MARC: a Multimodal Affective and Reactive Character

Matthieu COURGEON Jean-Claude MARTIN Christian JACQUEMIN
CNRS-LIMSI
BP 133, 91403 Orsay
FRANCE

courgeon@gmail.com martin@limsi.fr christian.jacquemin@limsi.fr

ABSTRACT
Affective interaction between a user and a virtual agent might depend on several parameters such as the capacities of the expressive agent to respond in real-time to user's actions, to render subtle signs of affective states (e.g. detailed head model, wrinkles, skin color / rendering, shadows), or to express individual differences in the expressive reactions. The combination of these three parameters of affective interaction remains a long-term challenge. MARC is a Multimodal Affective and Reactive Character that we designed in order 1) to enable experimental studies investigating the impact of these parameters, and 2) to be integrated in affective computing applications. In this paper we focus on the architecture of the MARC system. We also explain how we recently integrated image processing of user's facial expressions of emotions with FaceReader to demonstrate the real-time capabilities of MARC to mirror user's facial expressions. MARC is currently being used in several experimental studies and we plan to use it in several affective computing applications.

Categories and Subject Descriptors
I.2.11 [**Distributed Artificial Intelligence**]: Intelligent agents

General Terms
Experimentation, Human Factors.

Keywords
Affective interaction, expressive agent, rendering, real-time animation.

1. INTRODUCTION
Designing affective interfaces featuring expressive virtual characters raises several research questions. Several experimental studies compared different appearances of virtual characters but without a focus on the communication of affective states [1-8].

Other studies investigated the perception of expressions of complex emotions displayed by a virtual agent but most of the time without considering real-time continuous interaction between the user and the agent [9]. Finally, several researchers explored various mappings between user's input and the expression of affects but without rendering subtle signs of expressions (e.g. detailed head model, wrinkles, skin color / rendering, shadows) [10, 11] that are known to be relevant for the expression and perception of affective states [12].

Affective interaction between a user and a virtual agent might depend on several parameters such as the capacities of the expressive agent to respond in real-time to user's actions, to render subtle signs of affective states, or to express individual differences in the expressive reactions. The combination of these three parameters of affective interaction remains a long-term challenge.

MARC is a Multimodal Affective and Reactive Character that we designed in order 1) to enable experimental studies investigating the impact of these parameters, and 2) to be integrated in affective computing applications.

In this paper we focus on the underlying software and hardware architecture of the MARC system, and on the relations between rendering, animation and gestural interaction with the user through individual expressive profiles. We explain how we integrated image processing of user's facial expressions of emotions with FaceReader to demonstrate the real-time capabilities of MARC to mirror user's facial expressions.

2. ARCHITECTURE OF MARC
In this section we detail MARC's real-time architecture that we briefly introduced earlier [13] (see Figure 1).

2.1 Rendering
In order to display subtle signs of expressions in real-time, the rendering uses two GPU-based techniques called *shadow mapping* and *BSSRDF* (Bi-directional Sub Surface Reflection Distribution Function).

A shadow map is computed for cast shadows and is combined with a light diffusion function through the skin that simulates skin translucency.

The BSSRDF model [14] used in our system is a simplified version of the model proposed by d'Eon et al. [15] providing a good tradeoff between performance and realism.

