# User's Gestural Exploration of Different Virtual Agents' Expressive Profiles (Short Paper)

Matthieu Courgeon
courgeon@gmail.com

Jean-Claude Martin
LIMSI-CNRS
+33.1.69.85.81.04
BP 133, 91403 Orsay, France
martin@limsi.fr

Christian Jacquemin
Christian.Jacquemin@limsi.fr

## ABSTRACT
Designing affective user interfaces involving expressive characters raises several questions. The system should be able to display facial expressions of complex emotions as dynamic and realtime reactions to user's inputs. From a cognitive point of view, designers need to know how the user will perceive the dynamics of these facial expressions as a function of his/her input. We aim at evaluating if users can perceive different expressive profiles of a virtual character by manually controlling its expressions and observing its reaction to his/her input. This paper describes our platform that enables a virtual character to display blended facial expressions of emotions as realtime continuous reactions to users' gesture input. We explain the techniques underlying the computation of intermediate facial expressions of emotion, and their control in the 3D space PAD (Pleasure, Arousal, Dominance) using gesture input. Preliminary results of a perceptive study show the potential of such an approach for assessing the dynamics of the perception of emotional expressions during gesture interaction with virtual characters endowed with different expressive profiles.

## Categories and Subject Descriptors
H5.2 [**Information Interfaces and Presentation**]: User Interfaces—Input devices and strategies, Interaction styles, User-centered design.

## General Terms
Algorithms, Design, Human Factors.

## Keywords
Expressive agent, realtime interaction, facial expressions.

## 1. INTRODUCTION
Designing affective user interfaces involving expressive virtual characters raises several research questions. From a computer science point of view, the character should be able to display facial expressions of complex emotions as dynamic and realtime reactions to user's inputs.

**Cite as**: User's Gestural Exploration of Different Virtual Agents' Expressive Profiles (Short Paper), Matthieu Courgeon, Jean-Claude Martin, Christian Jacquemin, *Proc. of 7th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2008)*, Padgham, Parkes, Müller and Parsons (eds.), May, 12-16., 2008, Estoril, Portugal, pp. 1237-1240.


From a cognitive point of view, designers of virtual characters need to know how the user will perceive the dynamics of these facial expressions in relation with her/his own input.

There has been already a long history of computational models of facial expressions in virtual characters as well as psychological studies showing the importance of the dynamics of facial expressions [1]. Several animation techniques were proposed for talking heads involving model-based and image-based approaches [2]. In order to go beyond the display of individual basic emotions, models are defined for the facial display through so-called blends of emotion or nonarchetypal expressions [3-6]. Interpolation algorithms have been proposed for generating facial expressions for a continuum from pure to mixed emotions of varying intensity [6]. MPEG-4 is a standard for facial animation [7] that researchers use to specify both archetypal facial expressions and facial expressions of intermediate emotions [5, 8]. Rather than interpolating two expressions, models were also proposed that combine different areas of the face to display blends of emotions [3, 4]. Experiments were conducted to study individual differences in users' perceptions of blended emotions from virtual characters expressions [4, 9]. Layered models were defined for relating facial expressions of emotions on one hand, and on the other hand moods and personality traits using three different timescales [10, 11].

A few affective computing studies have combined the interaction between the user's input and the display of expressive signals by the system. Gesture input via a tablet is suggested as a relevant modality for realtime and interactive control of expressive speech [12]. An experiment used an anthropomorphic tangible interface to dynamically control the animation of a 3D face [13] and shows that the tactile control of a face can be experienced as a unique affective communication medium. The Sentoy enables the user to express emotion using tactile input [14]. All these studies suggest that gesture can be used as a means for exploring the space of facial expressions that a given virtual character can convey.

The PAD space (Pleasure, Arousal, Dominance) can be seen as a framework for the description and measurement of emotional states as well as temperamental dispositions to certain emotional responses (see [15] for an overview). The three dimensions of this model are: 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 is used both for experimental studies in Psychology (i.e. mapping of emotion terms onto the three dimensions [15]) and in computational models of emotions for virtual characters [10, 11, 16].

The above mentioned studies did not focus on the fine-grained dynamics of interaction between the user and an expressive character. In this paper we present the design of a software platform that enables virtual characters to display blended facial expressions of emotions as realtime reactions to users' input (Figure 1). We aim at mixed realistic and artistic applications [17] which require high quality rendering (detailed head model, wrinkles, and layered skin rendering techniques enabling the character to blush or to turn pale). We also use this platform to understand how users perceive and react to blends of facial expressions of emotions that a given virtual character is able to display. We are interested in the gesture input modality since it seems to be a relevant media for low-level realtime interaction in mixed reality applications.

