Analysing user physiological responses for affective video summarisation

ARTHUR G. MONEY and HARRY AGIUS*

Brunel University, School of Information Systems, Computing and Mathematics,

St John's, Uxbridge, Middlesex, UB8 3PH, UK

Abstract: Video summarisation techniques aim to abstract the most significant content from a video stream. This is typically achieved by processing low-level image, audio and text features which are still quite disparate from the high-level semantics that end users identify with (the 'semantic gap'). Physiological responses are potentially rich indicators of memorable or emotionally engaging video content for a given user. Consequently, we investigate whether they may serve as a suitable basis for a video summarisation technique by analysing a range of user physiological response measures, specifically electro-dermal response (EDR), respiration amplitude (RA), respiration rate (RR), blood volume pulse (BVP) and heart rate (HR), in response to a range of video content in a variety of genres including horror, comedy, drama, sci-fi and action. We present an analysis framework for processing the user responses to specific sub-segments within a video stream based on percent rank value normalisation. The application of the analysis framework reveals that users respond significantly to the most entertaining video sub-segments in a range of content domains. Specifically, horror content seems to elicit significant EDR, RA, RR and BVP responses, and comedy content elicits comparatively lower levels of EDR, but does seem to elicit significant RA, RR, BVP and HR responses. Drama content seems to elicit less significant physiological responses in general, and both sci-fi and action content seem to elicit significant EDR responses. We discuss the implications this may have for future affective video summarisation approaches.

Keywords: video summarisation, video content semantics, personalisation, physiological response, affective response

* Corresponding author. E-mail: harry.agius@brunel.ac.uk

1 Introduction

With the availability of digital video growing at an exponential rate, users increasingly require assistance in efficiently and effectively accessing digital video [23]. *Video summarisation* research helps to meet these needs by developing condensed versions of full length video streams by identifying and abstracting the most entertaining content within those streams. The resulting video summaries can then be integrated into various applications, such as interactive browsing and searching systems, thereby offering the user an indispensable means of managing and accessing digital video content [38, 39]. *Video summarisation techniques* produce summaries by analysing the underlying content of a video stream, condensing the content into abbreviated surrogates of the original semantic content that is embedded within the video [5]. Video summarisation research faces the challenge of developing such techniques which abstract personalised video summaries in step with what the individual user identifies with [15, 38, 60]. Given the complex, multimodal nature of video [17], recent video summarisation techniques have sought to enhance the levels of semantic abstraction and personalisation, by breaking from tradition and looking outside of the video stream, particularly to context and the user as potential sources for determining content significance.

In this paper, we concentrate on the user as a potential information source for video summarisation. Video is known to invoke strong emotional responses whilst the user views video content and user physiological responses are recognised as an effective means of measuring such emotional responses in the user [16, 18, 28, 29, 37, 43, 48, 54, 57]. Due to advances in sensor technology, user physiological responses can be measured in real time, requiring no conscious input from the user. Physiological responses provide an effective means of measuring changes in a users' affective state [3, 51], which is a generic term that refers to the user's underlying emotion, attitude, or mood at a given point in time [53]. Abstracting the affective content of video has received increased attention in recent years and several internal video content analysis techniques have been developed that attempt to automatically abstract affect-based semantics embedded within a video [13, 31, 32]. However, this is based on abstracting directly from the low-level features within the video stream, as opposed to abstracting affective-based semantics directly from the user. Indeed, affect-based semantics are now considered as a highly desirable level of semantic abstraction, that sit at the top of the semantic signification hierarchy [32]. Consequently, in this paper, we ask if data representing a user's physiological responses to video content could be used to form the basis of a video

summarisation technique, whereby the video summary presented to each individual user represents an amalgamation of the video segments from the stream that are most significant to them. To answer this question, we have undertaken a feasibility study that collected and analysed data representing a user's physiological responses to video content. Since processing of physiological response data is, however, a non-trivial task [46], a major step towards developing video summaries based on user physiological responses is to develop appropriate methods to process this data. We have therefore developed an analysis framework for effective processing and evaluation of user physiological responses to video content. Via the analysis framework, we evaluate the extent to which video content elicits significant user physiological responses from which video summaries may be derived.

The remainder of this paper is structured as follows. Section 2 reviews existing video summarisation and relevant physiological response research. In Section 3, the potential of using physiological response data as a video summarisation information source is considered. Section 4 presents the design and implementation of an experiment to collect data from users for 5 different video genres (horror, comedy, drama, sci-fi and action). Section 5 presents an analysis framework to process the users' physiological responses to video content and attach significance with regards the production of video summaries that are personal to them. In Section 6, the results of the experiment are presented. Finally, in Section 7, the significance of user's responses is evaluated and the implications discussed.

2 Existing Research

This section reviews existing relevant research within the field of video summarisation and within the field of physiological response which is relevant to video summarisation. The state of the art is then used to form the conceptual basis of our proposed approach in Section 3.

2.1 Video summarisation

Previously [42], we have surveyed video summarisation research and identified three types of techniques that can be used to generate video summaries. Figure 1 shows how these video summarisation techniques relate to video content. For the past two decades, the majority of research has focused on developing *internal video summarisation techniques*. These identify segments of video for inclusion in the video summary by analysing low-level features present within the video stream such as colour, shape, object motion, speech, or on-screen text.

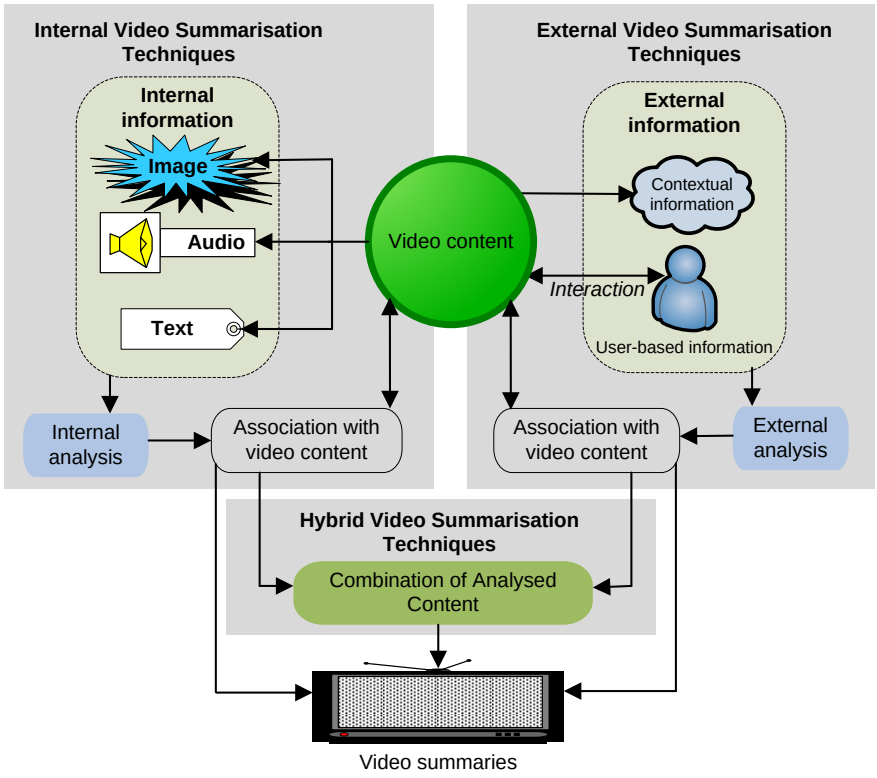

Figure 1: Internal, external and hybrid summarisation techniques.

For example, Shao et al. [52] summarise music videos automatically by using an adaptive clustering algorithm, and music domain knowledge to analyse content in the music track, while detecting and clustering shots in the video track. In contrast, motion activity, cut density and sound energy have been used to produce content-based excitement curves [30] and affect curves that probabilistically infer how the user's affective state might be changed by the video content [31]. Coicca and Schettini [12] analyse image features, in terms of the level of difference between two consecutive frames, resulting in a frame by frame measure of visual complexity. Cernekova et al. [11] present a fully automated technique that identifies shot boundaries by automated differentiation between shot fade-ins and fade-outs from pans and object and camera motion. Despite promising efforts, internal summarisation techniques intrinsically struggle to overcome the challenge of the *semantic gap* [55], that is, the disparity between the semantics that can be abstracted by analysing low-level features and the semantics that the user associates with and primarily uses to remember the content of a video. Moreover, personalised summaries (summaries that represent the most significant content to an individual user) do not tend to be achievable via the analysis of internal

information, since individual user information is not incorporated into the internal summarisation process at any stage.

External video summarisation techniques achieve more personalised video summaries by collecting and analysing information external to the video stream, notably *contextual information*, such as the time and location in which video was recorded, and *user-based information*, such as users' descriptions of content, and browsing and viewing activity. For example, Jaimes et al. [35] employ a high-level semantic analysis of basic manual annotations created by users, in conjunction with a supervised learning algorithm that determines a user's preference for particular content events based on their prior expressions of importance. Takahashi et al. [58] also use manual annotations to generate summaries of baseball videos, which include player information, key event information, e.g. 'plays of the ball', and information about the extent to which the user enjoyed specific events. Annotations are linked temporally to the original video to indicate significant events and individual players. As can be seen, external user-based information is often obtained by means of manual annotation, which is a costly and impractical solution for the end user, and consequently new external information sources are needed that can be used to develop more personalised video summaries whilst limiting required levels of user intervention.

