Predicting Audience Responses to Movie Content from Electro-Dermal Activity Signals

Fernando Silveira* **Brian Eriksson*** **Anmol Sheth*** **Adam Sheppard**†

*Technicolor †Stanford University

ABSTRACT

The ability to assess fine-scale user responses has applications in advertising, content creation, recommendation, and psychology research. Unfortunately, current approaches, such as focus groups and audience surveys, are limited in size and scope. In this paper, we propose a combined biometric sensing and analysis methodology to leverage audience-scale electro-dermal activity (EDA) data for the purpose of evaluating user responses to video. We provide detailed characterization of how temporal physiological responses to video stimulus can be modeled, along with first-of-its-kind audience-scale EDA group experiments in uncontrolled real-world environments. Our study provides insights into the techniques used to analyze EDA, the effectiveness of the different temporal features, and group dynamics of audiences. Our experiments demostrate the ability to classify movie ratings with accuracy of over 70% on specific films. Results of this study suggest the ability to assess emotional reactions of groups using minimally invasive sensing modalities in uncontrolled environments.

Author Keywords

EDA, biometrics, affective computing, user experiments, signal processing

ACM Classification Keywords

1.4.5 Ubiquitous and mobile computing: Ubiquitous and mobile devices - Empirical studies in ubiquitous and mobile computing

INTRODUCTION

Assessing the reaction of viewers to video content they consume is important for a wide variety of applications. Examples range from movie recommendation systems, where viewer ratings are used to profile their preferences [23], to market research, where content creators conduct surveys and focus groups with test audiences to predict the success of movie productions [27] or ad campaigns [14].

While these applications traditionally utilize explicit feedback of user responses provided via ratings and survey forms, this feedback is often constrained by numerous factors. For example, existing movie recommendation systems request viewers to provide only a single rating for the entire movie; survey forms are limited by space and their reliance on viewer memory; and focus groups are constrained by participation costs and time limitations. This makes it difficult to achieve fine-grained viewer feedback about the video content.

More recently, there has been growing adoption of wearable biometric sensors that enable capturing viewers response to content at a much finer granularity than what explicit techniques allow for. Biometric sensors like heart rate and electro-dermal activity monitors are increasingly being embedded in consumer electronic equipment like watches [1] and fitness devices [2] that continuously monitor the physiological responses of the user. These physiological signals provide a rich source of implicit feedback which can be used to infer viewer reactions at various granularities.

Unfortunately, direct inference of viewer opinion of video content using physiological signals is not straightforward and needs to address several important challenges. Physiological signals are often noisy and are impacted by stimuli that are not a part of the content, e.g., distractions in the environment. Additionally, the responses contained within the signals vary considerable based on the type of stimuli and also depend on the individual viewer's physiological and psychological state.

In this paper, we address the above described challenges and present the first end-to-end methodology that can use electro-dermal activity (EDA) signals of viewers watching video content and map it accurately to the self-reported explicit feedback provided by the viewers. This not only improves existing approaches to calibrate audience feedback, but also enables a range of new applications like indexing and search of personal video content, and movie recommendation systems that can propose movies that best match the physiological state of the user. To this end, the two main contributions of this paper are:

- An approach that can decompose raw EDA signals into responses that accurately pinpoint the time locations and intensities of viewer responses to the stimuli in the content. Our approach outperforms the best known technique in the research literature for this task [20] by a 2.8 factor in terms of detection rate.

- A machine learning framework that uses the EDA responses to accurately predict the explicit feedback provided by the viewer. While there is a large Affective Computing literature on mapping EDA to psychological variables such as valence and arousal [18] or emotion categories [19], predicting explicit feedback has a direct impact on the applications we highlighted above.

UbiComp'13, September 8–12, 2013, Zurich, Switzerland.

We design and evaluate our methodology around a series of field studies conducted both in a controlled research environment and in commercial theater facilities where moviegoers wear an off-the-shelf EDA sensor and provide explicit feedback through survey forms. Our experimental results show that, using EDA and demographics information (i.e., age and gender), for particular films we are able to predict the explicit ratings that a viewer will give to a movie with over 72% accuracy, which is a 31% improvement over predictions from demographics alone. These initial results point to the promising direction of using biometrics to assess audience emotional reactions.

The paper is organized as follows. The extensive prior work in the field of EDA and audience testing is reviewed in the following section. We then detail the setup of our biometrics field studies and introduce our novel user reaction inference methodology. The results of our experiments on real-world audiences are then presented. We conclude with an overview of the results from our real-world study and future work.

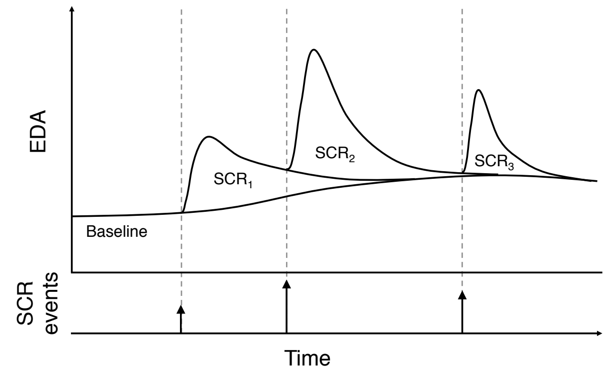

Figure 1. An example of EDA trace where the Skin Conductance Response (SCR) events are decomposed.

BACKGROUND AND RELATED WORK

Our primary focus is to characterize the changes in electrodermal activity (EDA) as audiences respond to the audio-visual stimuli in video. The goal is to develop a model that can accurately map the implicit EDA feedback to the explicit feedback provided by the viewers in the form of ratings and survey forms. To this end, we develop a framework that consists of three primary components: (1) the EDA sensors that viewers wear when watching video, (2) a novel EDA signal decomposition technique that can accurately detect and quantify EDA responses, and (3) a machine-learning model that can predict explicit viewer feedback at the end of a movie session.

