Evaluating Multimodal Affective Fusion using Physiological Signals

Stephen W. Gilroy, Marc Cavazza, Valentin Vervondel

Teesside University
Middlesbrough, UK
TS1 3BA

s.w.gilroy@tees.ac.uk, m.o.cavazza@tees.ac.uk, v.vervondel@tees.ac.uk

ABSTRACT

In this paper we present an evaluation of an affective multimodal fusion approach utilizing dimensional representations of emotion. The evaluation uses physiological signals as a reference measure of users' emotional states. Surface electromyography (EMG) and galvanic skin response (GSR) signals are known to be correlated with specific dimensions of emotion (Pleasure and Arousal) and are compared here to real time continuous values of these dimensions obtained from affective multimodal fusion. The results (both qualitative and quantitative) suggest that the particular multimodal fusion approach described is consistent with physiological indicators of emotion, constituting a first positive evaluation of the approach.

Author Keywords

Multimodal Interfaces, Affective Interfaces, Physiological Evaluation.

ACM Classification Keywords

H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems – *Evaluation/ methodology*.

General Terms

Design, Experimentation, Measurement.

INTRODUCTION

Affective interfaces rely on the detection and interpretation of emotional signals through a number of different modalities, creating new challenges in the field of multimodal interfaces. Projects such as SmartKom [44] have incorporated affective signals, such as facial expressions, as another input modality within traditional (e.g., task-based) multimodal interfaces, establishing affective input as an important component of modern multimodal interfaces [20].

Yet, even when the analysis of users' emotional states is the primary objective of an interface (rather than a sub-unit of a task-based interface), a multimodal approach is required in order to be successful. This is a logical consequence of the fact that the expression of a given affective state takes place through a range of non-verbal behaviours, including, but not limited to: posture and body motion, gestures, and affective speech parameters.

Multimodal corpora of affective expression (such as [1]) provide further evidence of the need to follow a more integrated approach. However, work dedicated to multimodal fusion in the context of affective interfaces is not as advanced as that relating to task-based multimodal systems. Most of the research to-date has investigated the fusion of low-level input signals to improve the recognition accuracy and robustness of a given emotional category [34, 43]. Little research has been dedicated to semantic approaches to multimodal affective fusion, in which a high-level affective interpretation of different modalities is integrated to capture complex affective states, even less so when these semantic representations depart from "universal emotions". Semantic considerations are a core part of traditional approaches to multimodal fusion [33], yet it is unclear whether discrete representations (in which semantic approaches are often realised) are the most appropriate for dealing with real-time affective states.

We have previously argued that multimodal affective fusion was justified at a semantic level in a range of practical applications, such as capturing representations of user experience. However, this is dependent on an appropriate underlying emotional model as a unifying representation across modalities, and we have suggested that dimensional emotion models in particular could be at the centre of the fusion mechanism [12].

The development of semantic approaches is hindered by a lack of reference models or ground truth, especially when departing from basic or "universal" emotions whose descriptive power in terms of user experience may be fairly limited. Richer methods of subjective evaluations such as FEELTRACE [41] are not employable in interactive scenarios, such as new media installations, where participants are constantly engaged with the interactive system.

This presents a major difficulty in evaluation, as multimodal semantic systems contain a number of empirical assumptions and incorporate multiple results from the affective computing literature (e.g., when defining the mapping between individual modality data and dimensional values of an emotional model).

In this paper, we evaluate a previously described multimodal affective fusion approach, described in [12], based on a dimensional model (Pleasure-Arousal-Dominance or PAD [28]), by using physiological signals which have been demonstrated to correlate strongly with the individual dimensions of this model. Our aim is to show that physiological signals associated with dimensions of Valence (Pleasure) or Arousal are consistent with the output of a PAD-based affective fusion mechanism.

RATIONALE

Traditional examples of affective interfaces have included the recognition of user emotions when engaging in multimodal dialogue with a computer [44], or measuring the affective state of a user engaged with an intelligent tutoring system [36]. More recently, there has been a growing interest in affective interfaces as a means to control computer games, by making these react to the emotional state of the user, often using physiological signals [39] or Brain-Computer Interfaces [29]. Despite their categorization as entertainment, most computer games remain task-based in essence, the player having to achieve specific objectives and complete “levels”.

However, affective interfaces have wider potential in the field of Art and Entertainment, as they can potentially be used to characterize the overall user experience, which is multimodal in nature. This type of application departs from traditional communication tasks and is one for which universal emotional categories (such as anger, joy or sadness) are insufficiently descriptive. On the other hand, dimensional models of emotion [28, 37], which support a continuous representational space, seem to offer a richer affective description of a changing user experience over time.

We have previously described a system to study how multimodal affective interfaces could be used to capture user experience, using a variety of unobtrusive channels such as video and speech [12]. Input modalities consist of both user attitudes (bodily movements and posture [17, 22], as well as more traditional non-verbal behaviour [14]), and affective content of speech utterances (considering both acoustic parameters and affective interpretation of specific keywords). These various modalities are fused via an