Hybrid video summarisation techniques use a combination of internal and external summarisation techniques. For example, external techniques can compliment internal techniques by providing additional levels of detail to reduce semantic ambiguity. In one hybrid approach [2], a worn video camera captures video content, while contextual information such as location, speed and acceleration is captured from worn GPS receivers. A conversation scene detection algorithm carries out low-level analysis on the video's audio track to identify interesting segments within the captured video content for inclusion in the summary. Rui et al. [50] automatically produce video summaries by using a training set of manually annotated videos which are then propagated to un-annotated video content through similarity matching of internal information. Likewise, Zimmerman et al. [67] produce video summaries by initially analysing internal information through shot and scene segmentation. Information about the video content is then sourced from the Web. Babaguchi et al. [4] combine analysis of image features with detailed contextual information sourced from sports websites about a particular soccer game. Xu et al. [66] also summarise soccer video, they use webcasting text sourced from the Internet to achieve real time event detection of live soccer coverage.

2.2 User physiological response

Advances in sensor technology have made it possible for physiological responses to be measured in real time, without requiring any conscious input from the user, and are well documented as providing a means for measuring changes in a user's *affective state*, which is the user's underlying emotion, attitude or mood at a given point in time [53]. Affective state is made up of: *valence*, the level of attraction or aversion the user feels toward a specific stimulus, and *arousal*, the intensity to which an emotion elicited by a specific stimulus is felt. Several physiological responses have been used to infer a user's affective state:

- *Electro-Dermal Response (EDR)* measures the electrical conductivity of the skin, which is a function of the amount of sweat produced by the eccrine glands located in the hands and feet. EDR is linearly correlated with the arousal dimension, hence the higher EDR value, the higher the arousal level and vice versa [7, 25, 26, 56].
- *Respiration amplitude (RA)* can also indicate arousal and valence levels, for example slow deep breaths may indicate low arousal and positive valence which in turn can be mapped onto candidate emotions such as relaxed or blissful state. Shallow rapid breathing may indicate high arousal and negative valence emotions such as fear and panic [6, 20, 45].
- *Respiration rate (RR)* has been used as an indicator of arousal. An increase in the number of breaths the user takes per minute can be an indicator of increased arousal, whilst lower number of breaths per minute can indicate lower levels of arousal [25, 26, 44].
- *Blood Volume Pulse (BVP)* measures the extent to which blood is pumped to the bodies extremities. This can serve as a measure of a user's valence, for example when a user is under duress or experiencing fear, blood flow is restricted in the user's extremities and, conversely, when a user is experiencing joy, the amount of blood pumped to the extremities is increased [10, 21, 33, 46, 49, 63].
- *Heart Rate (HR)* acceleration and deceleration has also been shown to be an indicator of valence [9]. Negative valence is signified by a greater increase in HR than positive valence [20, 27, 56, 61, 65].

Physiological response data is normally recorded in time series format, i.e. the data is a continuous data source recorded in temporal order. Maintaining the time evolving element when analysing physiological response data has

a key advantage over conventional summative assessment methods, such as questionnaires and interviews, because it gives rare insight into the granular shifts that occur in the user's affective state, which may not be measurable by other methods [36]. The drawback, however, is that it is a non-trivial task and time consuming to analyse said data effectively [46]. Some research does, however, capitalise on the time-evolving nature of physiological data. Scheirer et al. [51] develop and apply Hidden Markov Models (HMMs) to identify and model the time series data collected from multiple users interacting with a particularly slow computer game interface to identify periods of user frustration. Picard et al. [47] developed algorithms aid automatic feature recognition of one user's artificially induced physiological response over the course of 20 days. Ward & Marsden [64] monitored user physiological responses to a variety of Web pages, variance of physiological response was presented as percentage change from baseline, which indicated changes in user response over time. Although some existing affective computing studies offer useful insight into approaches for abstracting features from physiological data, the research presented in this field is still of an exploratory nature and methods are tailored to the specific needs of the research task in question.

3 Physiological responses for external video summarisation

It is known that video content elicits the above physiological responses in the user (i.e. EDR, RA, RR, BVP, HR) [8, 16, 37]. In recent years, it has become possible to measure users' physiological responses by means of wireless wearable sensors such as the SenseWear armband from BodyMedia. As a result, some multimedia metadata standards now allow for some limited affective description [1, 41] and, as was seen in Section 2.1, some internal video summarisation techniques have been developed that summarise video streams based on their affective content. With regards to processing and evaluating physiological response data for the production of video summaries, however, to the best of our knowledge there appears to be no research to date.

Consequently, in this study, we consider whether users' physiological responses may serve as a suitable external source of information for producing individually-personalised affective video summaries (via an external or hybrid technique). User responses are likely to be most significant during the segments of a video stream that have most relevance to that user, since these will tend to be the segments that have the most impact and are the most memorable; hence, it is these segments that are the foremost candidates for inclusion within a summarised version of the video stream. Figure 2 shows how physiological responses may be used within an external or hybrid

summarisation technique to produce personalised affective video summaries. It should be noted, however, that currently it is infeasible to map physiological responses onto a full range of specific emotions [51]. Hence we do not aim to summarise the emotions within a video, but rather evaluate the extent to which specific segments of video elicit significant physiological responses in the user and then use this information to establish the video segments that are likely to be of interest to the user and consequently included in the video summary.

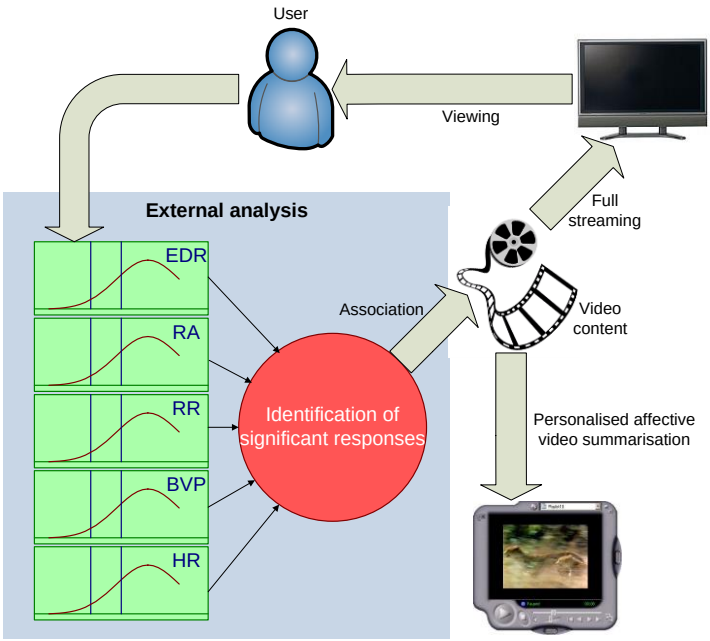

Figure 2: Physiological responses for an external video summarisation technique.

Initially, the user views the full video stream while physiological responses are captured and measured, and the most significant responses identified. After viewing the video content, the temporal locations of these significant video segments are then associated with the viewed video content and the video summary is created. The output is a personalised video summary that incorporates the video segments that elicited the most significant physiological responses in the user during viewing. This approach is specifically formulated to be applicable within the context of real-time viewing of video content, where analysis of user physiological responses is taken to directly reflect the user's personal experience of the video content at the time of viewing.

As a step towards establishing the feasibility of developing such affective video summaries, an analysis framework for the processing and evaluation of physiological responses to video content has been developed. In order to develop and demonstrate the framework, an experiment was designed and carried out to collect appropriate

physiological response data relating to a variety of video content. The design and implementation of the experiment is outlined in the next section. In Section 5, a framework for the processing and evaluation of user physiological responses to video content is presented.

4 Experiment for physiological data collection

We designed and carried out an experiment to collect user physiological response data to a range of video content including horror, comedy, drama, sci-fi and action. The data was then used to develop an appropriate technique to process and evaluate the users' physiological responses to the most pertinent sub-segments of each video, and evaluate the extent to which users respond significantly to these sub-segments. We used the five measures of a users' physiological response outlined in the previous section: EDR, RA, RR, BVP and HR.

4.1 Users participating in the study

A total of ten users took part in the experiment, with a mean age of 26. Six of the users were male. All users reported to have good eyesight and hearing and were all non-smokers. All users considered themselves to be experienced viewers of film, television and video.

4.2 Materials

To test a wide range of representative video content that is contemporary, popular, video content was chosen from both film and television domains, and from five different genres. Three films were chosen: *The Matrix* [59], *The Exorcist* [22], and *Lost in Translation* [14]. All three films were chosen from the IMDb (Internet Movie Database) Top 250 Films [34]. Two award-winning television programs were also chosen: *Lost* [24], and *Only Fools and Horses* [19].

Physiological data was recorded using the ProComp Infiniti system and BioGraph software produced by Thought Technologies. EDR was measured with a Skin Conductance Flex/Pro Sensor (SA9309M) connected to the middle phalanxes of the index and ring fingers. This sensor outputs a measure of skin conductance in micro-Siemens (μS).

Respiration was measured with a Respiration Sensor (SA9311M) which is an elasticised band connected around the thorax. The sensor outputs a relative measure of stretch. RR is calculated by measuring the peak to peak

interval of the stretch measure, output as the number of inter-beat intervals per minute. RA is measured by calculating the difference between the highest and lowest points in one breath. Figure 3 demonstrates how these relate to the raw respiration signal.