EDA and Emotional Reactions. EDA is typically recorded as the conductance between a pair of electrodes placed over a person's skin, near concentrations of sweat glands. An EDA signal is generally characterized by a slow frequency baseline component plus short-lived spike-like events denoted Skin Conductance Responses (SCRs) which often overlap with each other, as shown in Figure 1. A person's EDA has a well-known connection to the brain activation from emotional reactions to stimulus [15] which causes sudomotor neuron bursts [20] and results in sweat to be expelled from eccurine glands [30], finally causing conductance variations on the person's skin. The psychological connection between a person's emotional reactions and changes in EDA has been studied since the early 20^{th} century [5]; Detailed surveys have been done by Lang [24] and Dawson *et al.* [21].

Our current understanding of these phenomena has been recently extended by examining brain function via fMRI and skin conduction via EDA simultaneously [15], showing the activations in specific regions of the brain that result in variations in the EDA. In addition, micro-video recordings of sweat glands [30] clearly demonstrate that neuron firings result in variations in skin conductance. Since then, there has been extensive work in evaluating the connection between SCRs and numerous activities, including video game playing [26], performing arts viewing [7], everyday interactions [19], detecting stress [17], evaluating cognitive load [12], and determining perception changes due to mental illness [22] – to name only a few.

Identifying EDA Responses to Stimuli. Figure 1 illustrates challenges involved in characterizing SCR events from raw EDA sensor data. Specifically, true user neuron burst events are difficult to extract from EDA due to potentially overlapping events [20], attenuation of event amplitude for repeated stimulus [3], varying burst impulse functions [10], and underlying, slowly varying, skin conductance levels [8]. Various signal decomposition approaches have been proposed to combat these difficulties. Examples include a highly parametric sigmoid-exponential model [8], bi-exponential impulse responses [10], nonnegative deconvolution [20], and Variational Bayesian decomposition techniques [11]. All these prior techniques are limited by either computational complexity [11], inability to discover overlapping events [8], or a one-size-fits-all approach that is not robust to varying event durations [20, 10]. To defeat these issues, in this paper we present a matching pursuit-based methodology to extract relevant impulse information with low computational complexity and high adaptivity to changing physiological environments. Our approach inputs only the raw EDA signal and identifies both the time location and intensity of SCR events.

Predicting Explicit Film Feedback from EDA. Extracting the emotional states of a user through biometrics has long been the goal of the affected computing community (*e.g.,* [25] and [16]). In this paper, we consider the narrower problem of assessing user reactions to video stimulus using EDA. Initial studies of user engagement to films using skin conductance were done by Kaiser and Roessler [6], while the survey by Lisetti et. al. [9] contains a detailed review of prior research in this area. Of particular relevance to our research is prior work on using EDA to evaluate and classify users with respect to film viewing. Specifically, the work by Fleureau *et al.* [18] examined EDA and other physiological signals using Gaussian processes to predict positive and negative affect (*i.e.*, valence and arousal) to videos, and the work by Roth-

Figure 2. Affectiva Q Sensor.

well *et al.* [28, 4] detects emotional responses from EDA and correlates to film events. In contrast to these prior works focused on isolated experiments on individual users, this paper examines concurrent, audience-level evaluation of SCR events previously decomposed by a novel signal processing algorithm.

FIELD STUDY AND DATA COLLECTION

In this section we describe the field study we conducted in order to collect EDA data from viewers watching different types of audio-video content. The primary goal of the study was to measure EDA responses from audience at scale in an environment where distractions from external stimuli are minimal. To address this, we conduct our field study in commercial movie theaters where the participants are watching feature-length films. The controlled temperature, lighting and immersive nature of a movie theatre enables us to measure EDA responses that are mainly contributed by the stimuli in the film. In addition to EDA responses, we also collected explicit viewer feedback from the participants in order to learn models that map the implicit feedback in EDA responses to explicit feedback.

In the rest of this section we describe the sensors worn by participants, the data recorded by these devices, the content used as stimuli, and the characteristics of each study session.

Sensor Devices

Figure 2 shows the commercially available EDA sensor provided by Affectiva[1] that was worn on the palm by each study participant. Unlike medical grade EDA sensors that typically require wired connections and conductive gel to improve signal quality, the sensors we use are easy to wear and that enable us to setup a large group of study participants (between 20-30 participants) within a short time span (15-20 minutes). The participants were given no special instructions and told to act naturally.

In each experiment, we synchronize and pre-process all raw EDA traces. We synchronize the sensor clocks to a single computer prior to each recording session and we use that same computer's clock to record the beginning and ending times of the experiment session. The Q-sensor measures raw skin conductance levels at 32 Hz. Given the typical duration of user skin conductance responses and the computational complexity of prior techniques, we down-sample these signals to 4 Hz.

Data Collection

We perform two types of data collection experiments. While our goal is to analyze audience reactions to feature-length films, we initially perform a calibration experiment, where participants are monitored in isolation and are exposed to a short clip with controlled audio and image stimuli. After validating our ability to detect simple individual responses, we move on to experiments using feature-length films and simultaneous sensing of user groups.

Basic Stimulus Experiment

Calibration Study. Our first experiment examines individual responses to stimuli of varying levels of complexity. As seen in Figure 3, this experiment consists of a 220-second clip containing seven isolated stimulus events. The film begins with three sounds clips of a gun shot, a dog barking, and finally a baby crying. Then, the image of gun is shown to the users for 5 seconds, and then contrasted with an image of a kitten held on screen for the same amount of time. Finally, two short-duration (< 5 seconds) video clips of near-death experiences are shown, the first being a woman almost hit by an on-coming train, and then an attempt at "parkour" ending with the person falling face-first onto concrete. Before each stimuli, we insert silent intervals where nothing is played to the participant. We obtained EDA traces of nine individuals (6 male, 3 female, aged between 20 and 50 years old) who watched this basic stimulus video in isolation in a controlled laboratory environment.

