Towards Empathic Conversational Interaction

Micol Spitale ,
Politecnico di Milano, Italy
IBM, Italy
micol.spitale@polimi.it
Franca Garzotto
Politecnico di Milano, Italy
franca.garzotto@polimi.it

CHI'20,, April 25–30, 2020, Honolulu, HI, USA
ACM 978-1-4503-6819-3/20/04.
https://doi.org/10.1145/3334480.XXXXXXX

Abstract
In recent years, "computational empathy" has emerged as a new challenging research field. Computational empathy investigates how artificial agents can manifest empathic behaviours towards the user, and how they can elicit empathy during the human-agent interaction. Such "empathic agents" have the capacity to place themselves into the emotional position of a user (or another agent), and behave taking such emotional understanding into account. The paper explores a computational empathy approach in the context of conversational interaction, and presents an empathic conversational framework grounded on the empathy theory. The framework provides a conceptual tool for designing and evaluating empathic conversational agents. Overall, our research contributes to a deeper understanding of the role of empathy in conversational interaction.

Author Keywords
Empathy; Conversational Interaction; Artificial Agents; Framework; Computational Empathy

CCS Concepts
•Human-centered computing → HCI theory, concepts and models;

Introduction

Empathy is defined as the feeling by which one understands and shares another person's experiences and emotions, and it is known to facilitate the development of social relationships, affection, and familiarity among humans [6]. It is generally acknowledged that empathic behavior can be reinforced by appropriate verbal communication, through empathetic voice and phrases that are matched to the emotional state of the user [18]. "Computational empathy" is a new emerging research field, which investigates how artificial agents, such as robots or virtual agents, can manifest empathic behaviours towards the user and how they can elicit empathy in the user during interaction. Our research explores "computational empathy" in the context of conversational interaction. For a conversational agent an empathic behavior brings several benefit in term of length of interaction [2], reduction of user's stress and frustration [15, 3] and increased engagement [5]. In this paper, we conceptualize the main dimensions for modeling empathy during conversational interaction: behaviour, perception, regulation, and assessment. These dimensions are based on a literature review on the scientific research about empathy modeling. Our goal is to define a modeling framework for empathic conversational agents that helps researchers and designers to master the complexity of modeling empathy in this class of systems and designing their empathic conversational behaviors during the interaction with the users.

Related work

In recent years, many effort have been made for implementing empathy in artificial agents. Empathic agents can be seen as agents that have the capacity to place themselves into the position of a user's or another agent's emotional situation and respond appropriately [13]. Some researchers focused on modeling affective empathy of virtual agents, which are tutors on tablet or computer screens with abstract, cartoon-like or human-like appearance, called Embodied Conversational Agents (ECA) [11, 8]. [11] proposed a framework for modeling empathy in ECAs grounded on the empathy theory for which the empathic behavior is characterized by perception, recognition and expression of affective stimuli [13]. Other researchers investigated the empathic behaviors in human-robot interaction, [6] and [7], particularly among Social Assistive Robots (SARs). Some SARs can detect user emotions and expressed their emotions according to the user's tracked ones, such that user can perceive such robotic companions as "empathic".

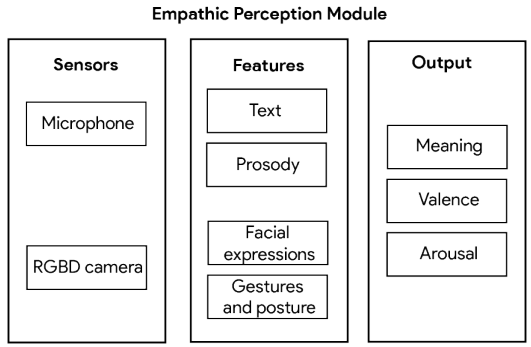

Figure 1: Empathic Conversational Perception Module.

