Electrodermal activity applied to violent scenes impact measurement and user profiling

J. Fleureau and C. Penet and P. Guillotel and C.H. Demarty

Technicolor R&I
Cesson-Sévigné, France
surname.name@technicolor.com

Abstract—Identifying violent scenes in a movie may be of high interest as soon as the associated content has to be shown to a specific audience, as children for instance. However, defining what a violent scene is as well as extracting the violent excerpts from the only audiovisual cue are two hard tasks. In this article, we propose a pilot study to evaluate the interest of an approach based on the use of the electrodermal activity (EDA) to address these problems in an objective and human-centered manner. Assuming a consumer context, we especially focus on the use of a commercial sensor to capture the EDA. A comparison with a more professional device is initially performed to validate the accuracy of the commercial sensor. Two main aspects of the violent scene impact measurement with the EDA are then discussed. To approach the violent scene detection problem, a methodology to correlate the EDA information with manual annotations of violent excerpts is first proposed. The user sensitivity to violence is then addressed and an approach to identify different profiles in the audience is qualitatively proposed. The advantages and limitations of such an approach are discussed and potential improvements are finally proposed.

Index Terms—Violent Scenes, Affective Computing, Film Annotation, Physiological Signals.

I. INTRODUCTION

Considering the ever increasing number of available multimedia creations and access platforms, the chance for a child to accidentally watch undesirable violent media is high and naturally raises the question of content control / filtering [1]. Addressing such an issue is thus of high interest in the context of media broadcasting / internet video distribution and may have important applications in a consumer environment.

So far, this detection of violent scenes in a professional framework is still manual, requires a lot of human resources and time, and is very sensitive to the definition of violence ([2], [3], [4]) that is initially adopted. This definition as well as its interpretation are very specific to one given person and, depending on these choices, the violent scenes may be either under or over-detected, leading to poor final results.

Alternative automatic strategies that are proposed in the literature are generally based on the use of machine learning algorithms trained on manually annotated databases of audiovisual features ([5], [6], [7]). Most of these approaches suffer from the lack of a common, consistent and objective database. For example, in [5], only four movies are used to detect actions. In order to cope with user sensitivity to violent content and to be as relevant as possible, several annotations should be created (one per user or group of user) which would imply a tremendous amount of work. To the best of our knowledge, the first attempt to build a consistent dataset and use it in a benchmark is the MediaEval 2011 Affect Task on violent scene detection in Hollywood movies [8] and used in [7] for example. The violence definition taken for this task is described in section IV. Moreover, the automatic detection is also limited by the nature of the information used for the violent scene detection. Indeed, such approaches generally make use of only audio and / or video features that are very low-level compared to the high-level nature of the feeling of violence.

As a consequence, it comes that at least three main central issues have to be addressed to make an acceptable automatic (or semi-automatic) violent scene detector: i) decide of a relevant definition of what is considered as violent, ii) have access to a sufficiently high-level description of the movie to be able to find violent excerpts, and iii) be as less invasive as possible for the end-user who should not be disturbed by any interaction.

To reach all of these three requirements, we propose here to study the interest of using the physiological signal called electrodermal activity (EDA, also known under the name of Galvanic Skin Response).

The latter measures the local variation of the skin electrical conductance in the palm area which is known to be highly correlated with user affective reactions ([9], [10]). This signal thus embeds high-level human-centered information and might be used to provide a user-specific feedback of what is currently felt as violent or not by the end-user. Moreover, consumer devices and especially armbands ([11], [12]) are available on the market to record this signal in a convenient and non-invasive way and may offer a nice possibility to seamlessly track the user state and accordingly censor some parts of the content.

Despite these obvious qualities, it also appears that: i) the EDA is sensitive to every kind of user affective reactions and not only to the reactions to violence and ii) the consumer devices to record the EDA are generally designed to be used in a sport or activity manager / tracker context that may be not sufficient to track the signal variations related to the affect changes.

The use of such a signal in the context of violence detection and with consumer sensor is thus not as clear as we could expect. That is why we propose here to evaluate the interest of the EDA measured with consumer devices for violent scene detection / annotation. In line with these considerations, we
also propose to explore the possibilities offered by such an input to build a profile of the sensitivity of one audience member to violence.

The paper is organized as follows. Section II describes related works existing in the literature. The methodology used in this paper to measure the impact of the content on the user from the EDA signal is described in section III. Then, the experimental parts including setup and results are respectively presented in section IV and V. Finally conclusions and perspectives are provided in Section VI.

II. RELATED WORKS

The electrodermal activity is one of the most frequently used psychophysiological evaluations in psychology research [10]. Making use of the EDA signal and more generally of physiological signals to measure the user reaction in the context of violence is thus definitively not new. A lot of studies from the literature have already examined the effects of film content by measuring physiological responses to filmed stimuli.

