

PAD-based Multimodal Affective Fusion

Stephen W. Gilroy
University of Teesside
Middlesbrough, UK

s.w.gilroy@tees.ac.uk

Marc Cavazza
University of Teesside
Middlesbrough, UK

m.o.cavazza@tees.ac.uk

Marcus Niiranen
VTT Electronics
Helsinki, Finland

marcus.niiranen@vtt.fi

Elisabeth André
University of Augsburg
Augsburg, Germany

andre@informatik.uni-augsburg.de

Thurid Vogt
University of Augsburg
Augsburg, Germany

thurid.vogt@informatik.uni-augsburg.de

Jérôme Urbain
Faculté Polytechnique de Mons
Mons, Belgium

jerome.urbain@fpms.ac.be

Maurice Benayoun
CiTu, Université Paris 1
Paris, France

mb@benayoun.com

Hartmut Seichter
HITLabNZ, University of Canterbury
Christchurch, New Zealand

hartmut.seichter@hitlabnz.org

Mark Billingham
HITLabNZ, University of Canterbury
Christchurch, New Zealand

mark.billinghurst@hitlabnz.org

Abstract

The study of multimodality is comparatively less developed for Affective interfaces than for their traditional counterparts. However, one condition for the successful development of Affective interface technologies is the development of frameworks for the real-time multimodal fusion. In this paper, we describe an approach to multimodal affective fusion, which relies on a dimensional model, Pleasure-Arousal-Dominance (PAD) to support the fusion of affective modalities, each input modality being represented as a PAD vector. We describe how this model supports both affective content fusion and temporal fusion within a unified approach. We report results from early user studies which confirm the existence of a correlation between measured affective input and user temperament scores.

1. Introduction

Affective expression in humans is naturally conveyed through multiple channels, and this has been used to make the recognition of emotional categories more robust and accurate in a variety of user interfaces [21, 25, 26, 30]. However, this innate affective multimodal nature has not always

been characterised in terms of the modalities themselves, defined as input channels possessing their own semantics. This can be in part because most work on affective fusion has taken place in the context of early fusion, including the search for an “ideal” feature set across modalities [40, 35], or in the context of improved robustness and classification within a pre-defined set of affective semantics, usually based on universal emotion categories, derived from the semantics of facial expressions. [27, 42, 8].

In this paper, we describe an approach to the multimodal fusion of affective input based on the identification of interaction modalities. To support our study, we have designed an experimental platform utilising a compatible digital arts installation.

Our starting postulate is that it is possible to analyse certain forms of spectator behaviour in terms of affective modalities, and that the overall reaction of the user to the installation can be described through the fusion of individual modalities.

Affective multimodality shows both commonalities and differences with “traditional”, information-based multimodality ([5, 34, 41, 13]). One major similarity is the coincidence of linguistic expression and “physical” expression. In traditional multimodality, physical expression tends to consist mostly of deictic gestures and/or symbolic ges-

tures. In affective multimodality, physical interaction tends to be represented through a range of non-verbal affective behaviour (as affective display is primarily non-verbal [28]) such as facial expressions[14], body movement [44], posture[19, 22], and gesture [4]).

One major difference between the two types of multimodality seems to be the way in which multimodal fusion is conceived. Theoretical frameworks are more advanced for informational fusion, both in terms of how modalities complement each other [12, 34], and in terms of the identification of specific problems such as temporal fusion [32, 33, 37]. On the other hand, there is a tendency for work on affective multimodal fusion has to look at increasing robustness in classification of basic categories rather than merging complementary information across channels. We are interested in exploring the combination of affective input across modalities to better capture user experience, in particular in the field of Art and Entertainment.

If we consider the ecological conditions of interactive media, we can identify linguistic and non-verbal modalities; amongst others, comments to other users, utterances, interjections and paralinguistic speech, physical interaction and non-verbal behaviour. These would give directions to identify modalities in each of these categories. A given modality would comprise a sensing channel whose semantics is determined by the possible mapping to an emotional model.

When dealing with aesthetic experiences there is no agreement on a well-defined set of emotional categories, nor can standard approaches such as appraisal-based methods be readily applied in the absence of a goal-oriented situation. On the other hand, dimensional models offer the level of genericity compatible with values attached to an aesthetic experience (through their dimensions) and an open space for mapping continuous values.