Feature-Length Film Experiments

The remainder of our experiments consists of audiences between 9 and 15 individuals simultaneously viewing a feature-length film. We use 3 different films produced in either 2011 or 2012, which we label using letters A through C[2]. We deliberately performed experiments with movies from different genres (*e.g.,* drama, thriller, foreign) to avoid limiting the scope of our conclusions to genre-specific phenomena. A summary of these 3 films and the experimental environment used is seen in Table 1. We note that while basic stimulus calibration study was performed in a relatively controlled laboratory environment, where participants were recruited in advance from among the local research staff, the remaining experiments were performed in commercial movie theaters where participants were solicited from the movies' regular audiences. These audience members were required to sign a consent form before participating in the study. While we have reason to believe that our observed data will be cleaner than signals observed on non-stationary subjects, the uncontrolled commercial movie theater environment potentially adds practical signal artifacts (*e.g.,* users eating and drinking while being sensed) that our methodology must consider.

In Table 2 we show the demographics of each screening.

In addition to the audience-wide EDA traces used for implicit audience feedback, we also ask participants to provide explicit feedback at the end of each movie screening. We do

[1]Affectiva - `http://www.affectiva.com/`

[2]The true names of these films are not used for internal reasons.

Movie	Genres	Runtime (min)	Release	Viewers	Location
A	Action, Crime, Thriller	130	2012	9	Theater
B	Drama	139	2012	10	Theater
C	Drama, Foreign	126	2011	15	Film Festival

Table 1. Summary of feature-length films used in experiments.

	Gender		Age			Rating				
Movie	Male	Female	$20-29$	$30-49$	> 49	1	2	3	4	5
A	5	4	4	3	2	0	0	6	3	0
B	4	6	4	3	3	0	0	2	3	5
C	7	8	7	5	3	0	0	3	5	7

Table 2. Demographics of feature-length film audiences.

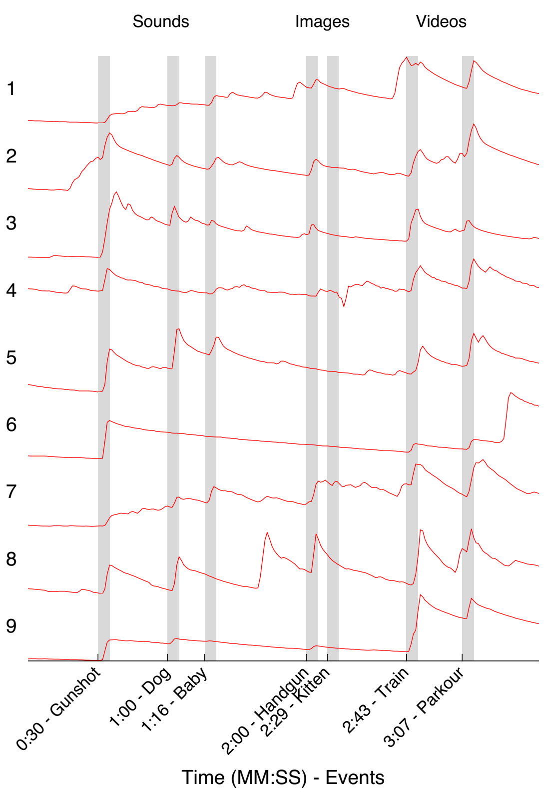

Figure 3. EDA traces of participants in the basic stimulus experiment. Stimulus events marked in vertical lines.

this in order to learn models that map the implicit feedback in EDA responses to explicit feedback. We design survey forms for each movie in which we ask participants to provide: (1) gender and age group, and (2) an overall rating for the movie in a 5-point scale. We left the interpretation of what this rating implies (*e.g.,* enjoyment, engagement, etc.) up to the user's discretion.

EDA DECOMPOSITION METHODOLOGY

In this section we present the adaptive decomposition methodology which process raw EDA traces to extract precise SCR events showing exactly *when* and *how much* the viewer responds to a stimulus. As shown in Figure 1, identifying the relevant SCR events from raw EDA is challenging as SCRs may overlap, have varying duration, and may sometimes not be correlated with the underlying stimulus (*e.g.,* if the viewer is distracted from the stimulus). Additionally, comparing EDA traces from multiple people is problematic due to varying levels of signal normalization, non-standard reaction impulse response magnitude, and differing susceptibility to react due to the deviations in subject psychology and physiology.

To address the above challenges we propose a novel signal decomposition approach that automatically adapts to the variations in subject physiology. We evaluate the performance of our approach using the basic stimuli experiment and compare our approach with the state-of-the-art EDA analysis technique of Benedek and Kaernbach [20] using the publicly available Ledalab software package[3]. Our results show that our methodology outperforms Ledalab in practically all operating points of the parameter space; For example, fixing the false alarm rate to 10%, Ledalab achieves a 25% detection rate of SCR events, while our approach detects over 70% of events – a 2.8 factor improvement.

Method Description

We begin by considering the slowly varying DC component of each viewer's signal. Often called the "tonic" signal, this component is due to physiological response to sweat saturation-levels of the person's skin and has little correlation with the underlying fine-scale user reactions we want to detect. Various prior techniques exists for removing this component [11, 20], but for the sake of simplicity, here we just subtract signal contribution related to the two coarsest-scale coefficients of a discrete-cosine transform (DCT). We denote the remaining high-pass, processed EDA signal as $\mathbf{x}$.

In contrast to the work of Benedek and Kaernbach [20], which fits a single canonical impulse response throughout the entire processed EDA trace, we decompose the signal using a large dictionary of feasible SCR shapes. The consideration of many different signal types, with varying durations and decay characteristics, allows us to better fit to the observed skin conductance.

