An Affective Model of User Experience for Interactive Art

Stephen W. Gilroy[1], Marc Cavazza[1], Rémi Chaignon[1],
Satu-Marja Mäkelä[2], Markus Niranen[2], Elisabeth André[3], Thurid Vogt[3], Jérôme Urbain[4],
Hartmut Seichter[5], Mark Billinghurst[5], and Maurice Benayoun[6]

[1]School of Computing, University of Teesside, Middlesbrough TS1 3BA, UK;
s.w.gilroy@tees.ac.uk, m.o.cavazza@tees.ac.uk, r.chaignon@tees.ac.uk
[2]VTT Electronics, Finland, satu-marja.makela@vtt.fi, markus.niiranen@vtt.fi
[3] University of Augsburg, Germany, andre@informatik.uni-augsburg.de, thurid.vogt@informatik.uni-augsburg.de
[4] Faculté Polytechnique de Mons, Department of Electrical Engineering, Belgium; jerome.urbain@fpms.ac.be
[5] HITLabNZ, New Zealand; mark.billinghurst@hitlabnz.org, hartmut.seichter@hitlabnz.org
[6]Citu, Université Paris 1 Panthéon-Sorbonne, mb@benayoun.com

ABSTRACT
The development of Affective Interface technologies makes it possible to envision a new generation of Digital Arts and Entertainment applications, in which interaction will be based directly on the analysis of user experience. In this paper, we describe an approach to the development of Multimodal Affective Interfaces that supports real-time analysis of user experience as part of an Augmented Reality Art installation. The system relies on a PAD dimensional model of emotion to support the fusion of affective modalities, each input modality being represented as a PAD vector. A further advantage of the PAD model is that it can support a representation of affective responses that relate to aesthetic impressions.

Categories and Subject Descriptors
H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems – *augmented reality, evaluation.*

General Terms
Theory, Design.

Keywords
Affective Computing, Augmented Reality, Multimodal Interaction, Interactive Art.

1. INTRODUCTION
User experience in Digital Arts and Entertainment cannot be reduced to basic, primitive emotions. In order to capture the user's affective state it is thus necessary to instantiate a sophisticated emotional model supporting the description of multiple states and the mapping between multiple categories. In turn, in order to instantiate such a model, we need to analyse users' expressions through multiple modalities and proceed to the real-time fusion of each modality input.

Fusion of affective modalities differs significantly from multimodal fusion as traditionally described in Human-Computer Interaction [1]. Our approach to multimodal affective fusion shares some of the problems of traditional multimodal fusion, such as the integration of affective data over time in an appropriate fashion (temporal fusion). On the other hand, a specific issue with affective multimodality concerns the relationship between individual modalities and their complementarity. Here the objective is not so much to reconstruct a "message" as to produce a single representation of affective input.

2. PAD MODEL AND INTERACTIVE ART
We are using Mehrabian's Pleasure-Arousal-Dominance (PAD) model [2] as a basis for an affective model of user experience. The PAD model measures emotional tendencies and response along three dimensions: pleasure-displeasure, corresponding to cognitive evaluative judgements; arousal-nonarousal to levels of alertness and physical activity; and dominance-submissiveness to the feeling of control and influence over others and surroundings. These three dimensions are sufficient for a general description of emotions that differentiates separate basic-emotion categories, and are able to distinguish between emotions that more common two-dimensional models conflate (e.g., anger vs. fear).

The continuous nature of this dimensional model is appealing, as it allows us to model intermediate states of affect that may not have an *a priori* label/categorisation. We use this model both for the interpretation of interactions of the users (affective input), but also to aggregate and integrate such input over time to represent the affective nature of user experience. Figure 1 shows a 3D representation of the affective space of the model, and our interpretation of the affective user input as a vector within that space. As interaction progresses this vector traces out a path (the dotted line) in the affective space that represents the overall user experience, and which controls the dynamic properties of the artwork.

