A Formal Model of Emotions for an Empathic Rational Dialog Agent

Magalie Ochs, David Sadek, and Catherine Pelachaud

Abstract

Recent research has shown that virtual agents expressing empathic emotions toward users have the potentiality to enhance human-machine interaction. To provide empathic capabilities to a rational dialog agent, we propose a formal model of emotions based on an empirical and theoretical analysis of the users' conditions of emotions elicitation. The emotions are represented by particular agent's mental states composed of beliefs, uncertainties and intentions. This semantically grounded formal representation enables a rational dialog agent to identify from a dialogical situation the empathic emotion that it should express. An implementation and an evaluation of an empathic rational dialog agent have enabled us to validate the proposed model of emotions.

1 Introduction

A growing interest in using embodied virtual agents as interfaces to computational systems has been observed in recent years. This is motivated by an attempt to enhance human-machine interaction. Such agents are generally used to embody some roles typically performed by humans, as for example a tutor [23] or a receptionist [1]. The expression of emotions can increase their believability by creating an *illusion of life* [5, 53]. Moreover, recent research has shown that virtual agent's expressions of empathic emotions enhance users' satisfaction [25], engagement [25], performance in task achievement [37], and the perception of the virtual agent [7, 39].

In our research, we are particularly interested in the use of rational dialog agents in information systems. Rational dialog agents are BDI-like agents based on a formal theory of interaction called the *rational interaction theory* [46]. Users interact with them using natural language to find out information on a specific domain. We aim to give such agents the capability to express empathic emotions toward users while dialoging. Our aim is to improve interaction using empathic agents [25, 7, 39, 37].

Empathy is commonly defined as the capacity to "put your-self in someone else's shoes to understand her emotions" [35]. To be empathic assumes one is able to evaluate the emotional dimension of a situation for another person. To achieve this goal, a rational agent should know in which circumstances which emotions may be felt. To endow a rational dialog agent with empathic capabilities, one way is to provide it with a representation of the conditions of users' emotions elicitation during dialog. Then, the agent can deduce the emotions potentially felt by the user during the interaction and consequently its empathic emotions. Indeed, an empathic agent should express the emotion that it thinks the user may feel.

Several computational models of emotions include a representation of emotions elicitation conditions (as for instance in [16, 44]). It enables one to determine which

emotions of the virtual agent are triggered during an interaction. Generally, researchers [16, 44, 20] use a specific cognitive psychological theory of emotion (mainly the OCC model [34]) to define the agent's emotions. In these approaches, the authors assume that the emotions that may appear during interaction with one or multi agents and their conditions of elicitation correspond to those described in the chosen theory. For an empathic virtual dialog agent, the emotions that should be modeled are those that may be felt by the user during the dialog.

Our formal model of emotions is based on an empirical and theoretical approach. An exploratory analysis of real human-machine dialogs that have led users to express emotions has been conducted to try to identify the characteristics of emotional dialogic situations. Combined with the descriptions of emotions in cognitive psychological theories, the types and the conditions of elicitation of emotions that may appear during human-machine dialogs have been defined. In this paper, we propose a formal model of emotions and more particularly of their conditions of elicitation. More precisely, in order to provide empathic capabilities to a rational dialog agent, we propose to formally represent emotions, that the users may feel during human-machine interaction, in terms of beliefs, uncertainties and intentions.

The paper structure is as follow. In Section 2, we first introduce theoretical foundations on empathy that enable us to highlight the capacity that a virtual agent should have to be empathic (Section 2.1). We then present some existing formal models of emotions (Section 2.2) and empathic virtual agents (Section 2.3). In Section 3, we present the conditions of elicitation of certain emotions that may be triggered during human-machine dialogs. Section 4 describes the formal model of emotions. In Section 5, we conclude by presenting the implementation of an empathic rational dialog agent based on the model of emotions proposed, and the results of an evaluation of the latter.

2 Background

In this section, after introducing the definition and the characteristics of empathy (Section 2.1) on which we base our work, we present different formal models of emotions (Section 2.2), existing empathic virtual agents (Section 2.2) and the methods used to construct them.

2.1 The concept of empathy

Definition of empathy. Although there is no consensual definition of empathy, it is commonly defined as the capacity to "put your-self in someone else's shoes to understand her emotions" [35]. Most of researchers agree with the fact that empathy is composed of two dimensions: an *affective* dimension and a *cognitive* one. In a cognitive point of view, empathy requires to take the perspective of another person. In other words, to empathize with other means to simulate in your own mind a situation experienced by another one, to imagine oneself instead of the latter (with the same beliefs and goals). This process enables one to understand the emotions potentially felt by another person. For instance, Bob can imagine that Fred is happy because he won a thousand dollars and instead of Fred, Bob would be happy. In an affective point of view, this process of simulation may lead one to feel an emotion, called *empathic emotion*. For example, Bob may feel happy for Fred[1]. In psychology theories, it is not clear if an

[1] We consider only empathic emotions *congruent* with the person's emotions (for instance, we do not take into account an emotion of joy elicited by the sadness of another person).

empathic emotion is equivalent to a non empathic one. In the OCC cognitive model of emotion [34], such emotions are distinguished (for instance *happy for* emotion is different from *happy*). We follow this approach. The authors of the OCC model describe only two types of empathic emotion: *happy for* and *sorry for*. However, empathic emotions are as rich as felt emotions; there are not only two [21]. Indeed, by empathy, someone may for instance feel fear for another person. Therefore, there exists as many types of empathic emotion as types of non empathic one.

The elicitation of empathic emotion. As said before, empathy is composed of a cognitive and an affective dimension. In some cases, only one dimension may appear. Indeed, as highlighted by Jorland [24], the process of empathy may elicit no emotion or emotion different from the person for whom one has empathy. One can understand another's emotions (cognitive dimension) without feeling an empathic one. Moreover, empathic emotion is not necessary similar to the emotion of the person for whom one has empathy. For instance, a person may feel joy for another one even if the latter does not feel joy. The elicitation of empathic emotion and its intensity depends on different factor. As highlighted in [36], people experience more empathic emotion with persons with whom they have some similarities (for example the same age) or a particular relationship (as for example a friendship). According to the OCC model [34], the intensity of the empathic emotion depends on the degree to which the person is liked and deserves or not deserves the situation. People tend to be more pleased (*resp.* displeased) for others if they think the situation is deserved (*resp.* not deserved). Therefore, the intensity of an empathic emotion may be different from the intensity of the emotion that the person thinks the other feels. For instance, Bob can imagine that Fred is incredibly happy because he won a thousand dollars but Bob is not very happy for him because he does not think that Fred deserves it [33].

Contrary to the phenomenon of emotional contagion, the perception of an individual's expression of emotion is not necessary to elicit empathic emotions. Indeed, empathic emotions may be triggered even if the person does not express or feel emotion [40]. For instance, one can feel empathy for an unknown person by the reading of her story in a newspaper. The emotional contagion corresponds to the elicitation of an emotion by the perception of another person's expression of emotion. For instance, one may feel sad because she sees another one cries. In the case of empathy, it is the mental simulation of a situation experienced by another one that leads to elicit an emotion [40].

Finally, to be empathic, first of all, a virtual agent should be able to identify which emotions a user may feel in a given situation. To endow a virtual agent with such capabilities, a suitable method consist in a formally representation of emotions elicitation based on mental attitudes. Then, if the virtual agent knows the user's mental attitudes, it may deduce the user's potentially felt emotions. In the next section, we present in more details formal models of emotions.

2.2 Formal models of emotions

Several researchers have already proposed formal models to represent conditions of emotions elicitation. According to appraisal theory of emotions [50, 34], mental attitudes, such as beliefs and intentions, are determinant in the elicitation of emotions. One's emotions depend mainly on her beliefs on the event just occurred and on the impact on her goals. Conditions of emotions elicitation may then be described by particular combinations of mental attitudes, *i.e.* by specific *mental states*. As highlighted

in [9], such formal representation of emotions present several advantages both to preserve a consistency between an agent's mental state and its emotions and to identify one's emotions based on her mental attitudes.

Castelfranchi [10] has proposed to describe the emotions of shame and envy in terms of beliefs and goals. Based on this work, these emotions have been formally described in [54]. In [15], the causes of emotions are represented by agent's beliefs on the state of its goals and on the responsible agent. Then, an agent has, for instance, a joy emotion if one of its goal is achieved. In [13], an emotion is elicited by the agent's belief that the probability to achieve its goals has changed. For instance, a negative emotion appears when an agent thinks that the probability to achieve one of its goals has just decreased. Jaques and Viccari [2] propose a BDI representation of emotions to infer the user's felt emotions during her interaction with a pedagogical virtual agent. However, this model is domain-dependant, usable only in the context of learning. In the work of Meyer [30], based on the communicative theory of emotions [31], emotions of joy, sadness, anger, and fear and the resulting emotional behavior are described by particular configurations of mental attitudes described in modal logic. For instance, joy is elicited when an agent's intention is completed and when the agent thought that this intention was feasible. In this model, the emotions are used as heuristics to identify the most appropriate action of the agent. In [3], a BDI formalization of the emotions described in the OCC model [34] is proposed. The emotions correspond to combinations of beliefs and desires. Distress, for example, is characterized by the fact that the agent believes a proposition true and desires the contrary. The particularity of this formalization is the use of the mental attitude of desire. However,some problems may appear. For instance, in the case of distress, while all the agent's desires are not satisfied, the agent feels distress.

