From Mediated Human-Human Interactions to Human-Virtual Character Interactions: Selection and Categorization of Expressive Gestures

Aurélie Zara[1,2], Valérie Maffiolo[1], Jean Michel Adam[2], Michel Dubois[3]

[1] France Telecom Orange Labs, 2 av. P. Marzin, 22300 Lannion, France
[2] Laboratoire d'Informatique de Grenoble, 385 rue de la Bibliothèque BP53, 38041 Grenoble cedex 9, France
[3] Laboratoire des Sciences de l'Education, Université de Grenoble 2, BP 47 - 38040 GRENOBLE Cedex 9 – France

{aurelie.zara, valerie.maffiolo}@orange-ftgroup.com ;
Jean-Michel.adam@imag.fr;Michel.Dubois@upmf-grenoble.fr

Keywords: expressive gestures, mediatised interaction, game, emotion.

Introduction. In order to design affective interactive systems, experimental grounding is required for studying multimodal expressions of emotion during interaction. Several studies have shown that expressions of anthropomorphic characteristics, such as cognitive states and processes (e.g. doubt, interest or thinking), are also necessary to enhance the believability of a virtual character [1][2]. Nowadays, a minority of these agents is able to express these characteristics and only through few modalities (e.g. gaze and facial expressions) [2].

Our aim is to select and categorize human gestures and facial displays, expressing emotions, cognitive states and processes, and several communicative acts (e.g. a nod expressing "yes"). The selected behaviours will be replayed by a virtual character to elicit fun and commitment of users implied in an interaction with the character. In this framework, we chose a corpus-based approach of human-human interactions in a game context.

Procedure. We defined a protocol called EmoTABOO2, which is an adaptation of the game Taboo. One of the two players had to guess a word described by the other player using his own speech and gestures, without uttering five forbidden words. Different strategies were used to elicit various emotional and expressive behaviours. A previous face to face version of this protocol is described in [3].

To be as closed as possible to a human-virtual character interaction, the recorded human-human interactions were mediated via webcams and microphones. One of the two players could see the upper body of his/her partner on a small screen. The latter was equipped with coloured markers placed on arm articulations, in order to generate the animation of the virtual character's gestures via a motion capture technique.

We selected two subjects for their extraversion and expressiveness, as "model" for the virtual character. Both of them had to interact with the same 4 interlocutors. The 8 pairs of players were recorded during one hour each, from three views: the face of the two players and the upper body of the model. A stereoscopic camera was also used in

order to retrieve the 3D position of model's markers. The videos were segmented into sequences corresponding to the guess of one word (about 83 seconds per word) and a selection was done to exclude sequences that contain only very few gestures. We finally obtained 4 hours of sequences to analyse (2 hours for each model).

Figure 1 Illustration of the collected corpus (model's upper body view (left), and its interlocutor's face view (right))

Corpus analysis. The analysis of the corpus consists in several steps:
− The segmentation of the upper body view sequences into gesture phrases or into a facial expression or head movement with no co-occurring gesture. Adaptators and hold phases occurring at the end of the gesture phrase are not taken into account.
− The evaluation by naïve judges of the expressiveness of the segmented multimodal behaviours, to exclude those with low expressiveness.
− The evaluation of the semantic dependency between co-occurring speech and multimodal behaviours.
− The categorization of the most expressive model subject's behaviours in terms of emotions, cognitive states and processes, and communicative functions [3] [4].
− The inference of the interlocutor's emotions and mental states replying to the behaviours of the model subject, and the evaluation of what these model subject's behaviours would induce on the judge if he or she was in the position of the interlocutor.
− The definition of a set of rules to select behaviours supposed to elicit fun and commitment of users that will interact with the character.
− The annotation of the selected behaviours with labels describing facial displays [4], and gestures [5].

Future directions. The selected multimodal behaviours will be generated for the virtual character with existing methods and software, and their impact on the human-Virtual character interaction will be evaluated using a Wizard of Oz protocol.

1.   Thomas, F. and Johnson, O.: Disney Animation.The illusion of life. Abbeville Press. 1981.
2.   Poggi I., Pelachaud C., de Rosis F., Carofiglio V. and De Carolis B.: GRETA. A Believable Embodied Conversational Agent. In O. Stock and M. Zancarano, eds., Multimodal Intelligent Information Presentation, Kluwer, 2005.
3.   Zara A., Maffiolo V., Martin J.C., Devillers L.: Collection and Annotation of a Corpus of Human-Human Multimodal Interactions: Emotion and Others Anthropomorphic Characteristics. In ACII, pages 464–475. 2007.
4.   J. Allwood, L., K. Cerrato, C. Jokinen , Navarretta, and P. Paggio, "The MUMIN annotation scheme for feedback, turn management and sequencing," presented at Second Nordic Conference on Multimodal Communication, Göteborg, Sweden, 2005.
5.   Michael Kipp, "Gesture Generation by Imitation - From Human Behavior to Computer Character Animation", Boca Raton, Florida: Dissertation.com, December 2004.