

An Emotionally Responsive AR Art Installation

Stephen W. Gilroy¹
S.W.Gilroy@tees.ac.uk

Satu-Marja Mäkelä²
Satu-Marja.Makela@vtt.fi

Thurid Vogt³
thurid.vogt@informatik.uni-augsburg.de

Marc Cavazza¹
M.O.Cavazza@tees.ac.uk

Markus Niiranen²
markus.niiranen@vtt.fi

Mark Billingham⁴
mark.billinghurst@hitlabnz.org

Maurice Benayoun
mb@benayoun.com

Rémi Chaignon¹
R.Chaignon@tees.ac.uk

Elisabeth André³
andre@informatik.uni-augsburg.de

Hartmut Seichter⁴
hartmut.seichter@hitlabnz.org

¹ University of Teesside, UK; ² VTT Electronics, Finland; ³ University of Augsburg, Germany; ⁴ HITLabNZ, New Zealand

ABSTRACT

In this paper, we describe a novel method of combining emotional input and an Augmented Reality (AR) tracking/display system to produce dynamic interactive art that responds to the perceived emotional content of viewer reactions and interactions. As part of the CALLAS project, our aim is to explore multimodal interaction in an Arts and Entertainment context. The approach we describe has been implemented as part of a prototype “showcase” in collaboration with a digital artist designed to demonstrate how affective input from the audience of an interactive art installation can be used to enhance and enrich the aesthetic experience of the artistic work. We propose an affective model for combining emotionally-loaded participant input with aesthetic interpretations of interaction, together with a mapping which controls properties of dynamically generated digital art.

1. INTRODUCTION

Affective Interfaces have developed as a major research topic in Human-Computer Interaction. These interfaces usually analyse user experience in a communication setting, their aim being to re-incorporate affective elements into that process (whether those emotions are detected or elicited by the interface). Comparatively less research has been dedicated to the affective aspects that result from interaction with digital media, in particular when the user experience is dependent on the aesthetic aspects. The aim of the CALLAS Project [1] is to develop multimodal affective interfaces in the context of new media and digital entertainment, including Digital Arts. In this paper, we describe research on the development of multimodal affective interaction with an Augmented Reality (AR) art installation.

In this context, AR achieves a unique combination of media display, real-world installation and sensor-based interactions. It thus constitutes a privileged environment to study user interaction with an artistic installation. The preservation of a real-world physical environment supports more natural user behaviour, whilst the incorporation of multimodal sensors (cameras, trackers, microphones) serves as a basis for developing multimodal affective processing, such as user attitude recognition, emotional speech recognition, and a range of non-verbal behaviour. Finally, as an artistic medium, AR provides both interactivity and the visual aesthetics of virtual elements. It can thus be used to experiment with “affective feedback loops”, in which the experience elicits affective responses from the user, which in turn

are analysed to modify the visual presentation of the installation. Beyond their potential to support artistic installations, such systems constitute similarly privileged test-beds for the development of multimodal affective interaction.

1.1 E-Tree: An AR Showcase

The original idea and brief for the AR Art installation has been created by Maurice Benayoun, a leading digital artist [2], whose previous works, such as “frozen feelings” or the “Emotion Vending Machine” have already explored the theme of emotion. He envisions an Emotional Tree (or “E-Tree”)—a virtual tree structure whose growth and evolution reflects the perceived affective response from the spectator throughout interaction (*e.g.*, in terms of interest or positive and negative judgement). The user experience is captured through dimensional models of emotion that are instantiated from multimodal input. In turn, the emotional models affect various parameters of tree growth via the underlying L-system used to generate a tree.

2. AFFECTIVE MODEL OF EXPERIENCE

Emotional models describe possible affective states, their causal relationships and patterns of expression. Usually a small number of possible emotional categories are posited based on the ability of various recognition techniques to detect distinct states (including human ability to recognise emotions in other humans). The most famous of these are probably those given by Paul Ekman, from research on universal facial expression: fear, anger, sadness, happiness and joy. A larger number of recognisable affective states can be expressed in words, but these can be cultural dependent, and in the case of English terms, can be explained in terms of just a few basic emotional terms (*e.g.*, by a variation in intensity or in a particular context).

