498.

[3] A. Coplan, P. Goldie, Empathy: Philosophical and psychological perspectives, Oxford University Press, 2011.

[4] Ö. N. Yalçin, Modeling empathy in embodied conversational agents, in: Proceedings of the 2018 on International Conference on Multimodal Interaction, ACM, 2018, pp. 546–550.

[5] F. B. De Waal, The 'russian doll'model of empathy and imitation, On being moved: From mirror neurons to empathy (2007) 35–48.

[6] S. Brave, C. Nass, K. Hutchinson, 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.

[7] H. Prendinger, J. Mori, M. Ishizuka, Using human physiology to evaluate subtle expressivity of a virtual quizmaster in a mathematical game, International journal of human-computer studies 62 (2) (2005) 231–245.

[8] M. Ochs, D. Sadek, C. Pelachaud, A formal model of emotions for an empathic rational dialog agent, Autonomous Agents and Multi-Agent Systems 24 (3) (2012) 410–440.

[9] S. H. Rodrigues, S. Mascarenhas, J. Dias, A. Paiva, A process model of empathy for virtual agents, Interacting with Computers 27 (4) (2015) 371–391.

[10] H. Boukricha, I. Wachsmuth, M. N. Carminati, P. Knoeferle, A computational model of empathy: Empirical evaluation, in: 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, IEEE, 2013, pp. 1–6.

[11] U. Hess, A. Fischer, Emotional mimicry: Why and when we mimic emotions, Social and Personality Psychology Compass 8 (2) (2014) 45–57.

[12] K. R. Scherer, T. Bänziger, E. Roesch, A Blueprint for Affective Computing: A sourcebook and manual, Oxford University Press, 2010.

[13] M. Schroder, E. Bevacqua, R. Cowie, F. Eyben, H. Gunes, D. Heylen, M. Ter Maat, G. McKeown, S. Pammi, M. Pantic, et al., Building autonomous sensitive artificial listeners, IEEE Transactions on Affective Computing 3 (2) (2012) 165–183.

[14] J. Bates, et al., The role of emotion in believable agents, Communications of the ACM 37 (7) (1994) 122–125.

[15] D. McNeill, Hand and mind: What gestures reveal about thought, University of Chicago press, 1992.

[16] J. Gratch, J. Rickel, E. André, J. Cassell, E. Petajan, N. Badler, Creating interactive virtual humans: Some assembly required, IEEE Intelligent systems (4) (2002) 54–63.

[17] R. Maatman, J. Gratch, S. Marsella, Natural behavior of a listening agent, in: International Workshop on Intelligent Virtual Agents, Springer, 2005, pp. 25–36.

[18] S. Duncan, Some signals and rules for taking speaking turns in conversations., Journal of personality and social psychology 23 (2) (1972) 283.

[19] G. Charles, Conversational organization: Interaction between speakers and hearers (1981).

[20] M. Skowron, Affect listeners: Acquisition of affective states by means of conversational systems, in: Development of Multimodal Interfaces: Active Listening and Synchrony, Springer, 2010, pp. 169–181.

[21] J. Gratch, N. Wang, J. Gerten, E. Fast, R. Duffy, Creating rapport with virtual agents, in: International Workshop on Intelligent Virtual Agents, Springer, 2007, pp. 125–138.

[22] G. Ball, J. Breese, Emotion and personality in a conversational agent, Embodied conversational agents (2000) 189–219.

[23] S. Brave, C. Nass, Emotion in human-computer interaction, The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications (2003) 81–96.

[24] T. Giannakopoulos, pyaudioanalysis: An open-source python library for audio signal analysis, PloS one 10 (12).

[25] A. Zhang, Python speechrecognition library (2017).
URL `https://github.com/Uberi/speech_recognition#readme.`

[26] Google, Google cloud speech api for python (2014-2017).
URL `https://cloud.google.com/speech-to-text/`

[27] D. McDuff, A. Mahmoud, M. Mavadati, M. Amr, J. Turcot, R. e. Kaliouby, Affdex sdk: a cross-platform real-time multi-face expression recognition toolkit, in: Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems, ACM, 2016, pp. 3723–3726.

