

Expressive Multimodal Conversational Acts for SAIBA Agents

Jeremy Riviere¹, Carole Adam¹, Sylvie Pesty¹, Catherine Pelachaud²,
Nadine Guiraud³, Dominique Longin³, and Emiliano Lorini³

¹ Grenoble University - LIG, Grenoble, France

{Jeremy.Riviere,Carole.Adam,Sylvie.Pesty}@imag.fr

² CNRS - Telecom ParisTech, Paris, France

{catherine.pelachaud}@telecom-paristech.fr

³ UPS - CNRS, IRIT, Toulouse, France

{nadine.guiraud,emiliano.lorini,dominique.longin}@irit.fr

Abstract. We discuss here the need to define what we call an agent conversational language, a language for Embodied Conversational Agents (ECA) to have conversations with a human. We propose a set of *Expressive Multimodal Conversation Acts (EMCA)*, which is based on the *Expressive Speech Acts* that we introduced in a previous work, enriched with the multimodal expression of the emotions linked to these acts. We have then implemented these *EMCA* for SAIBA-compliant ECA, and specifically for Greta. We were able to use Greta in experiments aimed at assessing the benefits of our language in terms of perceived sincerity and believability of an ECA using it to interact with a human user.

Keywords: ECA, Emotions, Interaction, Dialogue.

1 Introduction

Nowadays, human-computer interactions and agent-agent interactions are omnipresent, making dialogue a major research problem. Therefore, some recent works in the field of human-computer interaction have turned to dialogue annotation [2,1] for understanding dialogue, recognizing its structure and generating dialogues. In parallel, some other works have turned to natural language communication. With the advance of virtual agents, multimodal aspects have become essential in human-agent dialogue. In particular these agents are now able to express (their) emotions, but mainly in a non-verbal manner. We believe that it is important to closely link verbal and non-verbal aspects in order to improve the conversational capabilities of these agents. This is particularly true for what we call *complex emotions* (involving mental representations as cause, self, action ...) that are (in humans) mainly conveyed through language, and are therefore often neglected in current research. In this paper we thus focus on complex emotions such as guilt or reproach and propose a set of *Expressive Multimodal Conversational Act (EMCA)* that allows an ECA to express them in a multimodal manner: verbal and non-verbal.

We ground on the set of *Expressive Speech Acts* that we proposed in [5] and enrich them with the multimodal expression of their intrinsic emotion, thus allowing the automatic generation of the expression of the emotions linked to the act uttered by the agent. We believe that an agent able to not only verbally express sentence but also multimodally express the underlying emotions will be more sincere and believable to the user. To check our claim, we implemented our *MCL* in the standard SAIBA architecture [12], in particular in the Greta agent [10], and then conducted experiments with this agent.

2 Speech Acts to Express *complex emotions*

Starting from the formalization of counterfactual emotions proposed by [7], in [5], we defined *complex emotions* as based on one hand upon the agent counterfactual reasoning and on the other hand upon reasoning about responsibility, skills and social norms. We formalized eight *complex emotions* (rejoicing, gratitude, regret, disappointment, moral satisfaction, admiration, guilt and reproach) in terms of three types of logical operators representing the agent’s mental state: beliefs ($Bel_i\varphi$), goals ($Goal_i\varphi$) and ideals ($Ideal_i\varphi$); and one operator for the attribution of responsibility ($Resp_i\varphi$). We then used the operator $Exp_{i,j,H}\varphi$, representing what an agent i expresses to an agent j in front of group H , to define eight *Expressive Speech Acts*, each of which expresses one complex emotion. Each emotion is said to be “intrinsically” attached to each act. The figure 1 shows the set of *Expressive Speech Acts* according to the *complex emotions* they are expressing.