Discrete affective states are a rather impoverished way of describing a user experience, and don't take into account the wider notion of aesthetic judgement of a piece of art or entertainment. A better model might utilise a dimensional approach to affective response. Dimensional models posit the existence of an “emotion space” in terms of orthogonal components of affect. Common dimensional models have two or three dimensions, and usually include “arousal/intensity” and “valence” (positive and negative). The idea is usually to link the dimensions to measurable signals of affect, and to label points or regions within the model with affective states.

Dimensional models are appealing to us in the context of E-Tree, as we can map continuous values provided affective recognition components to properties of the artwork, giving a fine-grained representation of detected affective signals. They also give a common framework in which to combine multimodal input, provided we can express such input in terms of the dimensions of the chosen model.

However, we are also interested in the aesthetic aspects of experience, such as interest, exploration, approval, satisfaction and playfulness. We especially desire to integrate the interactions that can occur when a participant can directly manipulate parts of an installation.

2.1 The PAD model

We are still working on a richer model of experience based around the concept of “flow” suggested by Csikszentmihalyi [5] and refined by Novak and Hoffman [6], but for E-Tree, we are incorporating aesthetics and affective input into a single dimensional model. The model we are using is Mehrabian’s Pleasure-Arousal-Dominance (PAD) model [7]. It was designed to capture an individual’s tendencies to emotional reaction, but is also used as a way of rating consumer reactions to new products, and therefore already has elements of an aesthetical nature.

The dimensions in this model are Pleasure-Displeasure, Arousal-Non-arousal and Dominant-Submissive, and are rated on a normalised scale of -1.0 to +1.0, so, for example, in the Pleasure-Displeasure dimension -1.0 is extreme displeasure, 0.0 is neutral and 1.0 is extreme pleasure. It can be seen already that this dimension incorporates aesthetic and affective properties, as it can be used to rate the valence of affect (feeling positive or negative), and an aesthetic reaction (an object that is pleasing to look at, or an action that can be interpreted as positive). The Pleasure and Arousal dimensions correspond roughly to common valence/intensity models, while the Dominance dimension can be used to distinguish between similarly valenced emotions such as anger (dominant) and fear (submissive).

Our aim is not to produce distinct affective states as output (as the PAD model is often used for, after scoring feedback questions on the three dimensional scales), but quite the opposite, giving a way to combine a variety of discrete and continuous multimodal inputs into a single model of experience.

The model is useful in two main ways. First as a “fuzzy” categorical tool, where we can characterise points that are close together indicating a similar experience, which thus might evoke similar outputs, and as a way of integrating changes in mood and aesthetic appreciation over time, where a series of divergent inputs will cause the position in the model space to move towards a new interpretation.

2.2 Aggregation of Affective Response

The three dimensions of the PAD model provide us with three useful continuous values we can use to produce a display that both represent a large number of affective states and illustrates gradual (or sudden!) changes over time. As our “tree” grows, the existing branches remain in the configuration they were when created, illustrating the prevailing dimensional values at that time. New branches and future growth are determined by current values. The tree also has global properties that reflect the transient affective state.

We treat multimodal input as a signal of affect with a score in each of the dimensional of the PAD model. Details of how that is achieved for the inputs utilised in the current system are described in section 0. We give the system a “baseline” state, that in the absence of affective input, it will tend towards. This could reflect the latent “personality” of the installation. In the case of E-Tree the baseline is a neutral state (0.0, 0.0, 0.0). We have chosen a simple decay model, where the absolute PAD values (from 0.0 to 1.0 in each direction of each dimension) are halved each time step.

2.3 Aesthetics of Experience

In order to represent aesthetic aspects of experience in the PAD model, we provided a mapping from some of our early concepts of User Experience to the three dimensions of the model. As the user experience model develops it may no longer fit within the PAD model, and additional interpretations will be required for the artistic display.

