Affective Interface Adaptations in the Musickiosk Interactive Entertainment Application*

L. Malatesta [(1)], A. Raouzaiou [(1)], L. Pearce [(2)], K. Karpouzis[(1)]

[(1)]Image, Video and Multimedia Systems Lab., School of Electrical and Computer Engineering, National Technical University of Athens
9, Heroon Politechniou str., 15780, Zografou, Athens, Greece
{lori, araouz}@image.ntua.gr, kkarpou@cs.ntua.gr

[(2)]XIM Ltd,
Fountain Court, 2 Victoria Square,
St.Albans, Herts, AL1 3TF, UK
laurence@xim.co.uk

Abstract. The current work presents the affective interface adaptations in the Musickiosk application. Adaptive interaction poses several open questions since there is no unique way of mapping affective factors of user behaviour to the output of the system. Musickiosk uses a non-contact interface and implicit interaction through emotional affect rather than explicit interaction where a gesture, sound or other input directly maps to an output behaviour - as in traditional entertainment applications. PAD model is used for characterizing the different affective states and emotions.

Keywords: affective interaction, adaptive interaction, interactive entertainment, PAD model.

1 Introduction

Nowadays the design of Perceptual User Interfaces (PUIs) constitutes a significant challenge for many researchers. These interfaces are characterized by interaction techniques that combine an understanding of natural human capabilities (particularly communication, motor, cognitive, and perceptual skills) with computer I/O devices and machine perception and reasoning. They seek to make the user interface more natural and compelling by taking advantage of the ways in which people naturally interact with each other and with the world, both verbal and nonverbal communications. The main directives of this research area are that devices and sensors should be transparent and passive if possible, and machines should both perceive relevant human communication channels and generate output that is naturally understood. These goals require integration of technologies such as speech

* This work was conducted in the grounds of the programme PENED 2003 – ISTERA (ΠΕΝΕΔ 2003-ΥΣΤΕΡΑ) code 03ΕΔ853 which falls within the operational programme "Competitiveness" and is co-funded from the European Union – European Social Fund by 80% and from the Greek Public Sector – Ministry of Development – General Secretariat for Research and Technology by 20%.

and sound recognition and generation, computer vision, graphical animation and visualization, language understanding, touch-based sensing and feedback (haptics), learning, user modeling, and dialog management [9]. In the application presented in this paper, the interaction is focused on what is commonly referred to as affective interaction.

Berthouze et al. [3] identify three key points to be considered when developing systems that capture affective information: embodiment (experiencing physical reality), dynamics (mapping experience and emotional state with its label) and adaptive interaction (conveying emotive response, responding to a recognized emotional state).

Current work focuses on the third point of Berthouze, that of adaptive interaction. It is considered that embodiment and dynamics are sufficiently covered by the proposed architecture and the literature that supports each of the components used in the presented interface of entertainment. Adaptive interaction nevertheless poses several open questions since there is no unique way of mapping affective factors of user behaviour to the output of the system.

The proposed interactive application was chosen mainly for its requirement to entertain while using a non-contact interface and implicit interaction through emotional affect, rather than explicit interaction where a gesture, sound or other input directly maps to an output behaviour (as in traditional entertainment applications).

2 Affective Factors and Adaptive Interfaces

In a review on adapting interaction to affective factors [6], one of the key points is that personal goals are strongly linked to affective factors. Emotions control the process of goal achievement by informing the individual whether the monitored goals are compromised or have been achieved. So how are affective goals defined and which subset of these goals is relevant to entertainment interactions? According to Ford [7] there are three different kinds of within-person consequences that a person might desire: affective goals, cognitive goals and subjective organisation goals. Affective goals represent different kinds of feelings or emotions that a person might want to experience, or avoid. A well established psychological principle states that people are “intrinsically” motivated to seek and maintain an optimal level of arousal [8]. Ford classifies entertainment goals as a subgroup of affective goals. They represent a desire to increase one’s level of arousal by doing something that is stimulating or exiting, dangerous or simply different from one’s current activity. They are also defined as goals for avoiding boredom and stressful inactivity.