The continuous nature of the dimensions of this model will support fusion using analytical techniques. User input is analysed semantically for affective meaning, which is expressed in terms of the PAD dimensions. We can also utilise other underlying models of emotion that multimodal inputs employ.

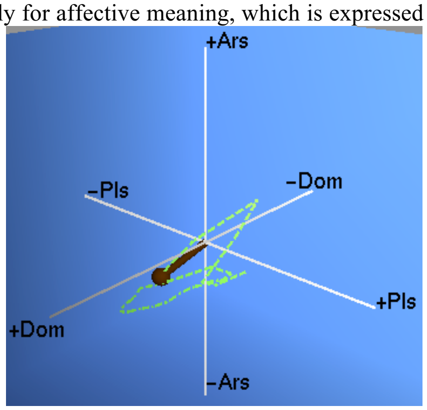

Figure 1. PAD vector representation moving through affective space.

We can map two-factor Valence-Arousal models (such as that underlying the circumplex model [3]) to Pleasure and Arousal dimensions, and also unipolar scales such as that underlying PANAS [4] (which can be mapped to two-factor bipolar models, as described in [5]).

Mehrabian [6] relates how reactive behavioural tendencies can be expressed in terms of PAD values, which supports the idea of mapping interpretation of interactions during an interactive experience to a PAD representation. In addition, the PAD model has already been used to assess user opinions and experience in the design of websites [7,8], and as the basis for computer-based agents that generate affective responses, which could applied to entertainment applications such as games [9,10].

3. E-TREE ARTWORK

E-Tree consists of a virtual tree which grows and branches in a naturalistic manner, from an initial cluster of small shoots to a larger, many-branched tree with tapering boughs and coloured leaves. The installation utilises a marker-driven Augmented Reality system, the ARToolkit, [11,12,13] that displays the naturalistic tree situated in the environment of the participants, following a "magic mirror" [14,15] paradigm for AR, using a 30' monitor.

The visual appearance of the E-Tree, a naturalistic tree structure, is defined by an L-system [16], and its growth is governed by rules that are modulated by the output of the Multimodal Affective interface. The artistic brief requires the E-tree to react to the spectators' affective response, perceived through: i) their interactions with the installation (e.g. manipulation of the AR marker serving as the E-tree base, spoken utterances aimed at the E-tree), ii) interactions between spectators (e.g. comments about the E-tree between spectators participating together), and iii) spontaneous reactions of spectators (e.g. face orientation and motion). The growth and branching of the tree serve to record a history of the user experience, as it changes over time. The "emotional" aspect of the E-tree is thus that it grows in a way that reflects *its* perception of the user response.

The speed of growth and branching of the tree are determined by Pleasure and Arousal, with negative values producing a small, stunted tree, and positive values producing a taller, bushier structure. The Dominance value determines the thickness of the branches and the size of the leaves. This is relating semantic aspects of affective description to naturalistic metaphors, as decided by the collaborating artist (MB). The colour of the leaves is also determined by a combination of Pleasure and Arousal. The effect of this is shown in figure 2.