Empathic Conversational Framework

We would like to propose an early framework for empathic conversational interaction grounded on the theory of empathy of [16] and inspired by the model proposed by [11]. During human-human conversation, alternatively we act as listener and speaker. We are able to actively listen to our companions and answer them expressing empathic behaviour and affective state. During the listening phase, the listener receives utterance signals as well as nonverbal

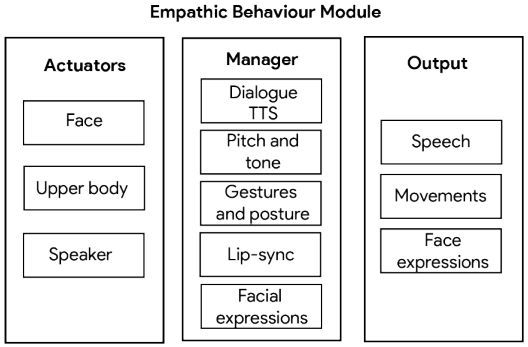

Figure 2: Empathic Conversational Behaviour Module.

cues from the speaker and may give nonverbal feedback in the meantime. Their roles exchange when the previous speaker stops talking. Then the next speaker produces verbal responses with appropriate body gestures based on what the speaker heard previously. This process keeps going on until the conversation ends [9]. For eliciting empathy into the users, the agent should on the one hand express the emotion in a natural way (speaker role), while on the other hand it should drive the conversation feeling how the user feels (listener role). Based on the empathy theory [13], any artificial agents should be equipped by three modules: Empathic Perception Module, what the artificial agent can track during the human interaction so it measures what the user feels (cognitive empathy); Empathic Behaviour Module, how the agent acts during the conversation in both listener and speaker roles; Empathic Regulation Module, how the agent can change its behavior according to the user reactions (affective empathy). In addition, an assessment method could be useful to measure the actual user empathy perceived during the interaction.

Empathic Perception Module

The perception module aims at tracking and measuring the users behaviours. To this purpose, the agent should be equipped by some basic sensors, such as a microphone and a RGB-D camera. From the former, the sound information are collected, such as the speech and prosody, and from the latter the visual data are gathered, such as facial expressions, and gestures and poses from the depth sensor data.

Four elements have been identify to characterize the perception module of the agent: text, speech prosody, facial expressions, and gestures and poses. For measuring these data, the agents need some dedicated modules. First of all, a face recognition module able to analyze the face and extract the Action Unit (AU) from the human face is required. Then, an algorithm should identify the gestures and the poses of the users, and classify them according to the Russels model [17, 12]. Finally, for the speech analysis we examine the content of the speech in term of meaning, the prosody of the speech, and the sentiment of the sentences of the speech to understand the affective state of the user. These affective data are processed in term of Valence, and Arousal scale [19], so that a space vector is extrapolated from the combination of the features extracted, and the meaning of the speech (see Fig.1). According to the user's role, the perception module has different purposes. If the user acts as listeners, the module's goal is to track the empathy elicited in the users. The real-time affective state of the user allows us to understand if the agent is eliciting empathy properly or not. Thank to this data, we can control and manipulate the behaviour of the agent, in term of both speech content and gestures to improve the elicitation process. For the user speaker role, the module aims to measure empathy expressed to understand how much the agent could be empathic with the user during the dialogue. The agent could react real-time according to the user in-

put with back-channel feedback, both in term of speech or movements [14, 10] as listener.

Empathic Behaviour Module
To provide believable behaviours and elicit empathy in the user, the agent should express an affective state by the means of its facial expression, the gestures, the pose, and besides the semantic, also the pitch of the speech. The Fig.2 depicts the architecture of the Behaviour Module and its components are the following: i) *Dialogue manager module*, which aims at answering properly to the user requests exploiting Natural Language Processing (NLP) algorithm; ii) *Facial expression module*, which aims is to regulate the emotions the agent expresses through its face according to the situation; iii) *Lip-sync module*, which goal is to simulate the human-speech synchronizing the lips movement with the words the agent speaks; iv) *Text-to-Speech module*, which convert the text into speech to be spoken; v) Gesture module, which is dedicated to the choice of the correct gesture for the purpose of the dialogue (listener or speaker role), and also its synchronization with the speech, to provide a natural communication during speaking; vi) Pitch and tone module, which regulate the speed and the tone of the speech according to the intention the agent wants to show. If the agent acts as listener or as speaker its goal is different. For the listener role, the behaviour is driven by the text of the sentence it has to speak. The facial expressions, the gestures and the prosody should be driven by the text analysis. The output of the text analysis is a 3D vector: a valence, arousal (from the sentiment and lexicon analysis) and meaning vector from the text analysis (NLP algorithm). This vector is the input for the behaviour modules which regulates the actions of the artificial agent. In the listener case, the Behaviour Module is dedicated to activate the back-channel feedback behaviour when required. It means giving constant feedback to the user to maintain the active listening phase.