The main aspect of experience we incorporate is interest. If a user interacts with an installation, or is seen to be studying it, we recognise this as an interest in the whole or a part of the installation. The combination of interest with traditional affective properties such as valence, leads to richer concepts such as having your attention held by some distasteful (like a horror film) and passively letting positive experiences unfold.

Aesthetic values like interest have their own semantic content that can be used separately (*e.g.*, providing something new if interest wanes), but can also be represented by affective components. Thus, in terms of the PAD model, we see interest as a combination of arousal and dominance. If a participant takes an active role or interest in something, it is provoking an arousal (whether intellectual or physical). A more intense studying of an object, or more participants taking notice can be characterised as an increase in dominance of that object (rather than the dominant feelings of the participant themselves). This expansion of affect beyond the user to the aesthetics of the installation can produce interesting feedback effects such as interest in an object leading to a display of dominance that causes the user to react in a more submissive way.

3. E-TREE SYSTEM ARCHITECTURE

The E-Tree system is divided into three parts: affective input capture, interpretation and aggregation, and display generation. User interaction with the AR system is captured and fed back to the system as additional affective input. These parts communicate via networking protocols (TCP and UDP) so they can be run on separate PCs. Figure 1 shows the main components in the architecture, indicating the grouping of related components into network-connected modules.

As an example of the interpretation process, consider the interpretation of a positive affective utterance (explained in detail in Section 3.2.1). The EmoVoice component will generate a TCP message with the text `PositiveActive`, and send it to the affective interpretation module. This is interpreted as indicative of a Pleasurable and Aroused emotional state, and an equivalent PAD score of {1.0, 1.0, 0.0}, which is sent to the aggregation component. The aggregation component looks at the current PAD scores, and determines a new set of values given the scores from the input (using an averaging function). If the current scores are say {0.6, -0.3, 0.2}, then the new scores will be something like {0.8, 0.2, 0.2}, indicating a large increase in arousal, and a

moderate increase in pleasure). The updated PAD score is sent to the display component over TCP. This component will then alter the properties of the L-system that generates the E-Tree. Its colour is updated to reflect the absolute current PAD values. The increase in arousal will cause the tree to look less “droopy”, while the increased pleasure will cause more branches to grow and the tree to grow faster.

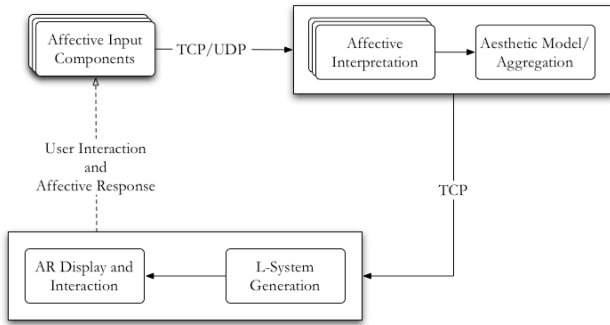


Figure 1: E-Tree system overview.

Feedback is provided in two ways. Firstly the audience’s affective reaction to the on-going generation of artwork will result in additional affective input. The artwork’s interpretation of input can be used to provide reinforcement of affect, subversion, and to react to perceived interest and boredom. Secondly a participant can directly manipulate parts of the installation that alter the display of the E-Tree. This will also generate the first kind of feedback as the participant reacts to the changes their interaction has produced.

Components are used in an “online” fashion—that is, they respond to affective input “as-it-happens”, though there may be a requirement for off-line training of the component before the use of the component.

3.1 Affective Input Components

Affective reactions to the installation are gathered by a variety of independently developed components that utilize various channels of input, such as speech, ambient noise and video input. The system currently utilises two affective recognition components— affective speech and face detection—as well as some early user experience analysis technology.

3.1.1 Affective Speech Classification (EmoVoice)

EmoVoice identifies affect conveyed by the voice. No semantic information is extracted—the recognition relies only on the acoustic signal. For the integration into the showcase, this has to be done in real-time, which hasn’t been fully attempted before.