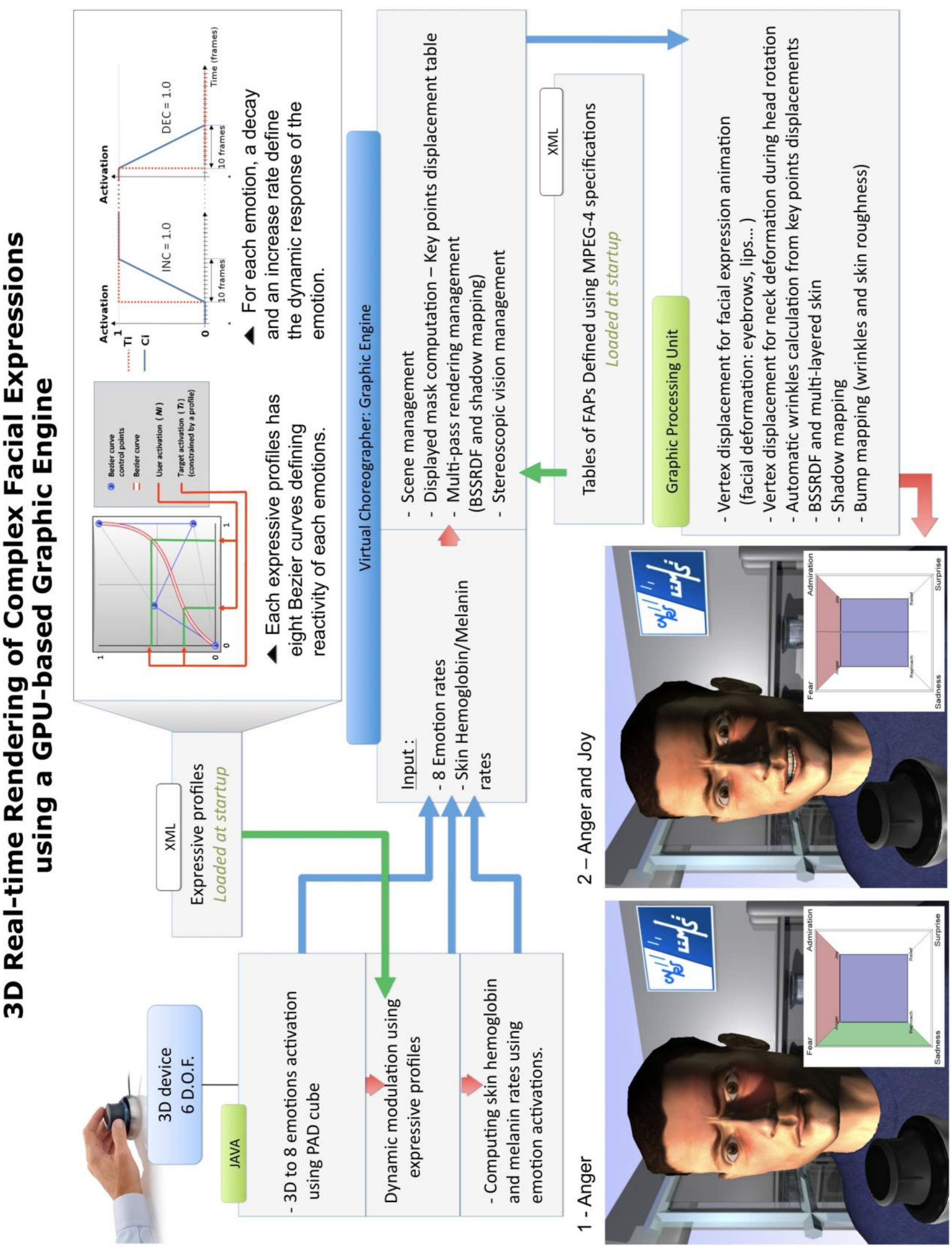

Figure 1. Marc's architecture.

We apply a BSSRDF function to the skin shadow map to simulate light diffusion through the skin. Skin rendering is computed by associating skin translucency and blurred cast shadows obtained from shadow maps. The vertex shaders applied to the face are performing GPU animations for facial expressions.

Vertex shaders also compute wrinkles activations from key point's displacements (e.g. the horizontal displacement of eyebrow key points triggers the activation of frown textures). These activations are transmitted to the fragment shader that uses them to activate bump mapping to create wrinkles on the face.

2.2 Animation

Animation is carried out through the Virtual Choreographer open source 3D engine (http://virchor.sf.net). Animation is encoded in VirChor using MPEG4 that ensures compatibility between different face models and facilitates the reuse of animation tables. Only a minimal computation of blend coefficients is made in the CPU and passed to the graphic card in order to overcome the bottleneck of data transmission through the graphic bus.

The animation is thus divided in two parts in order to enable a real-time interaction: an offline edition part and an online interactive part (see Figure 2). The design part is an offline edition process during which the static facial expressions are defined and from which the wrinkles maps are automatically computed. The interactive part involves real-time rendering and animation that rely on initial data (mesh, texture, key point weighting) and precompiled data from the first part (wrinkle bump maps).