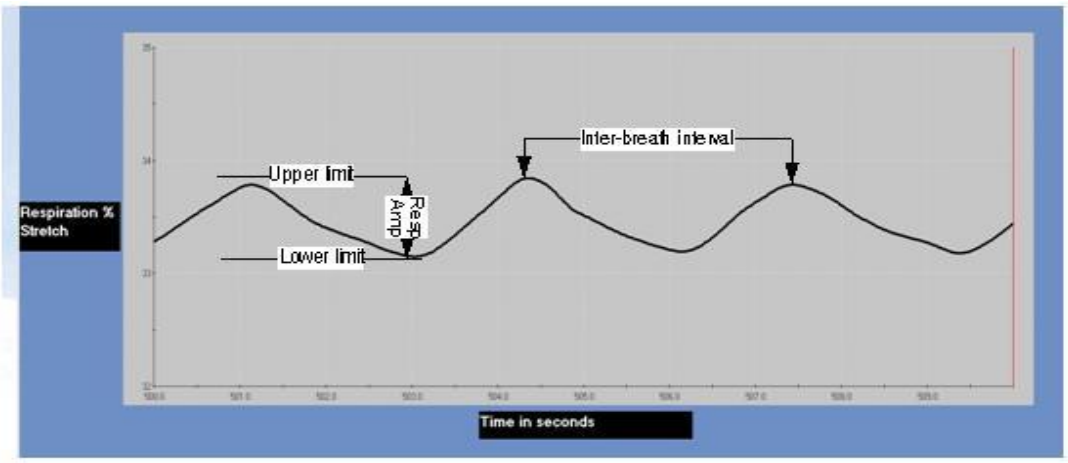

Figure 3: Respiration amplitude and inter-breath interval relative to raw Respiration signal

User's peripheral blood flow may be recorded by measuring levels of light absorption of blood through capillary beds in the finger via a HR/BVP Flex/Pro Sensor (SA9308M) connected to the peripheral phalanx of the middle finger. As blood volume increases in the fingers as a result of a heartbeat, so the light absorption levels increase, resulting. HR is derived by measuring the peak to peak interval of the light absorption measure and BVP is calculated by measuring the difference between the highest and lowest points in one heart beat. Figure 4 shows both HR and BVP and how they map onto the raw signal produced by the HR/BVP Flex/Pro Sensor.

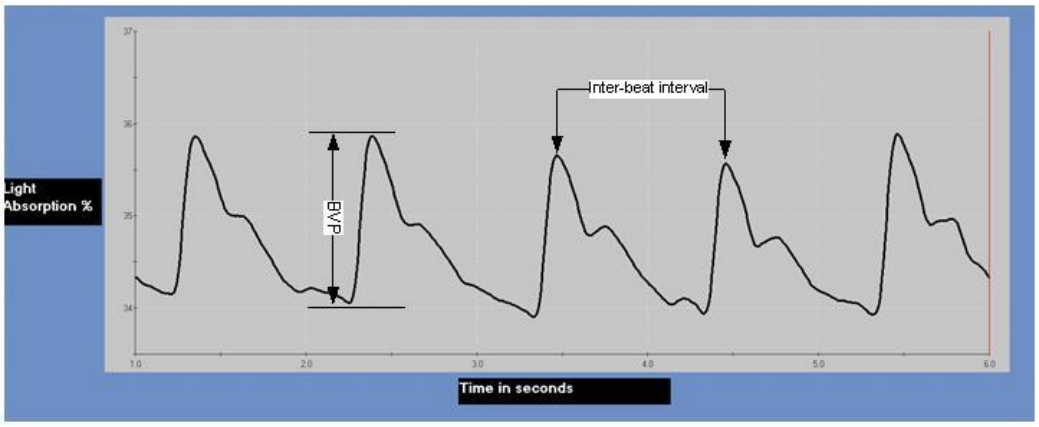

Figure 4: BVP and inter-beat interval relative to raw HR/BVP signal

4.3 Experiment design

One 15-18 minute video segment (VS) was chosen from each of the five full length videos respectively, as being representative of the content of each video. Short segments were chosen to ensure that the amount of content to be viewed by users, and the amount of sensing data generated were reasonable. For the purposes of later analysis, three video sub-segments (VSSs) from each of the VSs were chosen to represent the most pertinent segments of each VS. Each VSS was either 15 or 30 seconds in length, to best match the most pertinent content found within the video content.

The specific start and end points of each 15 and 30 second VSS were chosen in order to best capture the video content of the scene in question. However, since the physiological responses associated with each respective VSS would be later evaluated, it was important that the VSS start and end times fully incorporated the point of interest within the selected VSS itself. Some scenes took the full 15 or 30 seconds to unfold, and hence the point of interest for these scenes incorporated the full VSS duration. However, other scenes did not fit the pre-determined 15 or 30 second VSS duration precisely. In these cases, the excess time available within the 15 or 30 second time period was equally allocated to events leading up to the point of interest and those occurring after the point of interest. The former can be justified by the fact that events leading up to the point of interest are related to the events that occur during the point of interest. The latter case provides additional time after the point of interest to observe and include any delayed user physiological responses that may have occurred after viewing the point of interest.

It should be noted that the intention of this particular study was not to determine if the segments correlated with each user's own personal choices, since that would rely on the assumption that physiological responses could already be used to determine the significance of segments, for which there is no precedence in the research literature. The focus of this study was instead on whether physiological responses may be suitable for video summarisation at all (that is, whether they are even sensitive enough to distinguish between video segments) so that this could be established as a valid assumption for future research. Each VSS therefore represented a pertinent sub-segment of the video which essentially represented the highlights of each respective VS. Figure 5 shows how a VS relates to the full length video and how each VSS relates to the VS.

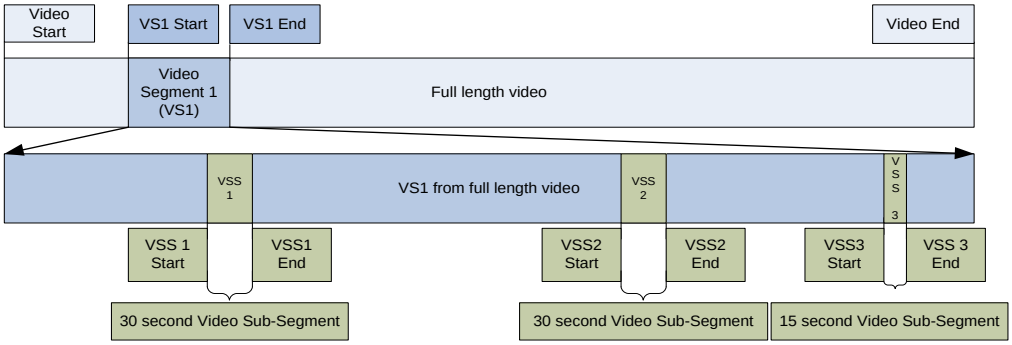

Figure 5: Full video, VS, and VSS segmentation

The duration of each VS varied slightly so that the start and end points of each VS coincided with the natural breaks in each video's narrative. Each user viewed three of the five VSs and users were not aware of the start and end times of each VSS. To control for order effects, VSs were allocated randomly to each user, but in a way that ensured each of the five segments had a reasonable number of viewings once all 10 experimental sessions had been completed.

4.4 Procedure

Upon arrival at the usability lab, users were shown to the viewing room, and asked to make themselves comfortable in the viewing chair. They were told they would be watching three segments of video, each between 15 and 18 minutes duration. They were told there would be an interval before the start of each VS during which relaxing music would be played. They were asked to relax and move as they wish during this period. Consent to connect the various physiological sensors was sought from each user prior to the commencement of the experiment. The sensors were then attached and users were asked to minimise movement of the hand that was attached to the sensor whilst the VSs were playing. This was in order to keep signal noise to a minimum. The BioGraph software for recording of the physiological responses was then started and checked for correct functionality. Finally, any questions users had were answered. During the experiment, physiological responses were monitored and start and finish times of each video segment were noted, to be used as VS start and finish markers at the analysis stage.

Two of the 30 VS viewings (both for The Exorcist) reverted to the interval music after approximately 12 minutes, instead of playing 18 minutes and 6 seconds. Response data relating to the 2 shortened segments were removed and not used at the analysis stage.

5 Framework for analysing physiological responses to video

In this section, we present a framework for evaluating physiological responses to video content, for personalised affective video summaries, based on our laboratory experimentation. The framework enables the significance of user responses to each respective VSS within each VS to be ascertained, an overview of which is presented in Figure 6, and is now summarised.

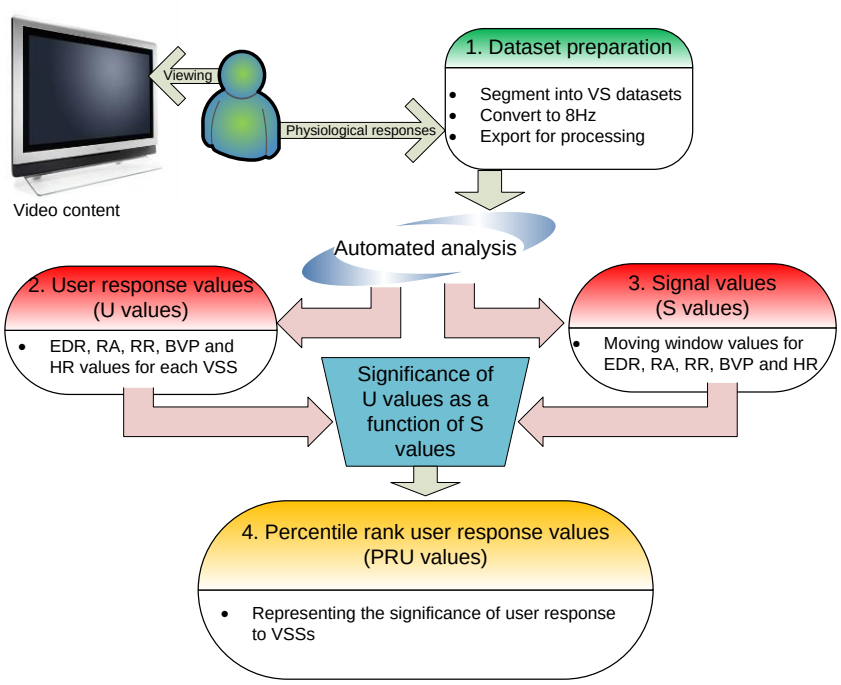

Figure 6: Framework for analysis of user physiological responses to video content

Initially, each set of experiment session data is segmented into VS datasets, relating to each user's viewing of each VS. Start and end points of each respective VSS within each VS dataset are tagged. Each VS dataset is then converted into an 8Hz format to reduce processing overhead.