Finally, formalization of emotions in terms of mental attitudes appears as useful method to describe conditions of emotions elicitation and to potentially infer user's emotions. However, in existing works, most of researchers address specific emotions based on particular psychological theories of emotions. In this paper, we go beyond by proposing a formal model of emotions based both on an empirical and theoretical approach. Before presenting our model, in the next section, existing empathic virtual agents are introduced.

2.3 Empathic virtual agents

Empathy in human-machine interaction can be considered in two ways: a user can feel empathic emotions toward a virtual agent (for instance in *FearNot!* [36]) or a virtual agent can express empathic emotions toward a user [16, 29, 44, 41]. In our research, we focus on the empathy of a virtual agent toward users.

Most of empathic virtual agents are based on the OCC model [34]. Consequently, only two types of empathic emotions are considered : *happy-for* and *sorry-for*. However, research in psychology suggests that the type of an empathic emotion toward a person is similar to the type of the emotion of the latter [21]. Indeed, by empathy, someone may, for instance, feel fear for another person. Therefore, there exist as many types of empathic emotion as types of non empathic one. Then, an empathic virtual agent should *feel* an empathic emotion of frustration for the user if it thinks the user is frustrated.

In [16], the *happy-for* (*resp. sorry-for*) emotion is elicited by the empathic agent when a goal of another agent (virtual agent or user) is achieved (*resp.* failed). The

empathic virtual agent has a representation of the other agent's goals. It deduces these goals from their emotional reactions. Consequently, the agent knows the other's goals only if they have been involved in an emotion elicitation. Therefore, the other agent's goals representation might be incomplete. In [44], the virtual agent expresses *happy-for* (*resp. sorry-for*) emotion only if it detects a positive (*resp.* negative) emotional expression of its interlocutor. The agent's empathic emotions are in this case elicited by the perception of the expression of an emotion of another agent. Identically, in [41], the virtual agent expresses empathy according to the user's emotions (frustration, calm or joy) recognized through physiological sensors. However, an empathic emotion can be elicited even if this emotion is not felt or expressed by the interlocutor [40].

Another approach consists in observing real interpersonal mediated interactions in order to identify the circumstances under which an individual expresses empathy and how it is displayed. The system *CARE* (*Companion Assisted Reactive Empathizer*) has been constructed to analyze user's empathic behavior during a treasure hunt game in a virtual world [29]. The results of this study are domain-dependent. The conditions of empathic emotion elicitation in the context of a game may not be transposable in another context (as for example the context in which a user interacts with a virtual agent to find out information on a specific domain).

Our method to create empathic rational dialog agent is based both on a theoretical and empirical approaches. It consists to identify through psychological cognitive theories of emotion and through the study of real human-machine emotional dialogs, the situations that may elicit emotions in users. In the next section, we present in more details our method to construct an empathic dialog agent.

3 The Conditions of Emotions Elicitation during Human-Machine Dialogs

An empathic rational dialog agent should express empathic emotions in situations in which the user potentially feels an emotion. To identify in which circumstances a user may feel an emotion during human-machine interaction, we have studied real human-machine dialogs, that have led a user to express emotion, in the light of cognitive appraisal theories of emotions.

According to the *cognitive appraisal theories* [50], emotions are elicited by a subjective interpretation of an event. An event may trigger an emotion only if the person thinks that it affects one of her goals [26]. According to Scherer [47], an emotion elicitation depends mainly on the *consequences of the event* on the individual goal (for instance a goal achievement or a goal failure), *the causes of the event*, the consistency between the current situation (*i.e.* the consequences of the occurred event on the individual's goals) and the situation expected by the individual, and the coping potential (*i.e.* the capacity of an individual to deal with a situation that has led to a threat or failed goal). Finally, the interpretation of an event depends principally on the individual's goals and beliefs (on the event, its causes, its real and expected consequences, and on her coping potential). That explains the different emotional reactions of distinct individuals in front of a same situation.

In a dialog context, an event corresponds to a communicative act [4]. Consequently, according to the appraisal theory of emotion [50], a communicative act may trigger a user's emotion if it affects one of her goals. To identify more precisely the dialogical situations that may lead a user to feel emotion, we have analyzed real human-machine

dialogs that have led a user to express emotions.

The analyzed dialogs have been derived from two vocal applications. The users interact orally with a virtual dialog agent to find out information on a specific domain (on stock exchange or on restaurants). First, the dialogs have been annotated with the label *negative_emotion* by two annotators[2]. The annotations have been done based on vocal and semantic cues of user's emotions. Secondly, these dialogs have been annotated with a particular coding scheme in order to highlight the characteristics of the dialogical situations that may elicit emotions in a human-machine context (for more details on the coding scheme see [32]). The analysis of the annotated dialogs has enabled us to identify more precisely the characteristics of a situation that may lead to a *negative* emotion elicitation in human-machine interaction (the results are described in details in [32]). These results have been combined with the descriptions of emotions from appraisal theory [47] in order to identify more precisely the types of emotion a user may feel during human-machine interaction and their conditions of elicitation.

According to appraisal theory of emotion, a positive emotion is generally triggered when a goal is completed. More precisely, if the goal achievement was expected, an emotion of **satisfaction** is elicited; while, if it was not expected, an emotion of joy appears [47]. In human-machine dialogs, a user's goal achievement corresponds to the successful completion of her intention[3]. Generally, the user expects that her intentions (underlying her communicative act) will be achieved. Therefore, we consider only the emotion of satisfaction. We suppose that the user may experience satisfaction when one of her intentions is completed. In the analyzed human-machine dialogs, it appears that a user's intention failure generally triggers a negative emotion. According to [47], if a situation does not match with an individual's expectations, an emotion of **frustration** is elicited. Consequently, the user may experience frustration when one of her intentions failed. An emotion of **sadness** appears when the individual cannot cope with the situation. On the other hand, if she can cope with the situation, an emotion of **irritation** is elicited [47]. We suppose that, in case of intention failure, the user may feel sadness when she does not know any other action that enables her to carry out her failed intention; while if an action can be achieved by the user to complete her intention, she may experience irritation. When the goal failure is caused by another person, an emotion of **anger** may be elicited. In the empirical dialogs analysis, this situation may correspond to a user's intention failure caused by the dialog agent due to a *belief conflict* (the agent thinks the user has an intention different from her own one). The user may experience anger toward the agent when a *belief conflict* with the dialog agent has led to a goal failure.

Of course, we cannot deduce the exact emotion felt by the user from this description of emotions. Other elements (as for example the mood, the personality, and the current emotions) influence the elicitation of an emotion. However, this approach enables us to provide the virtual agent with information on the dialogical situations that *may* trigger a user's emotion. In the next section, we present a formal representation of emotions that enables us to endow rational dialog agent with empathic capabilities.

[2]Unfortunately, the dialog corpus did not cover situations that have led users to express positive emotion.

[3]In our research, we have focused on the user's intentions that an agent can be deduced from the enunciation of a user's communicative act. An intention is defined as a persistent goal (for more details see [46]).

4 A Formal Model of Emotions

To give the capability to a rational dialog agent to identify user's potentially felt emotions during an interaction, the conditions of user's emotions elicitation presented in the previous section have to be formally described. In this section, after briefly introducing the logical framework of rational dialog agent, we present the formal representation of the emotions of satisfaction, frustration, irritation, sadness, and anger, the axioms of the model and examples of theorems which follow from the formalization.

4.1 The logical framework

In our research, we use a model of rational agent based on a formal theory of interaction (called the *rational interaction theory* [46]), and on a BDI-like approach [12, 43]. In the rational interaction theory, the agent uses the mental attitudes of belief, uncertainty, choice and intention to reason and act in its environment. The rational interaction theory is expressed by a logic of attitudes and events (or actions), formalized in a first-order modal language. Let us briefly outline the part of formalism used in this paper. The symbols $\neg$, $\wedge$,$\vee$, $\Rightarrow$, and $\Leftrightarrow$ represent respectively the logical negation, conjunction, disjunction, implication, and equivalence. The symbols $\exists$ and $\forall$ represent the existential and universal quantifier, ϕ and ψ formulas, i, j, and u schematic variables denoting agents (virtual or real), $type$ a variable representing a type of emotion, e, e', e'' sequences of events possibly empty. The mental attitudes of belief, uncertainty and choice are formalized respectively by the modal operator B, U, and C such as $B_i\phi$ can be read as "the agent i thinks that ϕ is true"; $U_i\phi$ means that "the agent i thinks ϕ true without certainty"; $C_i\phi$ can be read as "the agent i desires that ϕ be currently true". The composite modal operator of intention I is defined from the modal operators of choice and belief. The formula $I_i\phi$ means that "the agent i has the intention that ϕ be true".

