Embodied Conversational Agents and Influences

Vincent Maya and **Myriam Lamolle** and **Catherine Pelachaud**[1]

Abstract. In view of creating Embodied Conversational Agent (ECA) able to display individualized behaviors, we propose a taxonomy and a computational model of the influences, factors such as context environment, personality and culture may induce. Influences act not only on the type of the signals an agent conveys but also on the expressivity of the signals. Thus, to individualize ECAs, we consider not only the influences acting on the agent but also the notion of expressivity.

1 Introduction

We aim at creating **E**mbodied **C**onversational **A**gents (ECAs) able to display verbal and nonverbal behaviors. In particular, we wish to individualized ECAs; that is ECAs whose behaviors should be in harmony with factors such as their culture, their personality, their emotion. Their behaviors should also differ depending on whom they are talking to and their relation with their interlocutors, where the conversation takes place, etc. All these factors *influence* how ECAs express their goals and believes they aim at communicating. These behaviors vary at the level of their type of signals (a given gesture or a head movement) and of their expressivity (ample arm gesture or small head nod).

Our way of dealing with people, facts and events, of talking about them, of having feelings toward them are culturally dependent [1, 4]. Our personality affects also our decisions and our emotive reactions. But modelling the impact these complex factors have over the way an agent behave and communicate is extremely complex. In this paper, we aim neither at modelling what culture or personality mean, nor at simulating expressive animations. We limit our scope at representing influences that would modify the set of signals and the quality motions a particular agent will display to communicate a given meaning within a specific context. Therefore, we do not consider these factors according to the actions that they may have over the selection of goals, beliefs and/or emotions but according to their effects on the expressivity.

In order to animate **E**mbodied **C**onversational **A**gents, mark-up languages that associate behaviors to text components using tags have been widely used [3, 7, 5, 6]. At the level of the animation engine, the meaning of these tags has to be expressed in term of facial and body parameters.

To individualize the ECA, we create tools that allow the system to choose the signals expressing the communication acts according to agent's communicative characteristics, to the agent's behaviors characteristics, and to the influences that act on her at a given instant. We introduce at the various computational levels the notion of *expressivity*, a value that indicates the *strength* of the communication acts or of the signals (or the variation of this *strength*).

[1] LINC University of Paris VIII, France

2 Expressivity

We call expressivity the value that allows the system to relate strength to the communication act. The *communicative expressivity* corresponds to the degree of expressivity attached to a given meaning for a default agent. The *agent's expressivity* is related to the qualitative manner the agent uses to express herself. The *behavioral expressivity* represents the way that the considered agent expresses the tag meaning, taking into account her characteristics and the contextual factors that may modify her expressivity. The *signals expressivity* allows to choose the appropriate signal or to combine several ones, the system has to know the expressivity related to each signal. For example, it has to know that slight smile is less expressive than a large smile.

2.1 Expressivity and Influences

Influences may act on the selection of a non-verbal behavior to convey a meaning (i.e. on the choice of the signals), on the *expressivity* of this behavior (e.g. on their intensity level), in order to qualify it or to accentuate it, and on the communication strategies.

We differentiate several types of influences. The first type contains the *intrinsic* influences, i.e. the set of the conscious and unconscious habits of the agent that are reflected in the content of her discourse and that define her attitude and her behavior when she talks. These habits derive, among others, from personality, age, sex, nationality, culture, education and experiences.

We oppose to the intrinsic influences the *contextual* influences, that may increase or decrease the effects of the intrinsic factors, and may even cancel them. These influences group the external factors and the mental factors.The external factors refer to the environment setting such as the light conditions, the sound intensity, the spatial layout or the function of the conversation site. The mental and emotional factors represent the influences of the agent's mental state and of the social conventions respected by the agent. For example, a person does not talk and does not behave in the same way whether she is angry or not. Her relationships with her interlocutor modulate also her behavior: she does not behave in the same way with a friend, an unknown person, an employee, a child or a doctor [2].

In order to take into account this influences, from the *communicative expressivity*, and from the *agent's expressivity*, the system computes the *behavioral expressivity*. This value intervenes for the signals selection and modifies the quantity of movements related to these signals, their amplitude, their duration, their dynamism and/or their repetitiveness.

2.2 Signal expressivity and expression libraries

To be able to instantiate the *behavioral expressivity* into a set of expressive signals, the animation engine has to compute the appropriate

signal expressivity. The *expressions libraries* contain the expressivity associated to signals. We use fuzzy set in order to define domains where the signals are appropriately usable. The system may distort the signals (i.e. change their intensity values) in order to modulate their expressivity.

Behavior expressivity may be expressed not only through the signals, and their expressivity, but also by quantity of signals dispatched over modalities. We differentiate thus the *modal signal expressivity*, which concerns the qualitative parameters that determine the choice between several signals of a same modality with a same meaning, and the *inter-modal signal expressivity*, which is modelled by defining the functions that relates the behaviors across the modality, such as redundancy (i.e. expression of the same meaning with several signals of different modalities), complementarity (e.g. saying "*he goes to the stadium*", and complementing it with an iconic gesture that means "*he drives to the stadium*"), substitution (e.g. straight index over the mouth to mean silence), and masking (e.g. masking sadness by a smile).