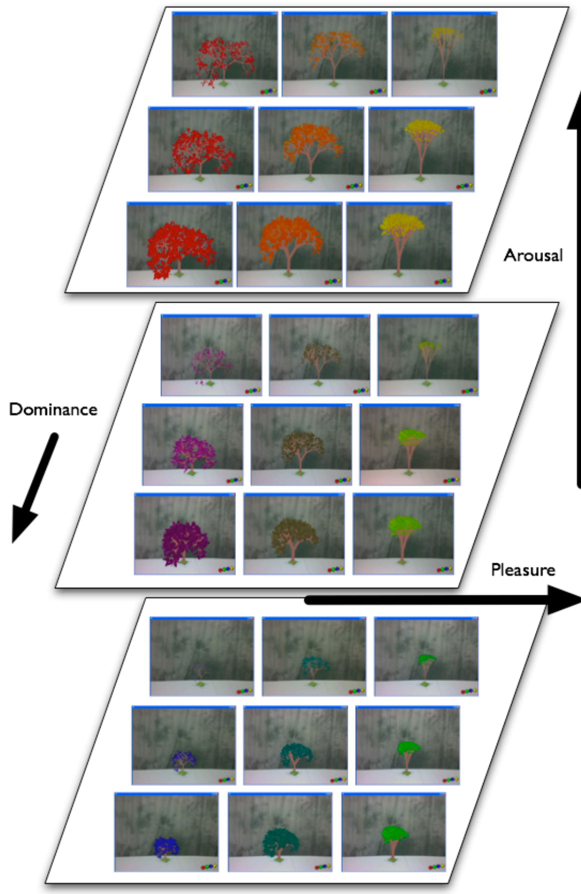

Figure 2. E-Tree growth and branching.

4. MULTIMODAL AFFECTIVE FUSION

There are two stages to our fusion approach. First, the output(s) of each input modality are mapped to a vector of PAD values. Secondly the PAD values for all modalities are combined to give a single PAD value, a point in the affective model space, which characterises the overall "mood" of an interactive experience.

Russell [17] characterises the placement of emotion terms in the circumplex model as vectors from the neutral state, and the intensity of the emotion the length the vector. We treat points in the PAD model space in the same way, so that points around the edge of the space represent intense emotions, while points closer to the centre are more neutral.

Each modality is represented a vector in the PAD space, with the direction indicating the emotional classification and the length representing the relative intensity of the emotion further weighted by a confidence score for the accuracy of recognition. For affective modalities that produce discrete output, such as emotional classification of speech utterances, such classifiers are mapped to an appropriate PAD vector.

The resulting affective state is calculated adding vectors from each modality. An example is shown in Figure 3. The red and blue vectors represent two modalities, both indicating generally positive affect. The black vector is the addition of these two vectors, representing the overall affect. The two modalities reinforce each other, so the resulting vector displays a greater intensity.

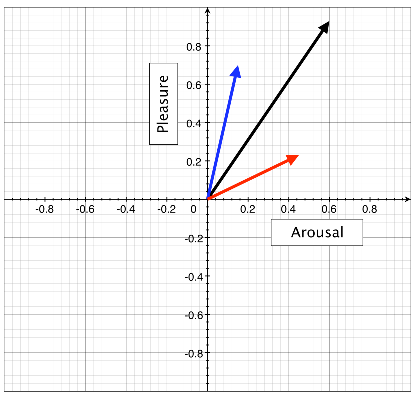

Figure 3. PAD vector summation.

When the two modalities are opposed, the addition of their vectors results in an overall vector of small magnitude, indicating a more neutral resulting affective state. This provides a way to reconcile conflicting emotional classifications from different modalities, without incurring the risk of arbitrary disambiguation.

4.1 Temporal Fusion

Affective Multimodal fusion shares with traditional multimodality the need for temporal fusion. The PAD vector representing the resulting affective state is calculated by summing the vectors for all modalities. However, the combined PAD vector is not immediately used to interpret user experience, but is combined with previous values for smoother state transitions. Over time, the vector will trace paths through the PAD model space as the resulting affective vector is modulated by on-going affective inputs, as shown in Figure 1 above.

The change in PAD vector is conceptualised as a vector between the points in the PAD space representing the old and new values. This is the direction in which the affective state is moving, and the length of this vector is the speed. We use this vector and the time since the last PAD update to determine the absolute change in PAD values.

Our PAD vector model of Multimodal Affective fusion provides a unified solution for the mapping of individual modalities, their fusion (including temporal fusion) and the exploration of complex affective states for which categorical models may not exist.

4.2 Mapping Modalities to PAD

We have derived methods of mapping a variety of modalities to PAD vectors. While the actual PAD values are specific to the implementations of affective components we are using and tailored to our example interactive artwork, we posit that these methods are applicable to alternative implementations as well as additional modalities.

As the first stage to incorporate aesthetic elements of the user experience we incorporate a higher-order modality of *interest*, which also has a PAD vector mapping in order to be integrated as an affective input, mapped to the Pleasure and Dominance elements, as we characterise interest as being independent of positive/negative judgements.

