Aesthetic Highlight Detection in Movies Based on Synchronization of Spectators' Reactions

MICHAL MUSZYNSKI, University of Geneva, Switzerland
THEODOROS KOSTOULAS, University of Geneva, Switzerland, Bournemouth University, United Kingdom
PATRIZIA LOMBARDO, University of Geneva, Switzerland
THIERRY PUN, University of Geneva, Switzerland
GUILLAUME CHANEL, University of Geneva, Switzerland

Detection of aesthetic highlights is a challenge for understanding the affective processes taking place during movie watching. In this paper we study spectators' responses to movie aesthetic stimuli in a social context. Moreover, we look for uncovering the emotional component of aesthetic highlights in movies. Our assumption is that synchronized spectators' physiological and behavioral reactions occur during these highlights because: (*i*) aesthetic choices of filmmakers are made to elicit specific emotional reactions (e.g. special effects, empathy and compassion toward a character, etc.) and (*ii*) watching a movie together causes spectators' affective reactions to be synchronized through emotional contagion. We compare different approaches to estimation of synchronization among multiple spectators' signals, such as pairwise, group and overall synchronization measures to detect aesthetic highlights in movies. The results show that the unsupervised architecture relying on synchronization measures is able to capture different properties of spectators' synchronization and detect aesthetic highlights based on both spectators' electrodermal and acceleration signals. We discover that pairwise synchronization measures perform the most accurately independently of the category of the highlights and movie genres. Moreover, we observe that electrodermal signals have more discriminative power than acceleration signals for highlight detection.

CCS Concepts: • **Human-centered computing → Social media**; • **Theory of computation → Pattern matching**;

Additional Key Words and Phrases: synchronization, dynamical systems, physiological signals, behavioral signals, aesthetic experience, aesthetic highlight detection, video summarization, affective computing

ACM Reference Format:
Michal Muszynski, Theodoros Kostoulas, Patrizia Lombardo, Thierry Pun, and Guillaume Chanel. 2010. Aesthetic Highlight Detection in Movies Based on Synchronization of Spectators' Reactions. *ACM Trans. Web* 9, 4, Article 39 (March 2010), 23 pages. https://doi.org/010.1145/3175497

Authors' addresses: Michal Muszynski, University of Geneva, 24 rue du General-Dufour, Geneva, 1211, Switzerland, michal.muszynski@unige.ch; Theodoros Kostoulas, University of Geneva, Switzerland, Bournemouth University, Poole House, Talbot Campus, Fern Barrow, Poole, Bournemouth, BH12 5BB, United Kingdom, tkostoulas@bournemouth.ac.uk; Patrizia Lombardo, University of Geneva, 24 rue du General-Dufour, Geneva, 1211, Switzerland, patrizia.lombardo@unige.ch; Thierry Pun, University of Geneva, 24 rue du General-Dufour, Geneva, 1211, Switzerland, thierry.pun@unige.ch; Guillaume Chanel, University of Geneva, 24 rue du General-Dufour, Geneva, 1211, Switzerland, guillaume.chanel@unige.ch.

1559-1131/2010/3-ART39 $15.00
https://doi.org/010.1145/3175497

1 INTRODUCTION TO AESTHETIC HIGHLIGHTS IN MOVIES

Aesthetic experience is one of the most substantial but also one of the poorly defined concepts in art. It can be defined as a special kind of relationship between a person and an artistic object in which a particular object absorbs the person's mind and overshadows other surrounding objects and events [56]. Aesthetic experience is also considered as being a subjective part of an artistic exposure and corresponds to a feeling of being engaged with a piece of art. Aesthetic experience is intended to be different from the everyday experience [22, 69], and to be a special state of mind in which attention of a person is focused on an artistic object while all other common objects, events, and everyday concerns are overshadowed. There are also different concepts of aesthetic experience, for example, an effortless mental energy flow induced by the awareness of agreement between incoming information and our goals [19, 20]. Futhermore, the concept of peak experience assumes that attention is fully focused on a particular object, while the object is seen as separated from its everyday purpose [57]. The concept of absorption refers to having episodes of amplified attention. It links the subjects' mental and executive resources [75]. Also, aesthetic experience is referred to creative processes occurring in art [47]. The creative action can happens when ambiguous concepts are assembled into a new whole object, for example, an old bicycle seat is mounted next to handlebars in the "Bull's Head" (Pablo Picasso).

An investigation of an inner affective state of a person, who is exposed to an artistic object, can provide insight into understanding of humans' engagement with art, humans' emotions and some features of artistic objects that affect a personal experience. People can be exposed to different pieces of art, such as paintings, sculptures, jewelry, images, music, video games, films. Aesthetic emotions as a part of aesthetic experience are elicited during an exposure to an artistic object and measurable in physiological and behavioral reactions of a person. Recent work has attempted to establish the link between aesthetic and everyday emotions [43]. Understanding spectators' affective responses to movies with aesthetic values is a challenge [74] and requires to model spectators' multimodal responses in a social context of watching movies.

The main focus of this paper is to understand spectators' responses to aesthetic highlights in full-length movies which correspond to scenes with high aesthetic values in terms of form and content. These scenes are constructed on purpose by the moviemakers in order to establish a connection between the spectators and the movie and to allow spectators to be engaged with the movie (including a high level of arousal). This research can make contributions to many applications, such as aesthetic scene detection, aesthetic scene design, video summarization and movie recommendation systems are able to predict the rating of aesthetic content.

In this paper we investigate spectators' responses to aesthetic highlights in a social context when spectators watch a movie together. We assume that spectators can display similar behaviors and have similar physiological reactions when they are watching a movie together because: (*i*) aesthetic choices of filmmakers are made to elicit specific emotional reactions (e.g. special effects, empathy and compassion toward a character, etc.) and (*ii*) watching a movie together causes spectators' affective reactions to be synchronized through emotional contagion [39]. For these reasons we take on gaining insight into synchronization among multiple spectators.

In order to uncover relations between an occurrence of aesthetic highlights in films and multiple spectators' affective states, we address the following research questions:

1. Do aesthetic highlights elicit emotions in movie audiences?

2. Can the level of synchronization among spectators' reactions be used to detect the different categories of aesthetic highlights?

3. If it is possible, which of these synchronization measures are the most reliable to efficiently detect aesthetic highlights?

Below we emphasize the main contributions of our paper, highlighting the novelty compared to our previous work [60]:
- We provide insight on emotional components of aesthetic highlights. We discover the direct link between emotional dimensions (arousal and valence space) and aesthetic highlights. There has been no previous work formally addressing the relationship between movie audiences' emotions and aesthetic highlights in movies.
- We investigate the relationship between different approaches to synchronization estimation, such as pairwise, group and overall synchronization measures to gain insight on multiple spectators' reactions. There have not been comprehensive and comparative studies on synchronization measures including multiple spectators' physiological and behavioral responses.
- We use the level of synchronization of movie audiences' electrodermal and acceleration measurements to detect aesthetic highlights. Then, we find the pairwise approach to synchronization that performs aesthetic highlight detection very efficiently compared to other measures through several movie genres.
- We create one of the largest database of aesthetic highlight annotations which will help to study movie audiences' responses to aesthetic content. This database consists of 30 full-length movies derived from 9 movie genres: action, adventure, animation, comedy, documentary, drama, horror, romance and thriller.

In section 2 we discuss related work and if there is a need to study synchronization measures in the context of processing of humans' affective states. In section 3 we detail architecture of our unsupervised highlight detection system and adaptation of synchronization measures to process spectators' physiological and behavioral signals. In section 4 we describe an annotation process of aesthetic highlights including evaluation of annotations in terms of evoking emotions (arousal and valence). In section 5 we present the results with their interpretation. In section 6 we provide discussion on the main results of our studies and the future work.

2 RELATED WORK

In the area of highlight detection many studies have been focused on analysis of audio-visual features of movies and videos. A fuzzy inference system to summarize the content of broadcast soccer videos using an on-demand feature extraction was implemented in [70]. Besides, a multi-task deep visual-semantic embedding model that automatically select query-dependent video thumbnails with regard to visual and side semantic information was developed [55]. An unsupervised learning of highlights from videos using generic deep learning features was proposed in work [78]. The approach is computationally efficient and accurate in characterizing both appearance and motion of objects in videos. Moreover, a pairwise deep ranking model was used to learn the relationship between highlight and non-highlight segments for video summarization [79]. Furthermore, research on detection of violent scenes in movies uncovered the relevant features: short time audio energy, motion component, and shot words rate for violent scene classification [33]. In other studies [12] the face, blood and motion information was integrated successfully to determine whether action scenes have violent content or not. Also, in recent work [28, 62] audio event and voice activity detection in movies were investigated using Bayesian networks and Long Short Term Memory recurrent neural networks, respectively. However, there are no studies have focused on the detection of aesthetic highlights in movies.

More than a decade ago, an initial attempt to detect affective events in video data applying Hidden Markov Models (HMMs) to color, motion and shot cut rate features was discussed [44]. Also, HMMs were used for detection of video affective content and audio emotional events in [77]. Furthermore, Support Vector Machine classifiers were applied to low-level audio-visual features to build a classification system of affective scene in movies [76]. Besides, affective video content was mapped

into an arousal-valence space using low level features from video data [38]. From the affective computing point of view, deep learning and transfer learning were applied in the context of emotion prediction in movies [3]. In [74], the authors carried out baseline studies on modeling a relation between a large set of low-level computational features (i.e., visual, auditory, and temporal) and perceptual stylistic, aesthetic and affective attributes of selected movie clips. Nevertheless, they did not investigate affective and aesthetic characteristics of full-length movies. This did not allow them to take into account the structure and the context of movie shots and scenes made on purpose by filmmakers.

In the meantime, another researchers have attempted to investigate emotion recognition in responses to multimedia content using electroencephalography (EEG) signals, peripheral physiological signals and facial expressions [52, 71] or movie genre was identified based on magnetoencephalography (MEG) signal recorded in a control environment [31]. In [72], the authors investigated affective ranking of movie scenes using physiological signals and content analysis, separately. However, they did not study fusion or integration of physiological or behavioral signals of multiple spectators.