2. PAD-based Fusion

The proposed fusion approach aims to support detection and integration of the spontaneous affective behaviour of users experiencing Arts and Entertainment. User behaviour is interpreted through a number of affective modalities, each defined in context through existing literature (see section 3). We use an interactive artwork to support our experiment, in the form of an installation which reacts to user perceived attitudes. The ‘Emotional Tree’ (e-Tree) [17] is an augmented reality graphical representation of a tree whose real-time growth patterns depend on spectators’ spontaneous attitudes.

This is illustrated by figure 1, where the interactions of spectators with an the eTree is captured by a number of modalities, which are fused as a overall representation of the emotional properties of the interactions. The installation supports multiple users, and in our experiments users were interacting in pairs: modalities have been adapted to

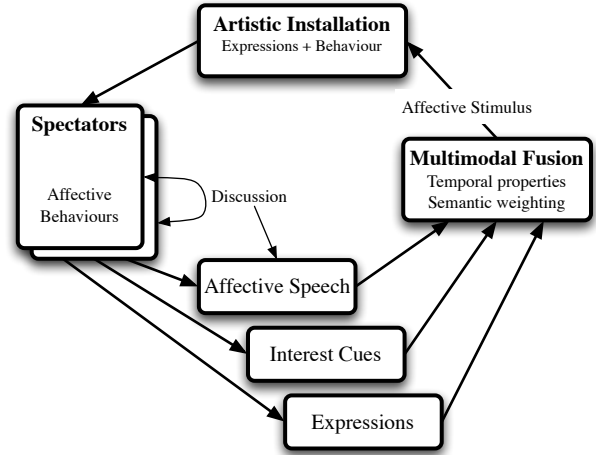


Figure 1. Affective multimodal fusion for interactive artworks.

be both subject-independent or to capture joint behaviour. The affective response is thus a measure of a global audience response.

2.1. Continuous interpretation of affect

Multimodal fusion is strongly dependent on the semantic representation used to represent modalities. In the case of affective multimodal Fusion, modalities should be represented in a unified emotional model. Continuous modelling of an affective response utilising dimensional models of emotion have been employed in emotion synthesis and visualisation, including “blended” facial expressions [2], and considerations of moods and temperaments in virtual agents [16, 11, 6]. Wassermann *et al.* [45] describe an interactive art installation that utilises a model of affect that combines both discrete emotions with a dimensional model of mood.

We wish to apply a similar continuous model, not through an appraising agent, but rather as a direct representation of the affective behaviours of spectators. This requires a way of representing and integrating *spectator* emotional information in a continuous manner. We model the affective user experience as a stimulus-response system inspired by Picard’s [38] suggested analogy of human affective response to systems with additive response with decay. Affective signals of low intensity close together add up to give a larger response, while in the absence of input, the response decays back to a baseline.

Each modality has a distinct relationship to affective events described by:

$$v(t) = ae^{-\lambda t} \quad (1)$$

where $v(t)$ is the affective response of that modality and a is the level of response at time t to an event that occurred

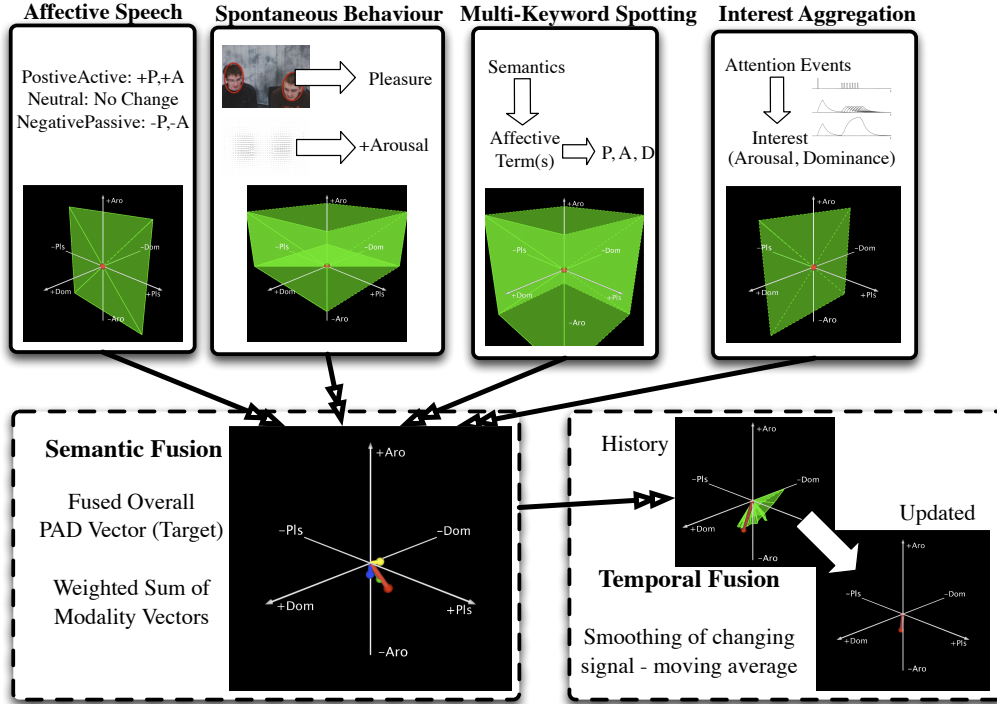


Figure 2. Framework for PAD affective fusion.