The mental state of the agent changes after the occurrence of an event. The notion of time is defined with respect to events and formalized through the operators $Feasible$ and $Done$. $Feasible(e, \phi)$ means that e can take place and if it does, ϕ will be true after that event. The formula $Done(e, \phi)$ means that e has just occurred and p was true before e ($Done(e) \equiv Done(e, true)$). Consequently, the remembrance of a belief about ϕ of an agent i before an event e is formalized by the following formula: $B_i(Done(e, B_i\phi))$. The formula $Agent(i, e)$ is true if and only if the agent i is the author of e. The operators B, C, $Feasible$, and $Done$ are based on a Kripke's possible-world semantic with, for each operator, an accessibility relation. The logic of the belief is KD45 (for more details on the logical framework see [45]).

4.2 The formal representation of emotions

Elicited emotions are emotions triggered by an event just occurred. They are related to a particular event and elicited because this event has positive or negative impact on an individual's goal[4]. An elicited emotion is generally defined by its *type*. We consider the types of emotion identified through our empirical and theoritical analysis: satisfaction, frustration, irritation, sadness, and anger (Section 3). An emotion is also characterized by an *intensity* that corresponds to the emotional impact of an event. Moreover, emotions may be directed toward someone in particular. For instance, one

[4]In our work, we focus on goal-based emotions. We do not consider emotions related to norms and values, such as pride or shame, which seem to not appear in the dialogical situations studied.

may feel anger against someone, himself, or happy for someone else. Given these characteristics, we propose to represent an elicited emotion by the following formula:

$$Emotion_u(type, i, e, \phi)$$

where u represents the agent who has the emotion, $type$ the type of the emotion (satisfaction, frustration, irritation, sadness, or anger), i the agent toward who the emotion is directed, e the event that has triggered the emotion, and ϕ the intention affected by the event. When u and i refer to the same agent, the emotion is not directed toward another agent. In our model, only the emotion of anger can be directed toward another agent (but the agent can also be angry against itself). The formula $Emotion_u(anger, i, e, \phi)$ means that an emotion of anger of the agent u against the agent i has been elicited by the event e that has affected the intention ϕ of u. The formula $Emotion_u(frustr, u, e, \phi)$ means that an emotion of frustration of the agent u has been elicited by the event e that has affected the intention ϕ of u.

An elicited emotion is defined by its conditions of elicitation which determine the type of the emotion, the agent who has the emotion, the agent toward who the emotion is directed, the triggered event, and the affected intention. The intensity of the emotion depends on the conditions of elicitation. We introduce in the following a formalization of the intensity variables used to compute the intensity of emotions.

The intensity of emotions. According to the appraisal theories of emotion [34, 50], the intensity of emotion depends on the subjective evaluation of the eliciting event. The intensity is defined as the emotional impact of the event [18]. It is determined by the values of variables called *intensity variables*, computed from the event's characteristics. Based on [34, 18], in our model, given the context in which a user dialogs with a virtual agent to find out information in a particular domain, we consider the following intensity variables:

- *The degree of certainty before the event concerning the intention achievement*: the degree of certainty of an information represents the probability for a person that an information is true. In the context of human-machine dialog, we are more particularly interested in the *degree of certainty* on the feasibility of an event to satisfy an intention. The higher is the degree of certainty, the more the user is certain to be able to satisfy her intention by the event. According to the OCC model [34], the unexpectedness is positively correlated to the intensity of emotion. Therefore, we suppose that the degree of certainty is negatively correlated to the intensity of negative emotion: the more the user was certain (before the event) to achieve her intention by the event just occurred, the higher is the intensity of the negative emotion in the case of an intention failure. In the same way, we suppose that the degree of certainty is positively correlated to the intensity of positive emotion: the more the user was uncertain (*i.e.* the lower is the degree of certainty) to satisfy her intention by the event just occurred, the higher is the intensity of positive emotion triggered by a satisfied intention.

  Formally, the degree of certainty of agent u concerning the feasibility of proposition p by sequence of events e is noted $deg\_cert(u, e, \phi) \in [0, 1]$.To represent the degree of certainty, we use a modal operator $P_u(\phi)$ which gives the probability that agent u allocates to proposition p. This operator is semantically defined in [46] as follows:

  $$M, w, v \models P_u(\phi) \text{ iff } Pr(M, w', v \models \phi | w' \in R_{B_u}(w))$$

The words are supposed equiprobable. The degree of certainty is then defined such as:

$$deg\_cert(u, e, \phi) = P_u(Feasible(e, \phi))$$

Consequently,

$$\begin{array}{l} B_u(\neg Feasible(e, \phi)) \text{ iff } deg\_cert(u, e, \phi) = 0 \\ U_u(\neg Feasible(e, \phi)) \text{ iff } deg\_cert(u, e, \phi) \in \ ]0, 0.5[\\ U_u Feasible(e, \phi) \text{iff } deg\_cert(u, e, \phi) \in \ ]0.5, 1[\\ B_u Feasible(e, \phi) \text{ iff } deg\_cert(u, e, \phi) = 1 \end{array}$$

The degree of certainty of agent u that intention ϕ is feasible by the sequence of events e equals to the probability that ϕ is reached by e ($P_{R_{B_u(w)}}(Feasible(e, \phi))$). Then, if the agent thinks event e is unable to satisfy ϕ, the degree of certainty is null. In the contrary case, it is equal to 1. If the agent is uncertain concerning the satisfaction of ϕ by e, the degree of certainty is in $]0.5, 1[$. If she is uncertain about the contrary, the degree of certainty is in $]0, 0.5[$.

- *The effort invested to try to complete the intention*: generally, a greater effort invested implies a more intense emotion [34]. We then suppose that the intensity of an emotion is proportional to the effort invested by the user.

  Formally, the effort done by agent u to try to satisfy intention ϕ (noted $effort(i, \phi)$) is represented by the number of actions done by u to try to satisfy ϕ:

  $$\begin{array}{l} \text{Given } \varphi(e_k) \equiv Unitary(e_k) \wedge Agent(u, e_k) \\ \wedge \exists e', e'' B_u \Big(Done(e'; e_k; e'', I_u\phi) \wedge \neg I_u\phi \\ \wedge \big([(e'' \neq Empty) \wedge Done(e'', I_u\phi)] \vee [(e'' = Empty) \wedge Done(e'', \neg I_u\phi)] \big) \\ \wedge Done(e_k; e'', I_u\phi) \Big) \\ effort(u, \phi) = card\{e_k | \varphi(e_k)\} \end{array}$$

  The effort represents formally the number of unitary events realized by the agent since she has the intention to satisfy ϕ and until she has not this intention anymore (*i.e.* she has satisfied it or she has decided to renounce). To represent the effort between 0 and 1, we introduce a maximum of effort (noted $effort\_max(i, \phi)$) that has to be fixed empirically. The latter corresponds to the maximum number of actions that agent u can do to achieve intention ϕ. We suppose that the *average effort* of agent u to try to complete ϕ (noted $effort\_average(u, \phi)$) is computed as follows: $effort\_average(u, \phi) = \frac{effort(u,\phi)}{effort\_max(u,\phi)}$.

- *the importance of the intention*: based on [16, 44], we distinguish the *importance for the user to achieve her intention* from *the importance not to have her intention failed*. The intensity of positive (*resp.* negative) emotion is proportional to the importance to achieve her intention (*resp.* not to have her intention failed). Typically, in the context of human-machine dialog, we can suppose that the intensity of a positive emotion elicited by the achievement of the user's intention

to be understood by the agent is less high than the intensity of negative emotion triggered by the fact that the agent does not understand her.

Formally, the importance for agent u to achieve intention ϕ is defined by the function $imp\_s(u, \phi)$. The importance not to have her intention failed is represented by the function $imp\_e(u, \phi)$[5]. The values of these functions are in $[0, 1]$. Depending on the application domain, they may be specified statically or dynamically. The values may be defined semantically as follows:

$$(1)\ imp\_s(u, \phi) = imp\_e(u, \phi)$$
$$(2)\ imp\_s(u, \phi) = 1 - imp\_e(u, \phi)$$

Indeed, the importance to satisfy an intention and the importance not to have an intention failed may be : (1) equal, or (2) negatively correlate depending on the intention itself. For instance, relation (1) can be used for the intention to have an article accepted in a journal, and relation (2) for the intention to be understood by an interlocutor.

- the *potential to cope in case of an intention failure*: research has not looked explicitly at the influence of *coping potential* on the intensity of an elicited emotion. We suppose that the intensity of a negative emotion is higher when the user does not know any action to complete her intention just failed. In other words, we assume that the intensity of a negative emotion is *conversely proportional* to the coping potential.