When talking of art and entertainment interfaces it is important to distinguish emotions represented in both of them from affective responses to them.

- Felt emotions / human affective responses to art/ entertainment. We refer to them as aesthetic emotions. By identifying and narrowing down this set of emotions in the case of interactive entertainment we can tackle the question how these emotions can be captured and interpreted so as to allow for affective interfaces to adapt accordingly.
- Perceived emotions in art/ entertainment/ interaction scenarios. How are

affective changes/ adaptations of the environment perceived? For example, how is a change in music is perceived?

Having this distinction in mind we are trying to find the appropriate emotional model to be used in an affective interface.

A typical method of characterizing affective states and emotions is to focus on the underlying, often physiologically correlated factors and map these onto distinct dimensions. This approach leads to spatial/ dimensional models for emotions. Several such two or three dimensional sets have been proposed. Sets of emotion dimensions include "arousal, valence and dominance" (known in the literature by different names, including "evaluation, activation and power"; "pleasure, arousal, dominance"; etc.). Recent research provides additional evidence for the existence of three dimensions and suggests there should be a fourth one as well: [4] report consistent results from various cultures where a set of four dimensions is found in user studies: "valence, potency, arousal, and unpredictability". Unpredictability is the dimension that monitors reactions related with surprise and high novelty.

The Pleasure-Arousal-Dominance (P-A-D) model [1] is one of the most discussed models in the field. “Pleasure” stands for the degree of pleasantness of the emotional experience, which is typically characterized as a continuous range of affective responses extending from “unpleasant” to “pleasant”. “Arousal” stands for the level of activation of the emotion, and it is characterized as a range of affective responses extending from “calm” to “excited”. “Dominance” describes the level of attention or rejection of the emotion.

Within the PAD Model, there are eight basic and common varieties of emotion, as defined by all possible combinations of high versus low pleasure (+P and -P), high versus low arousal (+A and -A) and high versus low dominance (+D and -D). Thus, for instance, Anxious (-P+A-D) states include feeling aghast, bewildered, distressed, in pain, insecure, or upset; hostile (-P+A+D) states include feeling angry, catty, defiant, insolent, and nasty; and exuberant (+P+A+D) states include feeling admired, bold, carefree, excited, mighty, and triumphant.

In principle, if the P-A-D dimensions are continuous, this model is able to generate an infinite number of emotional states. In the case of adaptive interfaces and entertainment there is no need to account for this wide range of emotional states. It makes sense to limit the set of states to the ones related to the entertaining experience. One way of achieving that is by eliminating one of the dimensions. The literature supports such an action since the dominance dimension is useful mainly in distinguishing emotional states that have similar “pleasure” and “arousal” values. For example, the dominance dimension helps discriminate “violence” from “fear”: “violence” has P-A-D values of (−0.50,+0.62,+0.38), and “fear” has P-A-D values of (−0.64,+0.60,−0.43) [1]. In our case in the monitoring of user experience dominance is of little importance since the affective states we aim to monitor are related with the aesthetic experience. A popular model that supports such a reduction is a simplified version of the P-A-D model that only uses the P-A dimensions which was introduced by Lang [2].

3 Description of Santa Cecilia

As interaction design continues to evolve for a plethora of interface types, a question that remains open is how affective factors can be accounted for in such design approaches. Current work focuses on a design of an interface built for entertainment purposes, called Musickiosk (Fig. 1).

Musickiosk is an exhibit in the National Academy of Santa Cecilia (*Accademia Nazionale di Santa Cecilia* [12], one of the world's oldest music institutes) in Rome. It uses multiple modalities as input and synthesizes a composite output comprised of music and visual animation. MusicKiosk is a walk-through installation of four consecutive rooms. Each room is equipped with cameras that monitor user behaviour. It takes inputs from "Shelf Components", accessed via the CALLAS Framework, and uses these to drive a cartoon-based story on the screen, and to generate music [14]. The design goal is to adapt the interface to the user's affective reactions and thus deliver a more engaging and customised entertaining experience. The audience is encouraged to interact with the showcase through words, emotional speech and facial expressions. The perceived emotion of these inputs is used to influence the mood of the main cartoon character, the animated musicians and of the music generated by the kiosk.