Affect recognition in EmoVoice is a three-step process as illustrated in Figure 2. First, the acoustic input signal coming continuously from the microphone is segmented into chunks by voice activity detection (VAD), which segments the signal into speech frames with no pauses within longer than about 0.5 seconds. Next, from this speech frame, a number of features relevant to affect are extracted. The features are based on pitch, energy, Mel-frequency cepstral coefficients (MFCC) (also used for automatic speech recognition), the frequency spectrum, the harmonics-to-noise ratio, duration and pauses. The actual feature vector is then obtained by calculating statistics (mean, maximum, minimum, etc.) over the speech frame ending up with around

1300 features. A full account of the feature extraction strategy can be found in [8].

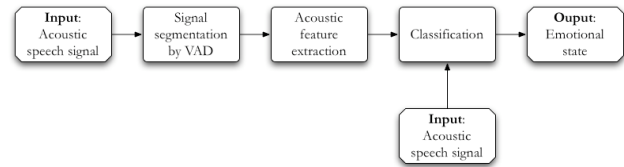


Figure 2: EmoVoice classification process.

In the last step, the feature vector is classified into an affective state by a Naïve Bayes classifier. This is a simple, but fast classifier which makes it suitable for a real-time recognition application, while its accuracy is not much worse than that of more sophisticated classifiers such as support vector machines.

As Naïve Bayes is a statistical classifier, it needs training data to be generated. Generally, it is best to have training data that is as similar to the application scenario as possible, especially since there is no general-purpose database with emotional speech available. So for the E-Tree showcase, 3 test speakers recorded 120 sentences in English simulating three affective states: positive-active, neutral and negative-passive.

3.1.2 Video Feature Extraction

The Video Feature Extraction component evaluates the number of faces in frame and tracks their movements in a live or recorded video stream. The component's functionality is divided into two parts. Face Detection tries to detect an initial set of faces for tracking, while Face Tracking keeps track of detected faces and provides their location information. Sample tracking output overlaid onto the source video is shown in Figure 3. Face Detection is based on the OpenCV library's (Open Source Computer Vision Library¹) object detection and Face Tracking uses OpenCV's object tracking functionality. The Video Feature component returns the number faces as well as the estimated facial area (ellipse) along the tilt of the ellipse.



Figure 3: Multiple face detection. The closer a viewer is to the camera the larger the area of the ellipse

The component is designed for real-time applications. In general the detection requires more processing power than the tracking. Tracking is therefore performed more often than detection to reduce the processing load of the component. The ratio of function calls is dependent on the computational power, with a bias towards detection.

¹ <http://www.intel.com/technology/computing/opencv/index.htm>

3.2 Interpretation and Aggregation

The affective input components each have their own particular data format for their output, and their own networking support. Each component therefore has a corresponding module on the receiving computer that receives these messages and transforms their content into an appropriate set of affective model scores ready for aggregation. We shall describe the workings of this for each input component in the system, as well as the aggregation mechanism that combines scores from all components.

3.2.1 EmoVoice Interpretation

As mentioned in section 3.1.1, we are using EmoVoice to separate utterances into three classes: neutral, positive-active and negative-passive. Choosing coarse, distinguishable classes improves the quality of the classification and our initial training has seen recognition rates of around 80%. This is enough for a convincing artistic representation, we do not require exact reproduction of affective states. In any case, aggregation smoothes out short-term divergences from overall trends.

The output of the component is a text string, dependent on the classification of the speech input—one of: “PositiveActive”, “Neutral” and “NegativePassive”. For networking, the receiving module connects to the EmoVoice component via a TCP connection (whose port can be configured in the component).

We characterize these classifications within the PAD model as lying on a line going from positive pleasure and arousal (positive-active) to negative pleasure and arousal (negative-passive) with neutral dominance (although the “active” portion could be seen as an indicator of interest, and therefore dominance). For the current implementation, we are not combining a large number of components, and so do not have a wide range of contributions to each dimension, we just take the extreme and middle values of the appropriate dimensions. This gives us the following PAD values: {1.0, 1.0, 0.0}, {0.0, 0.0, 0.0} and {-1.0, -1.0, 0.0}.