Static expressions are stored in XML files that contain key points displacements associated with each expression. Wrinkle maps are stored as normal map textures for bump mapping. These maps contain the maximum wrinkle activations of the associated expressions. The normal maps are computed directly in the offline face edition tool.

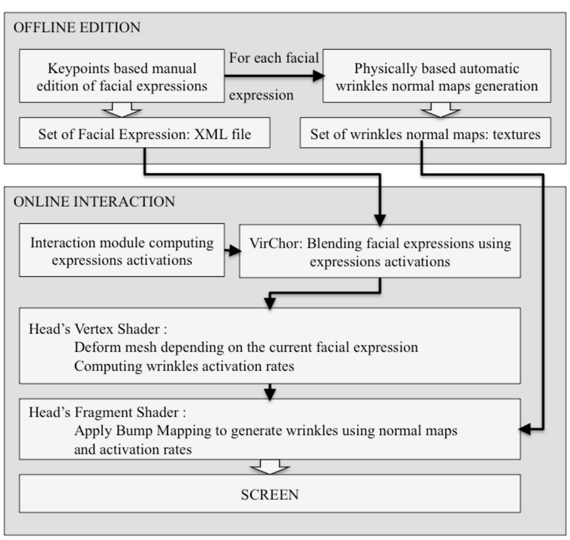

Figure 2. Real-time animation for affective interaction involving two steps: off line edition and online interaction.

Once facial expressions are modeled through key point displacement, a mass-spring system creates the wrinkles on the face based on a physical system. Wrinkles are stored as a deformation texture from which we obtain normal maps for bump mapping. When the physical system is stabilized, we compute the height map deviation that occurred between the original deformed mesh, without wrinkles, and the physically adapted mesh, with generated wrinkles. This height map is then used to generate the normal map of the current facial expression and saved into a texture.The precompiled data resulting of the offline design phase is used during the interactive phase: 1) the final animation computed at runtime results of multiple blends of predefined static expressions (achieved by VirChor and associated vertex shaders), and 2) wrinkles are automatically activated in GPU and rely on the normal maps produced by the skin models during the offline edition phase.

2.3 Affective interaction

2.3.1 PAD space

With respect to the internal representation of affective states, an interactive expressive agent such as MARC requires both categorical and dimensional representations of affects. Categorical representations enable to explicitly describe a wide range of situations that can occur in human-computer interaction and which are more easily described using labels. Dimensional representations fit well the continuous stream of data provided by input devices used in human-computer interaction.

MARC uses the PAD 3D-space for the internal representation of affective states. The three dimensions of the PAD model are [16]: Pleasure (i.e. positive versus negative affective state), Arousal (i.e. level of physical activation and/or mental alertness), and Dominance (i.e. feelings of control and influence over others and situations, versus feeling controlled and influenced by external circumstances). The PAD space has been used for experimental studies in Psychology (i.e. mapping of emotion terms onto the three dimensions [17]) and in computational models of emotions for virtual characters [18]. We use PAD in MARC since 1) the 3 dimensions enable to cover a rich set of emotions, and 2) its continuous dimensions fit well the continuous input provided by human-computer input devices. Affect categories can be placed at the 8 corners of the cube defined by the three dimensions or inside the cube using coefficients for each emotions [17].

2.3.2 Individual reaction to user's gesture

Figure 1 explains how user's gesture input on a 3D mouse is processed. In order to express individual differences in the expressive reactions, expressive profiles are defined to modulate MARC's reaction to user's action on the mouse. We define an expressive profile as a set of attributes for each affective state: an increment rate for the attack period, a decrement rate for the decay, and a Bezier curve for computing the final activation of the expressed emotion as a modulation of user's action on the joystick. Thus, from the user's actions on the 3D device, we compute a modulated target in the PAD space, and the activation of the expressed blend of emotional expressions moves toward this target, using the dynamics defined by the attack, decay and Beziers parameters.

Once computed, the 8 PAD activation values for the 8 corners of the PAD cube are sent to VirChor using UDP communication and VirChor computes the combined FAPs table for facial animation.