Fig. 1. MusicKiosk: an interface built for entertainment purposes

3.1. Scenario

The storyline was developed by Paola Pacetti, a children's author based at Santa Cecilia. It is inspired by a photograph in the museum of a 'Concerto Storico' which was held in honour of queen Margherita in the late 19th century. The historical photograph is used at the beginning and end of the animation.

The kiosk story centres on a boy in the Concerto Storico orchestra. He has arrived late for the special performance and needs the end user to help him find the concert. Without any physical contact with the computer the user can direct the boy to a series of rehearsal rooms – each featuring musicians playing a particular instrument group –

and finally to the stage door itself, by speaking the numbers on the backstage doors. Within each rehearsal room, as the user expresses their mood, the music created changes accordingly. The stage door is only unlocked once the boy has entered all three rehearsal rooms, and in so doing the user will have created a unique piece of music.

XIM created a set of 2D animated characters and scenes in order to realise this story as a non-linear game-like kiosk application. Drawings were based on the characters and scenery in the historical photograph. A variety of instruments were drawn and animated, which are played in the rehearsal rooms according to the prevalent mood of the user [13].

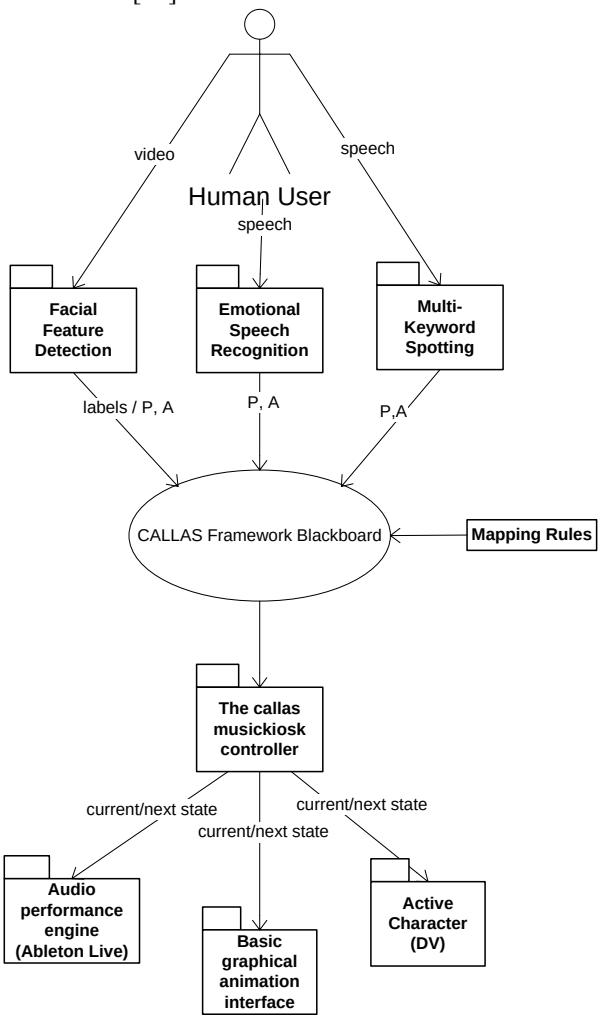

Fig. 2. MusicKiosk Architecture

As depicted in Figure 2 the user's behaviour is monitored by three components which in turn provide the input to the CALLAS Framework's blackboard. These components are a Multi-Keyword Spotting system detecting emotionally charged words and expressions, an affective speech detection system called EmoVoice([10]) and a facial feature detection component ([11]) which monitors the movement of points in the face of the user and with the help of a neural network classifies the user's expressions in a two dimensional space of activation/ evaluation.