In [54], the authors introduced a weighted mean galvanic skin response profile among spectators. Some efforts were made to create an affective profile of people who are exposed to movie content using a single modality, such as electro-dermal activity [30] or facial expression of viewers [42]. In [14], the authors also proposed to apply a physiological linkage to spectators' signals for determining highlights in movie scenes. Movie audiences could not interact among themselves because they were separately watching a movie without any social context. However recent studies on individuals' emotional responses and their physiological signals in the social context reported that emotional experiences are shared during watching emotional movies together [34].

Most recent work on aesthetic highlight detection in movies defined and estimated a reaction profile of spectators for identification and interpretation of aesthetic scenes [49], or estimated physiological and behavioral changes of spectators exposed to aesthetic content using the dynamic time warping [48, 50]. Furthermore, the manifold representation of multiple spectators' physiological signals was proposed to measure a level of synchronization among them [59], and the periodicity score was applied to physiological and behavioral signals to establish synchronization among groups of spectators [60]. Moreover, synchronization measurement has become an important tool for affective multimedia content analysis. It decodes information included in humans' physiological and behavioral signals [14, 31, 52, 54]. However, there is a lack of comparative study on synchronization measures for analysis of social interactions and highlight detection.

3 DETECTION OF AESTHETICS HIGHLIGHTS IN MOVIES

In this work we assume that physiological and behavioral responses of spectators in the context of watching movies together can be used to detect aesthetic highlights in movies. Electrodermal activity and acceleration measurements are selected because of two factors: (*i*) the utility and suitability of these signals for emotion and behavior assessments [30, 48, 54] (*ii*) the limitation of available resources (running a large scale experiments constrains the number of modalities that can be recorded).

In order to detect aesthetic highlights and gain insight on spectators' responses to aesthetic stimuli, we propose an unsupervised highlight detection system based on physiological and behavioral reactions of spectators watching movies together, as shown in Figure 1. It is composed of three parts: signal preprocessing, a synchronization estimation and detection based on a synchronization level. Filtering and time windowing are included in the signal preprocessing, while synchronization measures are used for the synchronization estimation and highlight detection.

We formulate highlight detection as a binary classification problem (highlight and non-highlight

class) to respond to our second and third research questions (see section 1). There is a possible overlap between different aesthetic highlights. A movie scene can contain more than one highlight, for example, spectacular moments and character development [49]. In this paper we focus our work on detecting the particular type of aesthetic highlights independently of those overlaps. We start

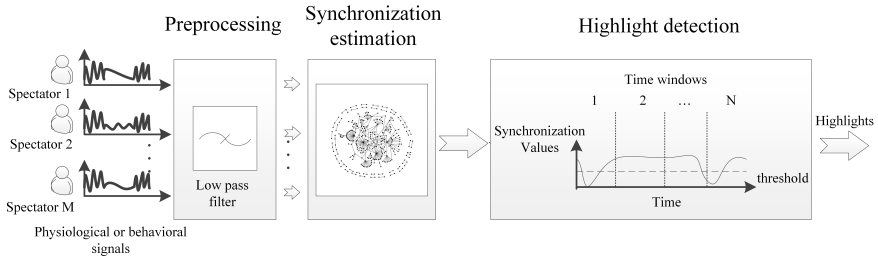

Fig. 1. The scheme of unsupervised highlight detection system based on synchronization among spectators' physiological or behavioral signals.

with the preprocessing of spectators' physiological and behavioral signals. Electrodermal activity and acceleration signals are filtered by a low-pass filter to remove noise and distortions. Then, time windowing is applied to each signals and a constant time lag between time windows is selected. The main component of our detection system is an estimator of a synchronization level among spectators that employs synchronization measures. To compute the amount of synchronization for each time window, we can use different synchronization measures. We expect that the value of synchronization increases when spectators jointly react to aesthetic scenes. The choice of a synchronization estimator is related to the type of analysis of synchronization which we attempt to carry out. We can analyze synchronization at pairwise (local descriptors), overall (global descriptors) and group (trade-off between local and global descriptors) levels to capture different patterns in multipule spectators' responses [23]. The other main limitation comes out with the properties of the aesthetic highlights (duration) and recorded electrodermal activity and accelerometer measurements.
Aesthetic highlights are determined for each time window based on the value of the estimated synchronization among spectators. If the value of a synchronization measure is higher (lower) than a threshold, we assign the time window to highlight (non-highlight) scenes [48, 49, 59]. The crucial issue of our unsupervised detection system is the choice of a threshold for a given synchronization measure which should be made with regard to performance of our system. We carry out the analysis of the receiver operating characteristic (ROC) curves and the areas under the ROC curves (AUC) to ovecome this and evaluate the overall performance of our system [29].

3.1 Synchronization measures

In this paper we attempt to investigate different synchronization measures with special emphasis on estimating of physiological and behavioral synchronization. This includes all constraints related to highlight detection e.g. sampling frequency of physiological and behavioral signals, the number of signals, the size of time window, the duration of highlights, etc. To understand different approaches to synchronization estimation for analysis of multiple spectators' signals, we divide synchronization measures into 3 classes: pairwise, group and overall measures. Pairwise measures establish synchronization between pairs of signals. Group measures can analyze clusters of signals in principle. Overall measures can simultaneously process an arbitrary number of signals. The obvious disadvantage of group and overall measures is that they are not able to provide local information on synchronous activities due to their global properties. On the other hand, pairwise

measures can only be successfully applied when some local activities are identified.
We underline that although we study a large variety of synchronization measures, it is impossible to include all existing synchronization measures in this paper. Moreover, novel measures are constantly being developed. Our choice of synchronization measures is supported by our previous work on this topic. We do not consider the basic pairwise synchronization measures, such as the correlation, Spearman's correlation, mutual information, Kolmogorov-Smirnov distance because they did not provide any plausible results for highlight detection [59, 60]. Also, it is worth pointing out that we are not able to estimate a covariance matrix by means of some pairwise synchronization measures, such as the dynamic time warping, the shape distribution distance, the nonlinear interdependence because of their properties. The dynamic time warping is not a normalized measure, the shape distribution distance and the nonlinear interdependence require distance measurements to neighbouring time windows.
In this section we briefly review three approaches to synchronization: pairwise, group and overall measures and we propose to apply them to spectators' physiological and behavioral signals in order to detect aesthetic highlights in movies. A level of a synchronization measure should reveal the synchronized reactions of spectators during watching a movie.
For spectators' electro-dermal and acceleration signals $\{x_i\}$ we consider time windows $\{x_i(l)\}$, $i = 1, ..., M$, $l = 1, ..., N$, where M is a number of spectators' signals and N is a number of time windows.

3.2 Pairwise synchronization

The key point of pairwise measures is to measure the amount of synchronization only at the local level, between two time series. When the number of signals is more than 2, the synchronization value is obtained by averaging synchronization values of all possible non-overlapping pairs of signals in a given time. We begin our review with mentioning about the Pearson correlation coefficient that is perhaps the most common measure for linear interdependence between two signals and the coherence function quantifies linear correlations in frequency domain (find the details [61]). There are some attempts to propose an extension of the correlation, such as correntropy coefficient [36], modifications of the partial coherence [23, 65]. Although the amplitudes of signals are statistically independent, their instantaneous phases can be strongly synchronized. This refers to phase synchronization [53]. Also, the Granger causality is considered as a part of synchronization measures that are derived from linear stochastic models of time series which extent linear dependencies between signals [6, 35]. Also, non-linear extensions of Granger causality have been proposed in [1, 13]. Several synchronization measures come from information theory [18]. The mutual information is perhaps the most well-known synchronization measures of them. To study nonlinear dependencies between time series, it has been calculated in time and time-frequency domain [2, 51]. Therefore, stochastic event synchrony characterizes a family of synchronization measures that quantifies the similarity between point process extracted from time-frequency representations of signals [24].
We introduce below the following pairwise synchronization measure: the dynamic time warping, the shape distribution distance and the nonlinear interdependence. In this paper, we use a mean value of a pairwise synchronization measure over all possible pairs of signals at a given time stamp l as the value of the synchronization measure [48, 50].

3.2.1 Dynamic time warping. Let us suppose there are two time windows $x_i(l)$ and $x_j(l)$, where $i, j = 1, ..., M$, $l = 1, ..., N$. In order to align these two signals, we create a matrix D_W which contains the Euclidean distances between pairs of samples from time window $x_i(l)$ and $x_j(l)$ [58]. A warping path W between two time windows is a set of matrix elements which creates a mapping between

them. The warping path W of the length p is defined as follows

$$W = w_1, w_2, ..., w_P, \tag{1}$$

where $w_1, w_2, ..., w_p$ are the elements of the matrix D_W.
The total cost $c_W(x_i(l), x_j(l))$ of the warping path W is expressed by

$$c_W(x_i(l), x_j(l)) = \sum_{p=1}^{P} w_p. \tag{2}$$

The optimal warping path between two time windows $x_i(l)$ and $x_j(l)$ is a warping path W^* that has a minimal total cost among all possible warping paths.
The Dynamic Time Warping (DTW) distance between two time windows $x_i(l)$ and $x_j(l)$ is the total cost of the warping path W^*, as follows [5]

$$d_{DTW}(x_i(l), x_j(l)) = c_{W^*}(x_i(l), x_j(l)). \tag{3}$$

The distance $d_{DTW}(x_i(l), x_j(l))$ is computed for each pair of time windows $x_i(l)$ and $x_j(l)$, where $i, j = 1, ..., M$, $l = 1, ..., N$. The computational cost of the dynamic time warping is $O(Nm^2M^2)$ and is bounded by the number M of signals, the number N and the size m of time windows.