Each user's average response is calculated for each physiological measure of each VSS within each VS, resulting in a set of user response values (*U values*). Moving average windows, of 15 and 30 second durations are then constructed for subject data for each VS dataset, relating to each physiological measure. This produces additional signal values (*S values*) for each users' physiological response measure to each VS. The significance of *U values* is derived and compared against the corresponding constructed *S values*. A percentile rank calculation is

carried out to establish a measure of the significance of *U values* as a function of the appropriate corresponding *S values*. The result is a normalised Percentile Rank value for each U value (*PRU value*). *PRU values* are presented on a scale of 0 to 1, representing the proportion of values in a sample that a specific value is greater than or equal to. Each stage of the analysis framework is now described in detail.

5.1 Dataset preparation

Initially, user data was segmented into datasets relating to each user's viewing of each VS. Each dataset consisted of data relating to the five physiological measures of EDR, RA, RR, BVP and HR respectively. Respective VS datasets were temporally tagged with the start and end points of each VSS. Tagged data files were then exported in CSV format at 8Hz sampling rate into Microsoft Excel.

5.2 U values

Individual U values were calculated for RA, RR, BVP and HR relating to each VSS. In total there were three VSS responses in each VS and five physiological measures for each VSS, totalling 420 U values. The U values for RA, RR, BVP and HR were calculated by averaging the respective physiological response values of each user for each VSS (EDR was treated separately). To indicate that the calculation was carried out separately for each respective physiological response measure, the notation $RA\_ \mid RR\_ \mid BVP\_ \mid HR\_$ is used. The VSS was either 15 or 30 seconds in length. The formula to calculate the U values is as follows:

$RA\_ \mid RR\_ \mid BVP\_ \mid HR\_$:

$$U_{jk} = \frac{1}{m_{vss_{jk}}} \sum_{i=1}^{m_{vss_{jk}}} C_{vssft_{jk}-(i-1)} \qquad \textbf{(1)}$$

where $m_{vss_{jk}}$ is the number of observations taken for vss_{jk}, which is a specific video sub-segment *k* in video segment *j*, $vssft_{jk}$ is the final point in time of the relevant physiological unit (RA, RR, BVP and HR) in the k^{th}

video sub-segment of the j^{th} video segment, and $C_{vssft_{jk}}$ is the value of the raw signal of the relevant physiological unit (RA, RR, BVP and HR) at time $vssft_{jk}$.

Unlike the other physiological measures, the EDR signal required detrending since its baseline varies in an unpredictable manner during the experimental session [51]. Evaluating absolute values of EDR signal throughout a session can be misleading. The detrended EDR signal reveals a more accurate representation of the fluctuations of the signal. The EDR signal was detrended using an approach similar to that of [62]. Detrended EDR values were calculated over a 15 or 30 second time period (to reflect the duration of the VSS). The reading at the start of the time period was subtracted from the highest value within the time period, which then represents the maximum level of deflection the stimuli has had within the specified time period. This constitutes a method for evaluating local fluctuations in the EDR signal regardless of unpredictable baseline variations. Hence U values for EDR were calculated as follows:

$$EDR\_U_{jk} = Max(C_{vssft_{jk}}, C_{vssft_{jk}-1}, C_{vssft_{jk}-2}, ...C_{vssft_{jk}-(m_{vss_{jk}}-1)}) - C_{vssft_{jk}-(m_{vssft_{jk}}-1)} \quad \textbf{(2)}$$

5.3 S values

The raw HR, BVP, RR, and RA measures for each VS dataset must be prepared so that the significance of U values can be derived as a function of their significance within each VS dataset. Hence the HR, BVP, RR and RA measures within each VS dataset were subjected to a 15 and 30 second moving average window, the output of which served as an appropriate dataset against which the significance of each users' response to each VSS could be measured. The calculation carried out on each of the measures produces a set of *S* values for each physiological measure (HR, BVP, RR and RA). S_t is the moving average, which is the mean of previous values prior to time *t*. It is calculated as follows:

$$RA\_ \mid RR\_ \mid BVP\_ \mid HR\_ :$$

$$S_t = \frac{1}{m}\sum_{i=1}^{m} C_{t-(i-1)} \qquad t = m, m+1, \ldots, T$$

(3)

where m is the number of observations taken in a specified time period (an 8Hz signal equates to m=120 in a 15 second period and m=240 in a 30 second period), C_t is the value of the raw signal of the relevant physiological unit (RA, RR, BVP and HR) at time t, and T is the index of the final observation point.

Equation (3) was calculated for m = 120 and m = 240, and has a smoothing effect on the signal: the longer the period over which the signal is averaged, the more the signal is flattened. Figure 7 is an example of a BVP signal in three states: raw, subjected to a 15 second moving average window, and subjected to a 30 second moving average window. As can be seen, the yellow line representing the 30 second moving average fluctuates substantially less than the original raw signal represented by the black line.

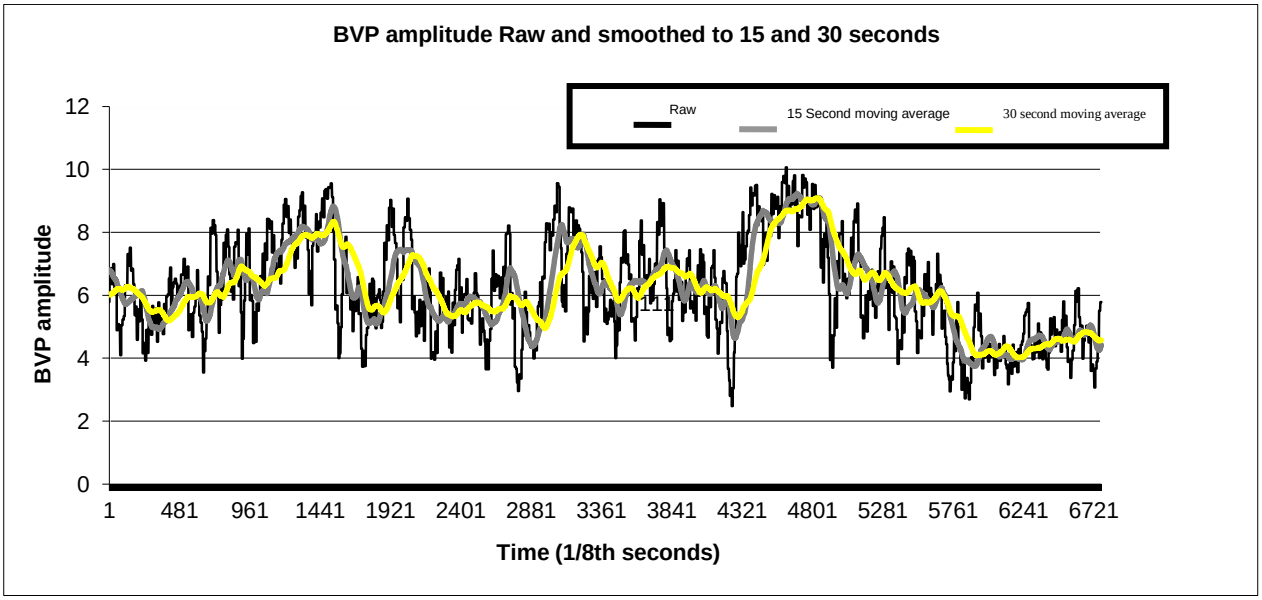

Figure 7 : BVP: raw, 15, and 30 second moving average

In order to mirror the calculation carried out in Equation (2) to derive EDR U values, the EDR signal for each VS dataset was detrended using a 15 and 30 second moving time window, subtracting the value at the start of the time period from the highest value within the time period. If the value at the start of the time window is greater than any subsequent values within the time window then the *S* value defaults to 0 reflecting no positive deflection during that period.

The EDR *S* values were calculated as follows:

$$EDR\_S_t = Max(C_t, C_{t-1}, C_{t-2}, ..., C_{t-(m-1)}) - C_{t-(m-1)}$$
$$t = m, m+1, ..., T \quad \textbf{(4)}$$

5.4 PRU values

Finally *U* values are calculated as *PRU* values. These also achieve a more meaningful measure of the users responses presented as values ranging between 0 and 1. The resulting *PRU* values were expressed as percentile rank values which can be used to infer the significance of a user's response as a function of each respective VS dataset. For example, a *PRU* value of 0.95 indicates a response that is equal to or greater than 95% of all responses in a dataset, a *PRU* value of 0.99 equates to a response equal to or greater than 99% of all responses in a dataset, and so forth. *PRU values* were calculated as follows:

$$EDR\_ \mid RA\_ \mid RR\_ \mid BVP\_ \mid HR\_ :$$
$$PRU_{U_{jk}} = \frac{n_e(S, U_{jk}) + n_b(S, U_{jk})}{n_w(S) - 1} \quad \textbf{(5)}$$

where $n_e(S, U_{jk})$ is the number of values in *S* equal to U_{jk}, $n_b(S, U_{jk})$ is the number of values in *S* less than U_{jk}, and $n_w(S)$ is the total number of values in *S*.