Section 2 of this paper describes the graphical rendering components that enable realtime and interactive display of the virtual character. Section 3 explains the techniques that underlie the display of blended facial expressions of emotion and its control using gesture input. Section 4 describes a preliminary perception study that shows how such a realtime interactive platform can be used for designing affective interfaces and for studying the perception that users have of various expressive profiles of virtual characters.

## 2. A PLATFORM FOR A REALTIME 3D INTERACTIVE AGENT

The purpose of this platform for real-time rendering of an interactive agent is to offer a cohesive and flexible framework that combines high quality and expressive visual rendering, and interactive control of facial expressions:

– *flexibility* results from the use of MPEG-4 visual animation encoding that enables various face models animation,

– *high quality visual rendering* is obtained through shader implementation of multi-layered skin model,

– *expressivity* is obtained by enhancing mesh animation with visual signs of emotions such as blushing and blemishing, frowning, and crows-foot wrinkles.

Animation is carried out through Virtual Choreographer (http://virchor.sf.net), an open source 3D engine. Animation is encoded in VirChor using MPEG4 that ensures compatibility between different face models and facilitates the reuse of animation tables (see [12] for more details). 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 next section explains how this realtime architecture is used to explore the perception through interactive gesture control of virtual characters endowed with different expressive profiles.

**Figure 1. The space of facial expressions that a virtual character can display is explored by a user with a joystick.**

## 3. GESTURE EXPLORATION OF EXPRESSIVE PROFILES

Our goal is to investigate how a user can perceive the expressive profile of a virtual character during realtime interaction using low-level gesture classical devices such as a joystick or similar devices that are well-known to users. Our purpose is to design applications in which the user should be able to perceive and react on implicit visual-only facial expressions of complex emotions (game, mediated communication, …). Three dimensions enable to navigate a rich set of emotional expressions.

The PAD space [15] is relevant to our research goals for practical reasons since we want to study user's exploration of a continuous space of blended expressions of emotions, and its three dimensions can be easily mapped onto a 3D gesture device. In our experimental approach, it is the user who decides when and where to move the state of the agent in the PAD space. We selected 8 affective state labels to represent the corners of this 3D cube: *fear*, *distress*, *anger*, *reproach*, *joy*, *relief*, *satisfaction*, and *admiration*. We interpret a point in this 3D space as a blend of these 8 emotions. Figure 2 presents the software architecture that we implemented and that we detail in this section. Via a joystick, the user moves a 3D point in the PAD space. The 3D coordinates *(x, y, z)* captured from the joystick are mapped onto 8 PAD activation values *(N1, …, N8)* for each of the corners of the PAD space. In order to test if the user is able to perceive different emotional expression profiles, we implemented different filters which transform these 8 coordinates into 8 values *(C1, …, C8)*. These 8 modified activation values are sent to the VirChor graphical engine which maps them to FAPs, which are finally sent to the graphical processing unit for rendering.

### 3.1 Implementation

Using the graphical platform described in the previous section, we defined static FAPs tables for each of the 8 expressions of emotion at the corners of the PAD cube. These tables were stored into XML tables which are loaded as weighting coefficients in GPU at compile time. We were willing to use low level gesture interaction in order to study the impact of realtime interactivity on the perception that the user has over the expressive profile of the agent. Through the three dimensions of a joystick (including joystick vertical rotation), the user can explore the PAD space of emotional expressions.

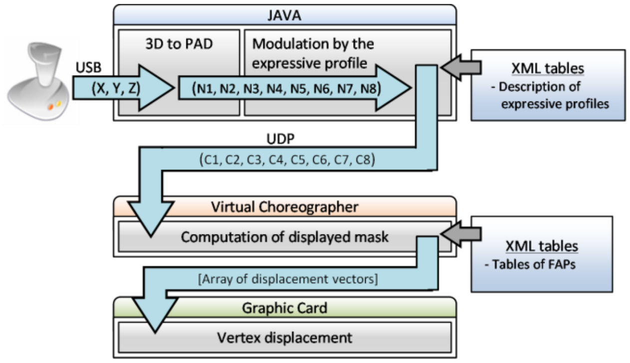
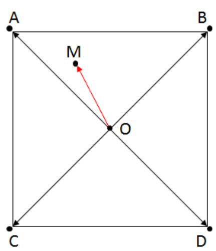



**Figure 2. Overview of the software architecture.**

The initial position of the joystick is (0, 0, 0) and is associated with the neutral expression in the PAD space. Figure 3 illustrates the mapping of a 2D point *M* onto a 4D point *N: (N1, N2, N3, N4)* in the emotions space defined by four vectors (*E1*, *E2*, *E3*, *E4*). In order to compute *N*, we use a function named *Act* for Activation. In the example of Figure 3, emotion A has a high activation, emotion B has a low activation, and emotions C and D have no activation (we do not enable any negative activation of an emotion expressed with a facial expression).