3.2.2 Shape distribution distance. Time-delay coordinate embedding is used in analysis of dynamical systems [73]. This method embeds a scalar time series into an m-dimensional space to reconstruct the state space trajectory of a dynamical system. For each sample x_i, $i = 1, 2, 3, ..., \mathcal{N}$ of time series $\{x_i\}$, a representation of the delay-coordinate embedding can be expressed as the following vector X_i which consists of m components

$$X_i = [x_i, x_{i+\tau}, x_{i+2\tau}, .., x_{i+(m-1)\tau}], \tag{4}$$

where τ is an index delay and m is an embedding dimension. Theoretical discussion on the choice of these parameters is out of the scope of our paper. The index delay and the embedding dimension are selected based on the duration of aesthetic highlights. Diffusion maps of time-delay coordinate embedding provides a new low dimensional parameterization that is able to capture the changes in physiological and behavioral signals. When diffusion maps are applied [17], an affinity metric $K(x_i, x_j)$ is defined between pairs of the samples x_i and x_j based on their representation in time-delay coordinate X_i and X_j, respectively. Then, we only take into account a collection $\mathcal{M}$ of samples x_i to define the following kernel

$$K(x_i, x_j) = e^{\frac{-||X_i - X_j||}{\epsilon}}, \tag{5}$$

where ϵ is a scale parameter of the affinity metric (the parameter is selected based on the mean distance between points in the m-dimensional space) and $i, j = 1, 2, 3, ..., \mathcal{M}$, $\mathcal{M} < \mathcal{N}$. We can consider the collection $\mathcal{M}$ as nodes of an undirected symmetric graph, where two nodes x_i and x_j are connected by an edge with the affinity weight $K(x_i, x_j)$. We pursue the construction of a Markov chain on the graph nodes by normalizing the kernel $K(\cdot, \cdot)$. Let K be the kernel matrix, and let $P = D^{-1}K$ be the corresponding transition matrix, where D is a diagonal matrix with elements $D_{ii} = \sum_{j=1}^{\mathcal{M}} K(x_i, x_j)$. In sequence, we calculate P_t analogues to P, where $P(x_i, x_j)$ is the probability of transition in a single step from node x_i to node x_j. In addition, we define $P_t(x_i, x_j)$ as the transition probability in t steps from node x_i to node x_j. This introduces us to a definition of the diffusion distance $D_t(x_i, x_j)$ between pairs of samples, defined by [17]:

$$D_t(x_i, x_j) = \sqrt{\sum_{q=1}^{\mathcal{M}} (P(x_i, x_q) - P(x_j, x_q))^2 w(x_q)}, \tag{6}$$

where $w(x_q)$ is a normalization weight. Intuitively, two points are similar when many short paths with large weights connect them. It is proven that the diffusion distance $D_t(x_i, x_j)$ can be computed using the eigenvalues $\{\lambda_i\}$, that tend to 0 and have a modulus strictly less than 1, and the corresponding eigenvectors $\{\varphi_i\}$ of the transition matrix P [17]. Let $\Phi_t(x_i)$ for some $t \geq 0$ be the diffusion maps of time series samples $\{x_i\}, i = 1, 2, 3, ..., \mathcal{M}$ into Euclidean space $\mathbb{R}^s$ that is expressed by

$$\Phi_t(x_i) = [\lambda_1^{2t}\varphi_1(x_i), ..., \lambda_s^{2t}\varphi_s(x_i)], \tag{7}$$

where $s \in \{1, 2, ..., \mathcal{M} - 1\}$ is the new space dimensionality.

It is shown that the diffusion distance between samples x_i and x_j is equal to the Euclidean distance in the diffusion map space, as follows [17]

$$D_t(x_i, x_j) = ||\Phi_t(x_i) - \Phi_t(x_j)||. \tag{8}$$

In this paragraph we present a geometric framework which computes the amount of synchronization between a pair of spectators' physiological or behavioral signals. The idea is to measure the similarity between local shapes of reconstructed signal manifolds [59]. To capture the unique local geometric properties of a signal manifold, we introduce the local shape cumulative distribution function $F_{x_i}^{\sigma}(\delta)$ of pairwise diffusion distances for each sample x_i and its delay samples $x_i, x_{i+1}, ..., x_{i+\sigma}$ defined by

$$F_{x_i}^{\sigma}(\delta) = \int 1_{\tilde{D}_t(x_i, x_{i+q}) \leq \delta} \mathrm{d}\mu, \tag{9}$$

where $q \in \{1, \sigma\}$, μ is a counting measure and $1_{\tilde{D}_t(x_i, x_{i+q})}$ is an indicator function with respect to a delay sample on manifolds. Moreover, σ should be chosen to obtain enough a number of samples required for density estimation ($\sigma = 50$). Besides, $\tilde{D}_t(\cdot, \cdot)$ is the cosine distance in the diffusion maps space that can be derived from the Euclidean dot product. Normalization is advantageous to the local shape distribution, as follows

$$\mathcal{F}_{x_i}^{\sigma}(\delta) = \frac{F_{x_i}^{\sigma}(\delta)}{F_{x_i}^{\sigma}(\infty)}. \tag{10}$$

For two time series $\{x_i\}$ and $\{y_i\}$, the synchronization measure that is named Shape Distribution Distance (SDD) is derived from computing the Kolmogorov-Smirnov distance between two local shape distributions of their manifold representations for each time step i that is defined

$$S_{\sigma}(x_i, y_i) = \max_{\delta} |\mathcal{F}_{x_i}^{\sigma}(\delta) - \mathcal{F}_{y_i}^{\sigma}(\delta)|. \tag{11}$$

If two signals are the same $S_{\sigma}(x_i, y_i)$ is equal to 0. The complexity of the shape distribution distance is $O(M^2N^3)$ and is bounded by the number M of signals and the number N of time windows.

3.2.3 Nonlinear Interdependence. The concept of the nonlinear interdependence comes from studies on generalized synchronization that evaluate the interdependence between signals in a reconstructed state space domain [67]. The nonlinear interdependence measures the geometrical similarity between the state space trajectories of two dynamical systems. Time-delay embedding is applied to two time series $\{x_i\}$ and $\{y_i\}$ $i, j = 1, ..., \mathcal{N}$ to reconstruct the trajectories analogous to shape distribution distance [73]. The mean square Euclidean distance of each sample x_i to its K nearest neighbours x_r, $r = 1, ..., K$ in the delay-coordinate embedding is

$$R^K(x_i) = \frac{1}{K}\sum_{r=1}^{K}(X_i - X_r)^2, \tag{12}$$

and the mean squared Euclidean distance conditioned by the equal time partners of the K nearest neighbours of y_i is

$$R^K(x_i|y_i) = \frac{1}{K}\sum_{r=1}^{K}(X_i - Y_r)^2. \tag{13}$$

The nonlinear interdependence (NI) measure is defined as [66]

$$S^K(x_i|y_j) = \frac{R^K(x_i)}{R^K(x_i|y_i)}. \tag{14}$$

The number of nearest neighbours should be selected to accurately estimate an average distance a point to its nearest neighbours ($K = 50$). To make the nonlinear interdependence symmetric, we consider $S^K(y_i|x_i)$ and we then average these parameters. When two time series are synchronized (desynchronized), the value of the nonlinear interdependence is close to 1 (0). Searching K nearest neighbours of N time windows for calculating the nonlinear interdependence among all possible pairs of signals M can be found in $O(M^2KNlogN)$ time.

3.3 Group synchronization

A group measure intends to be a trade-off between pair and overall approaches to synchronization. It can capture synchronization among groups of signals based on the connectivity of signal clusters. This approach to synchronization contains a multivariate measure which ascribes a single value to groups of signals in comparison with pairwise or overall measures [60]. We detail below how we can adapt the periodicity score to measure synchronization among groups of spectators.

3.3.1 Periodicity Score. Here we detail the usage of the periodicity score to measure synchronization of signals [63, 64]. First, we map spectators' physiological or behavioral signals to a geometric framework of real Grassmann manifolds by applying the reduced singular value decomposition (RSVD) to their short time Fourier transform (STFT). We analyze time windows $\{x_i(l)\}$, $i = 1, ..., M$, $l = 1, ..., N$ of spectators' signals as a sequence of points encoded on the Grassmann manifold preserving their intrinsic dependencies. Next, we associate a level of the periodicity score with the synchronized spectators' physiological and behavioral signals.

STFT. We apply STFT to given time windows $x_i(l)$, $i = 1, ..., M$, and we yield $x_{t,f}^{i,l}$ in the time and frequency domain, where t is a time frame index and f is a frequency band index. Each time window $x_i(l)$ is split into segments with an overlap of 50% to apply STFT in this paper. Let $S_x^{i,l}(t, f)$ be the squared magnitude of the STFT, as follows

$$S_x^{i,l}(t, f) = ||x_{t,f}^{i,l}||^2. \tag{15}$$

RSVD. Then, we map time windows $\{x_i(l)\}$ of all signals on Grassmann manifolds to recover the intrinsic dependencies among them [15]. The real Grassmann manifold $G(k, n)$ parametrizes all k dimensional subspaces of the vector space $\mathbb{R}^n$. A sequence of corresponding matrices $S_x^{i,l}(t, f)$, $i = 1, ..., M$ can be mapped to the points on the manifold $G(k, n)$ using RSVD. If we compute RSVD of matrix $S_x^{i,l}(t, f)$, as follows:

$$S_x^{i,l}(t, f) = U^i \Sigma^i {V^i}^T, \tag{16}$$

then the columns of the $n \times k$ orthogonal matrix U^i are a non-unique basis for the column space of $S_x^{i,l}(t, f)$. Thus, U^i can be used to represent the matrix $S_x^{i,l}(t, f)$, and can be identified with a point on the Grassmann manifold $G(k, n)$. Once the time windows are mapped to a sequence of points on $G(k, n)$, the pairwise distances between these points can be found using a function of the angles between subspaces.