[3] Ledalab Matlab Toolbox - `http://www.ledalab.de/`

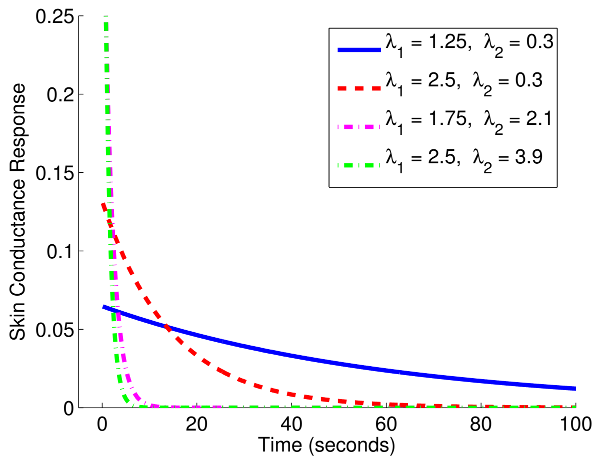

Figure 4. Four example dictionary signals.

The specific dictionary basis functions we consider can be parameterized by:

$$\mathbf{d}_{\lambda_1,\lambda_2,t_0}(t) = \begin{cases} \lambda_1^{-\lambda_2(t-t_0)} & t \geq t_0 \\ 0 & t < t_0 \end{cases}. \tag{1}$$

Such that λ_1 relates to the geometric decay of the impulse, λ_2 is the log-linear decay slope, and t_0 is the response start. From empirical examination of EDA signals, we construct the signal dictionary, $\mathbf{D}$, using all signals $\mathbf{d}_{\lambda_1,\lambda_2,t_0}(t)$ for:

$$\lambda_1 \in \{1.1, 1.25, 1.5, 1.75, 2, 2.5, e\}, \tag{2}$$

$$\lambda_2 \in \{0.3, 0.5, \ldots, 3.7, 3.9\}. \tag{3}$$

Selected examples from this constructed dictionary can be seen in Figure 4 for $t_0 = 0$.

To represent each EDA trace from this large collection of dictionary signals requires solving a standard linear inverse problem. Unfortunately, ordinary least squares approaches will require infeasibly large amounts of memory for large dictionaries, and also destroy the inherent desired sparsity of the SCR event process. We avoid these limitations by using an orthogonal matching pursuit [13] technique to greedily resolve the set of dictionary components that best describe the observed EDA trace.

Specifically, this matching pursuit procedure begins with the high-pass processed EDA signal, $\mathbf{x}$, a signal component dictionary, $\mathbf{D}$ constructed using Equation 1, and an empty constructed dictionary $\widehat{\mathbf{D}} = \{\}$. We first determine the single dictionary component ($\widehat{\mathbf{d}} \in \mathbf{D}$) that best fits the observed EDA signal:

$$\widehat{\mathbf{d}} = \arg\max_{\mathbf{d}\in\mathbf{D}} \left|\mathbf{d}^T\mathbf{x}\right|. \tag{4}$$

This dictionary component is then added to the constructed dictionary $\widehat{\mathbf{D}} = \{\ \widehat{\mathbf{D}}\quad \widehat{\mathbf{d}}\ \}$, and the contributions of this dictionary component are removed from the observed EDA signal, creating a new residual signal:

$$\mathbf{r} = \mathbf{x} - \widehat{\mathbf{D}}\left(\widehat{\mathbf{D}}^T\widehat{\mathbf{D}}\right)^{-1}\widehat{\mathbf{D}}^T\mathbf{x}. \tag{5}$$

This process is then repeated using the residual signal (*i.e.,* setting $\mathbf{x} = \mathbf{r}$) for a specified number of iterations.

Once the desired number of iterations complete, we obtain a collection of dictionary components that fits to the observed signal. Using standard least squares, we calculate the best coefficient vector, β, such that the observed EDA signal is represented by a combination of elements from the constructed dictionary, $\mathbf{x} \approx \widehat{\mathbf{D}}\beta$, where the amplitude of the non-zero elements of β correspond to the intensity of user reactions.

In summary, for each EDA trace, the adaptive decomposition approach returns, $\{\mathbf{t}_i, \mathbf{s}_i\}$, the set of time offsets (*i.e.,* the time-start of each SCR event) and the coefficient amplitude of SCR events (*i.e.,* the intensity of the SCR event), respectively. In the following section, we show how the decomposed SCRs accurately capture viewer reactions to content stimuli.

Method Evaluation

To evaluate the performance of the adaptive decomposition on EDA traces, we examine observations from the basic stimulus experiment. The benefits of using these simple stimulus traces is the ability to cleanly extract which time segments have stimulus and which do not. We compare against the nonnegative deconvolution approach of Benedek and Kaernbach [20] using the Ledalab software package. Similar to our adaptive decomposition technique, the Ledalab software analyzes EDA traces and returns both the location and amplitude of the SCR events. In addition to this analysis technique, we also examine a naive approach of using the raw unprocessed EDA signal energy, demonstrating the accuracy using no processing of the observed signals.

For the nine users viewing the basic stimuli, we look to distinguish between the six stimulus events and five defined silence events where a black screen and no sound are presented to the user, with the start time of these events marked in Figure 3. We then defined a time window size starting at the event marker (here we use 8, 10, and 15 second windows) and captured the sum coefficient energy (or raw signal energy) between the start of the event and the end of the time window. Classification was performed by thresholding these extracted energy values to classify if the time window relates to a stimulus event (greater than the threshold) or a non-stimulus event (less than the threshold), accuracy of this classification was then assessed given the true stimulus and silence interval labels from Figure 3.