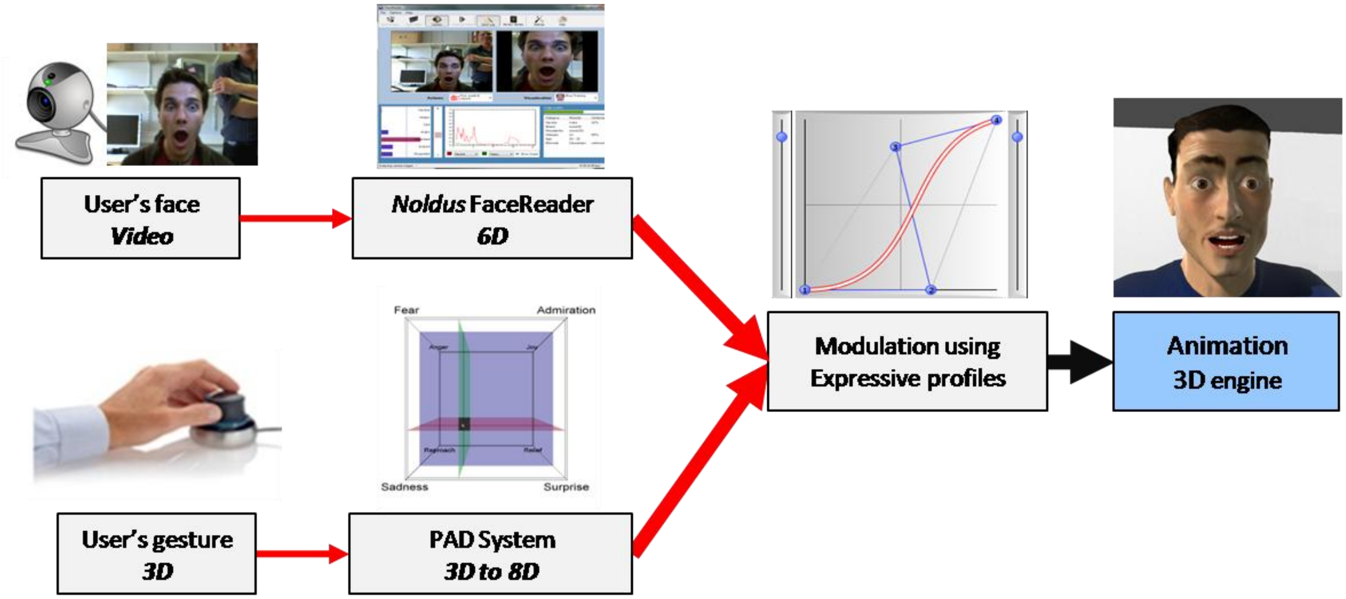

Figure 3. From user's gesture control or facial expression to facial expression of emotion by the MARC agent.

2.3.3 *Mirroring user's facial expression*

We are also considering other input modalities that we can plug into our system to investigate different mappings between input and output. We integrated image processing of user's facial expressions of emotions with FaceReader to demonstrate the real-time capabilities of MARC to mirror user's facial expressions (see Figure 3).

The classification of facial expressions is performed by FaceReader in 3 steps [10]. First, the position of the face is estimated using the Active Template Method [33].

Next, using the Active Appearance Model (AAM) which is trained on a set of annotated images, an artificial face model is synthesized that describes both the locations of key points in the face, as well as the texture of the face. An "appearance vector" can then describe new face models as deviations from the "mean face". Finally, in order to classify emotional expressions, a neural network is trained with the appearance vector as input and the seven affective states as output (happy, angry, sad, surprised, scared, disgust, neutral) [19].

FaceReader produces incrementally a log file that contains the recognition rate for these seven six affective states. This log file is processed by MARC continuously at runtime. The different recognition rates are combined to display a blended expression of emotion (see Figure 4).

3. CONCLUSION

Affective interaction between a user and a virtual agent might depend on several parameters such as the capacities of the expressive agent to respond in real-time to user's actions, to render subtle signs of affective states, or to express individual differences in the expressive reactions.

We presented the MARC agent that we designed in order 1) to enable experimental studies investigating the impact of these parameters, and 2) to be integrated in affective computing applications. We explained how user's input (gesture on a 3D device or facial expression of emotion) is mapped onto facial expressions of complex emotions.

MARC is currently being used in several experimental studies about the perception of affective states and evaluation of human-computer interfaces (for example by comparing different input devices for controlling the facial expressions). We plan to use it in several affective computing applications related to virtual arts and education.