4.2.1 Speech Input

There are a number of modalities that can be derived from spoken input, such as affective interpretation of vocal features, speech understanding, and categorisation of paralinguistic speech. We have implemented PAD mapping for an emotional speech classifier and a keyword-spotting system. In our installation, speech is elicited by having spectators interact with the installation in pairs, which prompts them to comment to each other about the system, as they explore its behaviour.

Speech can be analysed for affective features, such as prosody, pitch, energy and speed and trained to sort utterances into categories where these features are clustered. We aim to match such categories to points in the PAD model space. Our speech classifier [18] has been trained with the categories *PositiveActive*, *Neutral* and *NegativePassive*, based on the theory of positive and negative affect (as used in the PANAS scales), and as mentioned earlier, a PANAS based model of affect can be mapped to a two-factor bipolar model. We apply that approach, mapping *PositiveActive* to a vector that is positive in the pleasure and arousal dimensions, *NegativePassive* to a vector that is negative in the pleasure and arousal dimensions, and *Neutral* to the zero vector.

Speech can also be analysed on a semantic level, and we use a keyword spotting system to recognise a predefined set of keywords and keyphrases from speech utterances, independent of speaker, which we sort into semantic categories, such as *speed* or *approval*. Each category has a unit PAD vector, which is scaled by the relative intensity of the meaning of the utterance. For example, in the category *Speed*, the PAD vector is along the Arousal dimensional axis (as the arousal dimension is derived from activity), and scaled as follows:

$$STOPPED\,(-0.6) < SLOW\,(-0.3) < FAST\,(0.3) < RAPID\,(0.5)$$

4.2.2 Video Analysis

There is also a variety of video analysis that can produce affective output such as recognition of Ekmanian facial expressions, gestures and full-body movements. In e-Tree, video is used to capture users' attitudes and interest, through face orientation, distance to the installation and average movements. We utilise a simple, but fairly robust face-tracking system that produces face detection and localisation, facial geometry information as well as an indication of optical flow. A representation of the optical flow of two moving faces in a video image is shown in Figure 4.

As optical flow measures movement, we interpret optical flow signals as indication of arousal. More flow indicates higher levels of arousal. We smooth the measurement of flow by taking a moving average. We combine multi-directional movement into a single optical flow vector, the magnitude of which gives us a value for the amount of movement in the video frame, and thus the level of arousal.

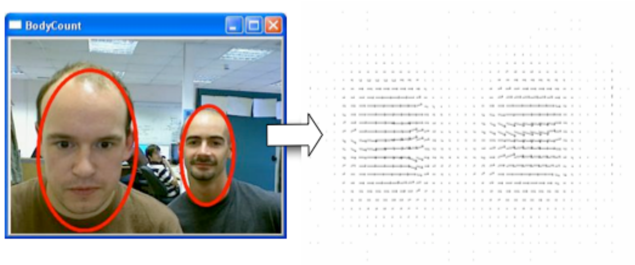

Figure 4. Analysing optical flow.

Each frame of video also generates a set of geometry details describing ellipses outlining each detected face, which can be seen overlaid onto the corresponding video input in Figure 4. The facial area is interpreted as an indication of pleasure based on an assumption that a person will come closer to the tree if they are pleased by it, and move away when displeased, and so the area of their face will change.

We consider that the more people are looking at the artwork the greater the level of interest is. Faces are tracked frame-to-frame and if a new face appears and stays in approximately the same place, it is detected as a new person and produces an increase in interest. If a face is lost for more than 10 consecutive frames it produces a decrease in interest decrease is produced, under the assumption that a person has left the frame.