at time $t = 0$. The rate of decay is determined by λ . This allows the calculation of the contribution of that modality at an arbitrary time in the experience.

Affective interpretations of user interactions act as the stimulus signal for the model of experience, and while we do not explicitly incorporate the temperaments of individual spectators as an *a priori* measure, the patterns of response over time should reflect such temperaments (by definition).

By modelling decay, activation and saturation properties separately based on the semantics of each modality, the overall experiential response is determined by adding the weighted affective responses of each modality at any point in time. This assumes the general time-invariance of emotional reaction and response decay - effects such as habituation and mood complicate the human response, but we have ignored them in this initial model, for simplicity. The temporal nature of this integration (from the decay of an affective response) mitigates the potential conflicts between modalities, as new emotions “replace” each other. In addition in a pair/group setting, we are indeed combining potentially discordant affective reactions - if one person is dominant, it will skew the on-going representation towards them, and some modalities are likely to be involved in turn-taking behaviour between spectators (*e.g.*, speech). Additional possible “conflict” between positive and negative is dealt with the consideration of the underlying commonali-

ties between dissimilar emotional interpretations which are revealed by utilising a dimensional representation of affective state.

2.2. Using dimensional models of affect

We aim to enhance each modality with appropriate semantics of affective behaviour, but our model requires a universal characterisation of affective content. A categorical approach does not directly support such a model, as, for example, how would “anger” be changed by the addition of “surprise”? We therefore appeal to dimensional descriptions of affective state, which support continuous, universal characterisations of emotional input, that we can relate mathematically.

The most universal model would seem to be Mehrabian’s Pleasure-Arousal-Dominance (PAD) model which measures emotional tendencies and affective states along three dimensions: pleasure-displeasure, arousal-nonarousal and dominance-submissiveness. The continuous nature of such a dimensional model is appealing, as it allows us to model intermediate states of affect derived from our modalities that may not have an *a priori* label/categorisation, as well as modelling group responses and response decay.

PAD supports the integration of existing other affective representations that may be useful for a modality, including expression of discrete emotions, provided a compati-

ble interpretation of that expression can be derived (such as the mapping of linguistic descriptions of emotional state to PAD values in [39]).

Figure 2 illustrates our method of using PAD as a universal representation of the affective signals from each modality. The particular output of a modality is given a vector combination of P,A, and D values situated in a 3-D emotional space. A point in this space is the location of a particular affective state, and the on-going changes in affective reaction is represented by the movement through this space over time. The position of the overall affect at any particular time is calculated by the sum of the PAD vectors of each modality (thus summing the affective inputs as described earlier).

As shown in figure 2, different modalities will have outputs that cover different parts of the affective space, depending on the affective semantics of that modality. This will affect the possible combined affective representations possible. The combined representation is weighted in two ways. Firstly, a modality can have an interpretation that is only appropriate for one or two dimensions, in which case it is not integrated in the others. Secondly, the relevancy of a particular modality is dependent on the types of modality available. These are represented in a set of weights for each dimension in each modality, the weighting for a modality determined using input from the literature, subsequently refined through calibration experiments. The individual dimensional weightings are calculated from the relative global weightings of each modality that includes a particular dimension, though future experimentation may provide different values for the dimensions within a modality.