Figure 5. Recognition of a blend of anger and disgust in user's facial expression by image processing (right) and facial expression of this blend by the MARC agent (left).

4. REFERENCES

[1] Van Vugt, H. C., Konijn, E. A., Hoorn, J. F., Keur, I. and Eliëns, A. Realism is not all! User engagement with task-related interface characters. Interacting with Computers, 19, 2 (2006), 267-280.

[2] Baylor, A. L. and Kim, Y. Pedagogical Agent Design: The Impact of Agent Realism, Gender, Ethnicity, and Instructional Role. City, 2004.

[3] Garau, M., Slater, M., Vinayagamoorhty, V., Brogni, A., Steed, A. and Sasse, M. A. 2003. The Impact of Avatar Realism and Eye Gaze Control on Perceived Quality of Communication in a Shared Immersive Virtual Environment. . In Proceedings of the SIG-CHI conference on Human factors in computing systems (Fort Lauderdale, FL, USA, April 5-10, 2003).

[4] Dehn, D. M. and van Mulken, S. The impact of animated interface agents: a review of empirical research. International Journal of Human-Computer Studies, 52 (2000), 1-22.

[5] Beun, R. J. Embodied conversational agents: Effects on memory performance and anthropomorphisation. City, 2003.

[6] Koda, T. and Maes, P. Agents with faces: the effects of personification of agents. The British HCI Group, City, 1996.

[7] Mc Breen, H. and Jack, M. Evaluating humanoid synthetic agents in e-retail applications. IEEE Transactions on Systems, Man and Cybernetics, 31, 5 (2001), 394-405.

[8] Wonisch, D. and Cooper, G. Interface Agents: preferred appearance charactersistics based upon context. City, 2002.

[9] Buisine, S., Abrilian, S., Niewiadomski, R., Martin, J.-C., Devillers, L. and Pelachaud, C. 2006. Perception of Blended Emotions: from Video Corpus to Expressive Agent. In Proceedings of the 6th International Conference on Intelligent Virtual Agents (IVA'2006) (Marina del Rey, USA, 21-23 august, 2006). Springer.

[10] Melder, W., A., , Truong, K. P., Uyl, M. D., Van Leeuwen, D. A., Neerincx, M. A., Loos, L. R. and Stock Plum, B. 2007. Affective multimodal mirror: sensing and eliciting laughter. In Proceedings of the Proceedings of the international workshop on Human-centered multimedia (Augsburg, Bavaria, Germany, 2007). ACM.

[11] D'Alessandro, C., D'Alessandro, N., Le Beux, S., Simko, J., Cetin, F. and Pirker, H. 2005. The speech conductor : gestural control of speech synthesis. In Proceedings of the eNTERFACE 2005. The SIMILAR NoE Summer Workshop on Multimodal Interfaces (Mons, Belgium, July 18-August 12, 2005).

[12] Ekman, P. and Friesen, W. V. Unmasking the face. A guide to recognizing emotions from facial clues. Prentice-Hall Inc., Englewood Cliffs, N.J., 1975.

[13] Courgeon, M., Martin, J.-C. and Jacquemin, C. 2008. User's Gestural Exploration of Different Virtual Agents' Expressive Profiles. In Proceedings of the 7th International Conference on Autonomous Agents and Multiagent Systems (Estoril, Portugal, 12-16 May, 2008). ACM.

[14] Donner, C. and Jensen, H. W. A Spectral BSSRDF for shading human skin. Eurographics Symposium on Rendering 409-417, 2006.

[15] d'Eon, E., Luebke, D. and Enderton, E. Efficient rendering of. human skin. City, 2007.

[16] Mehrabian, A. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. Current Psychology, 14 (1996), 261-292.

[17] Russell, J. A. and Mehrabian, A. Evidence for a three-factor theory of emotions. Journal of Research in Personality, 11 (1977), 273-294.

[18] Gebhard, P. 2005. ALMA - A Layered Model of Affect In Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'05) (Utrecht, 2005).

[19] van Kuilenburg, H., Wiering, M. and Den Uyl, M. 2005. A model based method for automatic facial expression recognition. In Proceedings of the European Conference on Machine Learning (ECML' 05), (2005).