FlurMax: An Interactive Virtual Agent for Entertaining Visitors in a Hallway

Bernhard Jung and Stefan Kopp

Artificial Intelligence and Virtual Reality Lab
University of Bielefeld
http://www.techfak.uni-bielefeld.de/ags/wbski/

Abstract. FlurMax, a virtual agent, inhabits a hallway at the University of Bielefeld. He resides in a wide-screen panel equipped with a video camera to track and interact with visitors using speech, gesture, and emotional facial expression. For example, FlurMax will detect the presence of visitors and greet them with a friendly wave, saying *"Hello, I am Max"*. FlurMax also recognizes simple gesturing of the by-passer, such as waving, and produces natural multimodal behaviors in response. FlurMax's behavior selection is controlled by a simple emotional/motivational system which gradually changes his mood between states like happy, bored, surprised, and neutral.

1   Introduction

In the AI & VR Lab in Bielefeld, a multimodal virtual agent, *Max*, is under development. Max has been employed as human-like interlocutor in various applications involving a CAVE-like VR installation [1]. This contribution presents a new application of Max, *FlurMax* (German Flur: hallway), where he is visualized on a wide-screen panel located in a fairly well frequented hallway, next to the door of our lab (see Fig. 1). Below the panel, a video camera is mounted that provides color images of the hallway area as seen from Max's perspective. The agent's task is to continuously observe his environment and, based on his
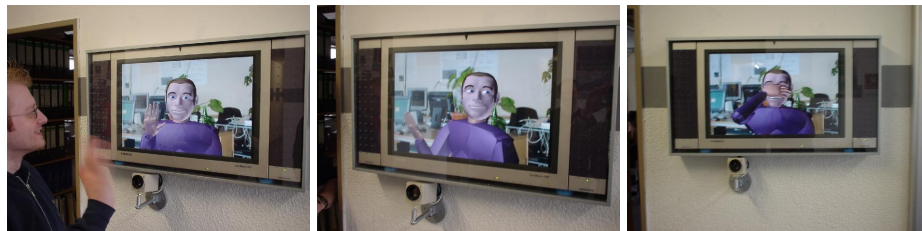

Fig. 1. FlurMax installation: Max waving back at a by-passer (left); Max follows the by-passer's position with his eyes (center); a bored yawn (left).
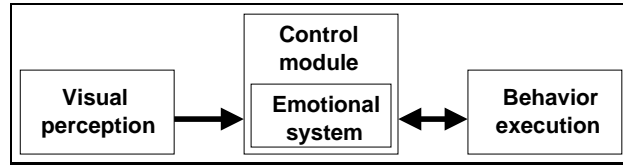
Fig. 2. Main components of FlurMax's software architecure

visual perception, to entertainingly interact with persons passing by or standing in front of the panel. Max thus acts as a kind of receptionist whose presence and non-obtrusive communicative behavior shall contribute to an overall friendly and creative atmosphere at the entrance to our lab.

2 Software Architecture

FlurMax's software architecure is composed of three main parts (see Fig. 2): (1) a visual perception component, (2) a central control module, and (3) a behavior execution component.

The visual perception component processes the video data from the camera. To ensure an overall reactive agent behavior, only real-time capable image recognition techniques are employed. Image analysis first scans the video data for skin-colored areas. The highest regions are then classified as faces and tracked over time. That way, the visual perception is able to discriminate between different persons (as long as no overlaps of face regions in the image occur). The positions of all visible faces are continuously sent to the control component. In addition, fast movements of small skin-colored regions, suddenly raised to a certain height, are interpreted as hand-waving and reported as separate events. To maintain a reliable perception over longer time periods, the image recognition component adapts itself to moderate changes of the overall lighting conditions.

The control component receives events from the visual perception component and schedules FlurMax's behavior routines for reacting to changes in his environment. One main task of the control component is to direct the agent's gaze to appropriate target positions. Every time a new set of positions is delivered, the control component keeps track of the number of people currently present and decides who the agent should look at. In the majority of cases, Max tracks the face of a particular person; sometimes, Max switches his attention by chance to another person. In addition, verbal utterances like *"Hello, I am Max"* are randomly scheduled to welcome a newly spotted person.