The overall value of the current affective representation from a set of modalities $v_0 \cdots v_i$ (taking into account dimensional weighting) with PAD vector weights of $w_0 \cdots w_i$ at time t is represented by:

$$PAD_{rep}(t) = \sum_{i=0}^n v_i(t) \cdot w_i \quad (2)$$

In order to provide a smooth transition of the changing affective stimulus for the installation and to reflect the influence of mood and historical interactions, the current affective stimulus value is not immediately set to representation vector, but rather a difference is calculated as a new direction vector, from the PAD vector at time t_0 to the new vector calculated in equation 2, which is scaled by a time-dependent function s . The function s is a sigmoid function that scales the change linearly with the amount of time passed, but with appropriate minimum and maximum. This means that the approach is smooth, but that small changes are incorporated straight away, while there is never a large “jump” if an extreme difference is calculated.

$$PAD_{diff}(t) = (PAD_{rep}(t) - PAD(t_0)) \quad (3)$$

$$PAD(t) = PAD(t_0) + (PAD_{diff}(t) \cdot s(t)) \quad (4)$$

3. Fusion Modalities

In this section we describe the modalities we have identified which could capture spontaneous affective reactions in an Art and Entertainment context. Their semantics map onto appropriate PAD dimensions, together with temporal properties to enable a continuous PAD-based output as required by our fusion method.

3.1. Spontaneous Emotional Interaction

The first step to our description of modalities is to identify non-linguistic, or physical, channels of expression. Posture has been analysed in terms of affective dimensions [23, 22], and bodily movements have previously been shown to have quantitative relationships to emotion [44]. In the first instance we have decided to limit ourselves to head and upper body motion, which is compatible with a whole range of *lean forward* installations. The corresponding sensor is a video channel performing head localization for multiple user together with orientation and size/distance estimation.

This modality is thus not concerned with the identification of gestural “utterances” (neither fixed body postures nor facial expressions) per say but rather a continuous evaluation of the affective aspects of interactions. Because of this continuous nature, this modality does not have any intrinsic temporal decay properties, as frame-by-frame updates are considered equivalent to instantaneous for our purposes. Our modalities are conceived within a framework of “active experience” [31], where participants are bodily involved in a dynamic situation (*i.e.*, a lean-forward entertainment), so engagement is a positive goal. We consider engagement as a positive reaction and withdrawal as a negative one. Berthouze et al. have shown a correlation between body movements with affective meaning and participating in engaging interaction. The notion of approach/avoidance and corresponding prototypical body movements has been shown to be related to affective experience [10, 15] and Carver [7] frames this in terms of feedback of positively and negatively valenced emotions. Hilman *et al.* [19] also found a correlation between postural leaning forwards and backwards in response to affective pictures.

Based on the above findings, we propose to interpret the posterior-anterior movement of tracked faces in terms of a continuum of pleasure-displeasure, assuming that as participants move towards and away from a central object of interaction, the approach/avoidance processes will cause them

to settle in positions that affect their affective reaction. Assuming the camera is set up at that object, this will cause approach to be seen as larger head areas, and avoidance as smaller areas. We therefore map the average area (converted to distance) of detected faces to the pleasure-displeasure affective dimension of the PAD model. Some assumptions about the average size of a face at a neutral distance are made, where the participants are close enough to view an interactive experience, but to engage would have to come closer.

We also posit that in our group situation the amount of overall gross body movements relate to arousal, and map a combined measure of activity to the arousal dimension. We derive this from measurements of optical flow in video of participants. Utilising the fact that close-up, small movements have more velocity in terms of flow than those far away from the camera, active participants in the experience have a greater influence on the measure of arousal, while still responding to interested individuals in the background.

3.2. Affective Speech Analysis

We consider integrating speech as an affective modality based not on its role as a component of recognition of a discrete emotion [9, 42, 40, 27], but as a continuous commentary of the interactive experience. This can be from multiple persons either talking to each other or directed at an object of interaction. Thus, we are not focusing on the communicative purpose of an combined affective reaction, but of integration of spontaneous speech acts over the course of an experience.

A dimensional approach has previously been used to distinguish acted emotional speech [36], and relates some of the acoustic features of speech to the PAD dimensions. However the majority of classification approaches still utilise categories of emotion. Indeed, the emotional acoustic speech classification underlying this modality [43] is also category based, and has been shown to only be sufficiently accurate at distinguishing emotional aspects of speech when applied to a small number of categories and to a whole utterance (rather than examining word-by-word prosody).

Our problem is then how to relate this to a dimensional approach and how to address the temporal relevancy of speech acts. Our solution was to reason about categories not in terms of tradition discrete emotions, but rather as regions of dimensional affective space. Pereira [36] has already demonstrated the distinctness of dimensional affective characterisation of speech (with caveats about particularly acoustic features), so we posit that making our categories characteristic combinations of PAD values they will be as valid as discrete emotional categories.

Additive approach depends on frequency of utterances, so we use it to moderate conflicting sequences of positive

and negative statements. Neutral statements are ignored, rather than pulling value back to neutral, letting signal decay naturally. PostiveActive/NegativePassive classification is based on psychological model of depression/joy, found to be most easily distinguished in speech, so we map into the Pleasure-Arousal plane of PAD model.