Converting *U values* into *PRU values* is a key step in our framework. Since each individual user's physiological responses have unique baselines and response ranges it is necessary to standardise the scale on which

physiological responses are evaluated. Some existing studies cater for the varying baselines and range of responses by normalising the response data. For example, some researchers [40, 64] adopt a normalisation calculation of (*signal – baseline*)/*range* to process the physiological response data. Where *signal* is one of the five physiological measures, *baseline* is either a value measured at rest before the experiment or the mean of the total response for that experimental session, while *range* is the upper and lower limits of the signal for the experimental session. Normalising data in this way converts all physiological response data into values between 0 and 1 (1 being the upper limit of the range and 0 being the lower limit). However, the values between 0 and 1 do not indicate the significance of a response within a dataset. *PRU values*, on the other hand, provide normalised response values for each user but also provide a measure of how significant the users' response is to each VSS.

6 Experiment results

The results in this section have been calculated using the analysis framework presented in the previous section. In Section 6.1, individual responses to a range of VSSs are reviewed. The VSSs with the highest overall percentage of significant response are selected for review with one VSS from each of the VS datasets. In Section 0, individual responses are aggregated for each physiological measure for each VS and a comparison is made for each respective video content genre. While any proportion of the sample could be used, for the purposes of this study, we used 30% of the sample since this was found to provide a range of responses sufficient for discerning significant candidate video segments. Thus, we considered significant responses to VSSs as those equal to or greater than the 85th percentile or equal to or less than the 15th percentile for RA, RR, BVP and HR, and those equal to or greater than the 70th percentile for EDR. This reflects the fact that EDR percentile rank values represent the level of positive deflection of the signal, and hence values less than or equal to the 15th percentile do not necessarily represent significant responses, but rather reflect little or no significant response.

6.1 Individual user responses to VSSs

Table 1 presents all the PRU values calculated for each user's physiological responses to each VSS within each VS and serves as a point of reference throughout this section. Significant responses to VSSs are highlighted by shaded boxes: the lighter shading represents significant responses in the 85th percentile or above for RA, RR, BVP and HR

Table 1: PRU response values for each users' EDR, RA, RR, BVP and HR for VSs 1..5 and VSSs 1..3

		VS1:Action/Sci-fi					VS2:Horror/Thriller					VS3:Comedy					VS4:Drama/Action					VS5:Drama/Comedy				
VSS1	**No.**	**EDR**	**RA**	**RR**	**BVP**	**HR**	**No.**	**EDR**	**RA**	**RR**	**BVP**	**HR**	**No.**	**EDR**	**RA**	**RR**	**BVP**	**HR**	**No.**	**EDR**	**RA**	**RR**	**BVP**	**HR**	**No.**	**EDR**	**RA**	**RR**	**BVP**	**HR**
	1	0.978	0.510	0.104	0.656	0.773	2	0.839	0.094	0.487	0.663	0.155	1	0.020	0.365	0.486	0.174	0.234	2	0.000	0.462	0.281	0.049	0.551	4	0.556	0.742	0.581	0.077	0.483
	2	0.801	0.247	0.519	0.990	0.606	4	0.884	0.284	0.964	0.704	0.096	3	0.012	0.537	0.210	0.952	0.093	4	0.915	0.463	0.793	0.087	0.898	5	0.055	0.774	0.335	0.213	0.062
	3	0.958	0.647	0.656	0.172	0.915	5	0.000	0.548	0.177	0.726	0.255	5	0.835	0.885	0.525	0.294	0.813	6	0.027	0.842	0.405	0.805	0.606	7	0.122	0.222	0.811	0.459	0.093
	6	0.996	0.772	0.428	0.232	0.692	9	0.667	0.438	0.524	0.506	0.241	6	0.791	0.777	0.221	0.995	0.030	7	0.000	0.066	0.831	0.120	0.064	8	0.715	0.250	0.854	0.210	0.786
	7	0.999	0.168	0.939	0.848	0.586	10	0.952	0.869	0.798	0.382	0.659	8	0.744	0.817	0.044	0.700	0.663	8	0.657	0.222	0.842	0.964	0.662	9	0.000	0.497	0.463	0.237	0.301
													10	0.743	0.905	0.039	0.955	0.773	9	0.000	0.054	0.520	0.924	0.526	10	0.607	0.275	0.381	0.555	0.749
	% Sig.	100.00	0.00	40.00	20.00	20.00		60.00	40.00	20.00	0.00	20.00		66.67	33.33	33.33	50.00	33.33		16.67	33.33	0.00	83.33	33.33		16.67	0.00	16.67	16.67	33.33
	VSS Tot. % Sig.					36.00						28.00						43.33						33.33						16.67
VSS2	1	0.000	0.366	0.869	0.448	0.000	2	0.852	0.004	0.876	0.298	0.326	1	0.679	0.274	0.559	0.140	0.211	2	0.999	0.035	0.993	0.386	0.976	4	0.526	0.045	0.994	0.720	0.602
	2	0.869	0.871	0.162	0.383	0.409	4	0.636	0.347	0.231	0.884	0.179	3	0.846	0.173	0.254	0.900	0.241	4	0.865	0.850	0.122	0.493	0.339	5	0.610	0.363	0.913	0.667	0.938
	3	0.722	0.741	0.091	0.720	0.167	5	0.879	0.204	0.660	0.102	0.929	5	0.799	0.757	0.505	0.331	0.345	6	0.913	0.809	0.580	0.238	0.984	7	0.000	0.320	0.131	0.600	0.457
	6	0.809	0.075	0.576	0.571	0.177	9	0.990	0.297	0.860	0.095	0.940	6	0.000	0.226	0.501	0.775	0.044	7	0.870	0.759	0.243	0.505	0.534	8	0.557	0.751	0.296	0.398	0.522
	7	0.806	0.478	0.307	0.588	0.267	10	0.991	0.264	0.488	0.076	0.829	8	0.990	0.875	0.402	0.681	0.939	8	0.947	0.178	0.596	0.470	0.056	9	0.972	0.944	0.953	0.424	0.929
													10	0.003	0.556	0.206	0.567	0.933	9	0.848	0.899	0.040	0.168	0.327	10	0.840	0.274	0.616	0.745	0.422
	% Sig.	80.00	40.00	40.00	0.00	20.00		80.00	20.00	40.00	80.00	40.00		50.00	16.67	0.00	33.33	50.00		100.00	50.00	50.00	0.00	50.00		33.33	33.33	66.67	0.00	33.33
	VSS Tot. % Sig.					36.00						52.00						30.00						50.00						33.33
VSS3	1	0.506	0.670	0.791	0.080	0.391	2	0.919	0.008	0.996	0.010	0.021	1	0.000	0.166	0.911	0.639	0.937	2	0.647	0.695	0.195	0.879	0.060	4	0.718	0.571	0.150	0.217	0.126
	2	0.506	0.852	0.498	0.113	0.098	4	0.958	0.011	0.931	0.052	0.064	3	0.390	0.099	0.949	0.035	0.597	4	0.000	0.575	0.084	0.600	0.140	5	0.793	0.300	0.685	0.923	0.102
	3	0.702	0.301	0.367	0.070	0.376	5	0.992	0.774	0.051	0.151	0.457	5	0.007	0.361	0.252	0.930	0.650	6	0.071	0.117	0.767	0.429	0.347	7	0.000	0.145	0.580	0.407	0.698
	6	0.000	0.206	0.655	0.534	0.007	9	0.995	0.142	0.921	0.391	0.674	6	0.691	0.943	0.120	0.141	0.897	7	0.000	0.639	0.265	0.811	0.536	8	0.931	0.102	0.711	0.133	0.496
	7	0.500	0.057	0.835	0.128	0.443	10	0.895	0.046	0.566	0.013	0.651	8	0.510	0.932	0.224	0.037	0.991	8	0.778	0.795	0.264	0.487	0.382	9	0.364	0.138	0.559	0.584	0.138
													10	0.000	0.295	0.449	0.055	0.091	9	0.000	0.195	0.103	0.641	0.191	10	0.871	0.701	0.432	0.125	0.043
	% Sig.	20.00	40.00	0.00	80.00	40.00		100.00	80.00	80.00	60.00	40.00		16.67	50.00	50.00	83.33	66.67		16.67	16.67	33.33	16.67	33.33		66.67	50.00	16.67	50.00	66.67
	VSS Tot. % Sig.					36.00						72.00						53.34						23.34						50.00
	VS % Sig.	66.67	26.67	26.67	33.33	26.67		80.00	46.67	46.67	46.67	33.33		38.89	33.33	27.78	55.56	50.00		44.46	33.33	22.22	33.33	38.89		38.89	27.78	33.33	22.22	44.44
	VS tot. % Sig.					36.00						50.67						41.11						34.45						33.33

No. = Participant number

Significant response in the 85th percentile or above for RA, RR, BVP, HR, and in the 70th percentile or above for EDR

Significant response in the 15 th percentile or below

and significant responses in the 70th percentile for EDR. Darker shading represents significant responses in the 15th percentile or below for RA, RR, BVP and HR. To follow is a brief description of the results. The VSSs with the highest overall percentage of significant response are discussed which are: VS1: VSS1, VS2: VSS2, VS3: VSS3, VS4: VSS2, and VS5: VSS3. Table 2 provides a brief description of the content of each VSS.