  Formally, we notice $potential\_cope(u, \phi)$ (in $[0, 1]$) the potential to cope of agent u in case of the failure of intention ϕ. To compute the value, the following formula is proposed:

$$potentiel\_cope(u, \phi) = max\,\{deg\_cert(u, e, \phi)\} \text{ for all } e \in Evt^{+}$$

  Consequently,

  if $\forall e B_u(\neg \text{Feasible}(e, \phi))$ then $potentiel\_cope(u, \phi) = 0$
  if $\exists e U_u Feasible(e, \phi) \wedge \forall e' \neg B_u Feasible(e', \phi)$
  then $potentiel\_cope(u, \phi) \in]0.5, 1[$
  if $\exists e U_u(\neg Feasible(e, \phi)) \wedge \forall e'(\neg B_u Feasible(e', \phi) \wedge \neg U_u Feasible(e', \phi))$
  then $potentiel\_cope(u, \phi) \in\ ]0, 0.5[$
  if $\exists e B_u Feasible(e, \phi)$ then $potentiel\_cope(u, \phi) = 1$

  The potential to cope equals to the maximum degree of certainty of agent u concerning the feasibility of ϕ by a sequence of events (Evt^{+} represents the set of sequences of primitive events). The potential to cope equals null if the agent believes that no sequence of events enables her to satisfy her intention that just failed, and 1 in the contrary case. If the agent is uncertain concerning the existence of events to satisfy ϕ, the potential to cope is in $]0.5, 1[$. If she is uncertain that event does not enable her to satisfy ϕ, the potential to cope is in $]0, 0.5[$.

[5]Functions of second-order modal logic are used because first-order modal logic does not enable us this writing.

The intensity variables implied in the computation of intensity differ depending on the valence of emotions (positive or negative). For instance, the intensity of positive emotion does not depend on the potential to cope contrary to the intensity of negative one. In the same way, the importance not to have an intention failed influences the intensity of negative emotion whereas the importance to satisfy an intention has an impact only on the intensity of positive one. Consequently, we introduce two intensity functions. The one for the positive emotions, noticed $f\_intensity\_pos$ is defined such as:

$$f\_intensity\_pos(u, e, \phi) = \\ (1 - deg\_cert(u, e, \phi)) * effort\_average(u, \phi) * imp\_s(u, \phi)$$

The intensity of positive emotions decreases conversely to the degree of certainty of the agent concerning the feasibility of her intention by the event $(1 - deg\_cert(u, e, \phi))$. The intensity is proportional to the average effort done by the agent ($effort\_average(u, \phi)$) and to the importance to satisfy her intention ($imp\_s(u, \phi)$). The function to compute the intensity of negative emotions , noted $f\_intensity\_neg$, is defined such as:

$$f\_intensite\_neg(u, e, \phi) = \\ deg\_cert(u, e, \phi) * (1 - potential\_cope(u, \phi)) * effort\_average(u, \phi) * imp\_e(u, \phi)$$

The intensity of negative emotions is proportional to the degree of certainty of the agent ($deg\_cert(u, e, \phi)$), to the average effort of the agent ($effort\_moyen(u, \phi)$), and to the importance not to have her intention failed ($imp\_e(u, \phi)$). The intensity decreases conversely to the potential to cope ($1 - potentiel\_reaction(u, \phi)$).

The intensity of both positive and negative emotions is between 0 and 1. Following the approach proposed in [19, 14, 44], in the functions, a multiplication between the parameters is used to ensure that when one of the parameters is null, the intensity of the emotion is null.

Remark. The proposed intensity functions do not enable one to compute the exact intensity of the emotion felt by the user during the interaction. However, it can be used by an empathic rational dialog agent to approximate the importance of an emotion potentially felt by the user.

The formalization of the conditions of emotions elicitation based on the empirical and theoretical analysis presented Section 3, is presented in the following.

Formal representation of the elicited emotions. Based on the analysis of human-machine dialogs (Section 3), the achievement of a user's intention may lead to an emotion of satisfaction. Consequently, formally, an emotion of *satisfaction* of agent u is triggered by event e that has affected intention ϕ of u (noted $Emotion_u(satisf, u, e, \phi)$) when event e has lead to the achievement of intention ϕ:

$$Emotion_u(satisf, u, e, \phi) \equiv^{def} B_u\big(Done(e, I_u\phi \\ \wedge (U_u Feasible(e, \phi) \vee B_u Feasible(e, \phi))) \wedge \phi\big) \\ \text{with } c = f\_intensity\_pos(u, e, \phi) \qquad \text{(Def.Satisf)}$$

This formula means that (1) u thinks e just occurred ($B_u(Done(e))$), (2) before e, u had the intention ϕ ($I_u\phi$), (3) and believed (without being necessarily certain) that e

would enable the achievement of ϕ ($U_u Feasible(e, \phi) \vee B_u Feasible(e, \phi)$), (4) after the occurrence of e, u thinks ϕ is true.

As highlighted Section 3, user may experience frustration in case of intention failure. Formally, an emotion of *frustration* of agent u is triggered by event e that has affected intention ϕ (noted $Emotion_u(frustr, u, e, \phi)$) when event e has lead to the failure of ϕ:

$$Emotion_u(frustr, u, e, \phi) \equiv^{def} B_u\big(Done(e, I_u\phi \wedge (U_u Feasible(e, \phi) \vee B_u Feasible(e, \phi))) \wedge \neg\phi\big)$$
$$\text{with } c = f\_intensity\_neg(u, e, \phi) \qquad \text{(Def.Frustr)}$$

This formula means that (1) u thinks e just occurred ($B_u(Done(e))$), (2) before e, u had the intention ϕ ($I_u\phi$), (3) and believed (without being necessarily certain) that e would enable the achievement of ϕ ($U_u Feasible(e, \phi) \vee B_u Feasible(e, \phi)$), (4) after the occurrence of e, u thinks ϕ is false ($B_u(\neg\phi)$).

Moreover, if the user can cope with an intention failure, irritation may appear (Section 3). Formally, an emotion of *irritation* of agent u is triggered by event e that has affected intention ϕ (noted $Emotion_u(irrit, u, e, \phi)$) when event e has lead to the failure of intention ϕ (*i.e.* to an emotion of frustration) which u thinks to be able to achieve by another way:

$$Emotion_u(irrit, u, e, \phi) \equiv^{def} Emotion_u(frustr, u, e, \phi) \wedge \exists e'(U_u Done(e', \phi) \vee B_u Done(e', \phi))$$
$$\text{with } c = f\_intensity\_neg(u, e, \phi) \qquad \text{(Def.Irrit)}$$

This formula means that an emotion of irritation is triggered by an event e in respect to an intention ϕ of u when (1) e has lead to an emotion of frustration in respect to the same intention ϕ (*i.e.* e has lead to the failure of ϕ), and (2) the agent u thinks (without being necessarily certain) an event e' should enable u to achieve ϕ ($\exists e'(U_u Done(e', \phi) \vee B_u Done(e', \phi))$).

An emotion of sadness is elicited, in case of an intention failure, when one has few potential to cope with the situation. Formally, an emotion of *sadness* of agent u is triggered by event e that has affected intention ϕ (noted $Emotion_u(sad, u, e, \phi)$) when event e has lead to the failure of intention ϕ (*i.e.* to an emotion of frustration) that u does not think to be able to achieve by another way:

$$Emotion_u(sad, u, e, \phi) \equiv^{def} Emotion_u(frustr, u, e, \phi) \wedge \forall e'\big(B_u(\neg\text{Feasible}(e', \phi)) \vee (\neg B_u Feasible(e', \phi) \wedge \exists e'' U_u(\neg Feasible(e'', \phi)))\big)$$
$$\text{with } c = f\_intensity\_neg(u, e, \phi) \qquad \text{(Def.Sad)}$$

This formula means that an emotion of sadness is elicited by e when (1) e has triggered an emotion of frustration ($Emotion_u(frustr, u, e, \phi)$), and (2) u thinks that no other event enables her to achieve ϕ ($\forall e'(B_u(\neg\text{Feasible}(e', \phi)) \vee (\neg B_u Feasible(e', \phi) \wedge \exists e'' U_u(\neg Feasible(e'', \phi))))$).

Following the analysis of human-machine dialogs (Section 3), it appears that the user may feel an emotion of anger against the virtual agent when an intention failure is caused by the virtual agent due to a belief conict. Formally, the emotion of *anger* of agent u against agent i about intention ϕ after the occurrence of event e (noted $Emotion_u(anger, i, e, \phi)$) is triggered by the failure of the intention ϕ caused by agent i because of a *belief conflict* on another intention ψ:

$$Emotion_u(anger, i, e, \phi) \equiv^{def} Emotion_u(frustr, u, e, \phi)$$
$$\wedge B_u\big(Done(e, \neg I_u\psi \wedge B_u(\neg B_i(I_u\psi)) \wedge B_i(I_u\psi))\big)$$
$$\text{with } c = f\_intensity\_neg(u, e, \phi) \qquad \text{(Def.Anger)}$$

The formula $B_u(Done(e, \neg I_u\psi \wedge B_u(\neg B_i(I_u\psi)) \wedge B_i(I_u\psi)))$ represents a *belief conflict* corresponding to the situation in which the agent u thinks, after e, that the agent i believes that u has another intention other than her own one. We suppose that, in this case, the failure of ϕ is due to this *belief conflict*[6].

Remark. Given the proposed formalization, several emotions may be triggered at the same time. For instance, in the case of an intention failure, an emotion of frustration may go with an emotion of sadness or irritation. That is coherent with the research showing that a person generally feels, at a given time, not a single emotion but a combination of several emotions [38, 49, 48]. For instance, one can feel joy to have a new job but sad that this job is far from her family.