Let U^i and U^j be two k dimensional subspaces, we measure the similarity $d_{min}(U^i, U^j)$ of two points on the Grassmann manifold $G(k, n)$ by applying the minimum correlation distance [37]

$$d_{min}(U^i, U^j) = \sin\theta_k, \tag{17}$$

where $0 \leq \theta_1 \leq \theta_2 \leq ... \leq \theta_k \leq \frac{\pi}{2}$ are principle angles between two subspaces.

Periodicity Score. Finally, we introduce the basics of persistent homology: filtrations and persistence diagrams [32, 63, 64]. Once the sequence of $S_x^{i,l}(t, f)$, $i = 1, ..., M$ matrices is mapped to $G(k, n)$ and defines a metric space $(U = \{U^1, ..., U^M\}, d_{min}(\cdot, \cdot))$, we recall the definition of the Vietoris-Rips complex $Rips_\alpha(U)$ as the set of the simplices $[U^1, ..., U^q]$ such that $d_{min}(U^i, U^j) \leq \alpha$ for $i, j = 1, ..., q$. There is an inclusion of $Rips_\alpha(U)$ in $Rips_\beta(U)$ for any $\alpha \leq \beta$. The sequences of inclusions are called filtrations $Filt_\alpha(U)$. Persistence diagrams study the evolution of the topology of a filtration, and to capture properties of the metric which is used to generate the filtration. Existing connected components are merged for $0-$th homology, when α increases. Persistent homology tracks the birth (appearance) b and death (disappearance) d of all connected components. The maximum persistence $mp(dgm(x_i(l)))$ of a persistence diagram $dgm(x_i(l))$ is defined as follows [64]

$$mp(dgm(x_i(l))) = \max_{(b,d)\in dgm(x_i(l))} pers(b, d), \tag{18}$$

where $pers(b, d) = d - b$ for $(b, d) \in dgm(x_i(l))$, and as ∞ otherwise. Finally, we can provide the periodicity score $S(x_i(l))$ [64]

$$S(x_i(l)) = \frac{mp(dgm(x_i(l)))}{\sqrt{3}}. \tag{19}$$

The normalized maximum persistence $mp(dgm(x_i(l)))$ of a persistence diagram $dgm(x_i(l))$ can help us to quantify synchronization among signals because it is capable of measuring their intrinsic geometric dependencies. The periodicity score can measure synchronization among groups of signals based on the connectivity of signal clusters. When $S(x_i(l))$ equals 0, it means that we can not explore any structure in our data. If a value of $S(x_i(l))$ rises close to 1, we find some strong connectivity structure of data. The computational complexity of the periodicity score is $O(mNM^2)$ and is bounded by the number M of signals, the number N and the size m of time windows.

3.4 Overall synchronization

The overall approach to synchronization measurement simutaneously processes all the time series and considers them as components of a single interdependent system. This provides us the overall characterization of signal dependencies. The omega complexity is a synchronization measure which is derived from applying the principle component analysis to a covariance matrix [68]. Given M signals, the multivariate time series is viewed as a series of temporary maps whose sequence over time defines a trajectory of a dynamical system in a M dimensional space. The omega complexity evaluates in particular the complexity of a trajectory by means of examining its shape along the principle dimensions. S-estimators are an extension of omega complexity based on Shannon entropy. [40]. We detail below the concept of the family of the S-estimators with different estimators of a covariance matrix.

3.4.1 S-estimators. All signals can be viewed as the representation of a trajectory that can be modeled in a high-dimensional state-space. The dimensionality of the trajectory in the state-space can be assessed based on the principle component analysis of an estimated covariance matrix. The minimum entropy characterizes the situtation when a few normalized eigenvalues only are nonzero showing the high level of synchronization. Let $C^l = (C^l_{ij})$ be a covariance matrix in which C^l_{ij} is cross-dependence between time window $x_i(l)$ and $x_j(l)$, where $i, j = 1, ..., M$, $l = 1, ..., N$, such as correlation, phase synchronization (phase locking value), synchronization likelihood, windowed

mutual information, event synchronization, heat kernel, diffusion map and so on [21], [40].
The eigenvalue decomposition of C^l is

$$C^l v_u^l = \lambda_u^l v_u^l, \tag{20}$$

where eigenvalues $\lambda_1^l \leq \lambda_2^l \leq \ldots \leq \lambda_M^l$ are in increasing order and $v_u^l, u = 1, \ldots, M$ are corresponding eigenvectors. As the matrix C^l is a real and symmetric, all eigenvalues are real numbers, and the trace of C^l is equal to the number of signals M.
The S-estimator is proposed to measure synchronization among signals by means of the distribution of the eigenvalues of the covariance matrix C^l, as proposed in [10]

$$S_l = 1 + \frac{\sum_{u=1}^{M} \frac{\lambda_u^l}{M} log(\frac{\lambda_u^l}{M})}{log(M)}. \tag{21}$$

When all the signals are synchronized (resp. desynchronized), the value of the S-estimator is close to 1 (resp. 0). The computational complexity of the S-estimator is $O(NM^3)$ and is bounded by the number M of signals and the number N of time windows.

4 EXPERIMENT:ANNOTATION OF DATABASE

Our current experiment is the extension of work [48, 49, 59, 60]. The proposed structure of aesthetic highlights is chosen based on various film theories and experts' feedback on the annotation process [4, 7, 11, 25–27] and is shown in Figure 2. We can distinguish two general categories of aesthetic highlights: Form and Content. Form (highlights H1, H2) corresponds to manners in which subjects are presented in films e.g., adding special effects and playing music in the background. Content (highlights H3, H4, H5) covers the subjects presented in the films, such as developments of main characters' emotions, dialogues that motivate actions and tensions among characters as well as a specific theme development in a movie e.g., occurrence of events or conversations that result in mental or emotional strains of characters.
The "LIRIS" movie database was selected to be annotated with respect to aesthetic highlights, as

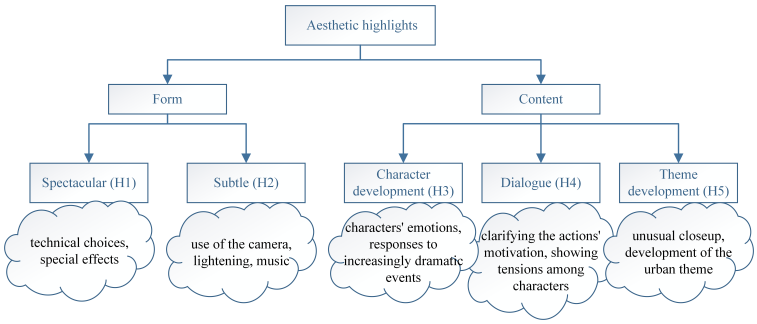

Fig. 2. Five categories of aesthetic highlights in movies [49, 50].

shown in Figure 2, because:
(i) There is a large amount of emotional and aesthetic scenes in these movies which affect spectators' affective states.
(ii) The movies in this dataset represent different movie genres, such as Action, Adventure, Animation, Comedy, Documentary, Drama, Horror, Romance and Thriller which effect differently various aesthetic experiences of spectators.

(*iii*) The physiological and behavioral reactions of 13 spectators watching the movies (30 movies, the total duration of the movies is 7 hours, 22 minutes and 5 seconds) in a darkened air-conditioned amphitheatre were collected using the Bodymedia armband sensors attached to their fingers [54]. Aesthetic highlights in the "LIRIS" database were annotated by an expert supported by one person

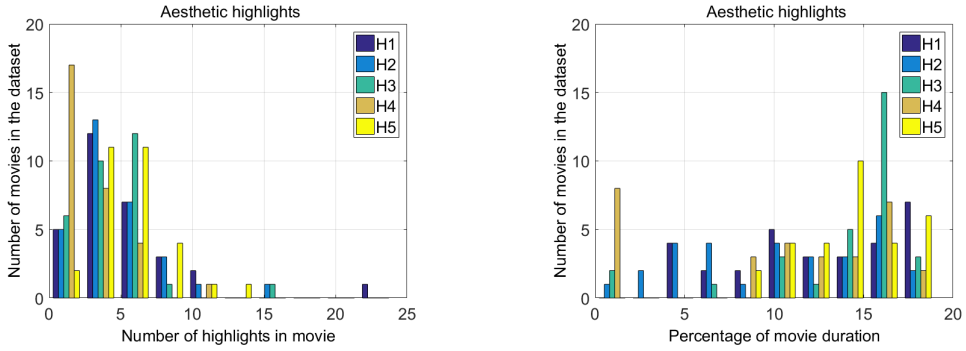

Fig. 3. Statistical analysis of aesthetic highlights annotated in the "LIRIS" database, the distribution of : a) the numbers of the particular highlight category per movie. b) the duration of the particular highlight category per movie.

with regard to form and content using an open-source annotation software [45], similarly to the previous work [49]. The annotations represent the objective assessment of the movies including 5 categories of aesthetic highlights as presented in Figure 2. We selected scenes with high levels of aesthetic values and emotions are constructed by moviemakers in a manner to establish the engagement between spectators and movies. The structure of the scenes enriches the enjoyment of watching the whole movie. A strong aesthetic experience can evoke the specific affective states of spectators. Figure 3 plots the distributions of the number of the aesthetic highlights per movies and the percentage of movie duration in the "LIRIS" dataset. We observe that there are no more than 25 highlights of a given type in a movie. The duration of these highlights is not longer than 20 % of a movie duration. That means that only particular scenes are considered as aesthetic highlights.

