# Towards Multi-modal Emotion Display in Embodied Agents

Anton Nijholt
Parlevink Research Group
Department of Computer Science
University of Twente,
Enschede, The Netherlands

*Abstract*

In this paper we explain why, given our background in research and development on interacting agents in virtual environments, we are joining emotion research. For that reason we first survey the research that is going on in our main virtual environment, the so-called Virtual Theatre. This theatre is a rather realistic virtual version of an existing theatre in our hometown. The theatre has been built using VRML (Virtual Reality Modelling Language) and in it we can find the usual locations: entrance hall, information desk, coffee stands, performance halls, stairs, lounges, stage, etc. This virtual world is accessible through Worldwide Web. Currently we are converting the VRML implementation to a Java 3D implementation. We are using this environment as a laboratory to develop and implement ideas about human-agent and agent-agent interaction in such visualized environments. The environment is meant to develop into a virtual interest community where people can represent themselves and can explore and interact, not only with each other, but also with community agents that have task and domain knowledge. Our aim is to build an inhabited world where is it difficult, if not impossible to distinguish agents that are somehow (semi-) controlled by humans and agents that act autonomously (in interaction with the environment and its other inhabitants). Obviously, research results can be used as well in a virtual environment where there is one-to-one communication between an (embodied) interface agent and a user. Several agents have already been introduced in this environment. A Java-based agent framework has been introduced to provide the protocol for communication between agents, also allowing the introduction of other agents. In fact, we can have a multitude of useful agents, where some just trigger an animation, others can walk around and others have built-in intelligence that allows them to execute certain actions based on interactions with visitors.

As mentioned, in our virtual environments we are introducing embodied agents that are (semi-) autonomous or represent humans, and that are able to interact with each other. We also may have the situation that a human user or visitor is not represented in a virtual environment but has direct verbal and non-verbal interaction with one or more embodied agents that are there to assist him or her. The embodied agents may find it necessary for their tasks to perceive emotions, to express emotions and to reason about and use emotions. Believability is an issue that has come from the film industry and that has drawn attention from researchers working on embodied agents. An agent is called believable, if some version of a personality shows in its behavior, especially in the interaction with a human. Main requirements for believability are: personality, emotion, self motivation, social relationships and consistency of expression. In addition to speech and language interaction, embodiment makes it possible to show facial expressions, body language, lip movements and gestures that support interaction. Embodiment allows - more than 'just' language - the expression of emotional behavior.

In our environments, we do not beforehand exclude any human activity or task from the activities of our embodied agents in our (future) environments. An agent may solve a problem individually, it may negotiate with others to solve a problem, it may be involved in a creativity demanding task, it may feel and express sympathy for others, feel friendship, or 'simply' fall in love with a visitor or an other agent. Clearly, the latter 'tasks' ask for some modelling of emotions, but that is also the case for the former. For example, well known are the observations presented in *Descarte's Error* by Antonio Damasio of human patients that suffer from frontal-lobe disorders. They seem to have normal intelligence, but when to have to make decisions, rather than considering some choices and associating maybe subjective feelings to alternatives, they seem to consider increasing amounts of choices, not being able to come to a conclusion or they are repeating wrong decisions without learning from it. Damasio's observations have added to the opinion that it can be useful to have a computer being able

to assign positive or negative emotional feelings to certain choices during certain problem-solving tasks. In a goal-oriented agent architecture this requires the integration of an emotion model in, e.g., a BDI (Believes, Desires, Intentions) agent.

If we return to the more obvious physical aspects of emotion that can be generated in embodied agents and that can be interpreted by humans and other embodied agents, then, in accordance with most authors, we distinguish a list of easily perceivable bodily components of emotions (facial expressions, voice intonation, gestures and movements, posture, and pupilary dilation) and a list of components less apparent to others (respiration, heart rate, pulse, temperature, perspiration, muscle action potentials and blood presure). Obviously, although input devices become available to feed values from the second list into the computer (or even, using haptic and tactile devices, to have embodied agents output and display emotions using some of these modalities), most of our research in the area of embodied agents, in addition to the more classical intelligence aspects, deals with the emotion elements of the first list.