In [13], a pilot study of galvanic skin response to motion picture violence is conducted and shows the correlation of such a signal with reaction to violent movies. In line with these works, a study of the effects of social class, gender, and personality on physiological responses to filmed violence is proposed in [14], whereas physiological responses are used to show the link existing between the sex-violence ratio of a viewer. Combining the events of a group of persons that watched the same content should provide a mean physiological profile highly correlated to a given annotation of the violent movie.

Numerous and more recent works generalize the use of physiological signals (including EDA) to predict the user affective reaction or state in the context of movie viewing. In [16], Soleymani et al. record physiological data to characterize the emotional content of video clips on the basis of correlation computations. In [17] a complete setup is proposed to estimate the affective state of a user watching emotion-eliciting movie selection using supervised classification techniques and noninvasive wearable computers. In [18] or [19], real-time approaches of the emotional state detection are presented. In [18], Fleureau et al. especially focus on the detection of affective events (i.e. fast increase of the emotional arousal) during a video viewing session and propose a methodology to label the affective stream of the audience member with these events and their associated valence.

So, even if a lot of material may be found in the literature justifying the interest of using the EDA in our context, as far as we know: i) no work addresses the use of such a signal in the context of violent scene annotation, and ii) few works make use of consumer devices to record the biological data. These points are now considered thereafter.

III. IMPACT MEASUREMENT FROM EDA

A. Methodology overview

The main object of this article is to provide elements to quantify the interest of using the EDA to help the objective annotation of violent scenes in a movie. We have previously explained that the EDA contains high-level information that may inform about the affective state of the audience member and especially about his reactions to violence. As shown in [18], it is possible to identify events in the emotional flow. These events are closely related to phasic changes of the EDA [10], i.e. fast changes in the skin conductance (see figure 1).

Practically, in the context of violent movies, these events should have strong correlations with the violent parts of the films or at least, to the excerpts that mainly impacted the viewer. Combining the events of a group of persons that watched the same content should provide a mean physiological profile highly correlated to a given annotation of the violent movie. This mean physiological profile could be very helpful for an annotator to identify on the time-line the parts that should be annotated as violent. In this case, the difficulties related to the subjectivity of the definition of violence could be partly reduced.

Such an approach will be conducted and described in sections IV and V.

B. Event detection

However, to build a mean physiological profile, one has firstly to identify the events of the emotional flow described here as phasic changes of the EDA. To do so, a very simple empirical approach is proposed that may be summarized by the following scheme:

- EDA answer recording during the movie viewing session
- Low-pass filtering to only keep the spectral content below 1Hz
- Identification of the locations of the local extrema of each minimum being followed by a maximum. For constant portions, the central position is retained. → \([Ids^{MIN}, Ids^{MAX}]\).
- Final construction of the “event” signal \(s_{EDA}^{Event}\) which is null everywhere excepted at the local maxima locations. In every local maxima location \(Ids^{MAX}(i)\), it is computed as

\[
s_{EDA}^{Event}(Ids^{MAX}(i)) = \sigma(Ids^{MAX}(i) - Ids^{MIN}(i))
\]

where \(\sigma\) is a sigmoid squash function comprised between 0 and 1 and which satisfies

\[\sigma(\xi) = 0.10 \text{ and } \sigma(2\xi) = 0.50\]

with,

\[
\xi = \mu_{1/2}(s_{EDA}^{MAX}(Ids^{MAX}) - s_{EDA}^{MAX}(Ids^{MIN}))
\]

if \(\mu_{1/2}\) designs the median operator.

Such an approach provides a non-linear transform of the original signal where each sample value is the confidence to have an EDA event at this location. The parameters of the squash functions are empirically set and provide visually coherent results on the captures that we performed (see Figure 1).
Fig. 1. Example of event detections applied to two EDA signals recorded with a consumer BodyMedia Armband sensor (left) and a research BIOPAC MP36RW amplifier (right). Top: EDA raw signals. Bottom: “event” confidence.

IV. EXPERIMENTAL SETUP

A. User tests
To be able to evaluate the interest of using the EDA following the strategy proposed in section III-A, a complete experimental setup has been designed. More precisely, five 20-minute movie excerpts, extracted from five movies belonging to the MediaEval 2011 Affect Task [8] database were selected, namely: “Save Private Ryan”, “Kill Bill 1”, “Reservoir Dogs”, “I am Legend” and “Billy Elliot”. We chose these movies as they belong to very different movie genres. Each excerpt contains both violent and neutral content.

These videos were shown in a random presentation order using a dedicated testing framework to 10 different subjects consisting of 8 males and 2 females aged between 20 and 40 years old. Each viewer was comfortably seated in a dedicated dark and quiet room. Headphones were used to isolate the subjects and increase their immersiveness.