Between blackboard and the system outputs resides the CALLAS Musickiosk controller. This is a controller application written in Java which manages all timing issues and is responsible for interpreting the blackboard events in the context of the application, mapping them to events in the audio and animation.

4 Adaptation of the Interface – State transition diagram

There is no unique way of mapping the overall pleasure, arousal values that the system calculates in each time cycle to a specific output. The idea is for the interface to adapt to the monitored dynamics of these dimensions. In order to achieve that in a systematic way we have put forward a set of rules. These rules are visualised in a state transition diagram (Fig. 3) and define the input conditions that fire specific changes in the environment. P and A stand for the two dimensions-pleasure and arousal- while the + and – are used when a predefined threshold is reached in the calculated values of these dimensions. Current design supports calibration of these thresholds such that their values are entered into a rule-mapping properties file that can be modified at run time. The rules are designed based on the fact that P-, A+ represents active disliking, P-, A- passive disliking, P+, A- serene liking and P+, A+ active liking. Interface adaptation is comprised of three actions:

- Change in the affective quality of music phrases played back
- Change in the facial expression of the virtual child
- Change in the number of instruments used to synthesise music playback.

These corresponding actions are triggered based the following rules:

- If arousal and/ or pleasure score less than the predefined threshold, neutral music phrases are played
- In the case of A+ the number of instruments is increased and A- leads to the decrease of the number of instruments used.
- In the case of significant changes in the pleasure variable the virtual child's facial expression mirrors this change (i.e. either negative, neutral or positive expressions are synthesised)
- In the case of active liking (A+, P+) joyful phrases are played
- Active disliking (A+, P-) triggers angry music phrases
- Serene/ passive liking (A-, P+) leads to sweet/ gentle music phrases being played back

- And finally if passive disliking (A-, P-) is recognised, melancholic music phrases are played.

These rules are applied in tandem meaning that the interface adapts in more ways than one depending on what changes are detected and in the current state of the system. For example if active liking is detected when neutral phrases are played and the virtual child has a neutral expression this will lead to a change in the child's facial animation to a happy one, an instrument will be added and music phrases will become joyful. In case no significant change is monitored in a cycle's duration the system remains in the same state. The overall interaction of the user with the system is bounded by a timeout limit. When the limit is reached the user is prompted to move to the next room.