Remark. The formalization of emotions are not sufficient to identify the exact user's felt emotions during the interaction. Indeed, only the emotion that *may* by triggered by the dialogical situations studied are considered. Other events external to the human-machine dialog may elicit emotions. However, the formalization of emotions described above can be used by an empathic rational virtual dialog agent to determine the type and the intensity of the emotion potentially felt by the user in certain dialogical situations.

The proposed formalization of emotions enables a rational agent to infer a user's potentially felt emotions given her beliefs, her uncertainties and her intentions. A rational dialog agent has a model of communicative acts based on the *speech acts theory* [4, 51] that provides it with the capabilities to deduce these user's mental attitudes underlying the communicative act performed. For instance, when the user asks an information, the rational agent deduces that the user has the intention to know the information and has the intention that the agent knows her intention to know the information. The agent infers also that the user believes that her intention will be achieved following the enunciation of the communicative act. Of course, the model of communicative acts does not enable the agent to infer all the user's mental attitudes. Moreover, the formalization of emotions does not cover all the emotions that may appear during human-machine interaction. However, the representation of emotions proposed enables us to provide to a rational agent some empathic capabilities in some dialogical situations. In the following, we introduce a formalization of empathic emotions.

[6]The logic used does not enable us to represent the causal relation between the belief conflict and the intention failure.

Formal representation of empathic emotions. An *empathic emotion* of agent i toward agent j is represented by a syntactic abbreviation noted $Emotion\_emp_{i,j}$. The formula:

$$Emotion\_emp_{i,u}(type, j, e, \phi)$$

means that agent i has an empathic emotion of a type $type$ toward another agent u. This emotion is triggered by an event e which has affected an intention ϕ of u. The variable j is introduced in order to represent an empathic emotion toward an agent and directed against another agent. For instance, the formula $Emotion\_emp_{i,u}(anger, j, e, \phi)$ represents an empathic emotion of anger of i toward u against j. To represent an empathic emotion of frustration of i toward u, we use the formula $Emotion\_emp_{i,u}(frustr, u, e, \phi)$.

An empathic emotion may be elicited when the agent thinks another agent has an emotion. The elicited empathic emotions are then defined as follow:

$$Emotion\_emp_{i,u}(type, c, j, e, \phi) \equiv^{def} B_i Emotion_u(type, c', j, e, \phi) \quad \text{with } c = f\_intensity\_emp_i(u, c', \phi) \quad \text{(Def.Emp)}$$

In others words, the fact that an agent i has an empathic emotion of a type $type$ toward the agent u and directed against the agent j because of an event e which has affected an intention ϕ, means that the agent i thinks that the agent u has an emotion of the type $type$ directed against j because of an event e which has affected one of her intentions ϕ. The intensity of the empathic emotion of agent i is computed with the function $f\_intensity\_emp_i$.

The intensity of empathic emotions. When an empathic rational dialog agent thinks the user feels an emotion, it may *have* an empathic emotion toward the latter. It depends on different factors as for instance the relation between the user and the virtual agent. Moreover, the intensity of the empathic emotion is not necessary similar to the intensity of the emotion that the agent thinks its interlocutor has. According to the OCC model [34], the intensity of an empathic emotion depends on the degree of appreciation of the interlocutor. However, it is not a necessary condition for the elicitation of empathic emotion. In other words, one can feel an empathic emotion toward someone she does not like. Generally, the intensity of an empathic emotion is proportional to the degree of appreciation. To represent the degree of appreciation between two agents i and j, the function $like_i(j)$ is introduced such as: the values of the function are in $[0, 1]$; if $like_i(j) \in$]0.5, 1] then agent i likes agent j and the closer the value is to 1, the more i likes j; if $like_i(j) \in$ [0, 0.5[then i dislikes j and the closer the value is to 0, the less i likes j; and if $like_i(j) = 0.5$ then i is neutral toward j. Moreover, the intensity of empathic emotion is higher if one thinks the person toward who she has an emotion, deserves the situation (or not deserve it in case of negative empathic emotion) [34]. The function $deserve_i(j, \phi)$[7] is introduced to represent the deservingness aspect of a property ϕ for agent j according to i. The values are between 0 and 1. $deserve_i(j, \phi) \in$]0.5, 1] means agent i thinks j deserves ϕ. The closer the value is to 1, the more i thinks j deserve it. On the contrary, if $deserve_i(j, \phi) \in$ [0, 0.5[then i thinks j does not deserve ϕ. If $deserve_i(j, \phi) = 0.5$ then i has no belief concerning the fact that j deserve or not ϕ. The value of this variable may depend on the probability to reach ϕ (if the probability is low - *resp.* high - the value of $deserve(\phi)$ is high - *resp.* low), social

[7] A function of second-order modal logic is used because firt-order modal logic does not enable us this writing

norms (related to the individual or social group), and lawful norms (defined by the laws). Finally, these values of appreciation and deservingness may be defined statically or dynamically (updated during the interaction). For instance, an emotion of anger of i against j can induce a decrease of the degree of liking i has for j. The function to compute the intensity of an empathic emotion of i, noted $f\_intensity\_emp_i$, is defined such as:

$$f\_intensity\_emp_i(u, c, \phi) = like_i(u) * deserve_i(u, \phi) * c$$

The intensity of empathic emotion of i toward u is proportional to the degree of liking of i toward u ($like_i(u)$), to the degree with which i thinks u deserves ϕ ($deserve_i(u, \phi)$), and the intensity of the emotion that i thinks u has (c). A multiplication between the parameters is used to ensure that when one of the parameters is null, the intensity of the emotion is null. Consequently, if i hates u ($like_i(u) = 0$), the intensity of the empathic emotion of i toward u is null.

4.3 Axioms and Theorems

Axioms. In the context of human-machine interaction, an empathic rational dialog agent should not adopt the intention that the user has negative emotions or does not feel positive ones. Consequently, the following axioms is imposed to our model:

$$\neg I_i Emotion_u(neg, k, e, \phi) \qquad \text{(A.1)}$$
$$\neg I_i(\neg Emotion_u(satisf, j, e, \phi)) \qquad \text{(A.2)}$$

By this way, the agent cannot act with the intention that the user has a negative emotion (represented by neg including frustration, irritation, sadness, and anger) or has not a positive emotion of satisfaction. These axioms guarantee that the agent is *well-intentioned*. We suppose that these axioms are applied for all empathic agents of the environment. By this way, an empathic agent cannot have positive (*resp.* negative) empathic emotion toward an agent which has positive (*resp.* negative) emotion because another agent has a negative (*resp.* positive) emotion.

Theorems. From our formal model of emotions, theorems have been proved. We present examples of them in the following.

First of all, a positive emotion of the rational dialog agent cannot be triggered because the agent thinks the user has a negative emotion (of frustration, irritation, sadness, or anger):

$$\vdash \neg Emotion_i\big(satisf, i, e, Emotion_u(neg, j, e_1, \phi)\big) \qquad \text{(T.1)}$$

In the same way, a rational dialog agent cannot have a negative emotion because the user has not a positive one (of satisfaction):

$$\vdash \neg Emotion_i\Big(neg, j, e, \big(\neg Emotion_u(satisf, u, e_1, \phi)\big)\Big) \qquad \text{(T.2)}$$

The rational dialog agent cannot have a negative emotion because the user has not a negative emotion:

$$\vdash \neg Emotion_i\big(neg, j, e, Emotion_u(neg, k, e_1, \phi)\big) \qquad \text{(T. 3)}$$

For these three theorems, proofs by contradiction follow from the definition of the emotions, the axioms A.1 and A.2, and from the distributivity of $\wedge$, B, and $Done$.

As corollary of the theorems T.1 and T.2, we have proved by contradiction the fact that an agent has not a positive (*resp.* negative) emotion does not imply that he has a negative (*resp.* positive) emotion:

$$\nvdash \neg Emotion_i(satisf, i, e, \phi) \Rightarrow Emotion_i(neg, j, e, \phi) \qquad \text{(C.1)}$$
$$\nvdash \neg Emotion_i(neg, j, e, \phi) \Rightarrow Emotion_i(satisf, i, e, \phi) \qquad \text{(C.2)}$$

Moreover, the elicitation of empathic emotions is consistent. Indeed, a same event cannot trigger both a positive and negative emotions with respect to a same intention:

$$\vdash \neg\big(Emotion_i(satisf, c, i, e, \phi) \wedge Emotion_i(neg, c', j, e, \phi)\big) \qquad \text{(T. 4)}$$

A proof by contradiction follows from the definition of emotions and from the following propriety of belief $B_i(p \wedge p') \Leftrightarrow B_i p \wedge B_i p'$.