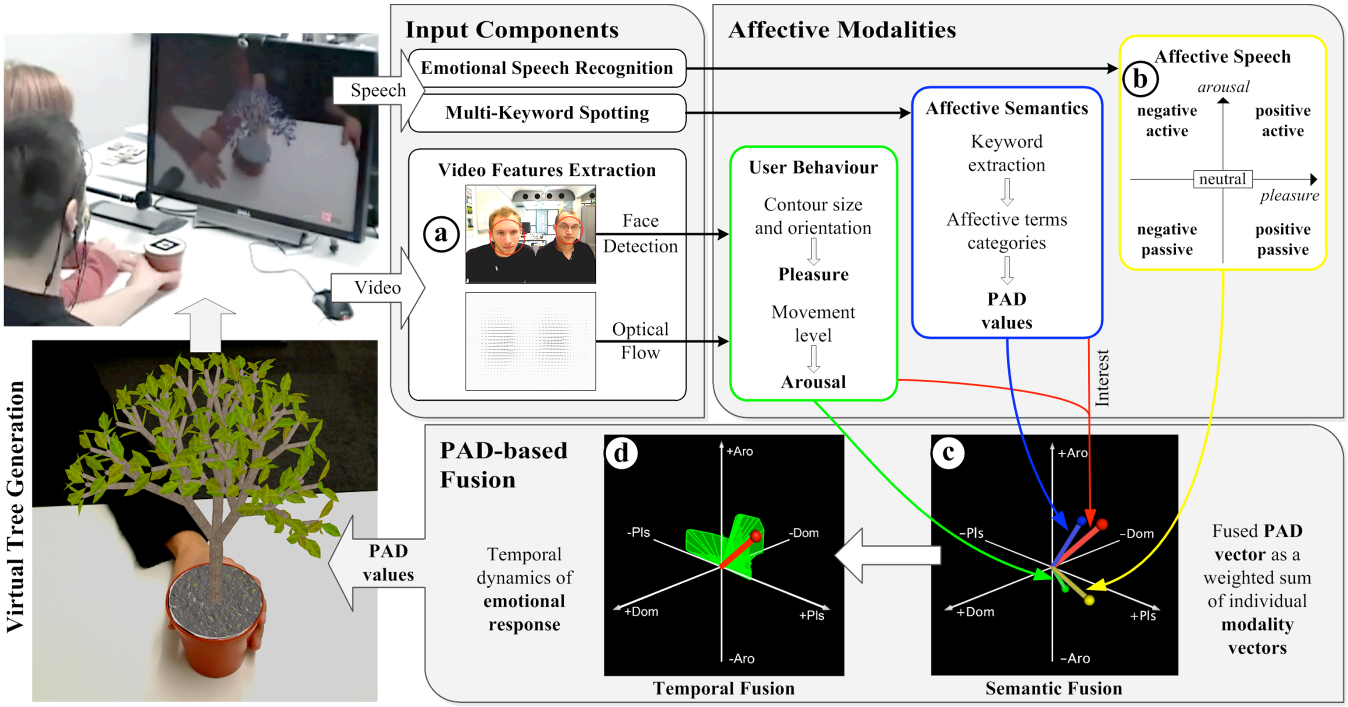

Figure 1. The affective multimodal installation. Input is gathered through cameras and microphones, and interpreted as a number of affective modalities. Upper body and face movement and orientation, affective keywords and affective speech characterizations are mapped to vectors in the PAD space. An overall vector is calculated as a weighted sum of modalities and is smoothed over time.

affective representation generic enough to capture the changing affective nature of the experience. The PAD model of Mehrabian [27] seems to provide such a representation, and there is a growing use of the PAD model in HCI research, in capturing user attitudes [25] and expressive movements [23], as well as simulating emotions in autonomous agents [11].

Interactive Installation

We have designed an interactive installation to explore the multimodal affective behaviour of users. The main requirement for such an installation is the ability to elicit sustained affective response from the users during an experimental session, expressed through multiple channels in a naturalistic manner. This has been achieved through an installation that dynamically responds to emotional input, providing an affective "feedback loop". A dynamic visualization is generated using the metaphor of a virtual tree whose growth and appearance reflect perceived user attitudes (Figure 1). Tree behaviour is not purely reactive, however, and incorporates some hysteresis in its response.

The tree does not reflect the exact pre-existing emotional state of one or other of the users, but uses affective interpretation to respond to perceived audience sentiment or aesthetic response to the installation as a whole. The interactive nature of the tree elicits responses from users and has been demonstrated to be successful throughout various experiments. This interactive approach is similar to the affective loop experiences described by Höök [18].

The installation is designed to elicit multimodal behaviour in the context of a situation that is predominantly one of observation rather than traditional communication. In other words, although the users may understand that the installation is responding, they are not engaging in task-related communication.

Through preliminary experiments, we have concluded that the best engagement is obtained with pairs of subjects rather than individuals, not least because they spontaneously communicate verbally with each other about the installation. Analysis of videos of pair interaction show that subjects spend ~90% of the time facing the display of the installation, interacting directly or commenting on it, with very little time spent interacting solely with each other.

From Input Devices to Affective Modalities

Each modality present in the installation consists of input from a sensing device (processing video or audio signals) together with affective semantics that determine the interpretation of input in terms of a dimensional model of emotion (PAD). Mapping parameters are derived from the literature and subject to experimental calibration.

For non-verbal behaviour (Figure 1a), we track interested parties through a face detection/tracking input system. Movements are related to affective dimensions via the theory of approach/avoidance [7, 17]. Approach is mapped to increasingly positive Pleasure, while avoidance is mapped to negative Pleasure. Since higher levels of movement activity are judged by observers to correspond to emotional states of higher arousal [9], we interpret the magnitude of optical flow in terms of the Arousal dimension.

Affective interpretation of speech is performed by two simultaneous analyses of audio from a microphone situated near participants. We have trained the EmoVoice system [45] to recognise characteristic speech in four quadrants of a 2-dimensional sub-set of PAD emotional space (Figure 1b). Multi-keyword spotting detects pre-defined affective keywords, whose semantic interpretation is provided as a PAD tuple, derived from a list of affective words by Russell and Mehrabian [38].

We define an additional modality to extract affective information from signs of user interest [16]. This interest measurement is characterised as an *active attentiveness*, corresponding to the Arousal and Dominance dimensions of the PAD model, and independent of the valence of a reaction. Other distinct aspects of user experience with an affective nature, such as frustration could be incorporated in a similar manner, mapped to an appropriate PAD representation.

Together, these modalities provide a complementary interpretation of behaviour in terms of the PAD emotional space. No one dimension is interpreted by a single modality, and each modality considers at least two dimensions.

PAD-based Multimodal Fusion

The PAD model has been defined as a generic representation of the emotional experience, in terms of felt emotion, affective evaluation of objects or situations and emotional temperament (a tendency to feel one emotion over another in response to a given situation). A wide range of descriptive affective labels have also been described in terms of representative values of PAD dimensions [38]. The canonical emotional meaning of stimuli can be derived from aggregation of PAD representations of measured affective responses, both in terms of the temperament of a single person and characteristic responses of multiple people.