Table 1. *Expressive Speech Acts* expressing *complex emotions*

$Exp_{i,j,H}(\dots)$	Expressive Speech Acts
$Goal_i\varphi \wedge Bel_i Resp_i\varphi$	$\stackrel{d\acute{e}f}{=} Exp_{i,j,H}(Rejoicing_i\varphi) \stackrel{d\acute{e}f}{=} \mathbf{Rejoice}_{i,j}\varphi$
$Goal_i\varphi \wedge Bel_i Resp_j\varphi$	$\stackrel{d\acute{e}f}{=} Exp_{i,j,H}(Gratitude_{i,j}\varphi) \stackrel{d\acute{e}f}{=} \mathbf{Thank}_{i,j}\varphi$
$Goal_i\neg\varphi \wedge Bel_i Resp_i\varphi$	$\stackrel{d\acute{e}f}{=} Exp_{i,j,H}(Regret_i\varphi) \stackrel{d\acute{e}f}{=} \mathbf{Regret}_{i,j}\varphi$
$Goal_i\neg\varphi \wedge Bel_i Resp_j\varphi$	$\stackrel{d\acute{e}f}{=} Exp_{i,j,H}(Disappointment_{i,j}\varphi) \stackrel{d\acute{e}f}{=} \mathbf{Complain}_{i,j}\varphi$
$Ideal_i\varphi \wedge Bel_i Resp_i\varphi$	$\stackrel{d\acute{e}f}{=} Exp_{i,j,H}(MoralSatisfaction_i\varphi) \stackrel{d\acute{e}f}{=} \mathbf{IsMorallySatisfied}_{i,j}\varphi$
$Ideal_i\varphi \wedge Bel_i Resp_j\varphi$	$\stackrel{d\acute{e}f}{=} Exp_{i,j,H}(Admiration_{i,j}\varphi) \stackrel{d\acute{e}f}{=} \mathbf{Compliment}_{i,j}\varphi$
$Ideal_i\neg\varphi \wedge Bel_i Resp_i\varphi$	$\stackrel{d\acute{e}f}{=} Exp_{i,j,H}(Guilt_i\varphi) \stackrel{d\acute{e}f}{=} \mathbf{FeelGuilty}_{i,j}\varphi$
$Ideal_i\neg\varphi \wedge Bel_i Resp_j\varphi$	$\stackrel{d\acute{e}f}{=} Exp_{i,j,H}(Reproach_{i,j}\varphi) \stackrel{d\acute{e}f}{=} \mathbf{Reproach}_{i,j}\varphi$

Thus, this *Expressive Language* links sharply *complex emotions* and speech acts, meaning you can’t compliment somebody without expressing the *complex emotion* of admiration. Indeed, according to the speech act theory, an ECA has to express the intrinsic emotion of the expressive acts to “successfully” perform

this act (*conditions of success*). We affirm that the multimodal expression of the *EMCA*'s intrinsic emotions expressed by an ECA increases its perceived sincerity and believability, under the human-ECA dialogue context and the cultural context.

3 Proposition: *EMCA* for SAIBA Agents

Our proposal consists in adding to the *Expressive Speech Acts* we defined the multimodal expression of each act's intrinsic emotion. *Expressive Speech Acts* thus become *Expressive Multimodal Conversational Acts (EMCA)*, containing the emotions intrinsically attached to each act, as well as their multimodal expression. We have implemented these *EMCA* into the SAIBA architecture to allow ECA SAIBA-compliant to express their *complex emotions*. We are only interested here in the expressive category of acts, whose aim is specifically to express emotions.

The SAIBA framework consists of three step process of intention planning, behavior planning and the virtual agent behavior realization. The FML (Function Markup Language [8]) and BML (Behavior Markup Language [12]) languages link the three modules. An xml file (the *lexicon* file) is used by the behavior planning module to convey patterns in multimodal signals before encoding into BML. Such patterns can be emotions, some performatives (like inform, offer or suggest) and other behavior. The *lexicon* file contains the list of these patterns as well as their multimodal expressions, which are mainly defined from psychological theories and video corpus analysis. Thanks to this *lexicon* file the behavior planning thus links the present patterns in the agent's communicative intention and their multimodal signals before encoding into BML and sending to the behavior realization module.

Our work consisted in offering a library of *EMCA* to SAIBA-compliant ECA. These ECA can thus choose the speech act matching with its communicative intention, thanks to the appropriate dialogue processing or reasoning module (which should have a set of rules based on our formalization). We have modified the *lexicon* file, where we have implemented our library of *EMCA*. Every act in the library is associated with its intrinsic emotion (see Section 2): concretely, this emotion is manifested by various multimodal signals that are used in the behavior planning module, and it will thus be expressed by the ECA alongside the conversational act.