5. CONCLUSIONS

Capturing user experience in a principled way is a major challenge for Digital Arts and Entertainment applications. Categories of user experience constitute a sophisticated form of affective categories, whose exploration is still at an early stage. Dimensional models have been previously mapped to existing affective categories, but could also be used to assist in the definition of novel states describing more complex user experiences. Popper, in his theorization of digital arts has suggested that in interactive digital installations, the interaction itself was a major component of a digital artwork's aesthetics [18]. In that sense being able to capture the affective content of interaction could be a way to gain insight into the specific aesthetic experience and even in the long-term to address it more directly to design more engaging installations. The approach we have introduced offers a promising framework for this exploration: whilst some of its components, such as PAD mapping, still maintain an empirical element, its principles are generic enough to be adapted to a wide range of interactive systems in support of the exploration of user experience.

6. ACKNOWLEDGMENTS

This work has been funded in part by the European Commission via the CALLAS Integrated Project. (ref. 034800, http://www.callas-newmedia.eu/)

7. REFERENCES

[1] Sharma, R. Pavlovic, V.I. Huang, T.S. 1998. Toward multimodal human-computer interface. Proceedings of the IEEE. Vol. 86(5), pp. 853-869.

[2] Mehrabian, A. 1995. Framework for a comprehensive description and measurement of emotional states. Genetic, Social, and General Psychology Monographs, 121, 339-361.

[3] Russell. J. A. 1980. A circumplex model of affect. Journal of personality and social psychology, 39. 1161-1178

[4] Watson, D., Clark, L. A., & Tellegen, A. 1988. Development and validation of brief measures of Positive and Negative Affect: The PANAS Scales. Journal of personality and social psychology, 54. 1063-1070

[5] Barrett, L.F. 1998. Independence and bipolarity in the structure of current affect. Journal of personality and social psychology, 74. 967-984

[6] 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

[7] Porat, T., Liss, R., Tractinsky, N.,2007. E-Stores Design: Influence of E-Store Design and Product type on consumers' emotions and attitudes. 12th International Conference, HCI International 2007, Beijing, China, July 22-27, 2007, Proceedings, Part IV, LNCS 4553, pp. 712-721.Springer-Verlag.

[8] Helfenstein, S. 2005. Product Meaning, Affective use and Transfer. Human Technology, 1. pp. 76-100, April 2005.

[9] Becker, C., Kopp, S., Wachsmuth, I., 2004. Simulating the Emotion Dynamics of a Multimodal Conversational Agent. Affective Dialogue Systems. Tutorial and Research Workshop, ADS 2004, Kloster Irsee, Germany, June 14-16, 2004. Proceedings. LNCS 3068, pp. 154-165. Springer-Verlag.

[10] Becker, C. Wachsmuth, I. 2006. Modeling Primary and Secondary Emotions for a Believable Communication Agent. International Workshop on Emotion and Computing, in conj. with the 29th annual German Conference on Artificial Intelligenz *(KI2006)*, Bremen, Germany, pp 31-34, 2006.

[11] J. Looser, R. Grasset, H. Seichter, M. Billinghurst. 2006. OSGART - A Pragmatic Approach to MR. In Industrial Workshop at ISMAR 2006, Santa Barbara, California, USA, October, 2006.

[12] H. Kato, M. Billinghurst. 1999. 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.

[13] Burns, D., Osfield, R.. 2004. Open Scene Graph. Proceedings of the IEEE Virtual Reality 2004 (VR'04), p. 265.

[14] Charles F., Martin O., Cavazza M., Mead S.J., Nandi A. and Marichal X., 2004. Compelling Experiences in Mixed Reality Interactive Storytelling. First International Conference on Advances in Computer Entertainment Technology, ACE 2004, Singapore, pp. 32-41.

[15] Fiala, M. ,2007. Magic Mirror System with Hand-held and Wearable Augmentations, IEEE Virtual Reality Conference, 10-14 March 2007, pp. 251-254.

[16] Prusinkiewicz, P., Lindenmayer, A.,. 1996. The Algorithmic Beauty of Plants. Springer-Verlag New York, Inc.

[17] J.A. Russell. 1989. Measures of Emotion. In *Emotion: Theory, Research and Experience*, ch.4, pp.83-111. Academic Press, Inc.

[18] Popper, F. 2007. From Technological to Virtual Art. Cambridge (Mass.), MIT Press.