Since the PAD space confers a unified semantics to the various affective modalities, each of which can be represented as a vector in the PAD space, we posit that fusion can be achieved through a linear combination of the individual modality vectors (Figure 1 c). The resulting overall vector characterises user affective response at a point in time. The aggregate user affective experience is then characterised by the trajectory of this vector in 3-D PAD space, over the length of an interactive session (Figure 1d). This is described in more detail in [12].

Temporal aspects play an important part in all types of multimodal fusion, and multimodal affective fusion is no

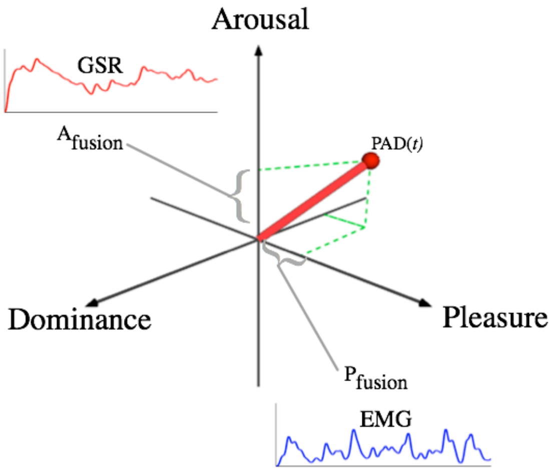

Figure 2. Using physiological signals as emotional references. Skin conductance (GSR) is compared to the projection of the fused PAD vector on the Arousal dimension (A_{fusion}), while EMG is compared to the Pleasure dimension (P_{fusion}).

exception. However, rather than the temporal patterns between modalities [32], the important aspect for affective input is the dynamics of modalities themselves. In the current system, these were inspired by a study of underlying psychophysiological processes: affective user experience incorporates feedback mechanisms, where a person's temperament and affective state affect future emotional responses [28, 35].

PAD-based multimodal fusion is also the key to utilising physiological ground truth indications of emotional feeling (through empirically discovered associations of PAD dimensions with physiological signals). The PAD model allows us to consider Pleasure, Arousal and Dominance separately as (almost) orthogonal dimensions[1], allowing comparison with separate physiological signals, supporting the evaluation of affective interpretation and fusion in terms of personal emotional response to an experience.

PHYSIOLOGICAL SIGNALS AS EVALUATION

The principle underlying our evaluation experiments is as follows: since specific physiological signals have been demonstrated to correlate strongly with individual dimensions of the PAD model, these can be used as reference values for representations based on those dimensions, within a hypothesis of orthogonality of the dimensions.

More specifically, surface electromyography (EMG) measurements of the *zygomaticus major* and *corrugator supercilii* facial muscles have been shown to be strongly correlated with positive and negative values of the P(leasure) dimension respectively [4, 14, 24]. Galvanic skin response, or skin conductance (GSR) has been shown to be strongly correlated with the A(rousal) dimension [2, 4, 24, 40].

Furthermore, because both types of signal are compatible with high frequency sampling, they can be used for dynamic value comparison, in the form of time series, over a whole experimental session. Correlation values given for session means in the studies by Lang [24] are 0.56 for *zygomaticus* EMG (zEMG), -0.90 for *corrugator* EMG (cEMG) and 0.81 for GSR (r=0.8 for GSR in [4] and r=0.81 in [40]).

While these properties have been assessed in use with discrete, separate stimuli, we seek to use them to explore the temporal dynamics of user interaction with a multimedia system. Rather than considering mean levels of activations in a period associated with an emotional stimulus, we attempt to match patterns of activation with similar patterns in constructed emotional representations from the fusion system, in the form of time series comparison, over a whole experimental session.

There does not appear to be a separate physiological correlate to the Dominance dimension, and a study by Oehme et al. [31], specifically considering PAD-based emotional representation supports this. Only very recently have new approaches to the measurement of Dominance been proposed [13].

The identified associations of physiological signals have a natural mapping onto the dimensions of the PAD model, illustrated in Figure 2. The basis of our evaluation is to compare the instantaneous value of P produced by our fusion method to the normalised instantaneous EMG value (calculated post-hoc), and similarly for A and GSR. The instantaneous value of P (and A) is produced by projecting the $PAD_{(t)}$ vector onto the P (and A) axis.

Because the fusion vector is the result of several modalities' contributions (see Figure 1c), its projection over the P and A axes retains this property. Therefore the comparison of

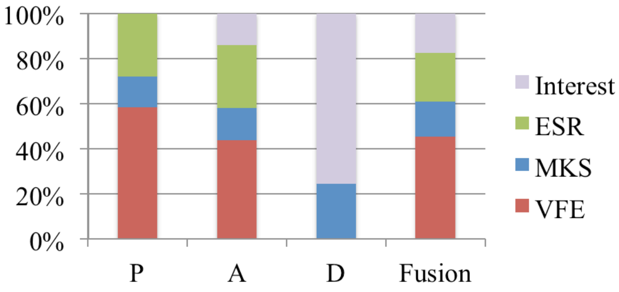

Figure 3. Contribution of each modality to each emotional dimension and to the overall fusion PAD vector. The distribution confirms the true multimodal nature of the interaction. The modalities are: Interest, Emotional Speech Recognition (ESR) by EmoVoice, Multi-Keyword Spotting (MKS) and Video Feature Extraction (VFE).

[1] Described by Mehrabian [28] as "nearly orthogonal" and showing "considerable independence".

P_{fusion} and EMG (as well as A_{fusion} and GSR) will actually evaluate the quality of fusion itself. This is further justified by the finding that i) three modalities contribute to P and all four contribute to A (Figure 3), and ii) upon inspection of the time dynamics of the various modality vectors in the PAD space, individual modality vectors all follow a rotating pattern, meaning no single one is permanently orthogonal to both the P and A axis. Figure 3 gives modality contributions in terms of percentage of aggregate P, A and D(ominance) value across all sessions, as well as the relative contribution to the overall fused PAD vector.