For example, let us consider the next turn in a dialogue with the user: "*I am disappointed in your behavior, It's not very nice of you.*". Here, the ECA's intention is to express that the ECA had a certain goal and that the human is responsible for not reaching this goal. In this case, the *EMCA complain*, whose intrinsic emotion is the *complex emotion* of *disappointment*, fits best the ECA's communicative intention. The first module translates this communicative intention in FML-APML¹, and completes it with a BML specification of the agent's utterance, as shown in the example below:

¹ Since FML is still under development, we will work here on the FML-APML specification proposed in [8].

```

<fml-apml>
<bml> <speech id="s1" language="english" voice="openmary" text="I am
disappointed in your behavior, It's not very nice of you." > </bml>
<fml> <performative id="p1" type="complain" > </fml>
</fml-apml>

```

In the FML-APML specification, the *performative* tag represents the *EMCA* that is linked with the utterance. The multimodal expression of the emotion associated with each *EMCA*, which was identified in FML by the *performative* tag, is now detailed in the *lexicon* file. For example, here is the *EMCA complain* in the *EMCA* specification general format:

```

<behaviorset name="complain">
<signal id="1" name="faceexp=disappointment" modality="face"/>

```

Here, the *complex emotion* of disappointment associated to the *EMCA complain* is multimodally expressed by its facial expression (signal 1). The behavior realization Module finally receives all the signals translated in BML, and it adapts them to animation rules and constraints: it checks coherence and absence of conflicts, before animating the verbal and non-verbal behavior of the ECA. So, the ECA has to express the emotion of disappointment to “successfully” complain (*i.e.* to seem sincere and believable to the user in the dialogue and cultural context).

Unlike current works [6,3], we established a link between acts and emotions from both the speech act theory [11] and our *complex emotion* definition. This approach enable a consistency of these acts and should improve the virtual agent’s sincerity and credibility. To confirm this hypothesis, we assessed the Greta agent’s sincerity and credibility when using our *EMCA* (see next section).

4 EMCA Evaluation: Application to Greta

The hypothesis that we want to test is the following: *adding non-verbal signals of the expression of the complex emotions coherent with the verbal signals expressed by an ECA increases its perceived sincerity and believability*. This hypothesis is to be considered through the human-ECA dialogue context, and the cultural context. We ran a first experiment with a subset of our *EMCA* and Greta, a SAIBA-compliant ECA ([10]). We have implemented our library of *EMCA* in the lexicon file of Greta (as described in 3), thus endowing this character with the ability to use our *EMCA*.

4.1 Test Scenario and Protocol

The protocol that we set up allowed us to test our hypothesis for two *EMCA* that express *complex emotions*: *to rejoice*, that expresses *rejoicing*, and *to apologise*, that expresses *regret*². In the setting of a dialogue between a user and a ECA (acted out by Greta) whose aim is to recommend movies, these two *EMCA*

² To apologise is to regret something which is against the human’s goal.

are expressed by the ECA in three conditions: a congruent condition, a neutral condition and a non-congruent condition. In the congruent condition, the ECA expresses the *EMCA* as it is defined in our language, *i.e.* with the intrinsic emotion matching the corresponding act (the regret emotion with an Apologise utterance, and the rejoicing emotion with a Rejoice utterance). In the neutral condition, the ECA only expresses the utterance without expressing its intrinsic emotion of the act. Finally in the non-congruent condition, the ECA expresses the utterance with the emotion that is the opposite (see table 1) of the intrinsic emotion of the act (an Apologise utterance is expressed with the Rejoicing emotion, while a Rejoice utterance is expressed with a Regret emotion).

In the first stage of the protocol, videos of two scenarios are proposed, each containing the expression by the ECA of one of the two tested *EMCA*. For each scenario, three videos are submitted to the subject (one per condition, *i.e.* altogether six videos) to be evaluated. Both scenarios imply a dialogue with Greta after the user supposedly comes back from the cinema where they saw a movie that had previously been recommended by Greta. In the first scenario, they loved the movie and has just thanked Greta for her good advice. The user is then asked to watch the three videos (in a predefined order that differs for each user), and to evaluate how Greta rejoices. Similarly in the second scenario, the user hated the movie and has just reproached Greta for her bad advice (*cf.* figure 1). The user is then asked to watch the three videos of the *EMCA to apologize* (also in a predefined order) and to evaluate how Greta apologises.