Aggregating the classification rate across all nine users, we show the detection and false alarm results of classification in Figure 5 for three different time window sizes (8, 10, and 15 seconds). We find that our adaptive decomposition approach performs significantly better than the competing nonnegative deconvolution method across all three window sizes. For example, for a window of 10 seconds in Figure 5-(Center), we find that the prior deconvolution method results in detection rate of 25% for a 10% false alarm rate, while the adaptive approach detects over 70% of the stimulus events for the same false alarm rate. Another insight from these results is that using the raw EDA traces results in classification accuracy that

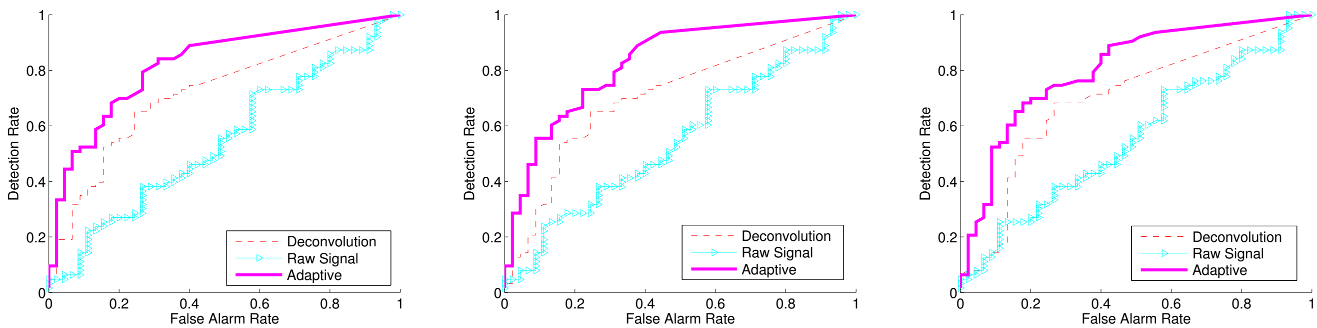

Figure 5. Stimulus classification accuracy results for calibration experiment (54 stimuli user-events and 45 non-stimulus user-events) with window size : (Left) 8 seconds, (Center) 10 seconds, and (Right) 15 seconds.

performs roughly no better than random guessing (*i.e.,* detection rate equal to the false alarm rate), showing the need for processing of the observed EDA signals.

We verify that these aggregated results are not biased by extreme good or bad performance on certain stimulus events. Separating out the classification rate for four of the stimulus events, we show the results in Figure 6 and Figure 7 for detecting these individual events against the five silence events. For the more extreme stimulus events of the gunshot sound and the train video (in Figure 6-(Left) and Figure 7-(Right)), we find that the adaptive decomposition method performs significantly better than both the raw signal and prior deconvolution approach. Meanwhile, we find the the dog barking sounds (in Figure 6-(Right)) shows the adaptive methodology performing significantly better than deconvolution for detecting seven of the nine users, and comparable results for the remaining two users. Finally, we find the photograph stimulus (in Figure 7-(Left)) to be the most difficult to distinguish for all techniques, with the adaptive methodology performance lagging behind the prior deconvolution method for detecting this stimulus event for two of the nine users.

EXPLICIT FEEDBACK ESTIMATION

In this section we describe the machine learning framework we develop in order to predict the *explicit* feedback provided in the form of movie ratings from the decomposed SCR events provided by our EDA decomposition based approach. The ground-truth data of ratings for the movie comes from the user surveys taken immediately following the film.

We compare the prediction accuracy of the EDA based approach to accuracy achieved by using the demographic information provided by the users. Specifically, we use the age and gender information provided by the study participants. Table 2 summarizes the 34 study participants along with their demographic information for the 3 films. While the comparison against demographic information may seem naive, feature-length films are produced and refined to target specific demographic groups [27], therefore we expected a large amount of correlation between demographics and the resulting user responses to the films.

Bagged Decomposition Classification

From the decomposition technique from the prior section, we obtain the time-stamp locations and coefficient values of the SCR events for each user of length T (where $T \gg N$). From this information, we now construct the $[N \times T]$-implicit user response matrix $\mathbf{S}$, such that the matrix element, $S_{i,t_{i,j}} = s_{i,j}$, user i's estimated response given the EDA decomposition at time j.

Figure 8 shows these responses as point intensities for two particularly relevant scenes from Movie A and Movie B. We observe that the SCR events are generally sparse and vary considerably in their intensities. Furthermore, due to the physiological differences across the different users, the SCR events may not be temporally aligned and could also consist of spurious events that may not be relevant to the stimuli in the film being watched.

To mitigate this inherent sparsity in the user response matrix, $\mathbf{S}$, we extract coarse-scale user response information by aggregating into a reduced number of time aggregated bins. Such that for each time bin we record a sum of the SCR coefficient energy for that time period. For the experiments here, we combine each user SCR events over the course of the entire stimulus into five equal-sized bins, denoting the aggregated $[N \times 5]$ user response matrix as $\mathbf{S}_A$.

Combining with the user demographic information, we obtain complete response matrix, $\mathbf{S}_C = [\ \mathbf{S}_A \quad \mathbf{C}\ ]$. Two characteristics are considered, gender and age, from which a $[N \times 2]$ matrix, $\mathbf{C}$, is constructed (where element $C_{i,1}$ is user i's gender and element $C_{i,2}$ is user i's age).