Function *Act* : $M \rightarrow N$, $\mathbb{R}^2 \rightarrow \mathbb{R}^4$

$$\begin{cases} N_1 = \max(0, (\vec{OM}\cdot\vec{OA}))\,/\,|\vec{OA}| \\ N_2 = \max(0, (\vec{OM}\cdot\vec{OB}))\,/\,|\vec{OB}| \\ N_3 = \max(0, (\vec{OM}\cdot\vec{OC}))\,/\,|\vec{OC}| \\ N_4 = \max(0, (\vec{OM}\cdot\vec{OD}))\,/\,|\vec{OD}| \end{cases}$$

**Figure 3. Computation of 4 emotional expression activations associated with a point M in a 2D square.**

The same principle applies when using a PAD 3D cube instead of a 2D square. We get 8 values from a three dimensional point. The function *Act* is defined as shown in Figure 4.


Function *Act* : $M \rightarrow N$, $\mathbb{R}^3 \rightarrow \mathbb{R}^8$

$$\begin{cases} N_1 = \max(0, (\vec{OM}\cdot\vec{OA}))\,/\,|\vec{OA}| \\ \dots \\ N_8 = \max(0, (\vec{OM}\cdot\vec{OH}))\,/\,|\vec{OH}| \end{cases}$$

**Figure 4. Computing the activation of 8 emotional expressions (right) corresponding to a point in a 3D PAD cube (left).**

Based upon our virtual character that reacts in realtime to user's gestures on a joystick, we have evaluated how users perceive different expressive profiles. Our hypothesis is that such real time and dynamic interaction enables the user to perceive the expressive profile defined for the agent. An expressive profile defined for the agent constrains user's action in the PAD cube. Six expressive profiles are defined along three dimensions: 1) Expressivity (Low / High), 2) Speed (Slow / Fast), and 3) Valence (Negative / Positive). These six expressive profiles are selected because of their potential to be perceived via low-level gesture interaction.

We define an expressive profile as a set of attributes for each edge/emotion of the PAD cube: 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 joystick, 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. The expressive profile assigned to the agent modulates the activation of a facial expression as follows. Let *Ni* be the user-defined activation of an emotion ***I***. Let the expressive profile of the current agent be defined using eight Bezier functions (one for each emotion) called $B_i : [0,1] \rightarrow [0,1]$. The activation target of this emotion for this expressive profile is computed as: *Ti* = *Bi*(*Ni*) (Figure 5).

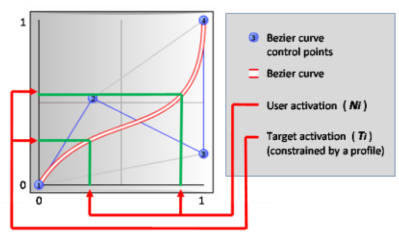



**Figure 5. Modulating user's activation of an emotion using a Bezier curve defined for a given expressive profile.**

The current modulated activations of the 8 emotions are called *Ci* (Figure 2). *Ci* depends on the T*i* target positions, and on the increase and decrease rates defining attack and decay. The increase rate *INC*, and decrease rate *DEC* have the same purpose, except that *INC* will be used if *Ci*<*Ti*, and *DEC* will be used if *Ci*>*Ti* (Figure 6). For example, if *Ci*=0, T*i*=1, and *INCi*=p, it will take 10/p frames for *Ci* to reach *Ti*. With a 100% increase rate, it will take at least 10 frames to stabilize *Ci*, and at 60 frames per second, 0.166 second, which is a very fast reaction. With a 1% ratio, it will take 1000 frames, which will last 16.66 second, which would be a slow reaction to user's action. Once computed, the 8 PAD activation values are sent to VirChor using UDP communication and VirChor computes the combined FAPs table for facial animation:

$$\text{Displayed FAPs table} = \sum_{i=1}^{8} C_i \times (\text{FAPs Table})_i$$

Where *C* contains the PAD values and the FAPs tables are the displacement tables loaded in VirChor at startup.