Table 2: Description of key VSS content

VSS	Synopsis
Action/Sci-fi: VS1: VSS1	Neo receives a phone call telling him that a group of men in black suits are looking to arrest him. His only hope of escape is via a window that is several storeys above the ground floor. This is a highly charged scene in which Neo nearly slips from the window ledge onto the busy city street below.
Horror/Thriller: VS2: VSS2	Regan has not been well for some time, since she started playing with an Ouija board she found in the basement. In this scene, she seems to be possessed and is screaming and being abusive to the doctor who has come to help her.
Comedy: VS3: VSS3	After being burnt to cinders by one of Del's sun-beds, Rodney gets even by arranging a hang-gliding lesson for Del, knowing full well that Del is scared of heights.
Drama/Action: VS4: VSS2	Sawyer is suspected of hiding medication that is needed to treat one of the sick members of the island community. Sayid has been asked to torture Sawyer, to find out where he is hiding the medicine.
Drama/Comedy: VS5: VSS3	On their night out together, Bob and Charlotte are drinking in a bar when a couple of locals jump over the bar and start to chase them with toy machine guns. Bob and Charlotte decide to play along and run from the night club.

All VSSs in *The Matrix* (VS1) elicited similar overall percentages of significant response (36%), thus VSS1 is selected for review here. EDR produced the most marked response to VSS1, which was significant in all 5 users. Four responses were well in excess of the 0.85, the lowest of which was 0.801 and the remaining four were 0.958 and above. For VS2: VSS2, EDR was significantly raised for 4 of the 5 users. User #10 had the highest significant rise in EDR with a value of 0.991. BVP was significant in four of the five cases. The general trend was a reduced BVP, with three of the four significant cases showing lowered BVP, indicating a constriction of blood to the peripherals. Interestingly, User #4 experienced significantly increased blood flow to the peripherals with a percent rank of 0.884. Although VS3: VSS3 elicited several significant responses, no EDR values were significant. RA was varied with three values in the lower 50th percentile, and 3 in the upper percentile. BVP was significantly reduced in 4 of the 6 users who experienced significantly constricted blood volume in their peripherals. User #3 showed the most significant constriction with a value of 0.035. HR tended to increase during this VSS with 5 of the 6 users showing some increase in HR. User #8 showed the most significant increase with 0.991.All users that viewed VS4: VSS2 produced significant EDR responses, the highest response was elicited in user #2 with a response

value of 0.999. Finally, VS5: VSS3 elicited significantly high EDR responses in four out of six users and significantly reduced HR values in four of six users.

6.2 Aggregate responses to video content genres

To examine the overall user physiological responses to the VSSs, EDR, RA, RR, BVP and HR response measures were aggregated, for each VSS, and compared. As stated in Sections 0 and 2.2, since this study was concerned with the feasibility of using physiological responses to video content as a means for developing video summaries, thus, the primary purpose of analysing the data was to establish whether user physiological response measures were sufficiently sensitive, and thus could be deemed significant for, the content of specific VSSs. It is therefore important to note that aggregated response values were not calculated with the primary aim of ascertaining general population trends. Nevertheless, since the VSS selections may be taken to be an approximation of the most pertinent video content, the results may provide an approximate indication of whether user physiological responses match pertinent video content.

Our results indicated that, in general, users responded significantly to VSSs, therefore indicating that physiological response measures seem to be sensitive enough to identify significant VSSs based on the user responses. Figure 8 is a summary of the results presented in Table 1, and illustrates the aggregated values of user responses to each of the five VSs viewed during trials. The values for '% significant response' (y/ axis) represent the percentage of all user responses that were considered to be significant (as presented in Table 1) grouped by video genre (Action/Sci-fi, Horror/Thriller, Comedy, Drama/Action, Drama/Comedy) and then grouped by physiological response type (EDR, RA, RR, BVP and HR). So as an example, the EDR '% significant response' value for Video1: Action/Sci-fi was 66.67%. This was calculated by totalling the number of EDR response values that were considered to be significant for all three VSSs (those in the 70^{th} percentile and above), and dividing this total by the total number of recorded EDR responses for Video 1: Action/Sci-fi. As can be seen in Table 1, from a total of 15 EDR responses for Video 1: Action/Sci-fi, 10 were significant, therefore 66.67% of all EDR responses for this VS were significant. For RA, RR, BVP and HR responses, Figure 8 also indicates the proportion

of '% significant response' that fell within the 85th percentile and higher (lighter colour), or the 15th percentile or lower (darker colour).

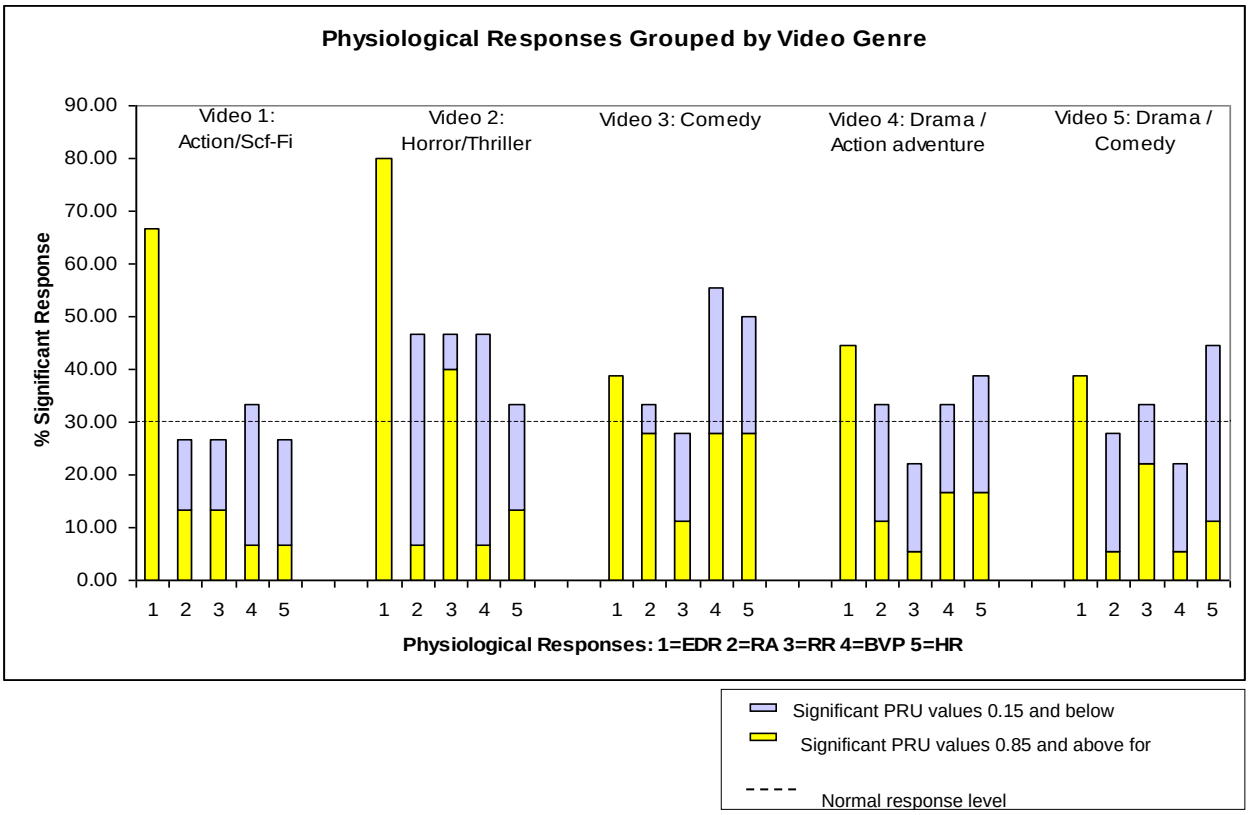

Figure 8: Physiological responses grouped by video genre

Since significant responses are considered to be those that are in the 15th percentile and below or in the 85th percentile and above for RA, RR, BVP and HR (totalling 30% of the whole sample), and those that are in the 70th percentile and above for EDR (totalling 30% of the sample), the expected proportion of significant responses that fall within these limits is 30%. In other words, regardless of the elicitation effects of a given VSS, 30% of RA, RR, BVP and HR user responses would be expected to fall either within the 15th percentile or below or within the 85th percentile and above; and likewise 30% of significant responses for EDR would be expected to fall within the 70th percentile and above. Therefore, a higher proportion than 30% of physiological responses falling within these limits indicate that users

responded in excess of what would normally be expected, thus indicating that these VSSs elicited higher than normal levels of physiological response in the user. To give an indication of the extent to which the proportion of physiological responses exceeded this 30% limit, Figure 8 includes a normal response level line set at the 30% level.

Evaluation of aggregate user responses to video content genres revealed significant differences in the types of physiological response they elicited by the different types of video content. Horror/Thriller (VS2) content elicited the highest aggregate EDR responses with 80% of all responses being significant. In addition, of the significant 44.44% BVP responses, 75% were in the lowest 15th percentile of responses, indicating that this content tends to cause a constriction of blood to the peripherals. A total of 46.67% of RR and RA responses were also significant for this content. 85.70% of these significant RR responses were in the top 15th percentile of responses, indicating that VSSs for this content cause significant increases in respiration rate, and 85.70% of significant RA responses were in the lowest 15th percentile, indicating that users take significantly shorter breaths whilst viewing this content. The nature of responses elicited by Comedy (VS3) content was also significant in several cases. This content seemed to elicit the smallest number of significant EDR responses, with 38.89% of all responses being significant. BVP and HR responses however seemed to be the most significant for this content, with 55.56% and 50.00% significant responses, respectively. Significant responses seemed to be evenly distributed between the highest and lowest 15th percentile. Comparing the responses to Comedy (VS3) with those of Horror/Thriller (VS2) content, there seemed to be a marked difference in the nature of physiological responses to the different types of content. Action/Sci-fi (VS1) only seemed to elicit a high percentage of significant EDR responses, with 66.67% of all responses being significant, aside from this, both Action/Sci-fi and Drama/Action (VS4) content did not seem to elicit a particularly high percentage of significant responses for any physiological measure and significant responses seemed to be evenly distributed across physiological response measures. Finally the Drama/Comedy (VS5) content seemed to have parallels with the VS3 (Comedy) content in that relatively high levels of significant HR responses (44.44%) and relatively low levels of significant EDR (38.89%) responses were elicited by this content.