The evaluation compares PAD fusion data obtained for one pair of subjects to physiological signals captured on just one subject from each pair. This is justified by previous comparison of subject experience on a small sample showing strong correlation of electrophysiological responses between subjects (correlation at matched time lag: $0.48 < r < 0.73$), confirmed via subjective emotional questionnaires for all pairs ($r = 0.75$ for Pleasure, $r = 0.56$ for Arousal, $p < 0.05$). Hence, in the experiments, fusion does not combine different individual, possibly diverging, experiences into an average across a pair: quite the contrary, the pair response is strongly correlated to both individual responses. An example trace showing the high correlation between subjects in a pair is shown in Figure 4.

As subjects spend most of the time watching and interacting with the tree, the correlation cannot just be explained by the communication between subjects in a pair or possible emphatic responses between them, but rather shared reaction to the installation itself.

EVALUATION SETUP

The physical setup of the installation takes the form of an Augmented Reality (AR) system implementing a “magic mirror” paradigm using a 30-inch LCD monitor. The use of an AR approach has a number of advantages for both interaction and display. In particular it defines a zone for interaction, delimited by the video cameras used both for motion analysis and AR, and supports the inclusion of elements of tangible interfaces in the form of AR markers attached to solid objects.

The setup is shown in Figure 5. Audio was recorded via the microphone used for speech capture, and an external video camera was used to record subject interactions and generated artwork visuals. Outputs from the fusion system were also recorded, both in terms of detected affective interaction events, and the contribution to the emotional representation from each modality used.

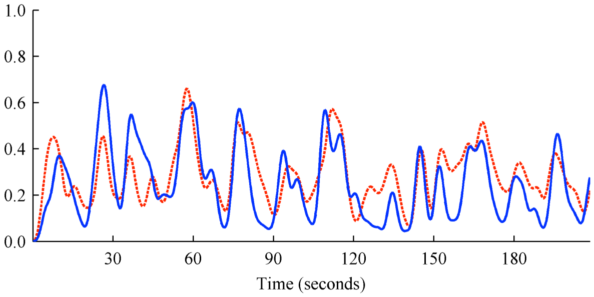

Figure 4. Correlation of zygomaticus major EMG signals between the two subjects in a pair (*r*=0.71), suggesting a shared affective experience.

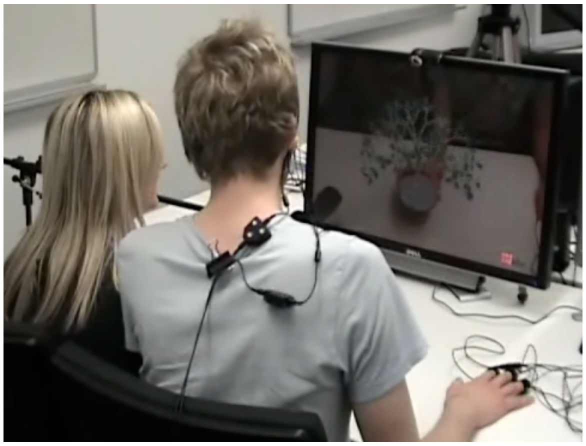

Figure 5. Experimental setup with two subjects interacting as a pair. Installation visuals are displayed on a 30-inch LCD screen, while audio and video signals are captured with the microphone and webcam (in addition to the AR camera).

Explicit instructions to subjects were kept to a minimum, to retain the exploratory aspects of the experience. Subjects were allowed to view a non-interactive session with no emotion-derived visual reactions during baseline physiological measurements. They were given an indication of the duration of the evaluation and examples of the broad range of explicit interactions understood by the system (including the fact that it would respond to ambient speech). However, users were not informed before sessions of any aspect of the underlying emotional model or the mapping between affective interpretation and installation visuals.

Pairs were allowed two sessions of around 3 minutes duration each—an empirically discovered engaging lifespan for the installation, which still allows the dynamics of interaction to play out. In particular, subjects were allowed to finish their last interaction before being interrupted.

Physiological signals were recorded using ProComp Infiniti™ data acquisition devices and sensors. GSR and EMG sensors were used, with recommended electrode placements as per Fridlund [10]. EMG sensors were placed in pairs over the specific muscles identified earlier. Ground electrodes were placed on the forehead below the hairline. EMG electrode placements are illustrated in Figure 6. EMG signals were smoothed using a 10-500Hz bandpass filter constructed in MATLAB.

21 single-wired pairs were evaluated in total, with two sessions of interaction each. This was considered a sufficient number of pairs—usability studies show effectiveness with a smaller number of subjects [19], and

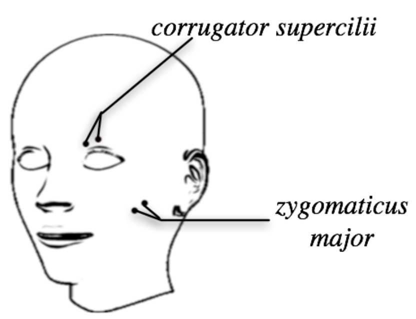

Figure 6. EMG electrode placements for the physiological measurement of the P dimension.

other typical studies of relationships between physiological signals and emotion have used between 15 and 33 [24, 6] subjects.

Of the 42 total sessions, 8 sessions encountered technical problems with the installation or the physiological sensors (mainly due to sensor artefacts or failure due to excessive movement dislodging sensors or cables). Sensor problems were identified by comparison of physiological baselines and normative reactions—artefacts were typically 4-8 times the magnitude of a maximal muscle response. Low amplitude noise and muscle movement artefacts were dealt with by smoothing the signal before analysis.

For each experimental session zEMG and GSR traces were normalized using baseline means and minima, as were the fusion-integrated Pleasure and Arousal output. Data are compared as time series with normalized values on a scale of 0.0 to 1.0.

Evaluative Principles

Although the emotional representation in the fusion system is integrative, it is still predicated on event-based emotional phenomena. We are thus interested in a match between the phasic properties of physiological signals and transitions of affective characterisations. The PAD emotional space is continuous and at any time the projection of the fusion PAD vector on the individual dimensions can be characterised as positive or negative, with a putative “neutral” region to allow for system variability/sensitivity.