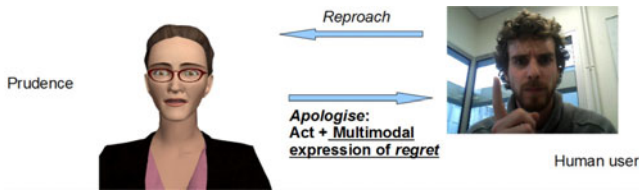


Fig. 1. Apologise *EMCA* in the congruent condition: Greta apologises in reply to user’s reproach

The videos are assessed on two criteria: sincerity (“Does Greta seem to express what she thinks?”) and believability (“Does the scene played by Greta look plausible?”) of the ECA. These criteria can each take four possible values on the same qualitative scale (Not at all - Rather not - Rather - Totally).

The second part of the protocol consists in a questionnaire whose aim is to collect the user’s impressions, mainly regarding the impact of the agent, and their subjective feelings. During this part, users can express themselves with their own words.

4.2 Results

23 users selected within students in science between 18 and 26 participated in the first experiment. To perform the statistical tests, we associated a score with

each qualitative value of the sincerity and believability criteria: 1 for “Not at all”, 2 for “Rather not”, 3 for “Rather”, and 4 for “Totally”. Since we evaluate 2 characteristics of 2 *EMCA*, we have 4 *dependent variables*: Sinc_Apologize, Believ_Apologize, Sinc_Rejoice, Believ_Rejoice. The only *independent variable* is the condition (congruent, neutral, non-congruent).

The first step was to check that the *dependent variables* follow a normal distribution. Secondly, the analysis of the results told us that the *EMCA* expressed in the congruent condition (C) have the best scores in both sincerity and believability. When Greta apologises under this condition, 65% of participants found Greta “rather” sincere (39%) or “totally” sincere (26%) and 70% of participants found Greta “rather” believable (52%) or “totally” believable (18%). Similarly when Greta rejoices under this condition, 74% of participants found Greta “rather” sincere (52%) or “totally” sincere (22%) and 78% of participants found Greta “rather” believable (52%) or “totally” believable (26%). This tends to confirm our hypothesis; we can thus assume that expressing the coherent emotion (C) with the act makes the ECA more sincere and believable to the user than expressing nothing (N) and expressing the opposite emotion (NC).

We then ran a one-way Analysis of Variance (ANOVA) test on each of our *dependent variables* with respect to the *independent variable*. The hypothesis H_0 we checked through this ANOVA test is the following: “There is no difference between the sincerity means (resp. the believability means) under the condition of congruence for the *EMCA Apologize* and *Rejoice*”. H_1 thus says that “there is a difference between the sincerity means (resp. the believability means) under the condition of congruence for the *EMCA Apologize* and *Rejoice*”. This ANOVA test allowed us to reject H_0 significantly in both sincerity and believability: for example, it showed a significant effect of the condition of congruence on Greta’s sincerity ($F(2.66)=22.80$ $p<0.05$) for the *EMCA Apologize*. A Tukey range test showed us that there is a significant difference between the sincerity (resp. believability) mean in the congruent condition (C) and the sincerity (resp. believability) means in the other conditions for both *EMCA Apologize* and *Rejoice*, while there is no significant difference between the non-congruent condition (NC) and the neutral condition (N) (see for ex. table 2).

Table 2. Tukey range test of the Sinc_Apologize variable: the difference between the congruent condition (C) and the neutral(N) and the non-congruent (NC) conditions is significant

Groups	Difference	Statistic	Probability
C - NC	1.478	q = 9.290	0.0000
C - N	1.478	q = 9.290	0.0000
NC - N	-0.435	q = 2.732	0.1376

Finally, when asked “What do you think about Greta?” during the qualitative interview, a lot of users assessed the Greta’s personality. Despite our efforts to avoid this evaluation, Greta’s personality Prudence [9] has been described as

“austere” and “severe”. We do not know how it has influenced the results, but it appears that several users has mentioned the trust notion; Greta’s personality and aspect seems to be as important as its expression of coherent emotion to the user, as shown in [4].