As corollary of the theorems T.4, a same event cannot trigger both a positive and negative *empathic* emotions with respect to a same intention:

$$\vdash \neg\big(Emotion\_emp_{i,u}(satisf, j, e, \phi) \wedge Emotion\_emp_{i,u}(neg, k, e, \phi)\big) \quad \text{(C.4)}$$

Finally, theorems on the capacity of *introspection* of the rational dialog agent on its empathic emotions has been proved :

$$\vdash Emotion\_emp_{i,j}(type, c, k, e, \phi) \Leftrightarrow$$
$$B_i\big(Emotion\_emp_{i,j}(type, c, k, e, \phi)\big) \qquad \text{(T.5.1)}$$
$$\vdash \neg Emotion\_emp_{i,j}(type, c, k, e, \phi) \Leftrightarrow$$
$$B_i\big(\neg Emotion\_emp_{i,j}(type, c, k, e, \phi)\big) \qquad \text{(T.5.2)}$$

In other words, if the agent has (*resp.* has not) an empathic emotion, it believes that he has (*resp.* has not) this empathic emotion and *vice versa*. The proof of these theorems follow from the definition of emotions and from the following properties of belief $B_i(p) \Leftrightarrow B_i(B_i(p)))$ and $\neg B_i(p) \Leftrightarrow B_i(\neg B_i(p)))$

5 Implementation and Evaluation of an Empathic Rational Dialog Agent

Based on the model of emotions presented in the previous section, we have developed an empathic rational dialog agent. In the following, we first present the implementation of such an agent and we then expose the results of an evaluation study of this empathic rational dialog agent.

5.1 Implementation

5.1.1 The module of emotions

In order to create empathic rational dialog agent, a module of emotions[8] has been developed. It corresponds to a plug-in for the JSA (*Jade Semantics Agents*) agents.

[8]The module of emotion is available on the JADE website [22] under a LGPL license.

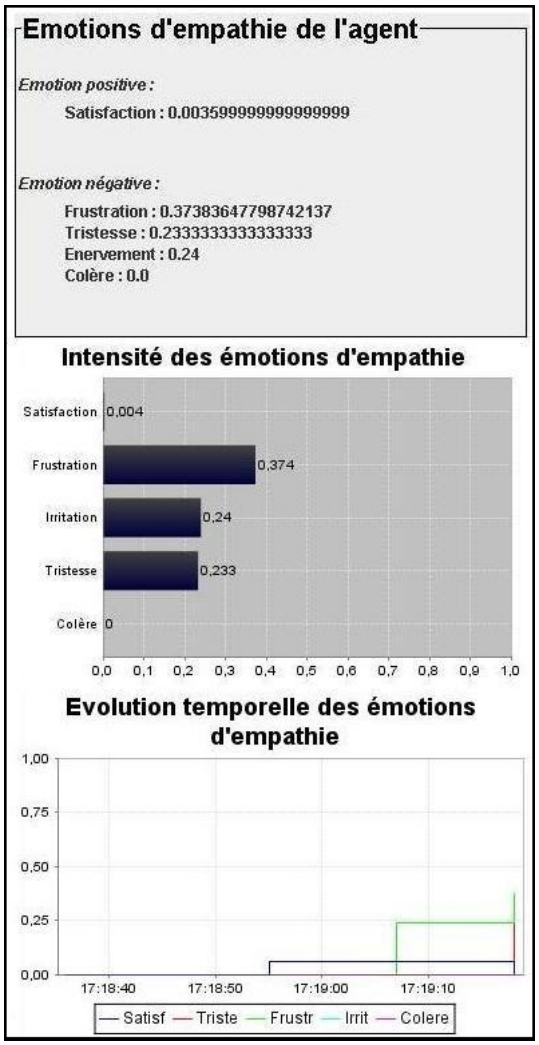

Figure 1: Screenshot of the graphic interface of the module of emotions

These agents are implemented from the JSA framework [28, 27], a plug-in of the JADE (*Java Agent DEvelopment Framework*) platform. The JSA framework[9] enables one to implement BDI-like dialog agent. The module of emotions is composed of a set of java classes to represent emotions, several methods for the emotions elicitation and for computing and updating the emotions intensity, and a graphic interface to visualize the agent's emotions and their intensity dynamic (Figure 1). Based on the speech act theory [4], the JSA agents use a model of communicative acts [46] to infer the user's beliefs and intentions concerning the dialog. For instance, when the user asks an information, the dialog agent deduces that the user has the intention to know the information and has the intention that the agent knows her intention to know the information. The agent infers also that the user believes that her intentions will be achieved following the enunciation of the communicative act. From these user's mental attitudes and given

[9]The JSA framework is open-source [22]

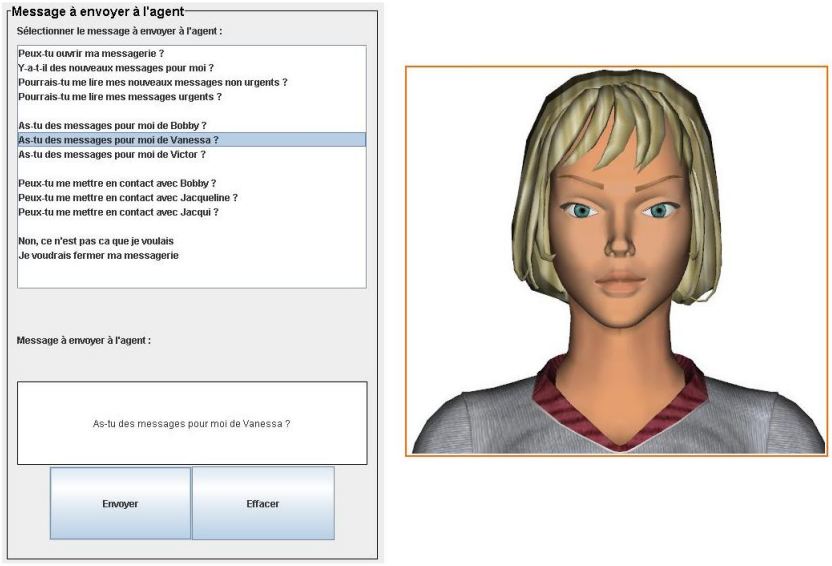

Figure 2: Screenshot of the interface of the ERDAMS

the formalization on empathic emotions elicitation, the agent computes its empathic emotions toward the user. For instance, if the agent believes that one of the user's intentions has just failed, the agent infers that the user potentially feels an emotion of frustration. Then, an empathic emotion of frustration is triggered. If the agent thinks that no other action enables the user to achieve this intention, an empathic emotion of sadness is elicited. The intensity of emotion is computed according to the values of the intensity variables introduced in the previous section. The intensity decreases when no emotion is elicited. The updating of the intensity when an emotion is triggered is inspired with the dynamic model of emotion proposed in [52].

5.1.2 ERDAMS: an Empathic Rational Dialog Agent in a Mail System

From the JSA framework and the module of emotions introduced above, a demonstrator of an empathic rational dialog agent (called *ERDAMS* : An Empathic Rational Dialog Agent in a Mail System) has been implemented. The user interacts with the ERDAMS to obtain information on her mails. She selects predefined sentences on the interface to dialog with the agent (Figure 2).

In order to display the empathic emotions, a 3D talking head developed in *Orange Labs* [8], is used (Figure 2). During the dialog, the module of emotions transfers to the talking head the type and intensity of the empathic emotion to display. The talking head adopts the facial expression corresponding to this emotion (Figure 3)[10]. To give the capability to the ERDAMS to answer the user's requests, different information on the user's messages (type of the message, sender, level of urgency, content, etc) are added to the ERDAMS' database. A module has been developed to translate the user's request from natural language to FIPA-ACL [17], the language used by the JSA agents to reason. Some values have to be fixed by the programmer to enable the ERDAMS

[10]Few research has been done on the expressions of empathy [11]. In our work, we suppose that the facial expression corresponding to an empathic emotion is similar to the one of an emotion of the same type.

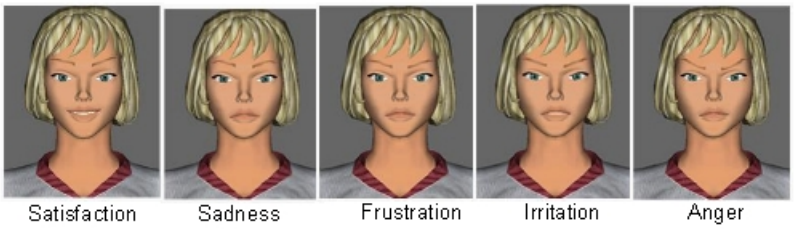

Figure 3: Facial expressions of the agent's empathic emotions

to compute the intensity of emotions: the degree of certainty of the user to achieve her intentions, the importance for the user that her intention is achieved and not failed, the effort maximum that can be done, and the agent's degree of liking and deservingness. These values should be defined depending on the application context, the type of the intentions and on the user's characteristics.

The ERDAMS has been used to evaluate the impact of an empathic virtual agent on the human-machine interaction, and more precisely on the user's perception of the agent. In the next section, we present the experimental protocol and the results of the evaluation.

5.2 Evaluation

Recent research has shown that virtual agents which express empathic emotions toward a user enhance human-machine interaction [7, 25, 37, 39, 42]. However, these few experimentations of emotional agents are mostly in the context of game [7, 25, 37, 42]. As highlighted by Becker *et al.* [6], the agent's expressions of emotions may be harmful to the interaction when they are incongruous to the situation. Today, no research seems to have explored the impact, on the interaction, of an empathic dialog agent used as information system. Therefore, an evaluation of the ERDAMS has to be done to assure the effect of the empathic virtual dialog agent on the interaction.