While the amplitude of reconstructed PAD representations is variable across the full range of -1.0 to +1.0, the most important property for EMG signal matching is whether it is positive or negative. We allow a general region of ±0.1 before considering a signal to be representative of a positive or negative affective reaction in the Pleasure dimension.

GSR signals needs to be calibrated with a neutral value in terms of user interaction. Negative Arousal values correspond to a relaxed or non-interactive state, with a neutral value indicating a normal amount of activity (constructed in our case from head movements and neutral statements). GSR baselines were taken while asking subjects to remain still and relaxed, so GSR is compared with the full range of Arousal values, meaning the Arousal dimension was normalised in analyses to be between 0.0 and 1.0.

While both GSR and EMG have phasic properties of interest, GSR is subject to changes in tonic arousal levels. While this can be adjusted for between sessions with new baselines, it cannot be automatically accounted for during an interactive session. Arousal and GSR recordings were therefore de-trended to remove changing patterns in baseline levels, and allowing us to concentrate on phasic properties.

The fusion system and components have a certain amount of delay in interpreting input signals, while the interval between samplings of overall PAD fusion is 500ms. In addition, while physiological recording was started at the same time as fusion during evaluation, this was performed manually, so there was human error in this synchronization. In order to accurately assess correlation between physiological signals and PAD emotion values, the two signals need to be synchronised by removing these lags. This was done by assessing cross-correlation between the signals in a narrow window (±6s) and finding the peak correlation. The lag at this correlation was then used to synchronise the two time series. Example cross-correlation

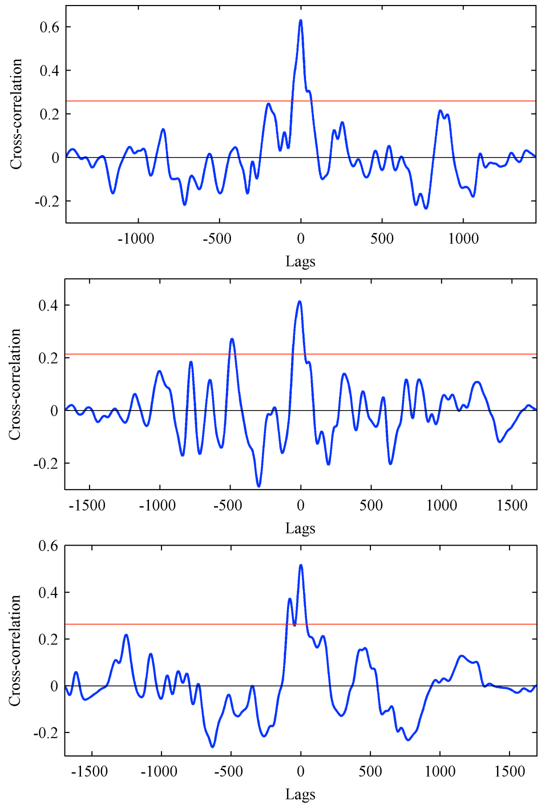

Figure 7. Cross-correlation examples for zEMG against Pleasure component of fusion showing significant correlation. Peak correlation occurs around the zero lag point, above the 95% confidence limit (horizontal line around 0.2).

peaks are shown in Figure 7. The true lag time of the system is expected to be around zero, and the mean calculated lag for experimental sessions was 18 samples, or 2250ms, with an approximately normal distribution ($-33 < lag < 47$).

There was no consistent correlation between corrugator activity (cEMG) and negative Pleasure component values from fusion within a session, while the highest correlation occurred during sessions where there was little activity overall. The particular nature of the experimental installation means that cEMG is poor at distinguishing between subjects who didn't participate or were disengaged, and those who engage in passive negative affective responses. Negative reactions manifested themselves as difficulty to engage with the installation or unease with the multimodal interface, confirmed by manual video analysis for subject pairs exhibiting the poorest correlations.

Analytical Considerations

Comparison of EMG and GSR to subject measurement of emotions while listening to music was performed by Klein [21], using cross-correlations as a measure of relatedness. An informal significance threshold for correlation of 0.35 was posited, without a formal justification, although confidence intervals are marked in the cross-correlation graphs.

Loeb et al. [26], when looking at synchronisation of muscle movements via EMG, suggest that peak cross-correlations of greater than 0.3 are non-random. They were unable to suggest a general quantitative test of significance, as their EMG signals display autocorrelation when looking at the timescale of individual muscle movements.

Miller et al. [29] looked at EMG synchronisation over longer timescales (85s), a similar order of magnitude to the study in this paper. They made an estimate of the 5% level of significance ($p < 0.05$) to be ±0.15, but did not have enough positive correlations to make a more accurate estimate, so selected correlations of above 0.25 as a conservative level of significance.

The traces in the current experiment are sampled and smoothed beyond the effects of autocorrelation present in short EMG bursts, and as they are also normalised, fit a normal distribution, indicating that classical statistical tests of significance can justly be applied to the cross-correlation [3, 5, 8].

The 5% level of significance (equivalent to a 95% confidence limit) is established for EMG cross-correlations, when comparing to other physiological data such as EEG [15] or the activity of neurons [42]. For our quantitative analysis, we therefore compare peak cross-correlations to 95% confidence limits in cross-correlation distributions. This is done individually for each session, and also taking the correlations for all sessions as a whole (possible because of the data normalisation).

RESULTS AND ANALYSIS

The strength of the correlation can be illustrated on a qualitative level by considering the event-based peaks in emotional expression in both fused PAD representation and physiological data. (Three representative examples are shown in Figure 8). It can be seen that peaks are generally aligned (taking into account variations in system lag), illustrating correlation in the nature of the signals. Qualitative examples are shown after overall lag-adjustment.

This also indicates that quantitative analysis will substantially underestimate the true correlation of the signals, due to the inherent difference in the relative magnitudes of the physiological and emotion signals. EMG magnitude shows less variance beyond a threshold of about 0.3, while PAD measures show much more variation. A possible explanation of this is that muscle signals are more discrete in nature than the continuous calculation of PAD. Details of a selection of peaks are shown in Figure 9, showing the proximity of related peaks, but also the greater variability in PAD values.