Our second speech modality focuses on the semantic content of utterances. It is not focussed on deictics or instructions, but rather the affective content, whether it is an appraisal, statement of feeling or motivation (without considering specific goals). We utilise exhortations (*e.g.*, “Come on!”), but do not support explicit instructions as affective interactions (though of course they still might be used in an interactive installation).

We relate words to PAD values through the use of affective lexicons, such as the linguistic references derived by Mehrabian and Russell.

The decay function instantiated for speech modalities, stands for both the transient emotional display in prosody and the small time-frame of relevancy for semantic references to an installation (although wider semantic interpretation could distinguish vocalisation of longer-term opinions/preferences).

3.3. Interest and Affective Attention

We are interested in modalities that capture extended aesthetics of interactive experience. We consider this to include notions such as curiosity, presence, novelty, engagement and flow. We suspect that capturing these kinds of response can give us insight into the unique properties of “entertainment”, with a relationship to the overall affective reaction (as described by Nakatsu *et al.* [31]).

Our first attempt at incorporating this into a affective processing modality addresses the notion of *interest*. Interest has been related to a trait-like variation of curiosity [1], so fits in with personality and affective traits used in the PAD model. We are more interested in a situational interest, where the motivational state during engagement with an interactive experience is affected by the attitudes and interactions with objects of interest [18].

We follow an approach that is similar to the “Attention Meter” [24] where we characterise certain interactions as indication of attention, adding events from different participants to get an overall rating of sustained interest. The interest we model as starting at zero and increasing upon detection of attention events. Lee *et al.* [24] describe the events that they correlate to interest in terms of head-tracking, including passing interest from heads moving rapidly across the frame, and have a model of resting attention, where stationary heads allow interest to increase, and lateral movement halts interest. We consider this more appropriate for passive interactions, as a sign of immersion, and look for more characteristic indications of interest, derived both

from head movements and speech content. We do utilise the notion of more attending persons increasing interest, and count new spectators as generators of interest, and leaving spectators as reducing interest.

We take a stationary spectator to be in a state of attentiveness, awaiting response, but not necessarily interested, so this does not generate any additional attention (after the initial registration). Head tilts and small lateral head movements generate small magnitude attention events. We also consider speech input in which we detect contextual keywords that relate to questions or instructions. Inquiring about an object or engaging directly with it through a command generates attention events. We model this attention-based situational interest as a short-term property, that is required to be sustained and decays back to zero, in a similar manner to speech-based modality input.

3.3.1 Affective Interpretation of Interest

When considering the emotional perception of content attended to, Turner and Silvia [20] show that pleasantness isn't a requirement for interest, and indeed that disturbing content can be interesting (yet not pleasant), while calming content isn't perceived as interesting. This supports the mapping of interest to the arousal dimension, but not the pleasure dimension. We contextualise interest in terms of attentional activity and the amount of engagement/control expressed in an active as opposed to passive experience. That is, interactive attention implies arousal, and also a measure of self-directed control, indicating a mapping to dominance as well.

4. Evaluation

While work still needs to be done to elicit and evaluate precise mappings within each modality, we tested our fusion approach using gross mappings (*e.g.*, that were perceived from the literature to be correct in direction if not exact magnitude), and a naive fusion configuration (with equal weighting for all modalities), as an initial evaluation of the concept. We utilised an example interactive artwork that was designed with such an approach in mind [17].

We discovered a fair amount of multimodal behaviour from participants, as shown in table 1, which also illustrates the consequences of the mappings for each modality. The Pleasure dimension was evenly split between speech and behaviour modalities, while user interaction was more important for Arousal measure, indicating a large amount of non-verbal interaction. Keywords were significantly less utilised, both due to limited vocabulary and recognition errors uncovered in the system.

We used Mehrabian's temperament scales [29] to assess user's tendencies for emotional response in each dimension of the PAD model. We wished to see if an application of

Table 1. Contribution of each modality to overall fused PAD values.

	Average	P	A	D
Affective Speech	20.95%	48.19%	10.37%	—
Spont. Behaviour	34.3%	51.08%	48.33%	—
Keyword Spotting	0.98%	0.73%	0.61%	1.16%
Interest	43.64%	—	40.69%	98.4%

our fusion model encouraged and interpreted affect that was consistent with user's tendencies. Users interacted in pairs, so one could expect the average amount of displayed affect to correlate with the average temperament score of each pair rather than the temperament of a given individual.