Although some of the bodily components of emotions are easily perceived, it does not mean that they are easily generated on an embodied agent, or easily correctly interpreted from a human (by a computer) or from a synthesized agent. Consider facial expressions and speech. In the case of facial expresions we can find the start of a reasonably comprehensive treatment in the literature. For example. Ekman and Friesen developed the Facial Action Coding System for scoring visually distinctive, observable facial movements, and they studied the relation between emotions and these facial movements. That is, they showed how emotion representations can be mapped on the contraction levels of facial muscle configurations. One of the results of their research was also that by looking at muscular actions it is possible to distinguish between genuine and fabricated emotional expressions, where fabricated may refer to referential expressions or, as a special case, mock expressions. Cultural rules, social norms and individual differences may also influence the display of emotions in the face. One other result we like to mention is that it is possible to generate a particular emotion in a human being if we can somehow generate the muscular configuration belonging to this particular emotion.

We can also find emotion research in speech recognition and synthesis. Different versions of different emotions can be recognized in human speech using a qualitative human analysis or a quantitative automatic analysis. In speech synthesis an emotional state should lead to the generation of emotionally colored speech as well. Speakers can sound angry, amused, bored or delighted, to mention a few aspects, and there are links between elements of prosody (e.g., pitch range, variation and contour, speech rate, voice quality and intensity) and emotional expression. As is the case in facial expressions (e.g. for spontaneous and mechanical smiles) one might even be able to distinguish between genuinely (spontaneously) generated expressions and acted expressions. Other aspects that can be taken into account include cross-cultural differences and context-sensitivity of the relation between speech and emotion. Because of these latter reasons, listeners can draw different conclusions about a speaker's emotional state. Another source of confusion might be the distinction between attitudes that reflect the speaker's behavior and that are also conveyed by prosodic patterns (being friendly, being rude) and emotions defined by an emotional state of the speaker.

Obviously, human beings do not only express emotions using facial expresions and speech only. Generally they have their emotions displayed using a combination of modalities that interact with each other. Facial expressions are combined with speech. There are not only audio or visual stimuli, but also audio-visual stimuli when expressing emotions. A smile gesture will change voice quality, variations in speech intensity will change facial expression, etc. Attitude, mood and personality are other factors which make interpretation and generation of emotional expressions even less straightforward. In addition we can have different intensities of emotion and the blending of different emotions in an emotional expression. We should consider combinations and integation of speech, facial expressions, gestures, postures and bodily actions. However, as mentioned, it should be understood that these are displays and that they should follow from some emotional state that has been computed from (virtual) sensory inputs for an (embodied) agent. A usual standpoint is that of appraisal theory, the evaluation of situations and categorizing arising affective states. Moreover, it should be understood that what exactly is said and what exacly is done in a social and emotional setting is not

part of the observations above. The importance of the meaning of words, phrases and sentences, uttered and to be interpreted in a specific context, is not to be diminished. Nevertheless, what is said does not always reveal what someone is feeling or thinking.

In this paper we present some of our research activities in the above mentioned areas. We survey some of our application areas, in particular the interaction with embodied conversational agents for several application domains: the building of customer trust in e-commerce environments, the stimulation of student motivation in educational environments, designing for friendship in conversational environments, etc. From more technical points of view, we have been working on a neural-network based model that makes it posible to talk about emotional state changes because of appraisals of events that an agent perceives in its environment. In order to design this model and to experiment with this model we had to design a much more simple environment than the environment which we discussed above. Events that are appraised in this simplified environment are, at least at first sight, far away from observing or deriving an emotion from generated speech, from a facial expression or from a bodily posture. Rather we have agents that look for water, apples and health and that have to deal with predators. Nevertheless, we assume that in the future it will become possible to use the model in our virtual environments inhabited by interacting embodied agents as well. Secondly, we are designing a fuzzy rule-based system to map representations of the emotional state of an animated agent onto muscle contraction values for the appropriate facial expressions. The implementation pays special attention to the way in which continuous changes in the intensity of emotions can be displayed smoothly on the graphical face. The rule system we have defined implements the patterns described by psychologists and researchers dealing with facial expressions of humans, including rules for displaying blends of expressions. Finally, we may pay attention to some of our modest attempts to design an embodied interface agent, using results from the areas of social psychology and personality psychology, that is meant to develop a friendship relation with a human using the interface.