For quantitative analysis, significance tests were applied as described in the previous section, against a 95% confidence limit. Peak correlations (r) were significant within-session for zEMG in 70% of sessions, with $-0.05 < r < 0.64$ (mean

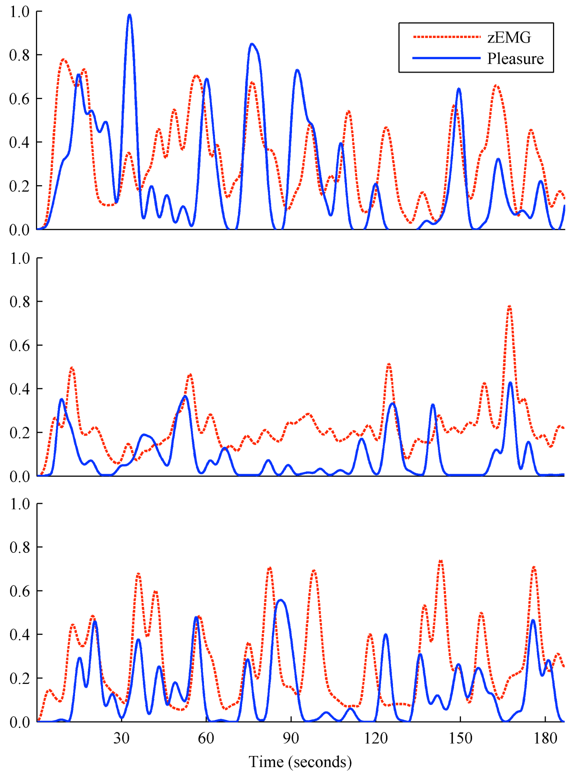

Figure 8. Qualitative analysis of the correlation between peak signals representing significance of zEMG and Pleasure component values obtained after fusion, for three sessions (*r=0.400*, *0.334* and *0.551*). This suggests a strong correlation between affective events, which is confirmed by video analysis.

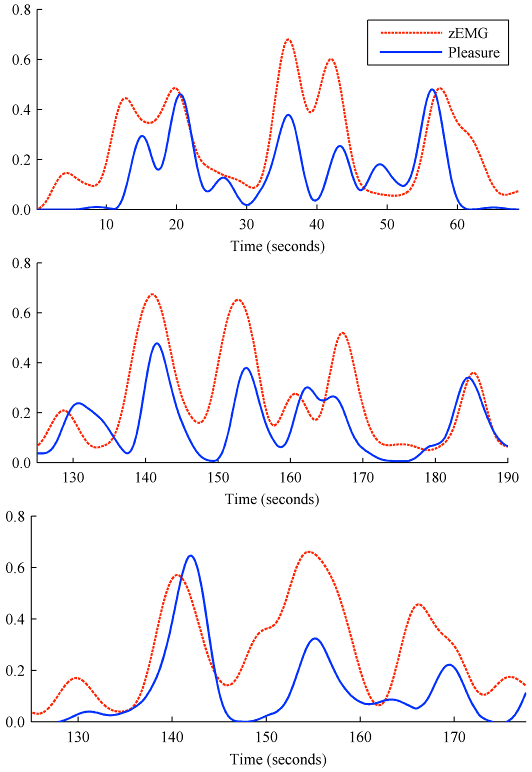

Figure 9. Qualitative analysis of peak alignment for three sessions. Peaks of Pleasure intensity, as determined by zEMG and by the Pleasure component of PAD-fused multimodal input, show very significant temporal alignment but differ in intensity, leading potentially to lower time series correlation values (see text for discussion). This is due to the intrinsic nature of EMG response.

$r = 0.315$), while 62% of sessions had peak correlation that was significant globally across sessions. For GSR, correlation was individually significant in 74% of sessions, with $0.04 < r < 0.54$ (mean $r = 0.290$), and 68% of sessions were globally significant.

This means that in the majority of cases, a significant portion of the variation in Pleasure and Arousal can be accounted for by the *continuous* physiological signal. With a 95% confidence limit, only 5% of sessions would be expected to be significant by chance. Comparing physiological and PAD data from different sessions gives no significant correlation, indicating that patterns of interaction are distinct in each session. This strengthens the argument that correlation is due to an inherent relationship between the two measures rather than an artefact of the similarity of interactive experience across all pairs. Given that correlation values underestimate the true correlation as mentioned above, a relationship between physiological signals and representative measure of emotion is well supported.

Considering the sessions that did not exhibit significant correlation, around half were very close to the confidence limit (within 0.02). An example of such a session is given in Figure 10. Video analysis of these sessions showed a small number of instances where conflicting reactions of subjects in, combined with the magnitude effects illustrated in Figure 9 could make the session potentially correlated in actuality.

We also analysed post-hoc subjective questionnaires of all pairs for which no significant correlation could be demonstrated. In a non-trivial number of cases (~18%) we found discrepancies between subjects suggesting they did not successfully share the same experience. As shared experience was one central hypothesis for our installation, this explains lack of correlation in these cases. Figure 11 shows the percentage of sessions that were correlated, uncorrelated without sufficient explanation and uncorrelated with accompanying subject differences.

CONCLUSION

The evaluation of Affective Multimodal Fusion is faced with very significant challenges due to the fragmented nature of data in literature, the richness and variability of signals that can be used as affective modalities, the rarity of multimodal corpora and their possible lack of genericity, beyond the applicative context in which they are collected.

We have proposed an approach in which real-time physiological data are used as ground truth for core dimensions of the emotional model into which Multimodal fusion takes place.