We consider the problem of inferring explicit user feedback information (*i.e.,* film ratings). To classify from the decomposed user responses, $\mathbf{S}_C$, we focus on using bagged classification trees [29]. Bagged classification trees allow us to learn an ensemble of simple tree classifiers over multiple subsamples of a held-out training set. Specifically, to classify a particular user's rating, we use leave-one-out cross validation such that the remaining users are held out as training data. From this collection of training data, a random subsample of training users are chosen and a single classification tree is learned with respect to that training subset ground truth. For example, we may learn that if the response energy in the first time bin is below a learned value, then the user will rate the

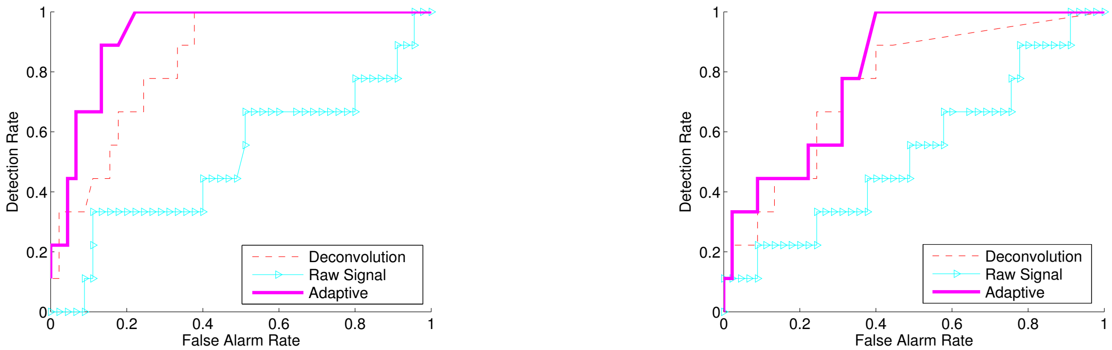

Figure 6. Stimulus classification accuracy for calibration experiment specific events (9 stimulus user-events and 45 non-stimulus user-events) using a 10 second time window : (Left) Gunshot sound, and (Right) Dog barking sound.

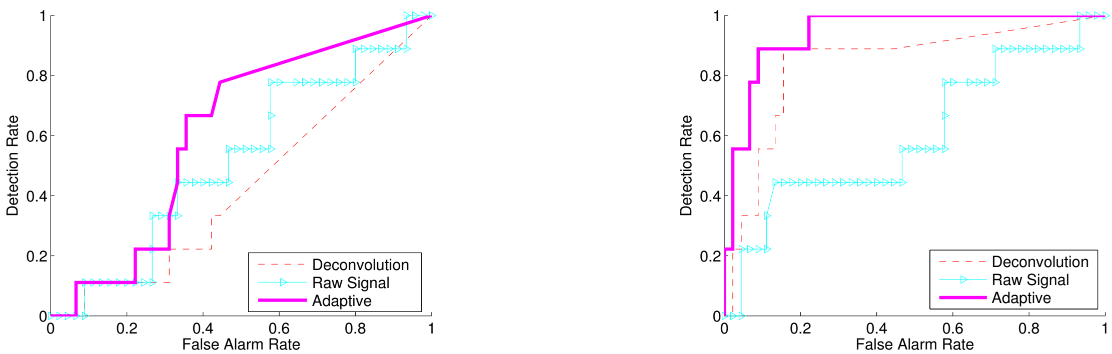

Figure 7. Stimulus classification accuracy for calibration experiment specific events (9 stimulus user-events and 45 non-stimulus user-events) using a 10 second time window : (Left) Photos, and (Right) Train video.

film poorly. At each iteration, in addition to the classification tree, weights with respect to the classification accuracy on the training set are learned. Finally, the specified test user data is used on a weighted combination of all the learned trees to classify the underlying explicit feedback for that user. We perform this bagged classifier approach on both the processed EDA data (the matrix $\mathbf{S}_C$) and the demographics-only information (the matrix $\mathbf{C}$).

Entire Film Rating Classification Experiment

We examine estimating user explicit responses to the question, "On a scale of 1-to-5, how would you rate the film?" using both the EDA-based methodology and comparing it to the demographics-only approach. Responses are registered between "1" and "5". From our **3** films and **34** total viewers, we find 11 users rating films as a "3", 11 users rating the film a "4", and 12 users rating the film a "5". No users rated any of the films either "1" or "2".

We specifically examine classifying users who rated the film a "4" or "5" versus users who rated the film a "3". Justification for examining these two specific classes of users is as follows: *(1)* - Real-world focus groups often place particular emphasis on finding users who rate items in the top two tiers (*i.e.,* rating content a "4" or "5" out of 5 [27], Chapter 2). And *(2)* - For a subset of the films under consideration, in addition to requesting their rating on a scale of one-to-five, we ask the user if they would recommend the film to other people. On this specific subset, we found that 87.5% of the users who rated the film a "4" and 93.33% of the users who rated the film a "5" would recommend that film to another. Meanwhile, only 44.44% of users who rated films a "3" would recommend to other people (with the remaining 55.56% of users saying they would not recommend the film). This change in recommendation rate shows a very clear delineation between users who rate films a "3" against users who ratings "4" or "5" (and few recommendation differences for users with the ratings of "4" and "5").

Using the extracted SCR responses using the adaptive decomposition methodology, the demographics information, and our learned bagged classification trees, we present the detection accuracy (with respect to classifying users who rated films a "4" or "5") for both the EDA-based bagged decomposition methodology and using only demographics information. The accuracy results compared with the underlying user surveys are shown in Table 3. From the table, we find significantly lower detection accuracy for both EDA-based and the demographics-only approach for Movie C. For the demographics-only approach, we find that gender and age groups does not correlate with the resulting user film rating, with the exception of the Age group 40-49 years old, where all three viewers rated the movie positively. Using the EDA-based bagged decomposition methodology, these results can