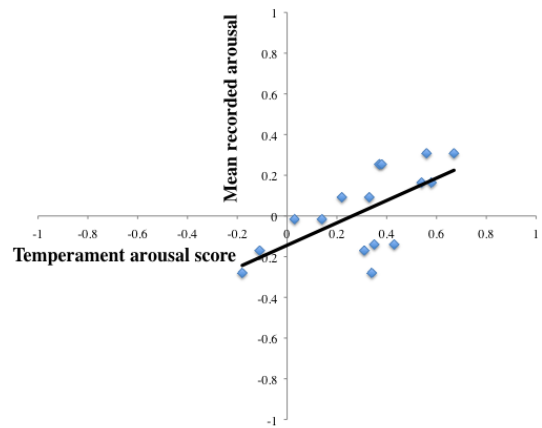


Figure 3. Correlation between fused PAD values and measured Temperament scores.

Figure 3 shows a correlation between temperament and recorded arousal values. The line of best fit shows a positive linear correlation of 0.55, covering 42% of variation, while still being statistically significant at the 5% confidence level ($p < 0.05$). p values for other dimensions were not significant.

Arousal is a dimension for which previous work has demonstrated accessibility to direct measurement via ANS responses [3]. We measured Galvanic Skin Response (GSR) to obtain an independent measure of arousal in users. Absolute values of normal GSR vary from person to person, and depend on external factors such as ambient temperature and time of day, so a baseline level of GSR was taken for each user being tested while they looked at a "neutral", unresponsive, state of the artwork without performing any interactions.

In order to compare relative GSR levels amongst all pairs, we normalised the mean GSR scores for interaction using the following formula:

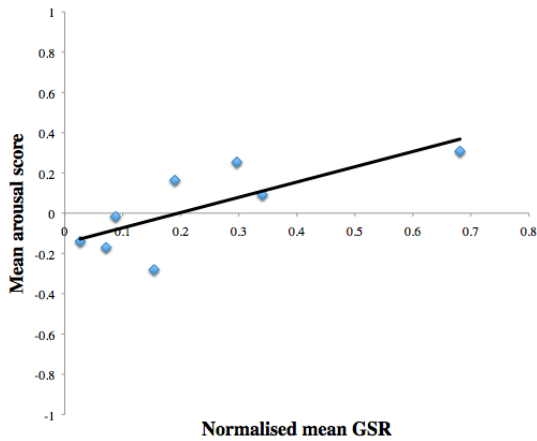


Figure 4. Correlation between GSR levels and fused PAD values.

$$GSR_{norm} = \frac{GSR_{mean} - GSR_{base}}{GSR_{max} - GSR_{min}} \quad (5)$$

Figure 4 shows the relationship between PAD-measured and interpreted levels of arousal. The coefficient of correlation (gradient) was 0.79 indicating a positive linear relationship between measured and interpreted levels of arousal. The line of best fit explains 61% of variation in arousal ($R^2 = 60.5\%$) and the gradient is statistically significant at the 5% confidence level ($p < 0.05$). Despite the small size of the sample, the correlation is encouraging and suggests that the arousal dimension is representative and its measure through Multimodal Affective fusion to be accurate.

5. Conclusion

Research in Multimodal fusion can be traced back to the observation of user interaction in the context of the emerging speech-based interfaces of the eighties [5]. In a similar fashion, a new range of application in the field of Art and Entertainment, which can benefit from capturing user emotions through various channels, calls for methods for real-time fusion of affective input across modalities [17]. We have described such a method, whose genericity is predicated on the possibility to map individual modalities onto a generic emotional model (in our case, PAD). Such a fusion method supports modality fusion without imposing assumptions on the status of modalities (e.g., complementarity or redundancy) as suggested by our evaluation of contribution across modalities. It also supports temporal fusion in a way that takes into account the specific temporal characteristics of individual inputs, thus enabling real-time fusion without temporal discontinuity. One additional benefit is that it enables multi-user input, an essential aspect of that type of application which is normally not considered in traditional research in affective interfaces.