3.2.2 Video Feature Interpretation

For the video component, the data we are given is an identifier for each face detected, together with numerical geometric data. Early versions of the component did not provide integrated networking facilities; data was delivered on standard output. To integrate this into our networked setup we wrote a small utility that reads in the standard input (the output from the component), performs some minor textual manipulation of the output, and sends the transformed output via a UDP connection (we are more concerned about speed than reliability, so we keep up with the frame rate of the detection, rather than relying on all information being captured). We are in the process of testing a new version that integrates UDP socket communication directly.

Again, we have a continuous stream of data, but at a much greater rate than for speech utterances (one message per frame of captured video). Our model for facial interest is that if a face does not move, then interest slowly fades away, but if it is moving smoothly, that is a sign of inspection. Random movement or unreasonably-sized ellipses are discarded as errors. Head tilt is seen as showing interest, while turning the head away (making the ellipse wider) is seen as losing interest. When a new face is detected in the frame, interest is increased, and when a face leaves the frame, interest is reduced. The size of the ellipse is an indication of the closeness to the camera/installation, and moving closer or further increases interest, but if moving towards the camera, it is also interpreted as a sign of pleasure or approval, and

when moving away, displeasure. The further away a face, the faster interest is assumed to fade away.

So we have two values to keep track of: *interest* that ranges from zero or minimal interest to some maximum, and *approval*, which maps onto the Pleasure dimension of the PAD model. Approval is only updated when the size of an ellipse changes, and is set as a function of the height of the ellipse. We do not want approval to overwhelm other pleasure measurements, so we weight it in overall aggregation to only have a quarter of the effect. Interest is more involved. What we do is quantize changes of interest as a small signal. When interest is increasing or decreasing, one of these signals is assumed to have occurred. If interest rapidly increases, many signals will happen, and interest will build up. This is achieved by adding the values of the small signals, but letting the overall value reduce over time (as opposed to the simple averaging function for PAD dimensions).

3.2.3 User Experience

The user experience analysis module is still in an early stage of development, and the recognition and modelling as seen in this showcase is an ad-hoc implementation of some of the ideas and models being developed.

We are working with markers that are recognised by the AR system, and our input is the distance between the markers (based on transformed camera co-ordinates), and the orientation of the markers. There are three markers in the showcase, one of which displays the E-tree on top of it, and two others. A participant is free to move any of the three markers. They can rotate the tree marker to see all sides of the tree, and this rotation is recorded as *interest* in the tree in the same way as face movement. The other two markers can be used to send additional signals (though the participant is not necessarily told what the markers represent). One marker represents positivity and the other negativity. The relative distance between the markers and the tree marker determines an overall Pleasure value, while the average of the two distances determines a Dominance value. By moving one marker closer than the other, pleasure or displeasure is indicated, and moving the markers away from the tree indicates submissiveness, and closer indicates dominance.

3.2.4 Aggregation

There are two types of aggregation we are performing. For our main PAD model, we take discrete values, which are a target value from a particular component. To determine the new value we take the difference between the current value and the desired value, then increase or decrease the current value by a fraction of that difference (we currently use half of the difference). So, if the current value of, say, Arousal is 0.2 and we receive a PositiveActive message, this has a desired value of 1.0. The difference is therefore 0.8, so we increase the arousal value by 0.4, to give a final value of 0.6. If we get successive signals, this process is repeated as they appear. So if we start off with a Pleasure score of 0.3 and receive a desired value of 1.0 from EmoVoice and 0.4 from Video. This is weighted by their importance, so that video only matters a third as much as voice: the desired value is $((1.0 * 3) + (0.4 * 1)) / 4 = 0.85$, with more weight to EmoVoice, and the new value will be $0.3 + ((0.85 - 0.3)/2) = 0.525$.

For interest we use a slightly different method. Interest signals build up if close together to give a larger and larger signal. We use a sigmoid function—as shown in Figure 4—to map these interest

signals to PAD dimensions. Low interest has little effect, while larger values level off to a maximum value in the appropriate PAD dimension. We weight interest half as much as EmoVoice in Arousal and fully in terms of Dominance (EmoVoice does not contribute to Dominance at the moment). This then feeds into PAD aggregation as described above.