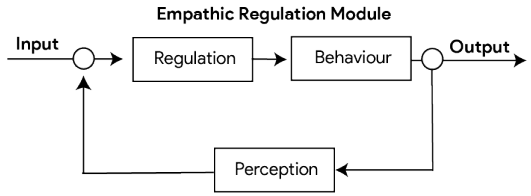

Figure 3: Empathic Conversational Regulation Module.

Empathic Regulation Module
To regulate the behaviour of the robot, the Regulation Module makes the decision of changing a specific behaviour or amplifying another one according to the perceived state of the user detected from the Perception module. The module controls the behaviour of the artificial agent to: i) keep the engagement of the user; ii) check if the empathy level elicited is above a threshold during speaking role; and iii) adjust the feedback behaviour during the active listening phase. This type of control is called feedback control system depicted in the Fig.3.

Assessment of elicited empathy in the user
Besides the quantitative measures of the user behaviours, some qualitative data could be collected to test the real empathy level elicited in the user. Typically for measuring an individual empathy level, the Empathy Quotient (QT) questionnaire is provided. It is a 60-item questionnaire designed to measure empathy in adults [1]. As our knowledge, there is a lack of standardized method which measures the real elicited empathy during human-computer interaction. We want to propose some new assessment techniques based on [4]. The user answer few questions that allow the researchers to understand if the user has been really involved and felt empathy during the interaction. In [4], the authors

just focused on the cognitive effort involved in a choice, and the results showed that people prefer not to choose the empathic option because of that. The same tasks can be used to test the empathy elicited in the user. For example, the user chooses 5 words to describe the agent. If the words are objective adjectives (i.e., physical descriptors) it implies that the agent wasn't able to elicit empathy in the user, vice-versa if the user selects some adjective related to the affective state, or the feelings, it means that the agent was able to establish a empathic relationship. The reported task is just one example which allows to test the effective empathy felt by the user during the conversational interaction.

Discussion and Conclusions

The expression, and elicitation of empathy in conversational interaction is a challenging research topic. This paper contributes to this field by proposing a conceptual framework for empathy to be applied in conversational interaction. We have described four aspects (grounded on the literature on empathy modeling) that highlight the role of empathy during conversational interaction, and correspond to four "modules" of a conceptual framework that would help the design of conversational human- computer interaction. We have pinpointed that these modules are useful for designing new empathic conversational technologies and for identifying open questions about the elicitation and expression of empathy in human-computer interaction. Our framework draws the attention on the role of empathy during a conversational human-agent interaction, independently from the agent "reification" (e.g., social robot, smart object, or software app). Several researches investigate some specific aspects of the single modules of the framework. For examples, many effort has been done in the scientific community for the automatic recognition of the emotion gestures in human, and many authors are still studying how to process the language and extrapolate the emotional content during a dialogue. The framework provides a conceptual tool to master the complexity of design and evaluation during the development of conversational interaction in its broader meaning. In addition, the framework gives some guidelines, and suggests an architecture of the agent which allows the researcher to have a complete overview of the implementation of the empathy module in a conversational human-agent interaction. Still, our framework is preliminary and has some limitations. In its present form, it might be too general, and might not be able to capture some design requirements for emphatic conversational interaction that are specific of some types of social agents, e.g., social robots or virtual reality agents, where emphatic properties should be customized to the very nature of the agent.