Our evaluation has comprised a qualitative element (the inspection of matched curves for fusion data and physiological signals, together with observation of subjects' videos) and a quantitative element in the form of time series correlation measures. Considering that the correlation between physiological signals and affective dimensions reported in the literature [4, 14, 26, 40] is itself not absolute, as well as the complexity of any experimental procedure involving multiple affective modalities, the high level of correlation observed over a large number of sessions is extremely encouraging, even more so if we take into account that the statistical analysis is sensitive to

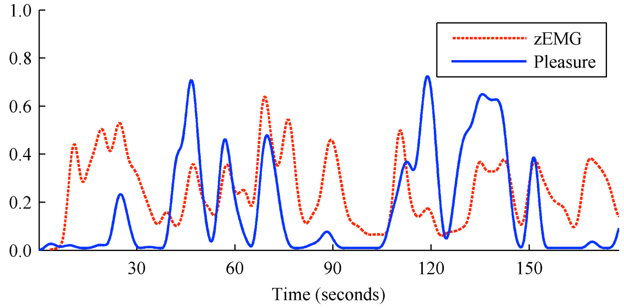

Figure 10. A plot of Pleasure intensity from zEMG against Pleasure component of PAD fusion for a non-significant session close to confidence limit. Differences in subject reactions account for the sections where Pleasure from fusion does not match zEMG,

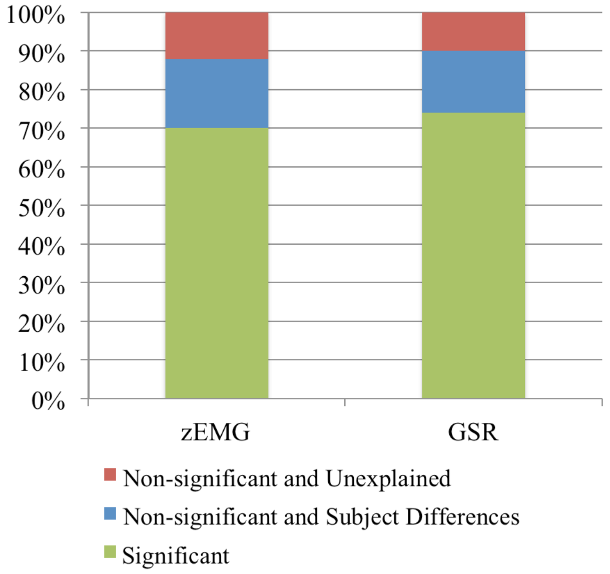

Figure 11. Illustration of percentage of sessions that were significantly correlated, showed unexplained lack of correlation, and that had a lack of correlation explained by subject differences in subjective questionnaires or video evidence.

differences in amplitude for the fusion data and the physiological signal, even when interaction events represented by P or A "spikes" are fully aligned across both curves (Figure 8).

ACKNOWLEDGEMENTS

Modality implementations were provided by members of the CALLAS project: ESR by Elisabeth André and Thurid Vogt at University of Augsburg, MKS by Jérôme Urbain at University of Mons and VFE by Markus Niiranen at VTT Electronics, Helsinki. The authors would also like to thank Jonathan Teutenberg at Teesside University for help with signal processing.

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

REFERENCES

1. Abrilian, S., Devillers, L., Buisine, S. and Martin, J.-C. 2005. EmoTV1: Annotation of Real-life Emotions for the Specification of Multimodal Affective Interfaces. HCII '05, (Las Vegas, USA, July, 2005).
2. Andreassi, J. 2006. Psychophysiology: Human Behavior and Physiological Response. Routledge.
3. Bartlett, M.S. 1935. Some Aspects of the Time-Correlation Problem in Regard to Tests of Significance. Journal of the Royal Statistical Society, 98(3), 536-543.
4. Bradley, M. M., Codispoti, M., Cuthbert, B. N. and Lang, P. J. 2001. Emotion and Motivation I: Defensive and Appetitive Reactions in Picture Processing. Emotion, 1, 3, (2001), 276-298.
5. Brockwell, P. J. and Davis, R. A. 1991. Time Series: Theory and Methods. Springer.
6. Cacioppo, J. T., Martzke, J. S., Petty, R. E. and Tassinary, L. G. 1988. Specific Forms of Facial EMG Response Index Emotions During an Interview: From Darwing to the Continuous Flow Hypothesis of Affect-Laden Information Processing. Journal of Personality and Social Psychology, 54, 4, (1988), 592-604.
7. Chen, M. and Bargh, J. A. 1999. Consequence of Automatic Evaluation: Immediate Behavioral Predispositions to Approach or Avoid the Stimulus. Personality and Social Psychology Bulletin, 25, 2, (1999), 215-224.
8. Ebisuzaki, W., 1997. A Method to Estimate the Statistical Significance of a Correlation When the Data Are Serially Correlated. Journal of Climate, vol. 10, 2147-2153.
9. Ekman, P. 1965. Differential Communication of Affect by Head and Body Cues. Journal of Personality and Social Psychology, 25, 5, (1965), 726-735.
10. Fridlund, A. J. and Cacioppo, J. T. 1986. Guidelines for Human Electromyographic Research. Psychophysiology, 23(5), 567-589.
11. Gebhard, P., ALMA – A Layered Model of Affect, Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'05), 29-36
12. Gilroy, S. W., Cavazza, M., Niiranen, M., André, E., Vogt, T., Urbain, J., Benayoun, M., Seichter, H. and Billinghurst, M. 2009. PAD-based Multimodal Affective Fusion. In Proceedings of the 3rd International Conference on Affective Computing and Intelligent Interaction (ACII 2009). IEEE Computer Society Press.
13. Glowinski, D., Coletta, P., Volpe, G., Camurri, A., Chiorri, C. and Schenone, A. 2010. Multi-scale entropy analysis of dominance in social creative activities. In *Proceedings of the International Conference on Multimedia* (MM '10). ACM, New York, NY, USA, 1035-1038.
14. Greenwald, M. K., Cook, E. W. and Lang, P. J. 1989. Affective Judgement and Psychophysiological Response: Dimensional Covariation in the Evaluation of Pictorial Stimuli. Journal of Psychophysiology, 3, (1989), 51-64.
15. Halliday, D. M., Conway, B. A., Farmer, S. F., Rosenberg, J. R., 1998. Using electroencephalography to study functional coupling between cortical activity and electromyograms during voluntary contractions in humans. Neuroscience Letters 241, 5-8.
16. Hidi, S. and Renninger, K. A. 2004. Interest: A Motivational Variable that Combines Affective and Cognitive Functioning. Motivation, Emotion and Cognition, 89-115, Lawrence Erlbaum Associates.
17. Hillman, C. H., Rosengren, K. S. and Smith, D. P. 2004. Emotion and Motivated Behaviour: Postural Adjustments to Affective Picture Viewing. Biological Psychology, 66, (2004), 55-62.