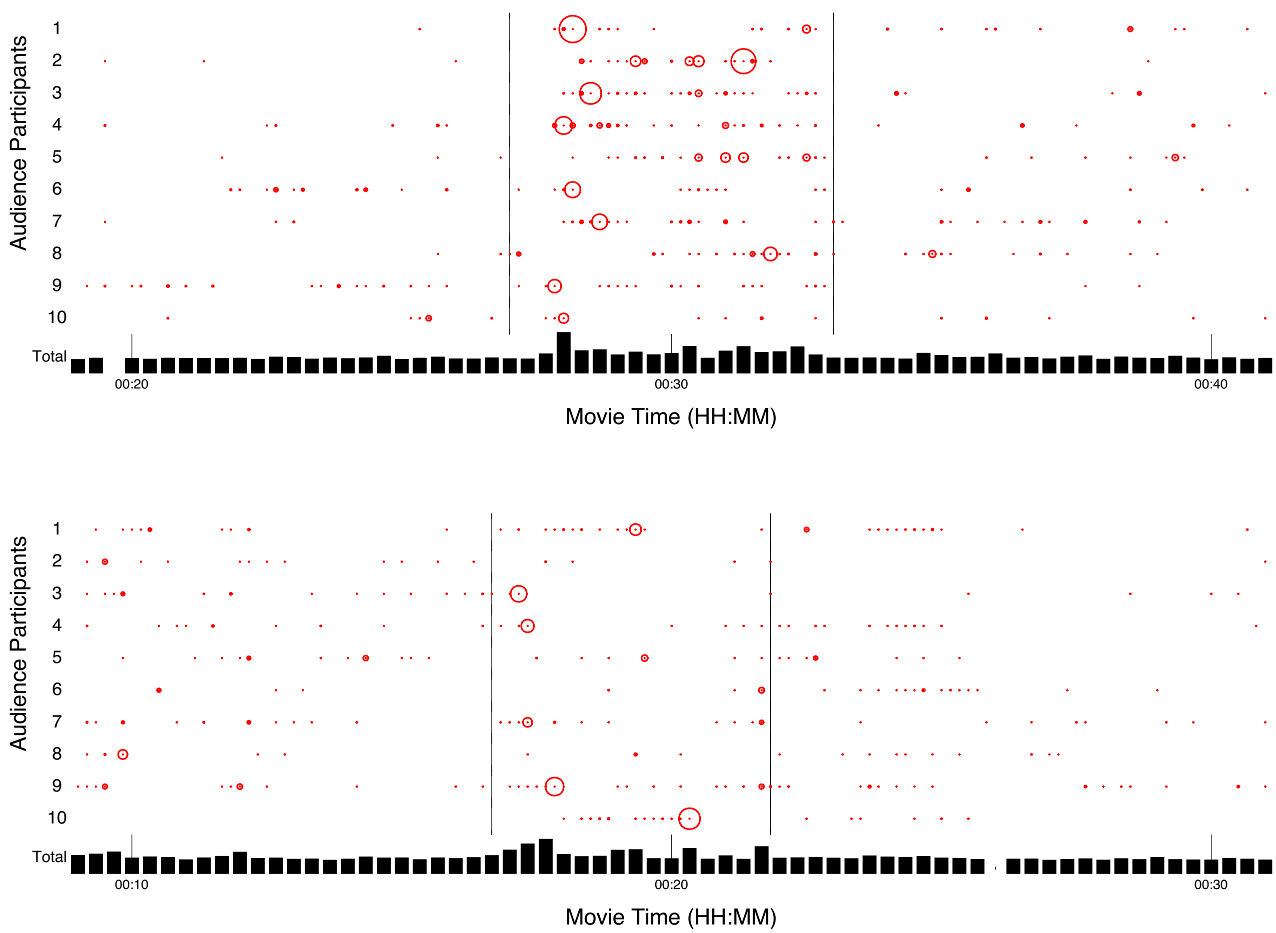

Figure 8. Responses for: (Top) a suicide scene in Movie C, and (Bottom) disaster scene in Movie B. Vertical lines indictate scene boundaries. Point size denote response intensity. Note that response intensities increases significantly, for individuals (top circles) and the aggregate (bottom bars).

be explained by the content of the stimulus. As partially denoted in Figure 8-(Top), this film was particularly violent and contained multiple scenes that while causing significant SCR events, did not correlate with user enjoyment and overall user rating. This demonstrates that the the EDA-based bagged decomposition approach may have difficulty on stimulus that are repellent or disgusting as in Movie C, but possibly including others. We leave automatic detection of these problematic genre- and content-specific stimulus as future work.

We find that the EDA-based technique has high classification accuracy of over 70% for Movies A and B. Meanwhile, for movies A and B, using only the demographics information results in detection accuracy of only 17% and 56% respectively.

CONCLUSIONS

The ability to evaluate viewer reactions to content has applications in areas ranging from advertisement to content production. To this end, we introduce an EDA-based framework for evaluating viewer responses to content. We present a novel signal processing methodology for decomposing user skin conductance response events from raw EDA signals, and show that this technique outperforms the current state-of-the-art on a calibration experiment of simple video stimulus. Finally, we use the estimated response events to classify explicit user ratings to feature films. For two specific films and simultaneous sensing of multiple users in an audience, we find that our biometrics framework allows for significantly more accurate classification compared with only using audience demographic information. However, for a third film, we did not find any benefit, limiting the general application of this approach until the method is further studied. Nevertheless, these results show potential for more efficient market research and improved recommender systems.

REFERENCES

1. Basis. `http://www.mybasis.com/`.

2. BodyMedia. `http://www.bodymedia.com/`.

3. A. Breska, K. Maoz, and G. Ben-Shakhar. Interstimulus Intervals for Skin Conductance Response Measurement. *Psychophysiology*, 48:1–4, 2010.

			Gender-Specific		Age-Specific		
Movie	Detection Technique	Overall Accuracy	Male Accuracy	Female Accuracy	Age 20-29 Accuracy	Age 30-39 Accuracy	Age 40-49 Accuracy
A	EDA Bagged Decomp.	73.33%	65.52%	82.98%	91.67%	70.15%	73.08%
	Demographics	17.14%	3.45%	34.04%	100.00%	8.96%	0.0%
B	EDA Bagged Decomp.	75.60%	77.52%	73.55%	59.49%	94.51%	70.00%
	Demographics	56.80%	43.41%	71.07%	82.28%	9.56%	51.25%
C	EDA Bagged Decomp.	23.14%	31.33%	17.00%	22.47%	20.17%	32.08%
	Demographics	25.71%	30.67%	22.00%	15.73%	26.05%	58.49%

Table 3. Summary of feature-length films experiment results.