[28] G. Bradski, The OpenCV Library, Dr. Dobb's Journal of Software Tools.

[29] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, I. Matthews, The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression, in: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, IEEE, 2010, pp. 94–101.

[30] S. R. Livingstone, F. A. Russo, The ryerson audio-visual database of emotional speech and song (ravdess): A dynamic, multimodal set of facial and vocal expressions in north american english, PloS one 13 (5) (2018) e0196391.

[31] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, M. Stede, Lexicon-based methods for sentiment analysis, Computational linguistics 37 (2) (2011) 267–307.

[32] S. M. Mohammad, S. Kiritchenko, X. Zhu, Nrc-canada: Building the state-of-the-art in sentiment analysis of tweets, in: Proceedings of the seventh international workshop on Semantic Evaluation Exercises (SemEval-2013), Atlanta, Georgia, USA, 2013.

[33] B. L. Omdahl, Cognitive appraisal, emotion, and empathy, Psychology Press, 2014.

[34] Ö. N. Yalçın, S. DiPaola, Evaluating levels of emotional contagion with an embodied conversational agent, in: Proceedings of the 41st Annual Conference of the Cognitive Science Society, 2019.

[35] M. H. Davis, Measuring individual differences in empathy: Evidence for a multidimensional approach., Journal of personality and social psychology 44 (1) (1983) 113.

[36] R. S. Lazarus, Progress on a cognitive-motivational-relational theory of emotion., American psychologist 46 (8) (1991) 819.

[37] K. R. Scherer, Appraisal considered as a process of multilevel sequential checking, Appraisal processes in emotion: Theory, methods, research 92 (120) (2001) 57.

[38] C. A. Smith, L. D. Kirby, Consequences require antecedents: Toward a process model of emotion elicitation.

[39] I. J. Roseman, Appraisal determinants of emotions: Constructing a more accurate and comprehensive theory, Cognition & Emotion 10 (3) (1996) 241–278.

[40] J. Broekens, D. Degroot, W. A. Kosters, Formal models of appraisal: Theory, specification, and computational model, Cognitive Systems Research 9 (3) (2008) 173–197.

[41] C. Lisetti, R. Amini, U. Yasavur, N. Rishe, I can help you change! an empathic virtual agent delivers behavior change health interventions, ACM Transactions on Management Information Systems (TMIS) 4 (4) (2013) 19.

[42] M. Thiebaux, S. Marsella, A. N. Marshall, M. Kallmann, Smartbody: Behavior realization for embodied conversational agents, in: Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 1, International Foundation for Autonomous Agents and Multiagent Systems, 2008, pp. 151–158.

[43] J. Briggs, Stomp messaging library for python (2018).
URL https://pypi.org/project/stomp.py/4.1.21/

[44] S. Kopp, B. Krenn, S. Marsella, A. N. Marshall, C. Pelachaud, H. Pirker, K. R. Thórisson, H. Vilhjálmsson, Towards a common framework for multimodal generation: The behavior markup language, in: International workshop on intelligent virtual agents, Springer, 2006, pp. 205–217.

[45] U. Bernardet, S.-h. Kang, A. Feng, S. DiPaola, A. Shapiro, A dynamic speech breathing system for virtual characters, in: International Conference on Intelligent Virtual Agents, Springer, 2017, pp. 43–52.

[46] U. Bernardet, S. DiPaola, A framework for exogenous and endogenous reflexive behavior in virtual characters, in: International Conference on Intelligent Virtual Agents, Springer, 2015, pp. 132–136.

[47] M. Nixon, S. DiPaola, U. Bernardet, An eye gaze model for controlling the display of social status in believable virtual humans, in: 2018 IEEE

Conference on Computational Intelligence and Games (CIG), IEEE, 2018, pp. 1–8.

[48] U. Bernardet, M. Saberi, S. DiPaola, Simulink toolbox for real-time virtual character control, in: International Conference on Intelligent Virtual Agents, Springer, 2016, pp. 456–459.