

Affective Embodied Conversational Agents for Face to Face Communication

Eva Cerezo, Sandra Baldassarri & Isabelle Hupont

Advanced Computer Graphics Group (GIGA), Computer Science Department,
Engineering Research Institute of Aragon (I3A), University of Zaragoza, Spain

ecerezo@unizar.es

KEYWORDS

affective interfaces, multimodal interaction, virtual agents, expression recognition, natural interaction

CONTENTS

Most research on social interfaces is related nowadays to the design of embodied conversational agents (ECAs). ECAs are agents that are visible in the interface sometimes just as an animated talking face, may be displaying facial expressions and, when using speech synthesis, with lip synchronization, and sometimes they have sophisticated 3D graphical representation, with complex body movements.

In order to obtain a more natural and trustworthy interaction, these virtual agents must be capable of responding appropriately to the users with affective feedback. Therefore, within the ECAs, our research is centered in interactive virtual agents that support multimodal and emotional interaction. The aim is so to establish more effective communication with the user as to broaden the number of potential users by making interaction with disabled users (for example hearing-impaired or paraplegics) and people of different ages and with different levels of education (people with or without a knowledge of computers) possible.

With all this in mind we have developed Maxine, a powerful engine to manage real-time interaction with virtual characters. The aim has been to make it easier to developers the inclusion of these agents in their applications. As pointed out, the consideration of emotional aspects has been a key factor in the development of our system. Special emphasis have been done in capturing the user's emotion through images and in synthesizing the emotion of the virtual agent through its facial expressions and the modulation of the voice.

The chapter will focus on two main aspects: the capture of the user emotional state from web cam images and the development of a dialog system in natural language (Spanish) taking also emotional aspects into account.

The facial emotional classification implemented (see Fig.1) is an effective method based on Ekman's emotional classification. From a set of distances and angles extracted from the user images and from a set of thresholds defined from the analysis of a sufficiently broad image database, the classification results are acceptable, and recent developments has enabled us to improve success rates. The utility of this kind of information is clear: the general vision in that is a user's emotion could be recognized by a computer, human computer-interaction would become more natural, enjoyable and productive. The computer could offer help and assistance to a confused user or try to cheer up a frustrated user and, hence, react in ways more appropriated than simply ignoring the user affective states, as is the case in most current interfaces.

Regarding the dialog's engine (see Fig.2), it is basically in charge of generating the answers to the user's questions and it is based on the recognition of patterns, to which fixed answers are associated (static knowledge). These answers, however, vary depending on the virtual character's emotional state, or may undergo random variations so that the user does not get the impression of repetition if the conversation goes on for a long time (dynamic knowledge). In the conversation emotions are taken into account at two levels:

1. The answer depends on the avatar's emotional state.

2. Besides, the emotional state of the virtual character may change during a conversation, depending on how the conversation develops. That is, if the conversation is on a topic that pleases the character, it gets happy; if it is given information it was not aware of, it is surprised; it is insulted, it gets angry; if it is threatened, it gets frightened, etc.

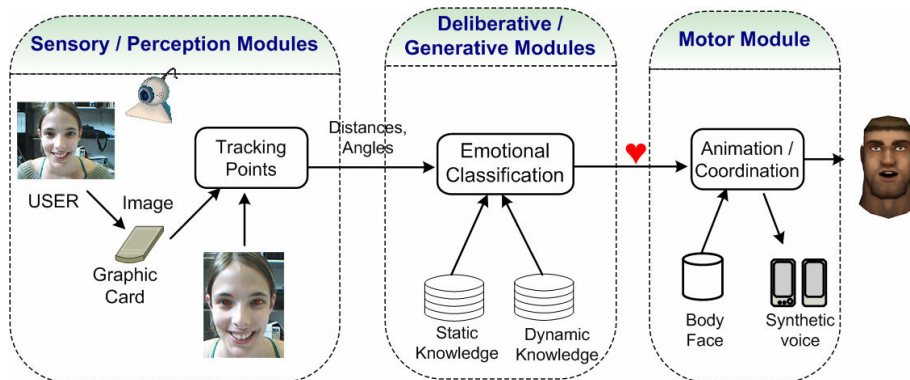


Figure 1 - Stages of the user-avatar image interaction process

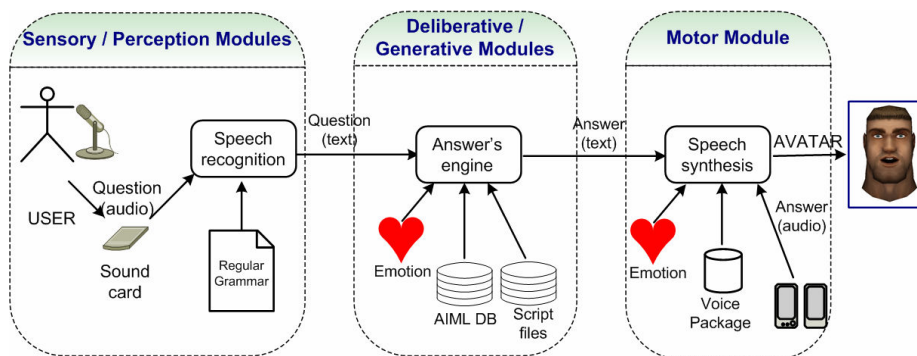


Figure 2 - Stages of the user-avatar voice communication process

Special attention has also been paid in adding an emotional component to the synthesized voice in order to reduce its artificial nature. Voice emotions also follow Ekman's ones and are modeled by means of modifying tone, frequency scale, volume and speed. To represent each emotion, fixed values are assigned to the parameters that enable the relevant emotion to be evoked. These emotional parameters were initially configured considering several studies and but the final values for each emotion have been assigned through voice assessment by users.

Maxine virtual agents have been already used as virtual presenters, domotic helpers and pedagogical agents. In this last application students have been asked to evaluate our embodied agents through questionnaires; the aim of this evaluation has been to assess the effectiveness of the "information provision" aspect of the message and to measure the perceptions about the agents. Evaluation results will also be commented.