5 RESULTS

5.1 Emotional Component of Aesthetic Highlights

In the "LIRIS" database, contiguous emotional annotations were collected from 10 French participants (7 female and 3 male). The continuous arousal/valence annotations were down-sampled by averaging over windows of 10 seconds with 1 second overlap to remove the noise and distortions. Then, these post-processed continuous annotations of arousal/valence were averaged again to create one signal, so called, "the mean of the arousal/valence self-assessments" [54].
Previous studies have confirmed that physiological signals of spectators are linked with their emotional states [46, 80]. In our studies we attempt to prove that there are emotional components of aesthetic highlights and aesthetic scenes are able to influence on the affective states of spectators. In order to evaluate it, we use meta analysis [8]. We relate the occurrence of these highlights in movies to felt emotions (level of arousal and valence) by the spectators for the "LIRIS" database. We consider the "LIRIS" database as a set of empirical experiments about the given topic: the level of emotions (arousal/valence) while watching aesthetic highlights in movies [8]. Effect-size indices are calculated over individual movies. The effect size is standardized mean difference that

is defined as the difference between mean values of continuous emotion annotations of highlight and non-highlight intervals divided by their pooled standard deviation. Positive values indicate a higher level of arousal/valence of highlight scenes in comparison with non-highlight scenes, whereas negative values of the effect size indicate a lower level.
To integrate the experiment results, statistical analysis demands the weighting of each effect size estimate as a function of its precision assuming a fixed-effect model [8]. In our studies we follow Cohen's benchmarks [16] for the interpretation of the practical significance of a weighted average effect size. We assume that the values around 0.2, 0.5, 0.8 can be interpreted as reflecting the effect of small, medium and large magnitude, receptively. The weighted average effect size of arousal/valence for the "LIRIS" dataset is reported in Table 1. The medium positive effect sizes (>0.5) of arousal is found for spectacular highlights H1, character development highlights H3 and theme development highlights H5. Also, the medium negative effect sizes (<-0.5) of valence is observed for spectacular highlights H1 and character development highlights H3.

Table 1. The weighted average effect size (fixed-effect model) of arousal and valence during aesthetic highlights over all the "LIRIS" database.

Emotions \Highlights	H1	H2	H3	H4	H5
Arousal	**0.76**	-0.23	**0.70**	0.04	**0.53**
Valence	**-0.64**	-0.11	**-0.55**	0.11	-0.22

To investigate a relationship between movie genres and the emotional component of the aesthetic highlights, we carry out the same meta analysis for each 9 movie genres. We infer that the direction and the power of the average effect size strongly depends on the movie genre for both arousal and valence, as shown in Tables 2 and 3. Strong emotional reactions are expected to be associated with spectacular highlights H1, such as using special effects, increasing saturation of colors, playing with lightening and camera location. The results from Tables 2 and 3 confirm our hypothesis. We observe the medium positive effect size (the high level) of arousal for drama, action, romance, adventure and large positive effect for horror. Moreover, we identify the large positive and negative effects of valence for documentary and horror movies, respectively.
Slow movements of cameras, lightening, shadowing and playing music in the background during subtle highlights H2 do not evoke strong emotional reactions among spectators, there is the medium negative effect size of arousal for action movies. Furthermore, the medium positive effect size of valence is only reported for horrors, unlike action and romance movies.
Following the main characters' development and tensions among them (character development highlights H3) can affect the emotional and physiological states of spectators. We find the medium positive effect size of arousal for comedy, adventure, documentary movies and the large positive effect size for the horror and animation movies. Also, we observe the high level of negative valence for animations (medium negative effect), and thriller, romance and horror movies (large negative effects), as illustrated in Tables 2 and 3.

Dialogues among main characters (highlights H4) in some specific movie genres are only able to elicit emotions. We indicate the low level of arousal in dramas (medium negative effect size) and the high level of arousal in comedies (large positive effect size) as well as the medium negative effect of valence for animation movies in Tables 2 and 3. We infer that dialogues carry the emotional tone of the genre. There are the low level of arousal (sad) for dramas and the high level of arousal (joy) for comedies. It may be caused by long duration of dialogues and spectators' emotions fade over time. Also, the main character are frequently ambiguous movie characters who could stimulate different reactions across the audience during the whole movies, as shown in Tables 2 and 3.

Table 2. The weighted average effect size (fixed-effect model) of arousal during aesthetic highlights calcuated per movie genre.

H \Genre	Drama	Animat.	Thrill.	Action	Comed.	Roman.	Advent.	Docum.	Horror
H1	**0.73**	0.49	0.22	**0.57**	0.34	**0.53**	**0.57**	-0.30	**1.06**
H2	-0.02	-0.18	-0.13	**-0.58**	-0.46	-0.38	0.11	-0.49	-0.42
H3	0.01	**0.92**	0.14	0.11	**0.65**	-0.04	**0.56**	**0.72**	**1.03**
H4	**-0.74**	0.10	0.25	-0.13	**0.92**	-0.26	0.15	-	-0.27
H5	0.32	**0.78**	**0.62**	-0.01	**0.76**	-0.16	0.19	-0.19	**0.65**

Table 3. The weighted average effect size (fixed-effect model) of valence during aesthetic highlights calculated per movie genre.

H \Genre	Drama	Animat.	Thrill.	Action	Comed.	Roman.	Advent.	Docum.	Horror
H1	-0.10	0.14	-1.04	0.19	0.13	-0.13	-0.48	**1.20**	**-1.13**
H2	-0.15	-0.10	-0.38	**-0.75**	-0.28	**-0.71**	0.08	0.23	**0.60**
H3	-0.16	**-0.75**	**-0.80**	-0.20	0.27	**-0.92**	0.11	0.45	**-1.01**
H4	0.07	**-0.55**	0.41	-0.02	-0.01	0.35	-0.43	-	0.39
H5	0.07	-0.17	**-1.02**	-0.40	**0.56**	-0.01	-0.11	**-0.69**	-0.48

Theme development highlights H5 incompletely overlap with the other types of aesthetic highlights, such as spectacular highlights H1 and character development highlights H3 because the development of a specific theme is often conjugated with the emotion development of main characters as their responses to dramatic events presented in a sublime manner. Also, we observe the medium positive effect size of arousal for the following movies genres: animation, thriller, comedy and horror, as presented in Table 2. In terms of valence, we remark the medium positive effect size for comedies and medium and large negative for thriller and documentary movies, as shown in Table 3.

5.2 Dependencies between Synchronization Measures

In order to find dependencies between different approaches to synchronization and evaluate their detection performance, we select the following synchronization measures: the nonlinear interdependence (NI), dynamic time wrapping (DTW), periodicity score (PS), shape distribution distance (SDD), S-estimators with different covariance matrices, such as correlation (S-COR), phase locking value (S-PLV), windowed mutual information (S-WMI), heat kernel (S-HK) and diffusion map (S-DM). To run all the synchronization measures over the "LIRIS" database, we use the following experimental settings. Physiological and behavioral signals of spectators are filtered by a third order lowpass Butterworth filter with cutoff frequency 0.3 Hz, and they are segmented into overlapping time windows with a time step and a window length equal 1 s and 5 s, respectively (some physiological signals are discarded due to the amount of artifacts). The values of all the mentioned synchronization measures are computed for each time window.

We calculate the Pearson correlation coefficient between those measures to gain insight into the dependencies between them [23, 41]. Statistical analysis requires to weight all correlation coefficients between all pairs of the synchronization measures over the different movies from the "LIRIS" database. To integrate the results and obtain the weighted average effect size (Pearson correlation coefficient), we utilize a fixed-effect model [8]. To interpret the practical significance of a weighted average effect size for the Pearson correlation coefficient, we assume that values around 0.1, 0.3, 0.5 can be interpreted as reflecting the effect of small, medium and large magnitude, respectively [16]. The values of thresholds are different in comparison to the standardized mean

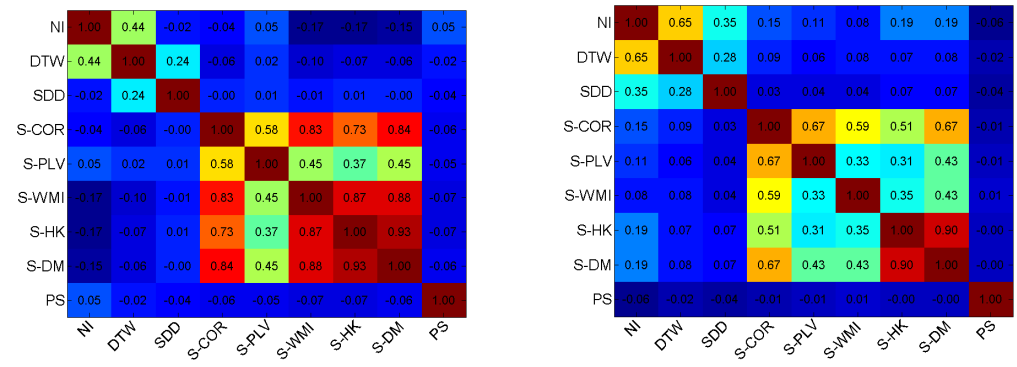

Fig. 4. The weighted average Pearson correlation effect size between the synchronization measures (red and blue color indicate strong correlation and anti-correlation, respectively). The synchronization measures are computed over spectators' a) electrodermal signals, b) acceleration signals.

difference effect size (see section 5.1).

As seen in Figure 4, we find that some synchronization measures are strongly correlated with each other independently of the processed modality (electrodermal and acceleration signals). Moreover, we can distinguish three families of synchronization measures: pairwise, group and overall measures. It becomes clear that all the S-estimators with the different estimators of the covariance matrix are dependent on one another (medium and large positive effect). Futhermore, we can emphasize the strong relations (small, medium and large positive effect) between all three pairwise synchronization measures: the NI, the DTW and the SDD. Interestingly, the PS measure seems to be mutually uncorrelated with the other measures. These results are in line with the other studies on synchronization applied to electroencephalograph signals for early diagnosis of Alzheimer's disease [23]. We find as well that some measures (pairwise measures or S-estimators) are strongly correlated or anti-correlated.

5.3 Aesthetic Highlight Detection

In this section we provide the results of aesthetic highlight detection per movie genre using the different approaches to synchronization estimation which are described in section 3. We detect the given category of aesthetic highlights (H1, H2, H3, H4, H5) based on the level of the estimated synchronization. If the value of the synchronization measure for a given sliding window is higher (lower) than a changing threshold, we assign the time window to the highlight class.