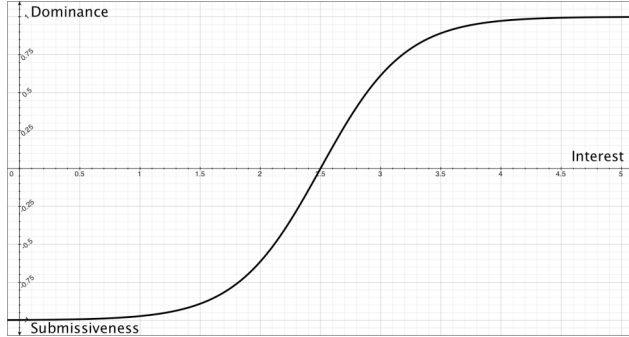


Figure 4: The sigmoid function for user interest. As interest level increases (x-axis), Dominance levels out at its maximum value (1.0).

3.3 E-Tree L-System Generation

The E-Tree has two main purposes. It interprets the current or recent state of the emotional model and it displays a history of past emotional state in the way it grows and branches, in the same way a living tree displays a history of the seasons and weather during its lifetime.

The AR component of E-Tree is implemented using the OSGART framework [9]. OSGART combines the ARToolKit [10] tracking system with the OpenSceneGraph [11] rendering engine. The tracking and therefore the positioning of the graphical overlay of the tree, is realised using physical markers with pre-defined black-and-white patterns. This allows 3D graphics to be overlaid relative to the markers, and with appropriate transformations so that perspective and size match the surrounding environment.

The E-Tree is generated using a custom L-System [4] which consists of rules that describe a recursively branching tree-like structure. Each branch segment in the tree is created from a single graphical model, modified differently depending on the properties of the branch it is part of. The tree is designed to dynamically update as the PAD model updates. Using a system of callbacks supported by the OpenSceneGraph API, the generation module tells the appropriate part of the tree to update itself when tree properties change.

The main properties of the tree that can change are growth rate, branching angle, and distribution of branches around the trunk or parent branch. The Pleasure dimension of the PAD model controls the overall growth and branching, positive values giving straight branches with regular branches, and negative values giving twisted growth with irregular, uneven branches. The Arousal dimension controls how fast the tree grows, and the “droop” of the branches, positive values give stiff branches and fast growth, negative values give “droopy” branches and slow growth. The Dominance dimension affects the thickness of branches as they grow, and also the overall size of the tree (scaling), dominant values increasing thickness and size, with submissive values producing thin branches and a smaller-sized tree.

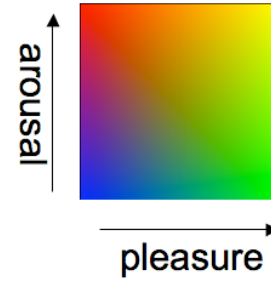


Figure 5: Colour as a combination of Pleasure and Arousal.

4. SAMPLE OUTPUT

In this section, we present samples of the visualisations produced by patterns of affective input.

4.1 Transient Emotions

The current values of the PAD model are displayed as transient tree properties that continuously update. These represent transient emotional states in contrast to longer term trends of affective response and interest.



Figure 6: A range of transient emotions, clockwise from top left: joy, anger, calm and sadness, with neutral in the centre.

The colour of the tree is based on the Pleasure and Arousal components, as show in Figure 5. This corresponds to quite natural interpretations of colour (*e.g.* anger as red, sadness as blue, joy as yellow, mellowness as green). The overall level of tropism (“droopiness”) is determined by arousal, and the thickness and overall size of the tree are determined by Dominance, as described in Section 3.2.3. Figure 6 illustrates a range of transient emotions as displayed by the E-Tree. Note that the emotions names are just labels for representative areas of the model space.