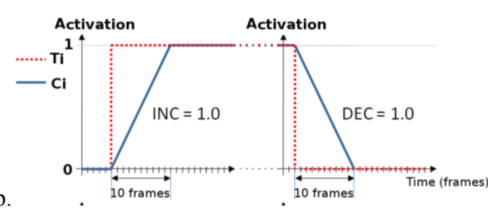



**Figure 6. Attack and decay rates for emotion dynamics**

## 4. CONCLUSIONS AND FUTURE DIRECTIONS

We presented a platform that we implemented in order to study the impact of realtime interaction on the perception of different expressive profiles of virtual characters. A preliminary evaluation study validated that subjects were able to perceive our specifications of several agents' profiles: the Low expressivity, High expressivity, Negative profiles were well recognized.

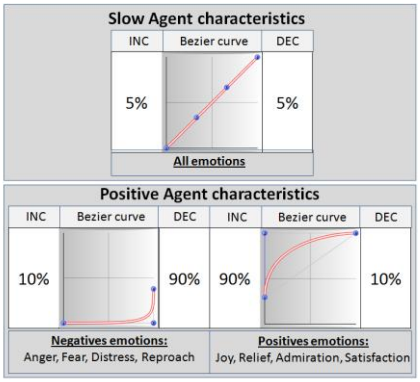



**Figure 7. The Slow agent features for all emotions a slow attack / decay, and an apex defined as a linear function of user's input. The Positive agent features for negative emotions a slow attack, and a curve which lets the user express negative emotions only when there is a large gesture.**

The results we presented here were collected with a small number of subjects. The experiment will be replicated with a larger collection of subjects and improved agents and protocol. Other media for 3D exploration will be evaluated.

# 5. REFERENCES

[1] Frank, M. G., Ekman, P. and Friesen, W. V. Behavioral markers and recognizability of the smile of enjoyment. Journal of Personality and Social Psychology, 64 (1993), 83-93.

[2] Bailly, G., Bérar, M., Elisei, F. and Odisi, M. Audiovisual Speech Synthesis. International Journal of Speech Technology. Special Issue on Speech Synthesis: Part II., 6, 4 (2003), 331-346.

[3] 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.

[4] Niewiadomski, R. A model of complex facial expressions in interpersonal relations for animated agents. PhD. dissertation, University of Perugia, 2007.

[5] Tsapatsoulis, N., Raouzaiou, A., Kollias, S., Cowie, R. and Douglas-Cowie, E. Emotion Recognition and Synthesis based on MPEG-4 FAPs. MPEG-4 Facial Animation. John Wiley & Sons, 141-167, UK, 2002.

[6] Albrecht, I., Schröder, M., Haber, J. and Seidel, H.-P. Mixed feelings: Expression of non-basic emotions in a muscle-based talking head. Special issue of Journal of Virtual Reality on "Language, Speech & Gesture", 8, 4 (2005), 201-212.

[7] Pandzic, I. S. and Forchheimer, R. MPEG-4 Facial Animation. The Standard, Implementation and Applications. John Wiley & Sons, LTD, 2002.

[8] Malatesta, L., Raouzaiou, A., Karpouzis, K. and Kollias, S. MPEG-4 facial expression synthesis. Special issue of the Journal on Personal and Ubiquitous Computing on 'Emerging Multimodal Interfaces' following the special session of the AIAI 2006 Conference. Springer. (2007).

[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] 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).

[11] Kshirsagar, S. 2002. A multilayer personality model. In Proceedings of the 2nd international symposium on Smart graphics SMARTGRAPH '02 (Hawthorne, New York, 2002). ACM Press.

[12] 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).

[13] Jacquemin, C. 2007. Pogany: A tangible cephalomorphic interface for expressive facial animation. In Proceedings of the 2nd International Conference on Affective Computing and Intelligent Interaction (ACII'2007), LNCS, vol. 4738 (Lisbon, Portugal, September 12-14, 2007). Springer.

[14] Paiva, A., Andersson, G., Hook, K., Mourao, D., Costa, M. and Martinho, C. Sentoy in FantasyA: Designing an affective sympathetic interface to a computer game. Personal Ubiquitous Comput, 6, 5-6 (2002), 378–389.

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

[16] Becker, C., Leßmann, N., Kopp, S. and Wachsmuth, I. 2006. Connecting feelings and thoughts - modeling the interaction of emotion and cognition in embodied agents. In Proceedings of the Seventh International Conference on Cognitive Modeling (ICCM-06) (Trieste, Italy, 2006). Edizioni Goliardiche.

[17] Martin, J.-C., d'Alessandro, C., Jacquemin, C., Katz, B., Max, A., Pointal, L. and Rilliard, A. 2007. 3D Audiovisual Rendering and Real-Time Interactive Control of Expressivity in a Talking Head In Proceedings of the 7th International Conference on Intelligent Virtual Agents (IVA'2007) (Paris, France, September 17-19, 2007).