4. A. F. Smeaton, and S. Rothwell. Biometric Responses to Music-Rich Segments in Films: The CDVPlex. In *Proc. of CBMI*, pages 162 –168, June 2009.
5. B. Sidis. The Nature and Cause of the Galvanic Phenomenon. *The Journal of Abnormal Psychology*, 5(2):69–74, 1910.
6. C. Kaiser, and R. Roessler. Galvanic Skin Responses To Motion Pictures. *Perceptual and Motor Skills*, 30:371–374, 1970.
7. C. Latulipe, E. Carroll, and D. Lottridge. Love, Hate, Arousal and Engagement: Exploring Audience Responses to Performing Arts. In *Proc. of ACM SIGCHI*, pages 1845–1854, April 2011.
8. C. Lim, C. Rennie, R. Berry, H. Bahramali, I. Lazzaro, B. Manor, and E. Gordon. Decomposing Skin Conductance Into Tonic and Phasic Components. *International Journal of Psychophysiology*, 25:97–109, 1997.
9. C. Lisetti, and F. Nasoz. Using Noninvasive Wearable Computers to Recognize Human Emotions from Physiological Signals. *EURASIP Journal on Applied Signal Processing*, 11:1672–1687, 2004.
10. D. Alexander, C. Trengove, P. Johnston, T. Cooper, J. August, and E. Gordon. Separating Individual Skin Conductance Responses in a Short Interstimulus-Interval Paradigm. *Journal of Neuroscience Methods*, 146:116–123, 2005.
11. D. Bach, J. Daunizeau, N. Kuelzow, K. Friston, and R. Dolan. Dynamic Causal Modeling of Spontenous Fluctuations in Skin Conductance. *Psychophysiology*, 48:1–6, 2010.
12. E. Haapalainen, S. Kim, J. Forlizzi, and A. Dey. Psycho-physiological Measures For Assessing Cognitive Load. In *Proc. of ACM UBICOMP*, pages 301–310, September 2010.
13. G. Davis, S. Mallat, and M. Avellaneda. Greedy Adaptive Approximation. *Journal of Constructive Approximation*, 13:57–98, 1997.
14. M. Gardner. Mood States and Consumer Behavior: A Critical Review. *Journal of Consumer Research*, 12(3):281–300, December 1985.
15. H. Critchley, R. Elliott, C. Mathias, and R. Dolan. Neural Activity Relating to Generation and Representation of Galvanic Skin Conductance Responses: A Functional Magnetic Resonance Imaging Study. *The Journal of Neuroscience*, 20(8):3033–3040, April 2000.
16. H. Gunes. Automatic, Dimensional and Continuous Emotional Recognition. In *International Journal of Synthetic Emotions*, volume 1, 2010.
17. H. Lu, D. Frauendorfer, M. Rabbi, M. Mast, G. Chittaranjan, A. Campbell, D. Gatica-Perez, T. Choudhury. StressSense: Detecting Stress In Unconstrained Acoustic Environments Using Smartphones. In *Proc. of ACM UBICOMP*, pages 351–360, September 2012.
18. J. Fleureau, P. Guillotel, and Q. Huynh-Thu. Physiological-Based Affect Event Detector for Entertainment Video Applications. *IEEE Transactions on Affective Computing,*, 3(3):379 –385, July 2012.
19. J. Healey, L. Nachman, S. Subramanian, J. Shahabdeen, and M. Morris. Out of the Lab And Into The Fray: Towards Modeling Emotion In Everyday Life. In *Proc. of Pervasive Computing*, pages 156–173, May 2010.
20. M. Benedek, and C. Kaernbach. Decomposition of Skin Conductance Data by Means of Nonnegative Deconvolution. *Psychophysiology*, 47:647–658, 2010.
21. M. Dawson, A. Schell, and D. Filion. The Electrodermal System. *Handbook of Psychophysiology*, pages 200–223, 2000.
22. M. Tarvainen, A. Koistinen, M. Valkonen-Korhone, J. Partanen, and P. Karjalainen. Analysis of Galvanic Skin Responses with Principal Components and Clustering Techniques. *IEEE Transactions on Biomedical Engineering*, 48(10):1071–1079, October 2001.
23. M. Tkali, A. Koir, and J. Tasi. Affective Recommender Systems: The Role Of Emotions In Recommender Systems. In *Proc. of ACM Decisions@RecSys*, October 2011.
24. P. Lang. The Emotion Probe: Studies of Motivation and Attention. *American Psychologist*, 50(5):372–385, 1995.
25. R. Calvo, and S. D'Mello. Affect Detection: An Interdisciplinary Review of Models, Methods and Their Applications. In *IEEE Transactions of Affective Computing*, volume 1, January-June 2010.

26. R. Mandryk, M. Atkins, and K. Inkpen. A Continuous and Objective Evaluation of Emotional Experience with Interactive Play Environments. In *Proc. of ACM SIGCHI*, pages 1027–1036, April 2006.
27. R. Marich. *Marketing to Moviegoers: A Handbook of Strategies and Tactics, Third Edition*. Southern Illinois University Press, 2013.
28. S. Rothwell, B. Lehane, C. Chan, A. F. Smeaton, N. O'Connor, G. Jones, and D. Diamond. The CDVPlex Biometric Cinema. In *Adjunct Proc. of Pervasive*, May 2006.
29. T. Hastie, R. Tibshirani and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*. Southern Illinois University Press, 2009.
30. T. Nishiyama, J. Sugenoya, T. Matsumoto, S. Iwawe, and T. Mano. Irregular Activation of Individual Sweat Glands in Human Sole Observed by a Videomicroscopy. *Automatic Neuroscience: Basic and Clinical*, 88:117–126, 2001.