5.2.1 Method.

In order to evaluate the impact of the empathic virtual dialog agent on the user, three versions of the ERDAMS have been developed:

1. *the non-emotional version* used as a *control condition* in which the virtual agent does not express any emotion;

2. *the empathic version* in which the virtual dialog agent expresses empathic emotions through its facial expressions during the interaction with the user. This version corresponds to the one described in the previous section. The conditions of emotions' expressions are those defined in our model of empathic emotions;

3. *the non-congruent emotional version* in which the virtual dialog agent expresses incongruous emotions to the situations of the interaction through its facial expressions. More precisely, the valence of the emotions expressed are the opposite of those in the empathic version. For instance, if, in a given situation, the virtual agent expresses an emotion of sadness in the empathic version, then, in the non-congruent emotional version, the agent expresses an emotion of satisfaction. In

other words, in this version, the agent expresses a positive (*resp.* negative) emotion when the user potentially feels a negative (*resp.* positive) emotion.

In the three versions, the interface of the system (Figure 2), the verbal behavior of the agent and its facial expressions are the same. Only the conditions of emotions elicitation vary.

Eighteen subjects (nine men and nine women) have participated to the experiment. The average age was 35 (standard deviation=11.86). No participant knew the research subject and the ERDAMS. During the test, each subject has performed three sequences of four or fives requests for each version of the ERDAMS. To achieve a request, the subject asked the virtual agent to execute an action by clicking on the corresponding sentence displayed on the interface (as for instance "Can you read me my new messages?").

In our research, we have focused on the effect of an empathic agent on the user's perception of the agent. The questionnaire to evaluate the user's perception of the virtual agent is composed of 15 affirmations: 11 regarding the virtual agent (as for example "She was pleasant") and 4 concerning more precisely her facial expressions (as for example "I have liked her facial expressions"). Finally, the perception of the following aspects of the virtual agent has been measured: pleasant, irritating, strange, compassionate, expressive, cold, jovial, boring, strict, cheerful, stressful, appreciation of the facial expressions, naturalness of the facial expressions, their perturbing aspect, and their exaggerated aspect[11]. The participants have indicated their agreement or disagreement for each affirmation by checking the box corresponding to their opinion on a Likert scale of 7 points (from 1 *not agree at all* to 7 *fully agree*). At the end of the test, each participant has received a gift token of 15 euros value. The test for each participant has not exceeded 40 minutes.

5.2.2 Results.

The results for each of the 15 quality factors evaluated have been analyzed separately. The distributions of the results are normal. An ANOVA of repeated measurements and a post-hoc test HSD-Tukey have been applied. In the following, these abbreviations are used to describe the different versions: *NE* for *Non-Emotional version* (no emotion displayed), *E* for *Empathic version*, and *NCE* for *Non-Congruent Emotional version*. The results are presented in Tables 1, 2 and 3. The first column indicates the studied quality factors and the first line the versions compared; the elements of the table (*i.e.* the intersection between one line and one column) correspond to the version in which the quality factor of the agent has been the best perceived (n.s. means non significant, *: $p < .05$, **: $p < .01$, ***: $p < .001$). For instance, in Table 1, the notation E** at the intersection of the line *NE-E* et the column *jovial* means that, in the empathic version, the virtual agent has been perceived significantly more jovial (with $p < .01$) than in the non-emotional one.

User's perception of the virtual agent's positive quality factors. The analysis of the results shows an effect of the version on the user's perception of the pleasant ($F(2,34)=20.597$, $p<.001$), compassionate ($F(2,34)=7.44$, $p<.01$), expressive ($F(2,34)=4.6790$, $p<.05$), jovial ($F(2,34)=12.246$, $p<.001$), and cheerful

[11] No definition of these adjectives has been provided to the subjects.

(F(2,34)=7.7887, p<.01) aspect of the virtual agent. When the virtual agent expresses empathic emotions (positive and negative), it is perceived more jovial, expressive and cheerful than when it does not express any emotion. Moreover, when the emotions are displayed in incongruous situations, the virtual agent is perceived less pleasant, compassionate, expressive, jovial and cheerful than when the same emotions are expressed by empathy (Table 1).

	NE-E	NE-NCE	E-NCE
pleasant	n.s.	NE**	E***
jovial	E**	n.s.	E***
expressive	E*	n.s	E***
cheerful	E**	n.s.	E*
compassionate	n.s.	n.s	E**

Table 1: Comparison of the user's perception of the agent's positive quality factors in the different versions

User's perception of the virtual agent's negative quality factors. The results reveal an effect of the version on the user's perception regarding the irritating (F(2,34)=15.409, p<.001), strange (F(2,34)=12.518, p<.001), cold (F(2,34)=5.1405, p<.05), and stressful (F(2,34)=11.679, p<.001) aspect of the virtual agent. The virtual agent is perceived as being more irritating, strange, cold, and stressful when it expresses emotions in incongruous situations than when it displays empathic emotions or no emotion (Table 2).

	NE - E	NE - NCE	E - NCE
irritating	n.s.	NCE***	NCE**
strange	n.s.	NCE***	NCE**
cold	n.s.	NCE*	NCE*
boring	n.s.	n.s.	n.s.
strict	n.s.	n.s.	n.s.
stressful	n.s.	NCE***	NCE*

Table 2: Comparison of the user's perception of the agent's negative quality factors in the different versions

User's perception of the virtual agent's facial expressions. The results of the experiment show a significant effect of the version on the user's appreciation of the agent's facial expressions (F(2,34)=19.324, p<.001), her perception of the naturalness aspect (F(2,34)=11.666, p<.001), the perturbing one (F(2,34)=14.880, p<.001), and the exaggerated aspect (F(2,32)=18.522, p<.001) of the agent's facial expressions. The facial expressions of emotions incongruous to the dialog situations are less appreciated than non emotional one or those expressed by empathy. The same facial expressions of emotion are perceived more naturalness and less perturbing and exaggerated when they are displayed in empathic situations that in incongruous ones (Table 3).

	NE - E	NE - NCE	E - NCE
appreciation	n.s.	NE***	E**
naturalness	n.s.	NE***	E**
perturbing	n.s.	NCE***	NCE**
exaggerated	n.s.	NCE***	NCE**

Table 3: Comparison of the user's perception of the agent's facial expressions in the different ERDAMS versions

5.2.3 Discussion.

Firstly, the evaluation enables us to compare the user's perception of the non emotional virtual agent and the empathic one. The results show that empathic expressions of emotions do not impair the user's perception of the agent. Indeed, it does not appear more irritating, strange, cold, or stressful when it expresses empathic emotions than when it displays no emotion. Moreover, the facial expressions of emotions are not perceived less naturalness, more perturbing or exaggerated than non emotional ones. Some significant differences have been observed. The virtual agent appears more expressive, jovial and cheerful when it expresses both positive and negative empathic emotions than when it displays no emotion.

On the contrary, the results reveal that the emotions expressed in incongruous dialog situations have a negative effect on the user's perception of the agent. Indeed, she perceives the virtual agent less pleasant, more irritating, strange, cold and stressful than when the agent expresses no emotion. The facial expressions of emotions in this case seem more exaggerated and perturbing and less naturalness in comparison with neutral facial expressions.

By comparing the user's perception depending on the agent's conditions of the expression of emotions, it appears that the global perception of the agent depends on the congruency between the dialog situations and the expressions of emotions. Indeed, when the emotional expressions are not congruent with the dialog situations, the agent is perceived more negatively. Moreover, the same facial expressions of emotions are perceived differently depending on the conditions of emotions' expressions. They seem less naturalness, more exaggerated and perturbing when they are not congruent with the dialog situations than when they are displayed by empathy.

To conclude, the results of the evaluation show that a virtual agent which expresses emotions in incongruous situations is perceived more negatively than one that does not express any emotions. Inversely, when the agent expresses emotions in the conditions described in our models of empathic emotions, it is perceived more positively than when it does not express any emotion. The expressions of emotions are, therefore, in this case, *appropriate* to the dialog situations. These results validate the conditions of empathic emotion elicitation defined in our model. They are relevant to determine which empathic emotions the agent should express in which circumstances in order to enhance the user's perception of the virtual agent.

6 Conclusion

References

[1] Integrating models of personality and emotions into lifelike characters. In Ana Paiva, editor, *Affective interactions: Towards a new generation of computer interfaces*, pages 150–165. Springer-Verlag, 2001.

[2] A BDI approach to infer student's emotions. In *the Proceedings of the Ibero-American Conference on Artificial Intelligence (IBERAMIA)*, pages 901–911, Puebla, Mexique, 2004. Springer-Verlag.

[3] Carole Adam, Benoit Gaudou, Andreas Herzig, and Dominique Longin. OCC's emotions: A formalization in a BDI logic. In J. Euzenat and J. Domingue, editors, *Proceedings of the International Conference on Artificial Intelligence: Methodology, Systems, Applications (AIMSA) 2006*, pages 24–32, Varna, Bulgarie, Jun 2006. Springer.

[4] J.L Austin. *How to do things with words*. Oxford University Press, London, 1962.