During the viewing of each clip, the EDA was captured using two kind of sensors. A research BIOPAC MP36RW amplifier was first used to capture the signal on the non-dominant hand at a 100Hz sample frequency. The two associated sensor electrodes were respectively placed on the second and third medial phalanges. Complementarily, a consumer BodyMedia Armband sensor was placed on the palm area of the other hand and recorded the skin conductance at a 32Hz rate. The entire recording process was automatically synchronized and controlled by dedicated operations.

B. Annotations
Two types of annotations were used for each movie excerpt. First, the MediaEval 2011 Affect Task annotations [8] were made by three persons in a pipeline way according to the following definition: violence is defined as “physical violence or accident resulting in human injury or pain”. With this definition, taken to be as objective as possible, several moments that might be considered violent are not considered violent, and vice versa. For instance, a dog fight is not considered violent as it does not involves human interaction. A person limping might also be considered as violent if it can be shown that he is in pain. The last important effect of this definition is that, in a gunfight for instance, only small segments showing someone being shot are considered violent as the shots that do not touch anyone do not result in human injury or pain. The second definition we took is less objective, as we asked one of the annotators to indicate which parts of the video excerpt he would not want an 8-year old child to see.

V. RESULTS AND DISCUSSIONS

A. Consumer versus professional devices
A first point that we would like to explore in this study deals with the use of commercial sensors to track the changes in the EDA related to affective reactions. In this context the consumer BodyMedia Armband sensor [11] has been chosen and compared to the BIOPAC MP36RW research equipment. The commercial sensor records the EDA using two dry electrodes embedded on the back of an armband whereas the second one makes use of dedicated silver electrodes which contact with skin is enhanced by the mean of an isotonic gel.

As it has been mentioned previously, the BodyMedia sensor is designed to be used in a sport or activity manager / tracker context which explains the armband packaging. The latter should be normally fastened on the wrist or on the upper part of the arm. However, as it can be qualitatively observed in figure 3, such a positioning leads to poor signal quality (in the context of psychophysiological measurements) whereas a positioning in the palm area or around the fingers (see figures 2 and 3) seems to provide coherent results.

Adopting the latter configuration, a deeper quantitative test has been then performed to have a better idea of the reliability of the commercial sensor data. To do so, for each subject and each movie, the events associated to each sensor have been computed following the method described in III-B (see figure 3). Taking the BIOPAC system as the reference, the
As it can be observed, the sensitivity of the commercial sensor tends to increase with the threshold value increase. In other words, events with higher confidence tend to be detected by both sensors. However the maximum sensitivity value do not exceed 0.7 which logically suggests that the research device is more sensitive than the commercial one. Regarding the specificity, the value seems to be quite independent of the threshold and is quite high (above 0.75). It thus suggests that the commercial device do not detect extra false event. As the total number of false events is much larger than the number of true events, the accuracy is thus very correlated to the sensitivity curve and naturally also shows good performances.

As a consequence, this first analysis tends to suggest that the BodyMedia sensor, when located at the appropriate place, provides relevant information and may be used as a consumer sensor for the further part of the study.

**B. Correlation with manual annotations**

We now address the study of correlation of the EDA signal with violent scene annotations. As suggested in III-A, for each movie, a mean physiological profile is computed by averaging the result of the event detection of each subject (following the method described in section III-B). More precisely and additionally, before the averaging step, the “event” signal of each subject is convolved with a Gaussian kernel characterized by a 5-second standard deviation. This step aims at temporally smoothing the answers of the different subjects in the final averaged profile. The values below 0.5 are then discarded (set to 0) to only preserve the most representative and common part of the mean profile.

The latter, obtained for each movie, is now compared with the two different manual annotations introduced in section IV.

1) MediaEval annotations: The five sequences annotated within the MediaEval 2011 Affect Task corpus are depicted in figure 5. As described in section IV this annotation has been done following a quite restrictive criteria. A global observation of the five sequences clearly demonstrates that correlations between the two signals exist, but also that significant differences are visible.

Concerning the good correlations that we can point out, one can especially highlight the “Billy Elliot” sequence where the only annotated violent scene corresponds to the fight between the father and his son, at the end of this excerpt. As one can see, in this sequence, the annotation and the physiological data match quite well from the beginning (quite neutral) to the end. Regarding the “I Am Legend” excerpt, the subjects seem to have jointly reacted at the end of the sequence (around 900s) which corresponds to a zombie attack, also annotated by the MediaEval annotators. In the “Reservoir Dogs” excerpt, the main annotations appear between 500s and 900s and corresponds to various parts of the torture of a policeman on his chair. Even if the physiological reactions are not exactly synchronized with those annotations, important biological answers may be nevertheless observed in this same part of the time-line. We could also finally add the same kind of comments regarding the “Kill Bill I” excerpt where two big fight scenes are labeled between 400s-550s and 650s-750s and also coincide with important physiological activities.