4.2 Growth as an Affective History

The growth of the E-Tree over time serves as a history of the affective input that was collected during an interactive session. Figure 7 shows a tree that was generated during a period of sustained Pleasurable and Aroused affect that also engendered high user interest. Figure 8 shows an E-Tree that was generated during a period of Displeasure and Non-arousal with less interest. Finally, Figure 9 shows an E-Tree which initially grew during a

period of positive influence (Pleasure and Arousal), and later during a period of negative influence (Displeasure and Non-Arousal).

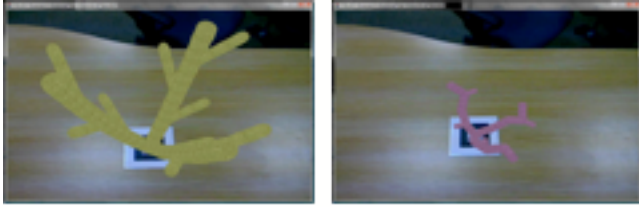


Figure 7: a) Tree growth under positive influence (left) and b) growth under negative influences (right).

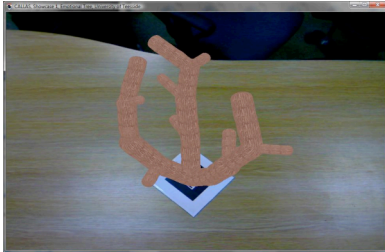


Figure 8: Tree growth under initially positive, then later negative influences.

5. CONCLUSION

While this is still work in progress we have, however, implemented various prototypes, integrating one or more components for the processing of affective modality. This work, beyond its technical application in the field of AR Art, contains several potential contributions to the field of affective interfaces. One of these consists in mapping dimensional models, not to traditional emotional categories, but to categories of user experience, such as interest and approval which are of an aesthetic nature. Another possible contribution lies in the exploration of how dimensional models can support the multimodal fusion of affective input.

We are working on integrating additional affective components into the system, as well as developing our model of user experience to capture more aesthetic properties. We are also developing a system to analyse user interactions (including through markers) in terms of these aesthetics properties.

6. ACKNOWLEDGEMENTS

We wish to acknowledge the support of the CALLAS project and contributions from CALLAS partners at VTT Electronics, the University of Augsburg, HitLABNZ and the University of Mons as well as the co-operation of Maurice Benayoun.

7. REFERENCES

- [1] Bertocini, M. and Cavazza, M., 2007. Emotional Multimodal Interfaces for Digital Media: The CALLAS Challenge. Proceedings of HCI International 2007.
- [2] Grau, O., 2002. Virtual Art: From Illusion to Immersion. MIT Press.
- [3] Picard, R. Affective Computing, MIT Press, 1997.
- [4] Przemyslaw Prusinkiewicz, Aristid Lindenmayer, 1996. *The Algorithmic Beauty of Plants*. Springer-Verlag New York, Inc.
- [5] Csikszentmihalyi, Mihaly, 1990. Flow: The Optimal Experience. Harper Perennial.
- [6] Novak, Thomas and Hoffman, Donna, 1997. Measuring the Flow Experience Among Web Users. Project 200, Vanderbilt University, Presented at Interval Research Corporation, July 2007.
- [7] Mehrabian, A. (1996). Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology: Developmental, Learning, Personality, Social*, 14, 261-292.
- [8] Thuriid Vogt and Elisabeth André. *Improving Automatic Emotion Recognition from Speech via Gender Differentiation*. In Proc. Language Resources and Evaluation Conference (LREC 2006), Genoa, Italy, 2006.
- [9] J. Looser, R. Grasset, H. Seichter, M. Billinghurst. *OSGART - A Pragmatic Approach to MR*. In Industrial Workshop at ISMAR 2006, Santa Barbara, California, USA, October 2006.
- [10] H. Kato, M. Billinghurst. *Marker Tracking and HMD Calibration for a Video-Based Augmented Reality Conferencing System*. Proceedings of the Second IEEE and ACM International Workshop on Augmented Reality (IWAR 1999), San Francisco, California, USA, pp.85-95, October 1999.
- [11] Don Burns, Robert Osfield. *Open Scene Graph*. Proceedings of the IEEE Virtual Reality 2004 (VR'04), p. 265, 2004.