5 Conclusions and Perspectives

In this paper we have presented a library of *Expressive Multimodal Conversational Acts*, that make a link between the expression of emotions through language and through other modalities. We formally defined and implemented this set of eight *Expressive MCA* in the SAIBA architecture. Two of them were evaluated through an implementation in Greta, during a first experiment that showed that they contribute to a better perceived sincerity and believability of ECA. A second phase of evaluation will concern all our library in longer dialogue scenarios involving several turns of speech of an ECA and a human user. In order for this library to be useful to other SAIBA-compliant ECA, we intend to complete it with *MCA* from the other categories: assertive (inform, tell ...), directive (ask, offer ...), commissive (promise, assure ...).

Acknowledgments. This work is supported by the French National Research Agency (ANR), project CECIL (www.irit.fr/CECIL), grant number ANR-08-CORD-005. We thank the statistician engineer and psychologist Nadine Mandran of the LIG Marvellig unit which is specialized in experimentation protocol.

References

1. Bunt, H.: The dit++ taxonomy for functional dialogue markup. In: Proc. of AMAAS 2009 Workshop "Towards a Standard Markup Language for Embodied Dialogue Acts" (2009)
2. Core, M.G., Allen, J.F.: Coding dialogs with the DAMSL annotation scheme. In: Proc. of the Working Notes of the AAAI Fall Symposium on Communicative Action in Humans and Machines, Cambridge, MA (1997)
3. Gebhard, P., Kipp, M., Klesen, M., Rist, T.: Adding the emotional dimension to scripting character dialogues. In: Rist, T., Aylett, R.S., Ballin, D., Rickel, J. (eds.) IVA 2003. LNCS (LNAI), vol. 2792, pp. 48–56. Springer, Heidelberg (2003)
4. Gong, L.: How social is social responses to computers? the function of the degree of anthropomorphism in computer representations. *Comput. Hum. Behav.* 24, 1494–1509 (2008)
5. Guiraud, N., Longin, D., Lorini, E., Pesty, S., Riviere, J.: The face of emotions: a logical formalization of expressive speech acts. In: Proc. of AAMAS 2011 (2011)
6. Lee, J., Marsella, S.: Nonverbal behavior generator for embodied conversational agents. In: Gratch, J., Young, M., Aylett, R.S., Ballin, D., Olivier, P. (eds.) IVA 2006. LNCS (LNAI), vol. 4133, pp. 243–255. Springer, Heidelberg (2006)
7. Lorini, E., Schwarzentruher, F.: A logic for reasoning about counterfactual emotions. *Artificial Intelligence* 175(3-4), 814–847 (2011)

8. Mancini, M., Pelachaud, C.: The fml-apml language. In: Proc. of the Workshop on FML at AAMAS 2008 (2008)
9. McRorie, M., Sneddon, I., de Sevin, E., Bevacqua, E., Pelachaud, C.: A model of personality and emotional traits. In: Ruttkay, Z., Kipp, M., Nijholt, A., Vilhjálmsson, H.H. (eds.) IVA 2009. LNCS, vol. 5773, pp. 27–33. Springer, Heidelberg (2009)
10. Poggi, I., Pelachaud, C., de Rosis, F., Carofiglio, V., Carolis, B.D.: Greta: A believable embodied conversational agent. In: Multimodal Communication in Virtual Environments, pp. 27–45 (2005)
11. Searle, J.R.: Speech acts: an essay in the philosophy of language. Cambridge University Press, London (1969)
12. Vilhjálmsson, H.H., Cantelmo, N., Cassell, J., E. Chafai, N., Kipp, M., Kopp, S., Mancini, M., Marsella, S.C., Marshall, A.N., Pelachaud, C., Ruttkay, Z., Thórisson, K.R., van Welbergen, H., van der Werf, R.J.: The behavior markup language: Recent developments and challenges. In: Pelachaud, C., Martin, J.-C., André, E., Chollet, G., Karpouzis, K., Pelé, D. (eds.) IVA 2007. LNCS (LNAI), vol. 4722, pp. 99–111. Springer, Heidelberg (2007)