18. Höök, K. 2008. Affective Loop Experiences - What Are They? In Persuasive Technology - PERSUASIVE 2008, LNCS 5033. 1-12, Springer-Verlag.
19. Hwang, W. and Salvendy, G. 2010. Number of People Required for Usability Evaluation: The 10±2 Rule. Communications of the ACM, 53(5), 130-133.
20. Jaimes, A. and Sebe, N. 2005. Multimodal Human Computer Interaction: A Survey. In Proceedings of the IEEE International Workshop on Human Computer Interaction in conjunction with ICCV 2005.
21. Klein, M.W., 2003. Psychophysiological and Emotional Dynamic Responses to Music: An Exploration of a Two-Dimensional Model. National Conferences on Undergraduate Research.
22. Kleinsmith, A. and Bianchi-Berthouze, N. 2007. Recognizing Affecting Dimensions from Body Posture. In Affective Computing and Intelligent Interaction - ACII 2007, LNCS 4738. Springer-Verlag.
23. Lance, B., Marsella, S., 2007. Emotionally Expressive Head and Body Movement During Gaze Shifts. Intelligent Virtual Agents, 7th International Conference, IVA 2007. LNCS 4722**,** Springer, 72-85.
24. Lang, P. J. 1995. The Emotion Probe: Studies of Motivation and Attention. American Psychologist, 50(5), (1995), 372-385.
25. Liu, H., Maes, P., 2004. What would they think?: a computational model of attitudes. Proceedings of the 2004 International Conference on Intelligent User Interfaces. ACM Press, 38-45.
26. Loeb, G. E., Yee, W. J., Pratt, C. A., Chanaud, C. M., Richmond, F. J. R. 1987. Cross-Correlation of EMG Reveals Widespread Synchronization of Motor Units During Some Slow Movements In Intact Cats. Journal of Neuroscience Methods, vol. 21, 239-249.
27. Mehrabian, A. 1972. Nonverbal Communication.
28. Mehrabian, A. 1996. Pleasure-Arousal-Dominance: A General Framework for Describing and Measuring Individual Differences in Temperament. Current Psychology, 14, (1996), 261-292.
29. Miller, L.E., van Kan, P.L.E, Sinkjaer, T., Andersen, T., Harris, G. D., Houk, J.C. 1993. Correlation of Primate Red Nucleus Discharge with Muscle Activity During Free-Form Arm Movements. Journal of Physiology, vol. 469, 213-243.
30. Nijholt, A., Tan, D., Pfurtscheller, G., Brunner, C., Millán, J. R., Allison, B., Graimann, B., Popescu, F., Blankertz, B. and Müller, K. R. 2008. Brain-Computer Interfacing for Intelligent Systems. IEEE Intell. Systems, 23(3), (2008), 72-79.
31. Oehme, A., Herbon, A., Kupschick, S. & Zentsch, E. (2007). Physiological Correlates of Emotions. Proceedings of the AISB Annual Convention, April 2-4, 2007, Newcastle upon Tyne, UK.
32. Oviatt, S., Angeli, A. D. and Kuhn, K. 1997. Integration and Synchronization of Input Modes During Multimodal Human-Computer Interaction. CHI '97, (New York, NY, 1997). 415-422.ACM Press.
33. Oviatt, S. and Cohen, P. 2000. Perceptual user interfaces: multimodal interfaces that process what comes naturally. Commun. ACM 43(3), (March 2000), 45-53. DOI=10.1145/330534.330538.
34. Pantic, M. and Rothkrantz, L. J. 2003. Towards an Affect-Sensitive Multimodal Human-Computer Interaction. Proceedings of the IEEE, 91, 9, 1370-1390.
35. Picard, R. 1997. Affective Computing. MIT Press.
36. Prendinger, H., Mori, H. and Ishizuka, M. 2005. Using Human Physiology to Evaluate Subtle Expressivity of a Virtual Quizmaster in a Mathematical Game. International Journal of Human-Computer Studies, 62, 2, (2005), 231-245.
37. Russell, J. A. 1980. A Circumplex Model of Affect. Journal of Personality and Social Psychology, 39, (1980), 1161-1178.
38. Russell, J. A. and Mehrabian, A. 1977. Evidence for a Three-Factor Theory of Emotion. Journal of Research in Personality, 11, (1977), 273-294.
39. Sakurazawa, S., Yoshida, N. and Munekata, N. 2004. Entertainment feature of a game using skin conductance response. In Proceedings of ACE 2004, 181-186.
40. Sánchez-Navarro, J. P., Martínez-Selva, J. M., Torrente, G. and Román, F. 2008. Psychophysiological, Behavioral and Cognitive Indices of the Emotional Response: A Factor-Analytic Study. The Spanish Journal of Psychology, 11(1), (2008), 16-25.
41. Schröder, M., Cowie, R., Douglas-Cowie, E., Savvidou, S., McMahon, E., Sawey, M. 2000. 'FEELTRACE': An Instrument for Recording Perceived Emotion in Real Time. In Proceedings of the ISCA Workshop on Speech and Emotion: A Conceptual Framework for Research (Belfast, 2000), 19–24.
42. Schwartz, A. B., Adams, J. L. 1995. A method for detecting the time course of correlation between single-unit activity and EMG during a behavioural task. Journal of Neuroscience Methods, 58(1-2), 127-141.
43. Sebe, N., Cohen, I., Gevers, T. and Huang, T. S. 2006. Emotion recognition based on joint visual and audio cues. In Proceedings of the 18th International Conference on Pattern Recognition (ICPR 2006). 1136-1139.
44. Wahlster, W., Reithinger, N. and Blocher, A. 2001. SmartKom: Multimodal Communication with a Life-Like Character. In Proceedings of Eurospeech 2001 (Aalborg, Denmark, September, 2001). 1547-1550.
45. Vogt, T. and André, E. 2006. Improving Automatic Emotion Recognition from Speech via Gender Differentiation. In Proceedings of the Language Resources and Evaluation Conference (LREC 2006).