Since the presented synchronization measures capture different dependences among signals, we decide to take an advantage of it. We follow a feature fusion approach and combine all the synchronization measure at a given time into one vector. Then, we use clustering based Gaussian mixture models to compute a probability of belonging to the highlight class (resp. non-highlight) for each window. If the probability for a given sliding window is higher (resp. lower) than a changing threshold, we assign the time window to the highlight class. Combining multiple measures and clustering them is named the CMMC approach. Identified labels are referred to the collected annotations (ground truth) and the true positive and false positive ratio are calculated to obtain the overall performance measured by the area under the constructed ROC curve (AUC).

In order to investigate the statistical significance of the results, we use the following validation.

Table 4. Performance (AUC) of our highlight detection system evaluated per category of aesthetic highlights and movie genre, different synchronization measures are applied to electrodermal signals of spectators. ⋆ stand for p-value < 0.05, † for p-value < 0.01 and ‡ for p-value < 0.001. We report p-value when we refer the performance of a measure to a random classifier (upper index of AUC performance) and when we find the groups of synchronization measures in the ranking significantly different in terms of perfomance (the upper index of an ordinal number of measure groups)

Gen. \H	H1	H2	H3	H4	H5
Drama	1. DTW (0.57) † CMMC (0.56)†	1. CMMC (0.56)* NI (0.55) *	1. SDD (0.58)‡ NI (0.57)‡	1. CMMC (0.61)‡ SDD (0.59)‡ DTW (0.58)‡ S-WMI (0.56)† S-HK (0.55)* NI (0.55)* S-DM (0.55)*	1. NI (0.58)‡ DTW (0.57)‡ CMMC (0.56)†
Animat.	1.S-DM (0.59)‡ S-HK (0.59)‡ S-WMI (0.58)‡ S-COR (0.55)*	1.* DTW (0.70)‡	1.‡SDD (0.72)‡	1.‡ CMMC (0.84)‡	1. NI (0.59)‡ DTW (0.59)‡
Thrill.	1. SDD (0.62)‡	1. SDD (0.62)‡ CMMC (0.62)‡ DTW (0.61)† S-HK (0.60)† S-WMI (0.57)* S-DM (0.57)*	1. S-DM (0.70)‡ S-HK (0.68)‡ S-WMI (0.66)‡ CMMC (0.65)‡ DTW (0.64)‡ S-COR (0.63)† S-PLV (0.59)*	SDD (0.74)‡	1. S-WMI (0.60)† S-HK (0.59)† S-COR (0.58)* S-DM (0.58)* S-PLV (0.58)*
Action	1.S-WMI (0.59)‡ NI (0.58)† S-HK (0.57)† S-DM (0.56)* DTW (0.56)* CMMC (0.56)*	1. S-HK (0.58)* NI (0.58)*	1. DTW (0.57)† CMMC (0.57)† SDD (0.55)*	1. SDD (0.61)‡	1. DTW (0.58)† CMMC (0.58)† NI (0.56)*
Comed.	1. DTW (0.63)‡ CMMC (0.63)‡	1. DTW (0.58)‡ CMMC (0.56)† S-DM (0.54)*	1.*SDD (0.62)‡	1. S-HK (0.56)‡ S-WMI (0.56)‡ SDD (0.56)‡ S-DM (0.54)† S-PLV (0.54)*	1.†SDD (0.63)‡
Roman.	1.‡SDD (0.82)‡	1. DTW (0.60)‡ CMMC (0.59)† NI (0.57)*	1. DTW (0.65)‡ CMMC (0.62)‡ NI (0.60)‡ SDD (0.57)*	1.‡DTW (0.77)‡ CMMC (0.77)‡	1. NI (0.60)‡ SDD (0.56)* S-COR (0.56)*
Advent.	1.*DTW (0.70)‡ CMMC (0.70)‡	1. DTW (0.58)† CMMC (0.58)† SDD (0.57)† NI (0.57)*	1. SDD (0.64)‡ CMMC (0.62)‡ DTW (0.61)‡	1. SDD (0.60)‡ NI (0.56)†	1.*SDD (0.67)‡
Docum.	1. CMMC (0.83)† DTW (0.75)* S-HK (0.72)* S-WMI (0.71)*	-	1. CMMC (0.97)‡ SDD (0.83)† DTW (0.81)† NI (0.78)*	-	Any Measures
Horror	1.*DTW (0.69)‡ CMMC (0.69)‡	1. SDD (0.67)‡ CMMC (0.62)† DTW (0.62)*	1.‡DTW (0.75)‡ CMMC (0.74)‡	1. SDD (0.61)‡ NI (0.59)‡ S-HK (0.57)† S-DM (0.57)† S-WMI (0.56)†	1. DTW (0.60)‡ CMMC (0.60)‡ NI (0.55)† S-HK (0.55)* S-WMI (0.54)* S-DM (0.54)*

Table 5. Performance (AUC) of our highlight detection system evaluated per category of aesthetic highlights and movie genre, different synchronization measures are applied to acceleration signals of spectators. ⋆ stand for p-value < 0.05, † for p-value < 0.01 and ‡ for p-value < 0.001. We report p-value when we refer the performance of a measure to a random classifier (upper index of AUC performance) and when we find the groups of synchronization measures in the ranking significantly different in terms of perfomance (the upper index of an ordinal number of the measure groups).

Gen. \H	H1	H2	H3	H4	H5
Drama	Any Measures	1.*SDD (0.63)‡	1. CMMC (0.57)‡ DTW (0.56)† NI (0.56)†	1.*DTW (0.62)‡ CMMC (0.59)‡	1.*NI (0.61)‡
Animat.	1. SDD (0.58)‡ NI (0.55)* CMMC (0.55)*	1.*SDD (0.74)‡	1. SDD (0.60)‡ NI (0.57)†	1. DTW (0.68)‡ CMMC (0.67)‡ SDD (0.64)‡	1. NI (0.64)‡ DTW (0.62)‡ SDD (0.61)‡ CMMC (0.59)‡
Thrill.	1. MCCM (0.61)† DTW (0.58)*	Any Measures	1. SDD (0.59)* S-WMI (0.59)*	1.*CMMC (0.70)‡ DTW (0.68)‡	1. SDD (0.66)‡ DTW (0.65)‡ CMMC (0.64)‡ NI (0.59)†
Action	1. SDD (0.63)‡ S-DM (0.58)† S-HK (0.57)† S-COR (0.56)* S-WMI (0.56)*	1. S-WMI (0.59)* CMMC (0.58)*	1. DTW (0.55)*	1. DTW (0.55)*	1. SDD (0.60)‡
Comed.	1. SDD (0.61)‡ CMMC (0.58)‡ NI (0.57)†	Any Measures	1. CMMC (0.56)‡ NI (0.55)† SDD (0.53)*	1. CMMC (0.56)‡	1.‡SDD (0.63)‡
Roman.	1.*SDD (0.70)‡	1. DTW (0.63)‡ CMMC (0.62)‡ SDD (0.59)† NI (0.57)*	1. NI (0.63)‡ DTW (0.59)†	1. S-COR (0.59)*	1. DTW (0.60)‡ CMMC (0.58)† NI (0.56)*
Advent.	1. SDD (0.58)† CMMC (0.56)*	Any Measures	1. DTW (0.63)‡ NI (0.61)‡ SDD (0.58)† CMMC (0.58)†	1. DTW (0.63)‡ CMMC (0.63)‡ SDD (0.60)‡	1. S-WMI (0.55)*
Docum.	1. SDD (0.81)†	-	1. SDD (0.78)* NI (0.77)* S-PLV(0.74)*	-	1. SDD (0.95)‡
Horror	1.‡SDD (0.62)‡	1. DTW (0.64)‡ SDD (0.64)† CMMC (0.64)†	1.†SDD (0.63)‡	1. DTW (0.57)‡ NI (0.57)† CMMC (0.55)*	1. SDD (0.58)‡

Firstly, we compute the areas under ROC curves (AUC) to evaluate the performance of our system, as well as the synchronization measures. Furthermore, we refer the performance of our system to the performance of a random classifier (AUC=0.5). Secondly, the synchronization measures that do not perform randomly are placed in rank order for each movie genre. Thirdly, multiple comparisons are made to find groups of measures that perform significantly better than others , such as 1st (the highest performance), 2nd and 3rd group of measures. All the statistical comparisons are made

using the two sided Bradley test at the significance level $\alpha = 0.05$ [9]. Low p-values indicate that there are large differences in the performance of the synchronization measures. Tables 4 and 5 show the detection performance (AUC) of all the synchronization measures are applied to spectators' physiological and behavioral signals. We only report the first group of synchronization measures that obtain significantly the best performance for the given category of the aesthetic highlights and movie genre.

In Table 4 the results illustrate the discriminative power of the synchronization measures to detect aesthetic highlights in movies based on spectators' electrodermal activity signals. Generally, we observe that the pairwise synchronization measures obtain the best performance in comparison with the group or overall approaches.

The DTW measure achieves the highest of performance for the following movie genre: animation, action, romance, documentary and horror, as well as the SDD measure reaches the best results for thriller, comedy, romance and adventure movies. Also, the NI measure could indicate these highlights with the highest performance in drama, action and romance movies. Moreover, the pairwise synchronization measures appear to have also the most discriminative power for detecting the particular type of aesthetic highlights. The DTW measure can be used to detect highlights H1, H2 and H3 unlike the DDS measure indicates highlights H3, H4 with the best performance. Besides, the NI measure can be applied to predict highlights H5.

Table 5 presents the detection performance, when synchronization measures are applied to the acceleration signals. We infer that the pairwise synchronization measures, especially the SDD, reach the best performance per movie genre and the type of aesthetic highlights. Moreover, the SDD measure obtains the best results for the movie genres: animation, action, comedy, adventure, documentary and horror, as well as, in terms of highlight type: highlights H1, H2, H3 and H5. The DTW measure performs the best for drama, thriller, action, romance movies, as well as for highlights H4. The NI can be used interchangeably with the DTW to detect these highlights in the specific movie genres, such as drama or romance, respectively. Also, it can replace the SDD to identify highlights H3.