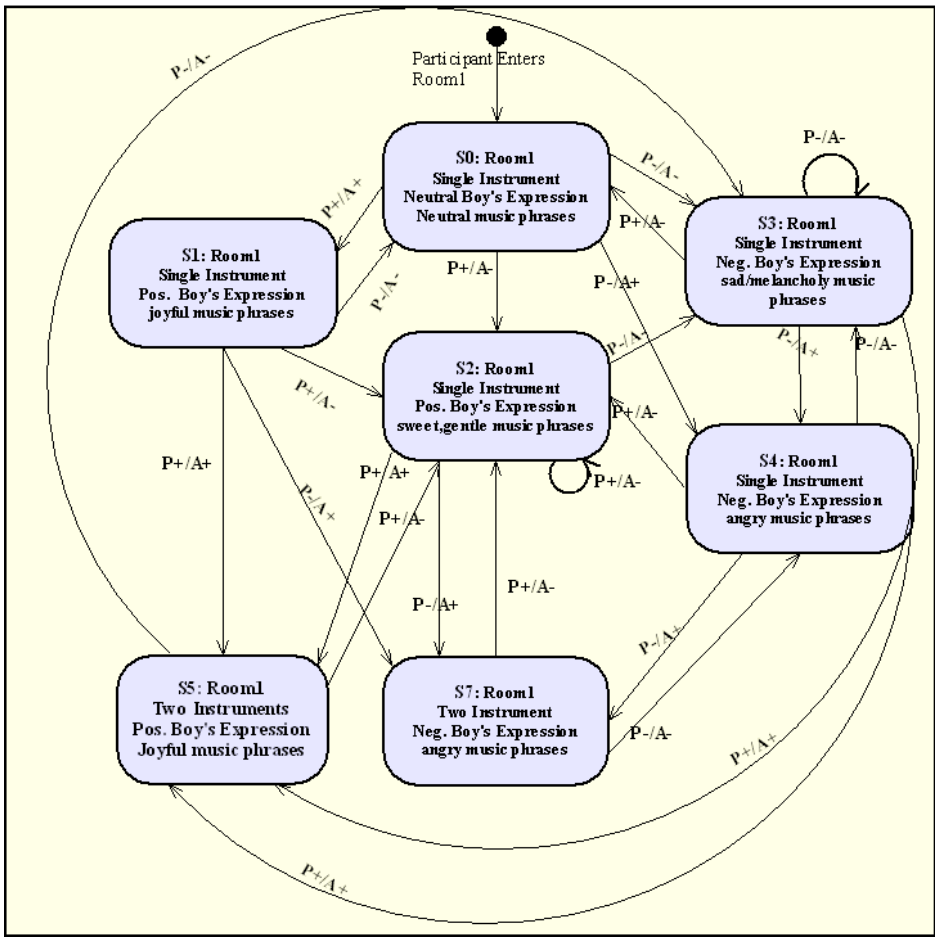

Fig. 3. State Transition diagram

5 Future work - Adaptive interface evaluation

Choices in the mapping of affective factors and adaptations of the interface remain to be tested empirically with extended user studies. Our overall aim is to enhance the user experience and provide a more natural response to perceived emotional expressions. In order to achieve this goal we plan to adopt a two fold approach on interface evaluation, one taking place during the course of the interaction and the other taking place after its completion. In the first case the user interacting with the system will be given the opportunity to evaluate interface adaptations on the fly. In the second case, detailed, time-stamped data logs generated by the kiosk will be compared with collected user feedback on the perceived environment adaptations through self report questionnaires.

References

[1] Mehrabian A., Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament, Current Psycho., vol. 14, no. 4, pp. 261–292, Dec. (1996).

[2] Lang P., Perspectives on Anger and Emotion. Lawrence Erlbaum Associates, pp. 109–134, (1993).

[3] Bianchi-Berthouze, N. and Lisetti, C. L, Modelling multimodal expression of user's affective subjective experience. *User Modelling and User-Adapted Interaction* **12**(1), 49–84 (2002).

[4] Fontaine, J., Scherer, K., Roesch, E., & Ellsworth, P. The world of emotions is not two-dimensional. Psychological Science: 18 (12), 1050-1057, (2007),

[5] Hudlicka, E. and McNeese M. Assessment of User Affective and Belief States for Interface Adaptation: Application to an Air Force Pilot Task User Modelling and User-Adapted Interaction 12: 1-47, (2002).

[6] De Rosis F., Towards Adaptation of Interaction to Affective Factors, User Modelling and User-Adapted Interaction, (2001).

[7] Ford M. E., Motivating Humans. Sage Publishing, Newbury Park, CA, 1992

[8] Berlyne D. E., Aesthetics and Psychobiology New York: Appleton-Century-Crofts, 1971.

[9] Turk, M. and Robertson, G. (2000) Perceptual user interfaces (introduction). Communications of the ACM, vol.43, pp. 32-34, ACM New York.

[10] J. Wagner, T. Vogt, E. André, "A Systematic Comparison of Different HMM Designs for Emotion Recognition from Acted and Spontaneous Speech", ACII 2007, pp. 114-125, 2007.

[11] S. Asteriadis, P. Tzouveli, K. Karpouzis, S. Kollias, "Estimation of behavioral user state based on eye gaze and head pose—application in an e-learning environment", Multimedia Tools and Applications, Springer, Volume 41, Number 3 / February, 2009, pp. 469-493.

[12] http://www.santacecilia.it/

[13] FP6 IP Callas (Conveying Affectiveness in Leading-edge Living Adaptive Systems), Contract Number IST-34800 – Music Kiosk Installation Showcase Description

[14] FP6 IP Callas (Conveying Affectiveness in Leading-edge Living Adaptive Systems), Contract Number IST-34800 – Framework Description