References

- [1] M. D. Ainley. The factor structure of curiosity measures: Breadth and depth of interest curiosity styles. *Australian Journal of Psychology*, 39:53–59, 1987.
- [2] I. Albrecht, M. Schröder, J. Haber, and H.-P. Seidel. Mixed feelings: expression of non-basic emotions in a muscle-based talking head. *Virtual Reality*, 8(4):201–212, 2005.
- [3] J. Andreassi. *Psychophysiology: Human Behavior and Physiological Response*. Routledge, 2006.
- [4] N. Bianchi-Berthouze and A. Kleinsmith. A categorical approach to affective gesture recognition. *Connection Science*, 15(4):259–269, 2003.
- [5] R. A. Bolt. “put-that-there”: Voice and gesture at the graphics interface. In *SIGGRAPH '80: Proceedings of the 7th annual conference on Computer graphics and interactive techniques*, pages 262–270, New York, NY, USA, 1980. ACM.
- [6] C. Breazeal. Affective interaction between humans and robots. In *ECAL '01: Proceedings of the 6th European Conference on Advances in Artificial Life*, pages 582–591, London, UK, 2001. Springer-Verlag.
- [7] C. S. Carver. Pleasure as a sign you can attend to something else: Placing positive feelings within a general model of affect. *Cognition and Emotion*, 17(2):241–261, 2003.
- [8] G. Castellano, L. Kessous, and G. Caridakis. Multimodal emotion recognition from expressive faces, body gestures and speech. In R. C. Fiorella de Rosis, editor, *Proceedings of the Doctoral Consortium of 2nd International Conference on Affective Computing and Intelligent Interaction*, September 2007.
- [9] L. Chen, H. Tao, T. Huang, T. Miyasato, and R. Nakatsu. Emotion recognition from audiovisual information. In *Proceedings of the IEEE Workshop on Multimedia Signal Processing*, pages 83–88, 1998.
- [10] M. Chen and J. A. Bargh. Consequences of automatic evaluation: Immediate behavioral predispositions to approach or avoid the stimulus. *Personality and Social Psychology Bulletin*, 25(2):215–224, 1999.
- [11] S. K. Christian Becker and I. Wachsmuth. Simulating the emotion dynamics of a multimodal conversational agent. In *Affective Dialogue Systems*, volume 3068 of *LNCS*, pages 154–165. Springer Berlin / Heidelberg, 2004.
- [12] J. Coutaz, L. Nigay, D. Salber, A. Blandford, J. May, and R. Young. Four easy pieces for assessing the usability of multimodal interaction: The CARE properties. In *Proceedings of INTERACT'95*, pages 115–120. Chapman & Hall, 1995.
- [13] C. Duarte and L. Carriço. A conceptual framework for developing adaptive multimodal applications. In *IUI '06: Proceedings of the 11th international conference on Intelligent user interfaces*, pages 132–139, New York, NY, USA, 2006. ACM.
- [14] P. Ekman and H. Oster. Facial expressions of emotion. *Annual Review of Psychology*, 30(1):527–554, 1979.
- [15] J. Förster and F. Strack. Influence of overt head movements on memory for valenced words: A case of conceptual?motor compatibility. *Journal of Personality and Social Psychology*, 71:421–430, 1996.