Detection of highlights H4 in animations and comedies only benefits from the basic combining multiple synchronization measures applied to spectators' electrodermal activity and acceleration signals, respectively. Generally, clustering concatenated synchronization measures does not outperform aesthetic highlight detection based on any single synchronization measures.

6 DISCUSSION AND CONCLUSIONS

In this work we extend our primary experiment [49]. It is conducted on a different database ("LIRIS") including 30 movies to gain insight into spectators' reactions to aesthetic highlights across different movie genres. Regarding our **first research question** we uncover that these highlights evoke some amount of emotions (arousal and valence level) that is strictly related to movie genres. Moreover, we propose an unsupervised architecture which is able to detect the aesthetic highlights in movies based on spectators' physiological and behavioral reactions. We investigate which approach to synchronization estimation, such as pairwise, group, overall measures obtains the best performance. In response to our **second research question**, the obtained results prove that the level of synchronization among spectators' electrodermal and acceleration signals in social settings has the discriminative power to detect the different categories of aesthetic highlights independently of movie genre and recorded modalities. Nevertheless, a general statement can not be made for different movie genres because of the small number of movies per genre and different movie duration. Furthermore, we infer from our analysis that all the pairwise synchronization measures are correlated with each other. Also, that is the case for all the overall synchronization measures. To study synchronization, we find that it is enough to evaluate a few measures derived from the

different families of synchronization measures instead of using all of them. Overall, we observe that the pairwise synchronization measures, such as the SDD (shape distribution distance) and the DTW (dynamic time warping) perform the best for the aesthetic highlight detection in movies independently of movie genre and highlight type, responding to our **third research question**. The group and overall estimation of synchronization perform unexpectedly at the lowest level. Also, the choice of covariance matrix estimator, such as the min correlation distance, correlation, phase locking value, windowed mutual information, heat kernel and diffusion map does not influence on performance. When rapid physiological and behavioral reactions are evoked, all the pairwise synchronization measures (the SDD and the NI) seem to take the advantage of including information on neighboring time windows unlike the estimation of covariance matrix. Moreover, the DTW is able to average the temporal reactions of spectators which vary in speed. These features of estimation allow them to suppress the oscillations of the values from one time window to another. In addition, we suppose that considering all spectators signals like one dynamic system suffers from rapid changes of social interactions among spectators through the whole movie, and may result in unstable behaviors of the dynamical system. Analysis of synchronization at the level of pairs could benefit from uncovering stables pairs of spectators through the majority part of a movie.

The electrodermal activity measurements appear to be more indicative for aesthetic highlight detection in the social context in comparison with the acceleration measurements as far as the signals are concerned. The main reason can be that aesthetic experience is associated with a high level of arousal with is depicted in physiological reactions of spectators. That is coherent with our finding that the annotated scenes contain a large amount of emotions (high level of arousal and valence) across the whole "LIRIS" database. Spontaneous rapid behavioral reactions could be expected to be evoked when spectators are exposed to very intensive stimuli e.g. spectacular killing people in a horror movie.

Generally, we observe that pairwise, group and overall synchronization measures are able to estimate synchronization among spectators' physiological signals when they are exposed to different aesthetic highlights that elicit the high level of arousal and valence, e.g. romance, action, adventure, horror movies, etc. This is not the case for the estimation of synchronization based on acceleration measurements, the pairwise synchronization measures only are plausibly capable of estimating the level of synchronization among behavioral responses of spectators.

Combining multiple synchronization measures into one vector does not significantly improve the performance of aesthetic highlight detection. There is a need to study fusion of multiple synchronization measures since they are defined in different manners and measure different nonlinear dependences between signals. This can be considered as one of the future directions of research on synchronization measures.

The main limitation of our work corresponds to the amount of available annotated data and the feasibility of running a large scale experiment in a cinema theater and using unobstructive and reliable sensors. In our studies we uncover that the estimation of synchronization among spectators from their physiological signals results in better performance of highlight detection than from their acceleration signals. However, this conclusion can be biased by the placement of sensors. The sensors were attached to spectators' hands when the experiment was conducted. Generally, spectators do not often make limb moves when they watch a movie.

Future work includes collecting more multimodal data in order to propose general architecture of a detection system. That allows us to apply more complex synchronization measures also between different modalities. In the future, we will possibly have access to cost-effective sensors that are capable of capturing the currently unavailable modalities, such as audio-video recording of movie audiences in a darkened cinema theater and spectators' physiological and behavioral signals. A comprehensive approach to understanding aesthetic experience also requires to explore movie

content combined with spectators' reactions. We will investigate integration of audio-visual movie attributes with spectators' physiological and behavioral signals. This can be beneficial for affective understanding of movies and aesthetic highlight detection.

ACKNOWLEDGMENTS

This work is supported by grants from the Swiss Center for Affective Sciences and the Swiss National Science Foundation.

REFERENCES

[1] Nicola Ancona, Daniele Marinazzo, and Sebastiano Stramaglia. 2004. Radial basis function approach to nonlinear Granger causality of time series. *Physical Review E* 70, 5 (2004), 056221.
[2] Selin Aviyente. 2005. A measure of mutual information on the time-frequency plane. In *Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on.* IEEE, IV–481.
[3] Yoann Baveye, Emmanuel Dellandréa, Christel Chamaret, and Liming Chen. 2015. Deep learning vs. kernel methods: Performance for emotion prediction in videos. In *Affective Computing and Intelligent Interaction (ACII), 2015 International Conference on.* IEEE, 77–83.
[4] André Bazin. 2004. *What is cinema?* University of California Press.
[5] Donald J Berndt and James Clifford. 1994. Using dynamic time warping to find patterns in time series.. In *KDD workshop*, Vol. 10. Seattle, WA, 359–370.
[6] Katarzyna J Blinowska, Rafał Kuś, and Maciej Kamiński. 2004. Granger causality and information flow in multivariate processes. *Physical Review E* 70, 5 (2004), 050902.
[7] David Bordwell, Kristin Thompson, and Jeremy Ashton. 1997. *Film art: an introduction.* McGraw-Hill New York.
[8] Michael Borenstein, Larry V Hedges, Julian Higgins, and Hannah R Rothstein. 2009. *Introduction to Meta-analysis.* John Wiley & Sons, Inc.
[9] Andrew Bradley. 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern recognition* 30, 7 (1997), 1145–1159.
[10] Cristian Carmeli, Maria G Knyazeva, Giorgio M Innocenti, and Oscar De Feo. 2005. Assessment of EEG synchronization based on state-space analysis. *Neuroimage* 25, 2 (2005), 339–354.
[11] Stanley Cavell. 1979. The World Viewed, enlarged ed. *Cambridge: Harvard Univer* (1979).
[12] Liang-Hua Chen, Hsi-Wen Hsu, Li-Yun Wang, and Chih-Wen Su. 2011. Violence detection in movies. In *Computer Graphics, Imaging and Visualization (CGIV), 2011 Eighth International Conference on.* IEEE, 119–124.
[13] Yonghong Chen, Govindan Rangarajan, Jianfeng Feng, and Mingzhou Ding. 2004. Analyzing multiple nonlinear time series with extended Granger causality. *Physics Letters A* 324, 1 (2004), 26–35.
[14] Christophe Chênes, Guillaume Chanel, Mohammad Soleymani, and Thierry Pun. 2013. Highlight detection in movie scenes through inter-users, physiological linkage. In *Social Media Retrieval.* 217–237.
[15] Sofya Chepushtanova, Michael Kirby, Chris Peterson, and Lori Ziegelmeier. 2015. An application of persistent homology on Grassmann manifolds for the detection of signals in hyperspectral imagery. In *Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International.* IEEE, 449–452.
[16] Jocob Cohen. 1988. Statistical power analysis for the behavioral sciences. Erlbaum. *Hillsdale, NJ* (1988).
[17] Ronald R. Coifman and Stéphane Lafon. 2006. Diffusion maps. *Applied and computational harmonic analysis* 21, 1 (2006), 5–30.
[18] Thomas M Cover. 1991. Elements of Information Theory Thomas M. Cover, Joy A. Thomas Copyright© 1991 John Wiley & Sons, Inc. Print ISBN 0-471-06259-6 Online ISBN 0-471-20061-1. (1991).
[19] Mihaly Csikszentmihalyi. 2000. *Beyond boredom and anxiety.* Jossey-Bass.
[20] Mihaly Csikszentmihalyi. 2014. *Toward a psychology of optimal experience.* Springer.
[21] Dong Cui, Xianzeng Liu, You Wan, and Xiaoli Li. 2010. Estimation of genuine and random synchronization in multivariate neural series. *Neural Networks* 23, 6 (2010), 698–704.
[22] Gerald C Cupchik, Oshin Vartanian, Adrian Crawley, and David J Mikulis. 2009. Viewing artworks: contributions of cognitive control and perceptual facilitation to aesthetic experience. *Brain and cognition* 70, 1 (2009), 84–91.
[23] Justin Dauwels, François Vialatte, Toshimitsu Musha, and Andrzej Cichocki. 2010. A comparative study of synchrony measures for the early diagnosis of Alzheimer's disease based on EEG. *NeuroImage* 49, 1 (2010), 668–693.
[24] Justin Dauwels, François B Vialatte, Tomasz M Rutkowski, and Andrzej Cichocki. 2007. Measuring Neural Synchrony by Message Passing.. In *NIPS.* 361–368.
[25] Bordwell David and Kristin Thompson. 1994. Film History: An Introduction. (1994).
[26] Gilles Deleuze. 1989. Cinema 2: The Time-Image, trans. Hugh Tomlinson and Robert Galeta. *London: Athlone* (1989).