7 Conclusions

This paper has examined whether physiological response data may serve as a potential information source that is external to the video stream, for developing personalised affective video summaries. An analysis framework for effective processing and evaluation of user physiological responses to video content has been presented and applied to user physiological response data. The framework assessed the significance of users' responses to specific video sub-segments (which represented the highlights of a video that may be included in a video summary) in a manner that allowed the significance of a user's response to particular video sub-segments to be determined. It also allowed the significance of users' responses to video genre contents to be determined. It was demonstrated that percentile rank calculations, such as those used in the proposed analysis framework, provide an appropriate means of measuring the significance of user response, which could serve as a means of developing techniques that summarise video content based on physiological response data. Based on the results of these calculations, it was shown that user physiological responses are a potentially valuable information source for video summarisation. Physiological response data seems to vary considerably between individual users; hence its unique nature indicates that video summaries based on this data have the potential for being highly personalised.

The fact that external information in the form of physiological responses can be collected automatically, without requiring any conscious input from the user, whilst potentially providing detailed information about the user's personal experience of the content, is of significant benefit since no additional effort is required from the user to collect this information. This is a valuable departure from the majority of current video summarisation techniques that achieve personalised summaries through requiring the user to manually input information relating to their experience of video content, which is costly in terms of the time and effort required.

Our research also appears to indicate that certain video genres may elicit significant physiological responses more consistently than others. This would have implications for how physiological responses may be automatically interpreted and video summaries automatically generated. Specifically, the following conclusions can be drawn regarding the nature of user responses to video sub-segments:

- Horror/Thriller content tends to elicit substantially higher levels of EDR compared with other content genres.
- User response to Horror/Thriller content tends to result in constricted BVP flow to the peripherals, as may be expected. This suggests the user is under duress or fearful during the viewing of VSSs of this content genre.
- Horror/Thriller content tends to elicit increased respiration rates and decreased respiration amplitudes in VSSs, eliciting significantly different to responses to Comedy content (which tends to cause increased respiration amplitudes in VSSs).
- Comedy content elicits a comparatively low percentage of EDR responses to VSSs. This is apparent in both Comedy and Drama/Comedy content.
- VSSs of Comedy content also seem to elicit the most significant BVP and HR responses, both of which seem to be evenly distributed between increased and decreased levels of BVP and HR responses.
- In general Drama/Action content elicits physiological responses evenly, i.e. no single physiological response seems to stand out as the most appropriate measure for content in these genres.
- Drama/Comedy content elicits the lowest amount of significant responses, perhaps reflecting the subtle and slow paced nature of this content. However, VSSs of this content do seem to elicit significant changes in HR and only a small number of significant EDR responses, a pattern which seems to be paralleled in Comedy content.

The implications of these results for video summarisation are numerous. There is a need to find new external sources of information to assist in the video summarisation process. Based on the findings that users seem to respond significantly to video content, new external summarisation techniques can be developed that incorporate physiological response data as a valuable and previously untapped source for analysis. User response data seems to vary considerably between users, and hence the unique nature of user responses indicates that video summaries based on physiological response data have the potential of being used to produce highly personalised video summaries, reflecting the individual's personal comprehension of the video content. Current external summarisation techniques require the user to manually input detailed descriptions of content in order to produce personalised summaries of video content, a task which is considered inconvenient and time consuming for the user. The fact that user physiological response data can be collected automatically, whilst providing detailed information about the individual's personal comprehension of the content, suggests that physiological response data is a valuable source of analysis, adding to the small number of external information sources that are currently used in external summarisation techniques. The framework presented in this paper provides a valuable foundation on which external summarisation techniques can set out to incorporate physiological response data into future techniques.

Other implications for video summarisation research arise from the fact that certain content genres do not appear to elicit particularly significant responses, and hence using physiological responses for video summarisation may be considered more appropriate for Horror/Thriller and Comedy content genres. Nevertheless, there did appear to be significant responses to specific VSSs within the other content genres and, when aggregating the results, Horror/Thriller and Comedy genres elicited significant responses more consistently. Hence, user physiological responses could be used to summarise video content for a range of content genres, but a better understanding of the ways in which the different content genres elicit responses in the users' is needed in order to effectively interpret the results and identify candidate segments for inclusion in a video summary.

Future directions for this research will be to further develop the framework so that VSSs can be identified automatically from the user response data. In addition, mapping users' physiological response data onto the affective dimensions of valence and arousal is likely to further assist in identifying VSSs that represent the most likely candidate segments for inclusion in a video summary and offer the user the option of viewing video summaries that consist of specific affective qualities. In the longer term, further research into other content genres such as sports, news and reality television shows is anticipated, with a view to developing a robust and in depth understanding of how users respond to all content genres.

8 References

[1] H. Agius, C. Crockford, A. G. Money, Geographic video content, in: Encyclopedia of Multimedia, B. Furht (Ed.). Springer, New York, NY, USA, 2006, pp. 257-259.

[2] K. Aizawa, D. Tancharoen, S. Kawasaki, T. Yamasaki, Efficient retrieval of life log based on context and content in: Proc. 1st ACM Workshop on Continuous Archival and Retrieval of Personal Experiences (CARPE '04), New York, NY, USA, 15 October, 2004, pp. 22-31.

[3] J. Allanson, S. H. Fairclough, A research agenda for physiological computing, Interacting with Computers, 16 (5) (2004) 857-878.

[4] N. Babaguchi, Y. Kawai, T. Kitahashi, Generation of personalized abstract of sports video, in: Proc. IEEE International Conference on Multimedia and Expo (ICME '01), Tokyo, Japan, 22-25 August, 2001, pp. 800-803.

[5] M. Barbieri, L. Agnihotri, N. Dimitrova, Video summarization: methods and landscape, in: Internet Multimedia Management Systems IV, Proceedings of SPIE, J. R. Smith, S. Panchanathan, T. Zhang (Eds.). SPIE, Bellingham, WA, USA, 2003, pp. 1-13.

[6] F. A. Boiten, N. H. Frijda, C. J. E. Wientjes, Emotions and respiratory patterns: Review and critical analysis, International Journal of Psychophysiology, 17 (1994) 103 - 128.

[7] M. Bradley, M. K. Greenwald, A. O. Hamm, Affective picture processing., in: The Structure of Emotion, N. Birbaumer, A. Öhman (Eds.). Hogrefe & Huber Publishers, Toronto, 1993, pp. 48-65.

[8] W. A. Brown, D. P. Corriveau, P. M. Monti, Anger arousal by a motion picture: A methodological note, American Journal of Psychiatry, 134 (1977) 930-931.

[9] J. T. Cacioppo, G. G. Berntson, D. J. Klein, K. M. Poehlmann, The psychophysiology of emotion across the lifespan, Annual Review of Gerontology and Geriatrics, 17 (1997) 27-74.

[10] N. R. Carlson, Psychology of Behaviour, 7th ed. Allyn and Bacon, Boston, MA, USA, 2001.

[11] Z. Cernekova, I. Pitas, C. Nikou, Information theory-based shot cut/fade detection and video summarization, IEEE Transactions on Circuits and Systems for Video Technology, 16 (1) (2006) 82-91.

[12] G. Coicca, R. Schettini, An innovative algorithm for key frame extraction in video summarization, Journal of Real-Time Image Processing, 1 (1) (2006) 69-88.

[13] C. Colombo, A. Del Bimbo, P. Pala, Retrieval of commercials by their semantic content: The semiotic perspective, Multimedia Tools and Applications, 13 (2001) 73-91.

[14] S. Coppola, Lost in Translation: Momentum Pictures, 2003.

[15] G. de Silva, T. Yamasaki, K. Aizawa, Evaluation of video summarization for a large number of cameras in ubiquitous home, in: Proc. 13th ACM International Conference on Multimedia, Singapore, 6-11 November, 2005, pp. 820-828

[16] B. H. Detenber , R. F. Simons, B. G., Roll 'em!: The effects of picture motion on emotional responses, Journal of Broadcasting & Electronic Media, 42 (1) (1998) 113-127.

[17] N. Dimitrova, Context and memory in multimedia content analysis, IEEE Multimedia, 11 (3) (2004) 7-11.