3 Agent's definition

In the input text in which the tags describe the communicative acts, these tags are defined for the *default agent*. We associate to each agent a behavioral profile, which specifies, on the one hand, the agent's expressivity, i.e. the intrinsic factors, and, on the other hand, the effects of the contextual factors.

To represent these intrinsic factors the tag <**AgentDefinition**> contains the element <**intrinsicFactors**> which associates a numeric value to the attributes *face*, *posture*, *gaze* and *gesture*. These values lessen or accentuate the expressivity of the tag meaning for the related modality.This intrinsic profile, given as input, is constant during a dialog session, although it may be surcharged within the tag $< agent >$.

The intrinsic behavioral profile indicates the effects of the *agent's expressivity* for each modality. In order to choose the modality (or modalities) that the agent uses for a given tag, we represent a hierarchy over the modalities for the agents. We associate to each modality (*face*, *gaze*, *gesture* and *posture*) a numeric value that represents their preferential level in this hierarchy. In case several modalities have the same hierarchical level the system considers the expressivity of all the signals of the concerned modalities to choose a signal at this level. This hierarchy is also used for the inter-modal signal expressivity, and, in particular, to express the redundancy.

4 Selection process

In order to indicate a specific contextual behavioral profile that models the effects of contextual factors, we introduce the coefficient *contextCoeff*. This coefficient aims only to lessen or to accentuate the *behavioral expressivity* expressed by these tags according to the influences. This coefficient is a very simplified modelization of the effect of contextual factors as they may not only act on the intensity of expressivity but also on masking an expression by another one (in some situation anger may not be shown and a polite smile may have to be displayed).

For each tag, the system has to decide the modality (*face*, *gesture*, *gaze* or *posture* one) to express the given meaning. For most cases, the decision is based on the modality hierarchy: among the modalities that have at least one expression which allows the system to represent the meaning, it choose the one with the highest priority and that is not used yet, in order to prevent conflicts. Some contextual factors may however modify this hierarchy. For example, for an agent that expresses her communication acts mainly by facial expression, the anger or the nervousness may incite her to use gestures more intensively.

As seen in the previous section, the system obtains the name of the selected signal from the expression library. Now, it has to compute the distortion to apply to this signal. This distortion allows us to obtain a widest range of expressions and to modulate the expressivity. Signals, even if they belong to a same modality, may vary in their ways to express expressivity variation. We consider several types of distortion: temporal, spatial or repetition. For facial expression, variation of expressivity can be expressed through variation of muscular contractions as well as variation of its temporal course; while when talking about gaze, expressivity variations may be related to factors such as length of mutual gaze or length of looking at the conversation partners; while when talking about gesture, it may be related to parameters such as the strength of a movement, its tempo, its dynamism or its spatial amplitude. Variation of expressivity may also be expressed by the rapid repetition of the same gesture (rapid head nods, fast beat gestures).

5 Conclusion

In this paper a taxonomy of the influences that represent factors such as the culture, the personality, the context has been described. We have also presented our computational model of expressivity. Expressivity may arise at several levels of the agent's specification. Behaviors are defined by the signals that composed them and by their expressivity. We believe this model is a first step toward the creation of individual agent, that is agent that exhibits personal behaviors.

REFERENCES

[1] R Brislin, *Understanding culture's influence on behavior*, Hartcout Brace College Publishers, New-York, 1993.

[2] B. DeCarolis, V. Carofiglio, M. Bilvi, and C. Pelachaud, 'APML, a markup language for believable behavior generation', in *Embodied conversational agents - let's specify and evaluate them!, Proceedings of the AAMAS'02 workshop*, Bologna, Italy, (July 2002).

[3] B. DeCarolis, C. Pelachaud, I. Poggi, and M. Steedman, 'APML, a markup language for believable behavior generation', in *Life-like Characters. Tools, Affective Functions and Applications*, eds., H. Prendinger and M. Ishizuka, 65–85, Springer, (2004).

[4] F. deRosis, C. Pelachaud, I. Poggi, V. Carofiglio, and B. 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. Special Issue on Applications of Affective Computing in HCI*, (in press).

[5] P. Piwek, B. Krenn, M. Schröder, M. Grice, S. Baumann, and H. Pirker, 'RRL: a rich representation language for the description of agents behaviour in NECA', in *Embodied conversational agents - let's specify and evaluate them!, Proceedings of the AAMAS'02 workshop*, Bologna, Italy, (July 2002).

[6] H. Prendinger, S. Descamps, and M. Ishizuka, 'Scripting affective communication with life-like characters in web-based interaction systems', *Applied Artificial Intelligence*, **16**(7-8), 519–553, (2002).

[7] VHML. Virtual human markup language. http://www.vhml.org.