REFERENCES

[1] Simon Baron-Cohen and Sally Wheelwright. 2004. The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of autism and developmental disorders* 34, 2 (2004), 163–175.

[2] Timothy W Bickmore and Rosalind W Picard. 2005. Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction (TOCHI)* 12, 2 (2005), 293–327.

[3] Winslow Burleson and Rosalind W Picard. 2004. Affective agents: Sustaining motivation to learn through failure and a state of stuck. In *Workshop on Social and Emotional Intelligence in Learning Environments*.

[4] C Daryl Cameron, Cendri A Hutcherson, Amanda M Ferguson, Julian A Scheffer, Eliana Hadjiandreou, and Michael Inzlicht. 2019. Empathy is hard work: People choose to avoid empathy because of its cognitive costs. *Journal of Experimental Psychology: General* (2019).

[5] Amy Coplan and Peter Goldie. 2011. *Empathy: Philosophical and psychological perspectives*. Oxford University Press.

[6] Berardina De Carolis, Stefano Ferilli, and Giuseppe Palestra. 2017. Simulating empathic behavior in a social assistive robot. *Multimedia Tools and Applications* 76, 4 (2017), 5073–5094.

[7] Berardina Nadja De Carolis, Stefano Ferilli, Giuseppe Palestra, and Valeria Carofiglio. 2015. Towards an empathic social robot for ambient assisted living.. In *ESSEM@ AAMAS*. 19–34.

[8] Fiorella De Rosis, Addolorata Cavalluzzi, Irene Mazzotta, and Nicole Novielli. 2005. Can embodied conversational agents induce empathy in users? *Virtual Social Agents* 65 (2005).

[9] Minjie Hua, Fuyuan Shi, Yibing Nan, Kai Wang, Hao Chen, and Shiguo Lian. 2019. Towards More Realistic Human-Robot Conversation: A Seq2Seq-based Body Gesture Interaction System. *arXiv preprint arXiv:1905.01641* (2019).

[10] Yuka Kobayashi, Daisuke Yamamoto, Toshiyuki Koga, Sachie Yokoyama, and Miwako Doi. 2010. Design targeting voice interface robot capable of active listening. In *Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction*. IEEE Press, 161–162.

[11] Robert R Morris, Kareem Kouddous, Rohan Kshirsagar, and Stephen M Schueller. 2018. Towards an artificially empathic conversational agent for mental health applications: system design and user perceptions. *Journal of medical Internet research* 20, 6 (2018), e10148.

[12] Fatemeh Noroozi, Dorota Kaminska, Ciprian Corneanu, Tomasz Sapinski, Sergio Escalera, and Gholamreza Anbarjafari. 2018. Survey on emotional body gesture recognition. *IEEE transactions on affective computing* (2018).

[13] 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), 11.

[14] Hae Won Park, Mirko Gelsomini, Jin Joo Lee, and Cynthia Breazeal. 2017. Telling stories to robots: The effect of backchanneling on a child's storytelling. In *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI*. IEEE, 100–108.

[15] Helmut Prendinger and Mitsuru Ishizuka. 2005. The empathic companion: A character-based interface that addresses users'affective states. *Applied Artificial Intelligence* 19, 3-4 (2005), 267–285.

[16] Stephanie D Preston and Frans BM De Waal. 2002. Empathy: Its ultimate and proximate bases. *Behavioral and brain sciences* 25, 1 (2002), 1–20.

[17] James A Russell and Albert Mehrabian. 1977. Evidence for a three-factor theory of emotions. *Journal of research in Personality* 11, 3 (1977), 273–294.

[18] Adriana Tapus and Maja J Mataric. 2007. Emulating Empathy in Socially Assistive Robotics.. In *AAAI spring symposium: multidisciplinary collaboration for socially assistive robotics*. 93–96.

[19] Gyanendra K Verma and Uma Shanker Tiwary. 2017. Affect representation and recognition in 3d continuous valence–arousal–dominance space. *Multimedia Tools and Applications* 76, 2 (2017), 2159–2183.