[5] Jospeh Bates. The role of emotion in believable agents. *Communications of the ACM (CACM)*, 37(7):122–125, 1994.

[6] Christian Becker, Ipke Wachsmuth, Helmut Prendinger, and Mitsuru Ishizuka. Evaluating affective feedback of the 3D agent max in a competitive cards game. In Jianhua Tao, Tieniu Tan, and Rosalind W. Picard, editors, *Proceedings of the International Conference on Affective Computing and Intelligent Interaction (ACII)*. Springer, oct 2005.

[7] Scott Brave, Clifford Nass, and Kevin Hutchinson. Computers that care: Investigating the effects of orientation of emotion exhibited by an embodied computer agent. *International Journal of Human-Computer Studies*, 62:161–178, 2005.

[8] G. Breton, C. Bouville, and D. Pelé. FaceEngine: A 3D Facial Animation Engine for Real Time Applications. In *Web3D Symposium*, Germany, 2000.

[9] V. Carofiglio and F. deRosis. In favour of cognitive models of emotions. In *the Proceedings of the Joint Symposium on Virtual Social Agents, conference on Artificial Intelligence and Simulated Behavior (AISB)*, pages 171–176, Hatfield, UK, 2005.

[10] Cristiano Castelfranchi. Affective appraisal versus cognitive evaluation in social emotions and interactions. In A.M. Paiva, editor, *Affective Interactions: Towards a New Generation of Computer Interfaces*, pages 76–106. Springer-Verlag, 2000.

[11] N. Chovil. Social determinants of facial displays. *Journal of Nonverbal Behavior*, 15:141–154, 1991.

[12] P.R Cohen and H.J Levesque. Intention is choice with commitment. *Artificial Intelligence*, 42(2-3):213–232, 1990.

[13] Fiorella deRosis, Catherine Pelachaud, Isabella Poggi, Valeria Carofiglio, and Berardina De Carolis. From Greta's mind to her face: Modelling the dynamics of affective states in a conversational embodied agent. *International Journal of Human-Computer Studies*, 59(1-2):81–118, 2003.

[14] T DuyBui. *Creating emotions and facial expressions for embodied agents*. PhD thesis, University of Twente, 2004.

[15] Michael G. Dyer. Emotions and their computations: Three computer models. *Cognition and Emotion*, 1(3):323–347, 1987.

[16] C. Elliot. *The Affective Reasoner: A process model of emotions in a multi-agent system*. PhD thesis, Northwestern University, 1992.

[17] FIPA-ACL. http://www.fipa.org/repository/aclspecs.html, 2002.

[18] N. Frijda, A. Ortony, J. Sonnemans, and G.L.Clore. *Emotion: Review of Personality and Social Psychology*, chapter The complexity of intensity: Issues concerning the structure of emotion intensity. Sage, 1992.

[19] J. Gratch. Socially situated planning. In *AAAI Fall Symposium on Socially Intelligent Agents - The Human in the Loop*, North Falmouth, 2000.

[20] Jonathan Gratch and Stacy Marsella. A domain-independent framework for modeling emotion. *Journal of Cognitive Systems Research*, 5(4):269–306, 2004.

[21] Jakob Hakansson. *Exploring the phenomenon of empathy*. PhD thesis, Department of Psychology, Stockholm University, 2003.

[22] JADE. http://jade.tilab.com/, 2001.

[23] W.L. Johnson, J.W. Rickel, and J.C. Lester. Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *International Journal of Artificial Intelligence in Education*, 11:47–78, 2000.

[24] G. Jorland. Odile Jacob, 2004.

[25] Jonathan Klein, Youngme Moon, and Rosalind Picard. This computer responds to user frustration. In *Proceedings of the Conference on Human Factors in Computing Systems*, pages 242–243, Pittsburgh, USA, 1999. ACM Press.

[26] Richard S. Lazarus. *Emotion and adaptation*. Oxford University Press, 1991.

[27] V. Louis and T. Martinez. *Developing Multi-agent Systems with JADE*, chapter JADE Semantics Framework, pages 225–246. John Wiley and Sons Inc, 2007.

[28] Vincent Louis and Thierry Martinez. An operational model for the FIPA-ACL semantics. In M. Pechoucek, C. Republic, D. Steiner, and S. Thompson, editors, *Proceedings of AAMAS (International joint Conference on Autonomous Agents and Multi-Agent Systems) Workshop on Agent Communication*, pages 1–14, Utrecht, The Netherlands, jul 2005. ACM Press.

[29] Scott McQuiggan and James Lester. Learning empathy: A data-driven framework for modeling empathetic companion agents. In P. Stone and G. Weiss, editors, *Proceedings of AAMAS*, Hakodate, Japan, may 2006.

[30] J.J.Ch. Meyer. Reasoning about emotional agents: Research articles. *International Journal of Intelligent Systems*, 21(6):601–619, 2006.

[31] Keith Oatley. *Best Laid Schemes, the psychology of emotions*. Cambridge University Press, 1994.

[32] Magalie Ochs, Catherine Pelachaud, and David Sadek. Emotion elicitation in an empathic virtual dialog agent. In *Proceedings of the Second European Cognitive Science Conference (EuroCogSci)*, 2007.

[33] Becky Lynn Omdahl. *Cognitive Appraisal, Emotion, and Empathy*. Lawrence Erlbaum Associates, 1995.

[34] A Ortony, G.L Clore, and A Collins. *The cognitive structure of emotions*. Cambridge University Press, United Kingdom, 1988.

[35] E. Pacherie. *L'empathie*, chapter L'empathie et ses degrés, pages 149–181. Odile Jacob, 2004.

[36] Ana Paiva, Joao Dias, Daniel Sobral, Sarah Woods, and Lynne Hall. Building empathic lifelike characters: the proximity factor. In A. Paiva, R. Aylett, and S. Marsella, editors, *Proceedings of AAMAS, Workshop on Empathic Agents*, New York, USA, aug 2004.

[37] T. Partala and V. Surakka. The effects of affective interventions in human-computer interaction. *Interacting with computers*, 16:295–309, 2004.

[38] Rosalind Picard. *Affective Computing*. MIT Press, 1997.

[39] R.W. Picard and K.K. Liu. Relative Subjective Count and Assessment of Interruptive Technologies Applied to Mobile Monitoring of Stress. *International Journal of Human-Computer Studies*, 65:396–375, 2007.

[40] Isabella Poggi. Emotions from mind to mind. In A. Paiva, R. Aylett, and S. Marsella, editors, *Proceedings of AAMAS, Workshop on Empathic Agents*, New York, USA, aug 2004.

[41] Helmut Prendinger and Mitsuru Ishizuka. The empathic companion: A character-based interface that addresses users' affective states. *International Journal of Applied Artificial Intelligence*, 19:297–285, 2005.

[42] Helmut Prendinger, Junichiro Mori, and Mitsuru Ishizuka. Using human physiology to evaluate subtle expressivity of a virtual quizmaster in a mathematical game. *International Journal of Human-Computer Studies*, 62:231–245, 2005.

[43] A. S Rao and M.P Georgeff. Modeling rational agents within a BDI-architecture. In *Proceedings of the International Conference on Principles of Knowledge Representation and Reasoning (KR)*, 1991.

[44] Scott Reilly. *Believable Social and Emotional Agents*. PhD thesis, Carnegie Mellon University, 1996.

[45] D. Sadek. A study in logic of intention. In *Proceeding of the 3rd Internatinal Conference on Principles of Knowledge Representation and Reasoning (KR'92)*, 1992.

[46] David Sadek. *Attitudes mentales et interaction rationnelle: vers une théorie formelle de la communication*. PhD thesis, Université Rennes I, 1991.

[47] K. Scherer. Criteria for emotion-antecedent appraisal: A review. In V. Hamilton, G.H. Bower, and N.H. Frijda, editors, *Cognitive perspectives on emotion and motivation*, pages 89–126. Dordrecht, Kluwer, 1988.

[48] K. R. Scherer, T. Wranik, J. Sangsue, V. Tran, and U. Scherer. Emotions in everyday life: Probability of occurence, risk factors, appraisal and reaction patterns. *Social Science Information*, 43(4):499–570, 2004.

[49] Klaus Scherer. Analyzing emotion blends. In A. Fischer, editor, *the Proceedings of the Conference of the International Society for Research on Emotions*, pages 142–148, 1998.

[50] Klaus Scherer. Emotion. In M. Hewstone and W. Stroebe, editors, *Introduction to Social Psychology: A European perspective*, pages 151–191. Oxford Blackwell Publishers, 2000.

[51] J.R. Searle. *Speech Acts*. Cambridge University Press, 1969.

[52] E. Tanguy, J. Bryson, and Willis P. A dynamic emotion representation model within a facial animation system. Technical report, Department of Computer Science; University of Bath; England, 2005.

[53] F. Thomas and O. Johnston. *The Illusion of Life: Disney Animation.* Disney Editions, 1981.

[54] Paolo Turrini, John-Jules Ch. Meyer, and Cristiano Castelfranchi. Rational agents that blush. In *Affective Computing and Intelligent Interaction*, 2007.