[18] P. Ekman, R. W. Levenson, W. V. Friesen, Autonomic nervous system activity distinguished between emotion, Science, 221 (1983) 1208-1210.
[19] M. Fletcher, Tea for Three, in Only Fools and Horses, Series 5: BBC Television, 1986.
[20] T. W. Frazier, M. E. Strauss, S. R. Steinhauer, Respiratory sinus arrhythmia as an index of emotional response in young adults., Psychophysiology, 41 (1) (2004) 75 - 83.
[21] N. Fridja, The Emotions. Cambridge University Press, Cambridge, 1986.
[22] W. Friedkin, The Exorcist: Warner Bros., 1973.
[23] M. Furini, V. Ghini, An audio-video summarisation scheme based on audio and video analysis, in: Proc. IEEE Consumer Communications and Networking Conference (CCNC '06), Vol. 2, Las Vegas, NV, USA, 8-10 January, 2006, pp. 1209-1213.
[24] T. Gates, Confidence Man, in Lost, Season 1: Channel 4 Television Corporation, 2004.
[25] P. Gomez, B. Danuser, Affective and physiological responses to environmental noises and music, International Journal of Psychophysiology, 53 (2) (2004) 93-103.
[26] P. Gomez, W. Stahel, B. Danuser, Respiratory responses during affective picture viewing, Biological Psychology, 67 (3) (2004) 359 - 373.
[27] M. K. Greenwald, E. W. Cook, P. J. Lang, Affective judgement and psychophysiological response: Dimensional covariation in the evaluation of pictorial stimuli., Journal of Pyschophysiology, 3 (1989) 51 - 64.
[28] J. J. Gross, R. W. Levenson, Emotion elicitation using films, Cognition and Emotion, 9 (1) (1995) 87-108.
[29] D. Hagemann, E. Naumann, S. Maier, G. Becker, A. Lürken, D. Bartussek, The assessment of affective reactivity using films: Validity, reliability, and sex differences, Personality and Individual Differences, 26 (1999) 627-639.
[30] A. Hanjalic, Adaptive extraction of highlights from a sport video based on excitement modeling, IEEE Transactions on Multimedia, 7 (6) (2005) 1114-1122.
[31] A. Hanjalic, L. Xu, User-oriented affective video content analysis, in: Proc. IEEE Workshop on Content-Based Access of Image and Video Libraries (CBAIVL 2001), Kauai, HI, 14 December, 2001, pp. 50-57.
[32] A. Hanjalic, L. Q. Xu, Affective video content representation and modeling, IEEE Transactions on Multimedia, 7 (1) (2005) 143-154.
[33] J. A. Healey, Wearable and Automotive Systems for Affect Recognition from Physiology., in Department of Electrical Engineering and Computer Science, vol. PhD. Cambridge, MA, USA: MIT, 2000, pp. 158.
[34] IMDb, Top 250 Films of All Time, http://www.imdb.com/chart/top, 2005.
[35] A. Jaimes, T. Echigo, M. Teraguchi, F. Satoh, Learning personalized video highlights from detailed MPEG-7 metadata, in: Proc. IEEE International Conference on Image Processing (ICIP 2002), Vol. 1, New York, NY, USA, 22-25 September, 2002, pp. 133-136.
[36] A. F. Kramer, Physiological metrics of mental workload: a review of recent progress., in: Multiple-Task-Performance, D. L. Damos (Ed.). Taylor & Francis, London, 1991, pp. 329-360.
[37] A. Lang, P. Bolls, R. Potter, K. Kawahara, The effects of production pacing and arousing content on the information processing of television messages, Journal of Broadcasting and Electronic Media, 43 (4) (1999) 451-476.
[38] M. S. Lew, N. Sebe, C. Djeraba, R. Jain, Content-based multimedia information retrieval: state of the art and challenges, ACM Transactions on Multimedia Computing, Communications and Applications, 2 (1) (2006) 1-19.
[39] Y. Li, S. Lee, C. Yeh, C. Kuo, Semantic retrieval of multimedia, IEEE Signal Processing Magazine, 23 (2) (2006) 79-89.
[40] R. L. Mandryk, K. M. Inkpen, Physiological indicators for the evaluation of co-located collaborative play, in: Proc. ACM Conference on Computer Supported Cooperative Work (CSCW '04), Chicago, IL, USA, 6-10 November, 2004, pp. 102-111.
[41] G. McIntyre, R. Göcke, The Composite Sensing of Affect, in: Affect and Emotion in Human-Computer Interaction. LNCS, vol. 4868, C. Peter, R. Beale (Eds.). Springer, Heidelberg, Germany, 2007.
[42] A. G. Money, H. Agius, Video summarisation: A conceptual framework and survey of the state of the art, Journal of Visual Communication and Image Representation) (accepted.
[43] F. Nasoz, K. Alvarez, C. L. Lisetti, N. Finkelstein, Emotion recognition from physiological signals for presence technologies, International Journal of Cognition, 6 (1) (2003) 1 - 32.
[44] D. Palomba, L. Stegagno, Physiology, perceived emotion and memory: responding to film sequences., in: The Structure of Emotion: Psychophysiological, Cognitive, and Clinical Aspects, N. Birbaumer, A. Ohman (Eds.). Hogrefe & Huber, Toronto, 1993, pp. 158-168.

[45] P. Philippot, C. Chapelle, S. Blairy, Respiratory feedback in the generation of emotion, Cognition and Emotion, 16 (2002) 605-627.
[46] R. W. Picard, Affective Computing. MIT Press, Cambridge, MA, 1997.
[47] R. W. Picard, E. Vyzas, J. Healey, Toward machine emotional intelligence: Analysis of affective physiological state, IEEE Transactions on Pattern Analysis and Machine Intelligence, 23 (10) (2001) 1175-1191.
[48] R. L. Piferi, K. A. Kline, J. Younger, K. A. Lawler, An alternative approach for achieving cardiovascular baseline: Viewing an aquatic video, International Journal of Psychophysiology, 37 (2000) 207-217.
[49] M. Power, T. Dalgliesh, Cognition and Emotion: From Order to Disorder. Psychology Press, Guildford, Surrey, 1998.
[50] Y. Rui, S. X. Zhou, T. S. Huang, Efficient access to video content in a unified framework, in: Proc. IEEE International Conference on Multimedia Computing and Systems (ICMCS '99), Vol. 2, Florence, Italy, 7-11 June, 1999, pp. 735-740.
[51] J. Scheirer, P. Fernandez, J. Klein, R. J. Picard, Frustrating the user on purpose: A step toward building an affective computer, Interacting with Computers, 14 (2002) 93-118.
[52] X. Shao, C. Xa, M. S. Kankanhalli, A new approach to automatic music video summarization, in: Proc. IEEE International Conference on Image Processing (ICIP 2004) Vol. 1, Barcelona, Spain, 24-27 Oct., 2004, pp. 625-628.
[53] H. A. Simon, Comments, in: Affect and Cognition, C. Sydnor, S. T. Fiske (Eds.). Lawrence Erlbaum Associates, Hillsdale, NJ, 1982, pp. 333-342.
[54] R. F. Simons, B. H. Detenber, J. E. Reiss, C. W. Shults, Image motion and context: A between- and within-subject comparison, Psychophysiology, 37 (2000) 706-710.
[55] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain, Content-Based Image Retrieval at the End of the Early Years, IEEE Transactions on pattern analysis and machine intelligence, 22 (12) (2000) 1349-1380.
[56] N. Steinbeis, S. Koelsch, J. A. Sloboda, The role of harmonic expectancy violations in musical emotions: evidence from subjective, physiological, and neural responses, Journal of Cognitive Neuroscience, 18 (8) (2006) 1380-1393.
[57] J. Suziki, N. Hiroshi, T. Hori, Level of interest in video clips modulates event-related potentials to auditory probes, International Journal of Psychophysiology, 55 (1) (2004) 35-43.
[58] Y. Takahashi, N. Nitta, N. Babaguchi, Video summarization for large sports video archives, in: Proc. IEEE International Conference on Multimedia and Expo (ICME 2005), Amsterdam, The Netherlands, 6-8 July, 2005, pp. 1170-1173
[59] The Wachowski Brothers, The Matrix: Warner Bros., 1999.
[60] B. L. Tseng, J. R. Smith, Hierarchical video summarization based on context clustering, in: Internet Multimedia Management Systems IV, Proceedings of SPIE, J. R. Smith, S. Panchanathan, T. Zhang (Eds.). SPIE, Bellingham, WA, USA, 2003, pp. 14-25.
[61] I. Van Diest, W. Winters, S. Devriese, E. Vercamst, J. N. Han, K. P. Van de Woestijne, O. Van den Bergh, Hyperventilation beyond fight/flight: respiratory responses during emotional imagery., Psychophysiology, 38 (6) (2001) 961 - 968.
[62] C. M. van Reekum, T. Johnstone, Psychophysiological responses to appraisal dimensions in a computer game., Cognition and Emotion, 18 (5) (2004) 663-688.
[63] H. Wang, H. Prendinger, T. Igarashi, Communicating emotions in online chat using physiological sensors and animated text, in: Proc. ACM Conference on Human Factors in Computing Systems (CHI '04), Vienna, Austria, 24-29 April, 2004, pp. 1171-1174.
[64] R. D. Ward, P. H. Marsden, Physiological responses to different Web page designs, International Journal of Human Compuer Studies, 59 (1-2) (2003) 199-212.
[65] W. M. Winton, L. E. Putnam, R. M. Krauss, Facial and autonomic manifestations of the dimensional structure of emotion, Journal of Experimental Social Psychology, 20 (1984) 195-216.
[66] C. Xu, J. Wang, K. Wan, Y. Li, L. Duan, Live sports detection based on broadcast video and Web-casting text, in: Proc. 14th ACM International Conference on Multimedia, Santa Barbara, CA, 23-27 October, 2006, pp. 221-230.
[67] J. Zimmerman, N. Dimitrova, L. Agnihotri, A. Janevski, L. Nikolovska, Interface design for MyInfo: a personal news demonstrator combining Web and TV content, in: Proc. IFIP TC13 International Conference on Human-Computer Interaction (INTERACT), Zurich, Switzerland, 1-5 September, 2003.