The control component utilizes a simple emotional/motivational system that lets the agent's mood gradually vary between happy, bored, surprised, and neutral depending, e.g., on the presence or absence of people to interact with. The emotional state influences the agent's behavior in several ways: First, the internal emotional state is quantitatively reflected in the animation of Max's facial

expression. Second, verbal utterances are modulated in prosody, a bad mood resulting in a lower average pitch, a narrowed amplitude of intonational variation and a slower speech rate. Third, and most importantly, the agent's emotional state affects the selection of behaviors. For example, periods lacking by-passer interaction increase the `bored` value, which eventually results in the scheduling of certain behaviors, such as Max leaning back, head cupped in hands, displaying frustrated inclination. Other behaviors resulting from Max being bored include taking deep breaths, yawning (see Fig. 1, right), head-scratching, and looking around as well as verbal complaints like "Nothing's up here" or even leaving the panel. In contrast, the presence of people increases Max's `happy` value by degrees, resulting in friendly facial expression, verbal utterances like "Have fun at work!" or "How are you?", and greetings involving a friendly wave (Fig. 1, left). To increase the variation of Max's behavior, the emotional system now and then amplifies the agent's `surprise` about visual perceptions, which is immediately reflected in a corresponding facial display. In addition, the dominant emotion is stochastically reduced at regular intervals.

The central control module schedules behaviors by transmitting XML-based action specifications as requests to the execution component. Such requests are formulated in the representation language MURML [3] that provides flexible means of specifying prosodic speech, synchronized gestures of a definable form, emotional expression, locomotion, and arbitrary keyframe animations of the agent's body and face. Upon deciding which behavior to execute, the control component draws a parameterized MURML specification from a library and adapts it by inserting required parameters (e.g. the target location for the `look-at`-behavior). Currently, the control component is implemented in CLIPS, a rule-based production system.

The behavior execution component is responsible for the real-time execution of action requests specified in MURML. The agent's underlying kinematic skeleton comprises 103 DOF in 57 joints, all of which are constrained to realistic joint limits. This articulated body can be driven either by applying keyframe animations or by means of a hierarchical gesture generation model that creates animations from specifications of spatiotemporal gesture properties [2]. The face of the agent is controlled by 21 muscles that are employed to animate lip-sync speaking movements, to create facial expression of emotion (examples shown in Fig. 3), or to perform arbitrary keyframe animations. Gesture generation and face animation are combined with a module for synthesizing prosodic speech in an overall production model for synchronized multimodal utterances (see [2]).

While the central control module is exclusively responsible for selecting and invoking primary behaviors, incessant low-level actions like eye blink and breath movements are controlled directly in the behavior execution component. To avoid interferences between behaviors, a mediator detects whether two behaviors are consuming conflicting body resources at the same time. In this case, the mediator removes the behavior with the lower priority value and the earlier start time. That is, lower-level actions as well as previous behaviors may be interrupted or skipped when a more recent, high-level behavior is to be executed.

Fig. 3. Facial display of various emotional states.

3 Discussion

FlurMax' personality was designed to be friendly and non-obtrusive to visitors. His facial animation capabilities, in principle, provide for a wide sprectrum of emotional expressions, including negative ones (see Fig. 3). Likewise, negative emotional states, like anger, sadness, and disgust, are actually represented by corresponding variables in his emotional system. However, such states are considered inappropriate for FlurMax's task and, thus, remain intentionally unused in the current implementation.

Generally, FlurMax is well received by visitors of our hallway. Especially striking is the attribution of personality and human-like qualities by persons who are, at least in principle, fully aware of his lack of speech understanding and higher-level cognitive capabilities. For example, it is quite typical that FlurMax encourages by-passers to engage in natural conversational behavior (*"Hello, Max!"*). Although FlurMax's conversational repertoire is currently rather limited, he has proven himself as quite successful entertainer of visitors to our lab waiting for a demonstration of his CAVE-based sibling Max [1].

Current work on FlurMax aims at a camera-based recognition of the by-passers' emotional states. Further work concerns the development of simple chatterbot-like capabilities based on typed natural language.

References

1. S. Kopp, B. Jung, N. Lessmann, and I. Wachsmuth. Max - a multimodal assistant in virtual reality construction. *KI - Künstliche Intelligenz, Special Issue Embodied Conversational Agents*, 2003. arenDTap Verlag, Bremen, to be published.
2. S. Kopp and I. Wachsmuth. Model-based animation of coverbal gesture. In *Proc. of Computer Animation 2002*, pages 252–257, Los Alamitos, CA, 2002. IEEE Computer Society Press.
3. A. Kranstedt, S. Kopp, and I. Wachsmuth. MURML: A multimodal utterance representation markup language. Technical Report 2002/05, SFB 360 Situierte Künstliche Kommunikatoren, Universität Bielefeld, 2002.