Despite those interesting correlations, it is however clear that some important differences exist between the physiological profile and the annotated data. It is for example the
case of the part located between 600s and 700s of the “I am Legend” clip. Here an important reaction may be observed and corresponds to the scene where the hero is forced to kill his own dog because he realized that it has been contaminated by the zombies. This scene is affectively strong but it is not in line with the definition adopted by the MediaEval annotators. At the end of the “Reservoir Dogs” excerpt (1000s), one can also observe a strong physiological response that is not correlated with annotations. This part corresponds to a scene where a blooding policeman is lengthily filmed. This is a disgusting scene that could be censored, but it is not violent according to the MediaEval definition. At the beginning of the “Kill Bill 1” excerpts (150s-200s), a strong biological answer is also visible. It corresponds to a live music show where a swinging rock and roll music peace is played. Here the EDA answer is relevant but not related to a reaction to violence. Finally, at the end of the “Saving Private Ryan” excerpt (around 1000s), the audience member has the feeling that the private Ryan could be finally found and thus a biological reaction is observed. It is here again an example of the fact that the EDA is strongly related to any kind of affective reaction and not only to reaction to violence.

2) Extended annotations: As described in section IV the extended annotation is less objective than the previous one. Here the annotator was asked to indicate which parts of the video excerpt he would not want an 8-year old child to see. Comparatively to the previous annotations, much more time periods have been annotated and, for instance, in the “Reservoir Dogs” excerpt, almost all the sequence have been censored (see figure 6). Previous events such as the dog death (“I Am Legend”, 600s) are now annotated and some violent scenes previously split in several discrete parts have been now gathered in large annotated blocks (“Kill Bill 1”, 400s-550s, 600s-750s or “Reservoir Dog”, 500s-1000s, “BillyElliot” 600s-700s). As a consequence, a better consistency with the biological responses may be noted in those parts.

However, reactions of pleasure or excitement detected by the physiological signals are still logically not annotated.

3) Discussion: As a consequence of all those previous observations, it appears that the EDA signal may be a useful tool to help for the annotation of violent scenes but it cannot be used alone.

On the one hand, the EDA provides the unconscious physiological reaction of a viewer, and not the subjective interpretation of an annotator. Therefore, the physiological responses may allow to preselect some part of the time-line that would be likely to contain violent content. Such parts can be quite accurately delineated and may prevent an annotator to hesitate about the fact to put or not an annotation. By using the extra EDA information, the annotation process could be thus sped up and qualitatively improved.

On the other hand, the skin conductivity changes are sensitive to various kinds of affective stimulations and not only to violence. Pleasure or excitement may cause EDA events and may introduce false positive in the detection process. As is, this only biological information is not sufficient to completely characterize a violent scene, and it may be completed by a human intervention or the use of another source of information (possibly another kind of biological input).

C. User profiles

In the previous section, we have shown how the EDA signal was correlated with the user emotional state when he is watching a movie content. The number of events detected in this physiological input is consequently, for a given user, an image of his own sensitivity (including his sensitivity to violence). A complementary interesting information that we address here deals with the possibility to derive some
profiles regarding the different users that we considered in our experimental setup.

To do so, we computed, for each user, the cumulated amount of events (see section III-B) for the all five excerpts he watched. The associated results are depicted in figure 7.

As one can observe, it appears that, despite the few number of subjects, some different profiles may be extracted from the 10 audience members involved in the user tests. For instance, subjects 4 and 6 are probably more sensitive than the others, whereas subjects 2, 7 and 8 are probably less expressive people.

Depending on those profiles, one could define different levels of violence to adapt the final censored parts to each individual.

VI. CONCLUSION AND PERSPECTIVES

In this article, we have evaluated the interest of using the EDA measured with consumer devices to annotate violent scenes and to build a profile of the sensitivity of a given viewer. After showing that, under certain conditions, a consumer sensor may be used to track the changes in the user affective state with a reasonable accuracy, we demonstrated, on the basis of concrete examples, how the EDA signal may be a useful tool to accelerate and enhance the process of violence detection.

More precisely, by identifying events in the skin conductivity changes, we pointed out that: i) this signal provides an objective measure of the user reaction that may help the annotator to accelerate, simplify and make more reliable and robust the detections of events, and ii) this signal makes possible the construction of a user sensitivity profile that may allow a personalization of the violence detection by, for instance, adapting the level of censor to apply to a given user.

However, this biological response, inherently sensitive to every kind of effective changes including excitement or pleasure, cannot be used alone and has to be completed with extra information to make a relevant violence annotation. This information could be found in the audiovisual cue or in other kinds of physiological inputs. Further works in line with these suggestions have now to be done to overcome this limitation.

REFERENCES