- [16] P. Gebhard. Alma: a layered model of affect. In *AAMAS '05: Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, pages 29–36, New York, NY, USA, 2005. ACM.
- [17] S. W. Gilroy, M. Cavazza, R. Chaignon, S.-M. Mäkelä, M. Niranen, E. André, T. Vogt, J. Urbain, M. Billinghurst, H. Seichter, and M. Benayoun. E-tree: emotionally driven augmented reality art. In *MM '08: Proceedings of the 16th ACM international conference on Multimedia*, pages 945–948, New York, NY, USA, 2008. ACM.
- [18] S. Hidi and K. A. Renninger. Interest, a motivational variable that combines affective and cognitive functioning. In D. Y. Dai and R. J. Sternberg, editors, *Motivation, Emotion and Cognition: Integrative Perspectives on Intellectual Functioning and Development*, pages 89–115. Laurence Erlbaum Associates, 2004.
- [19] C. H. Hillman, K. S. Rosengren, and D. P. Smith. Emotion and motivated behavior: postural adjustments to affective picture viewing. *Biological Psychology*, 66:51–62, 2004.
- [20] S. A. T. Jr. and P. J. Silvia. Must interesting things be pleasant? a test of competing appraisal structures. *Emotion*, 6(4), 2006.
- [21] A. Kapoor, W. Bursleson, and R. Picard. Automatic prediction of frustration. *International Journal of Human-Computer Studies*, 65(8):724–736, 2007.
- [22] A. Kleinsmith and N. Bianchi-Berthouze. Recognizing affective dimensions from body posture. In *Affective Computing and Intelligent Interaction - ACII 2007*, volume 4738 of *LNCIS*. Springer-Verlag, 2007.
- [23] A. Kleinsmith, P. R. D. Silva, and N. Bianchi-Berthouze. Grounding affective dimensions into posture features. In *Affective Computing and Intelligent Interaction - ACII 2005*, volume 3784 of *LNCIS*. Springer-Verlag, 2005.
- [24] C.-H. J. Lee, C.-Y. I. Jang, T.-H. D. Chen, J. Wetzell, T.-H. D. Shen, and T. Selker. Attention meter: a vision-based input toolkit for interaction designers. In *CHI '06: CHI '06 extended abstracts on Human factors in computing systems*, pages 1007–1012, New York, NY, USA, 2006. ACM.
- [25] C. L. Lisetti and F. Nasoz. Maui: a multimodal affective user interface. In *MULTIMEDIA '02: Proceedings of the tenth ACM international conference on Multimedia*, pages 161–170, New York, NY, USA, 2002. ACM.
- [26] L. Maat and M. Pantic. Gaze-x: adaptive affective multimodal interface for single-user office scenarios. In *ICMI '06: Proceedings of the 8th international conference on Multimodal interfaces*, pages 171–178, New York, NY, USA, 2006. ACM.
- [27] M. Mansoorizadeh and N. M. Charkari. Bimodal person-dependent emotion recognition comparison of feature level and decision level information fusion. In *International Conference on Pervasive Technologies Related to Assistive Environments - PETRA '08*, 2008.
- [28] A. Mehrabian. *Non-verbal communication*. Aldine-Atherton, 1972.
- [29] A. Mehrabian. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology*, 14:261–292, 1996.
- [30] W. A. Melder, K. P. Truong, M. D. Uyl, D. A. Van Leeuwen, M. A. Neerincx, L. R. Loos, and B. S. Plum. Affective multimodal mirror: sensing and eliciting laughter. In *HCM '07: Proceedings of the international workshop on Human-centered multimedia*, pages 31–40, New York, NY, USA, 2007. ACM.
- [31] R. Nakatsu, M. Rauterberg, and P. Vorderer. A new framework for entertainment computing: From passive to active experience. In *Entertainment Computing - ICEC 2005*, volume 3711 of *LNCIS*. Springer-Verlag, 2005.
- [32] L. Nigay and J. Coutaz. A generic platform for addressing the multimodal challenge. In *CHI '95: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 98–105, New York, NY, USA, 1995. ACM Press/Addison-Wesley Publishing Co.
- [33] S. Oviatt and P. Cohen. Perceptual user interfaces: multimodal interfaces that process what comes naturally. *Communications of the ACM*, 43(3):45–53, 2000.
- [34] S. Oviatt, A. DeAngeli, and K. Kuhn. Integration and synchronization of input modes during multimodal human-computer interaction. In *CHI '97: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 415–422, New York, NY, USA, 1997. ACM.
- [35] M. Pantic and L. J. Rothkrantz. Towards an affect-sensitive multimodal human-computer interaction. *Proceedings of the IEEE*, 91(9):1370–1390, September 2003.
- [36] C. Pereira. Dimensions of emotion meaning in speech. In *Proceedings of the ISCA Workshop on Speech and Emotion*, pages 25–28, 2000.
- [37] N. Pflieger. Context based multimodal fusion. In *ICMI '04: Proceedings of the 6th international conference on Multimodal interfaces*, pages 265–272, New York, NY, USA, 2004. ACM.
- [38] R. W. Picard. *Affective Computing*. The MIT Press, 1997.
- [39] J. A. Russell and A. Mehrabian. Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, 11:273–294, 1977.
- [40] N. Sebe, I. Cohen, T. Gevers, and T. S. Huang. Emotion recognition based on joint visual and audio cues. In *Proceedings of the ICPR International Conference on Pattern Recognition*, pages 1136–1139, 2006.
- [41] R. Sharma, V. I. Pavlovic, and T. S. Huang. Toward multimodal human-computer interface. *Proceedings of the IEEE*, pages 853–869, May 1998.
- [42] L. C. D. Silva and P. C. Ng. Bimodal emotion recognition. In *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, pages 332–335, 2000.
- [43] T. Vogt and E. André. Improving automatic emotion recognition from speech via gender differentiation. In *Proceedings of Language Resources and Evaluation Conference (LREC 2006)*, 2006.
- [44] H. G. Wallbott. Bodily expression of emotion. *European Journal of Social Psychology*, 28:879–896, 1998.
- [45] K. C. Wassermann, K. Eng, and P. F. M. J. Verschure. Live soundscape composition based on synthetic emotions. *IEEE MultiMedia*, 10(4):82–90, 2003.