[27] Gilles Deleuze, Hugh Tomlinson, and Barbara Habberjam. 1986. *The Movement-Image.* University of Minnesota.
[28] Florian Eyben, Felix Weninger, Stefano Squartini, and Björn Schuller. 2013. Real-life voice activity detection with lstm recurrent neural networks and an application to hollywood movies. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on.* IEEE, 483–487.
[29] Tom Fawcett. 2006. An introduction to ROC analysis. *Pattern recognition letters* 27, 8 (2006), 861–874.
[30] Julien Fleureau, Philippe Guillotel, and Izabela Orlac. 2013. Affective benchmarking of movies based on the physiological responses of a real audience. In *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on.* IEEE, 73–78.
[31] Pouya Ghaemmaghami, Mojtaba Khomami Abadi, Seyed Mostafa Kia, Paolo Avesani, and Nicu Sebe. 2015. Movie genre classification by exploiting MEG brain signals. In *International Conference on Image Analysis and Processing.* Springer, 683–693.
[32] Robert Ghrist. 2008. Barcodes: the persistent topology of data. *Bull. Amer. Math. Soc.* 45, 1 (2008), 61–75.
[33] Gabin Gninkoun and Mohammad Soleymani. 2011. Automatic violence scenes detection: A multi-modal approach. (2011).
[34] Yulia Golland, Yossi Arzouan, and Nava Levit-Binnun. 2015. The mere co-presence: synchronization of autonomic signals and emotional responses across co-present individuals not engaged in direct interaction. *PloS one* 10, 5 (2015), e0125804.
[35] Clive WJ Granger. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society* (1969), 424–438.
[36] Aysegul Gunduz and Jose C Principe. 2009. Correntropy as a novel measure for nonlinearity tests. *Signal Processing* 89, 1 (2009), 14–23.
[37] Jihun Hamm and Daniel D Lee. 2008. Grassmann discriminant analysis: a unifying view on subspace-based learning. In *Proceedings of the 25th international conference on Machine learning.* ACM, 376–383.
[38] Alan Hanjalic and Li-Qun Xu. 2005. Affective video content representation and modeling. *IEEE transactions on multimedia* 7, 1 (2005), 143–154.
[39] Elaine Hatfield, John T Cacioppo, and Richard L Rapson. 1994. *Emotional contagion.* Cambridge university press.
[40] Mahdi Jalili, Elham Barzegaran, and Maria G Knyazeva. 2014. Synchronization of EEG: Bivariate and multivariate measures. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 22, 2 (2014), 212–221.
[41] B Jelles, Ph Scheltens, WM Van der Flier, EJ Jonkman, FH Lopes da Silva, and CJ Stam. 2008. Global dynamical analysis of the EEG in AlzheimerâĂŹs disease: frequency-specific changes of functional interactions. *Clinical Neurophysiology* 119, 4 (2008), 837–841.
[42] Hideo Joho, Jacopo Staiano, Nicu Sebe, and Joemon M. Jose. 2011. Looking at the viewer: analysing facial activity to detect personal highlights of multimedia contents. *Multimedia Tools and Applications* 51, 2 (2011), 505–523.
[43] Patrik N Juslin. 2013. From everyday emotions to aesthetic emotions: towards a unified theory of musical emotions. *Physics of life reviews* 10, 3 (2013), 235–266.
[44] Hang-Bong Kang. 2003. Affective content detection using HMMs. In *Proceedings of the eleventh ACM international conference on Multimedia.* ACM, 259–262.
[45] Michael Kipp. 2010. Anvil: The video annotation research tool. Handbook of Corpus Phonology. (2010).
[46] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. 2012. Deap: A database for emotion analysis; using physiological signals. *IEEE Transactions on Affective Computing* 3, 1 (2012), 18–31.
[47] Arthur Koestler. 1970. The act of creation (Revised Danube Edition ed.). (1970).
[48] Theodoros Kostoulas, Guillaume Chanel, Michal Muszynski, Patrizia Lombardo, and Thierry Pun. 2015. Dynamic Time Warping of Multimodal Signals for Detecting Highlights in Movies. In *Proceedings of the 1st Workshop on Modeling INTERPERsonal SynchrONy And infLuence.* ACM, 35–40.
[49] Theodoros Kostoulas, Guillaume Chanel, Michal Muszynski, Patrizia Lombardo, and Thierry Pun. 2015. Identifying aesthetic highlights in movies from clustering of physiological and behavioral signals. In *Quality of Multimedia Experience (QoMEX), 2015 Seventh International Workshop on.*
[50] Theodoros Kostoulas, Guillaume Chanel, Michal Muszynski, Patrizia Lombardo, and Thierry Pun. 2017. Films, Affective Computing and Aesthetic Experience: Identifying Emotional and Aesthetic Highlights from Multimodal Signals in a Social Setting. *Frontiers in ICT* 4 (2017), 1–11.
[51] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. 2004. Estimating mutual information. *Physical review E* 69, 6 (2004), 066138.
[52] Eleni Kroupi, Jean-Marc Vesin, and Touradj Ebrahimi. 2013. Phase-Amplitude Coupling between EEG and EDA while experiencing multimedia content. In *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on.* IEEE, 865–870.

[53] Jean-Philippe Lachaux, Eugenio Rodriguez, Jacques Martinerie, Francisco J Varela, et al. 1999. Measuring phase synchrony in brain signals. *Human brain mapping* 8, 4 (1999), 194–208.
[54] Ting Li, Yoann Baveye, Christel Chamaret, Emmanuel Dellandréa, and Liming Chen. 2015. Continuous arousal self-assessments validation using real-time physiological responses. In *Proceedings of the 1st International Workshop on Affect & Sentiment in Multimedia*. ACM, 39–44.
[55] Wu Liu, Tao Mei, Yongdong Zhang, Cherry Che, and Jiebo Luo. 2015. Multi-task deep visual-semantic embedding for video thumbnail selection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 3707–3715.
[56] Slobodan Marković. 2012. Components of aesthetic experience: aesthetic fascination, aesthetic appraisal, and aesthetic emotion. *i-Perception* 3, 1 (2012), 1–17.
[57] Abraham H Maslow. 2013. *Toward a psychology of being*. Simon and Schuster.
[58] Meinard Müller. 2007. *Information retrieval for music and motion*. Vol. 2. Springer.
[59] Michal Muszynski, Theodoros Kostoulas, Guillaume Chanel, Patrizia Lombardo, and Thierry Pun. 2015. Spectators' Synchronization Detection based on Manifold Representation of Physiological Signals: Application to Movie Highlights Detection. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*. ACM, 235–238.
[60] Michal Muszynski, Theodoros Kostoulas, Patrizia Lombardo, Thierry Pun, and Guillaume Chanel. 2016. Synchronization among Groups of Spectators for Highlight Detection in Movies. In *Proceedings of the 2016 ACM on Multimedia Conference*. ACM, 292–296.
[61] Paul L Nunez and Ramesh Srinivasan. 2006. *Electric fields of the brain: the neurophysics of EEG*. Oxford University Press, USA.
[62] Cédric Penet, Claire-Hélène Demarty, Guillaume Gravier, and Patrick Gros. 2015. Variability modelling for audio events detection in movies. *Multimedia Tools and Applications* 74, 4 (2015), 1143–1173.
[63] Jose A Perea, Anastasia Deckard, Steve B Haase, and John Harer. 2015. SW1PerS: Sliding windows and 1-persistence scoring; discovering periodicity in gene expression time series data. *BMC bioinformatics* 16, 1 (2015), 257.
[64] Jose A Perea and John Harer. 2015. Sliding windows and persistence: An application of topological methods to signal analysis. *Foundations of Computational Mathematics* 15, 3 (2015), 799–838.
[65] Ernesto Pereda, Rodrigo Quian Quiroga, and Joydeep Bhattacharya. 2005. Nonlinear multivariate analysis of neurophysiological signals. *Progress in neurobiology* 77, 1 (2005), 1–37.
[66] Rodrigo Quian Quiroga, Jochen Arnhold, and Peter Grassberger. 2000. Learning driver-response relationships from synchronization patterns. *Physical Review E* 61, 5 (2000), 5142.
[67] Nikolai F Rulkov, Mikhail M Sushchik, Lev S Tsimring, and Henry DI Abarbanel. 1995. Generalized synchronization of chaos in directionally coupled chaotic systems. *Physical Review E* 51, 2 (1995), 980.
[68] Naomi Saito, Toshiaki Kuginuki, Takami Yagyu, Toshihiko Kinoshita, Thomas Koenig, Roberto D Pascual-Marqui, Kieko Kochi, Jiri Wackermann, and Dietrich Lehmann. 1998. Global, regional, and local measures of complexity of multichannel electroencephalography in acute, neuroleptic-naive, first-break schizophrenics. *Biological psychiatry* 43, 11 (1998), 794–802.
[69] Klaus R Scherer. 2005. What are emotions? And how can they be measured? *Social science information* 44, 4 (2005), 695–729.
[70] Mohamad-Hoseyn Sigari, Hamid Soltanian-Zadeh, and Hamid-Reza Pourreza. 2015. Fast highlight detection and scoring for broadcast soccer video summarization using on-demand feature extraction and fuzzy inference. *International Journal of Computer Graphics* 6, 1 (2015).
[71] Mohammad Soleymani, Sadjad Asghari-Esfeden, Maja Pantic, and Yun Fu. 2014. Continuous emotion detection using EEG signals and facial expressions. In *Multimedia and Expo (ICME), 2014 IEEE International Conference on*. IEEE, 1–6.
[72] Mohammad Soleymani, Guillaume Chanel, Joep JM Kierkels, and Thierry Pun. 2008. Affective ranking of movie scenes using physiological signals and content analysis. In *Proceedings of the 2nd ACM workshop on Multimedia semantics*. ACM, 32–39.
[73] Floris Takens. 1981. Detecting strange attractors in turbulence. *Dynamical Systems and Turbulence, Lecture Notes in Mathematics* 898 (1981), 366–381.
[74] Jussi Tarvainen, Mats Sjöberg, Stina Westman, Jorma Laaksonen, and Pirkko Oittinen. 2014. Content-based prediction of movie style, aesthetics, and affect: Data set and baseline experiments. *IEEE Transactions on Multimedia* 16, 8 (2014), 2085–2098.
[75] Auke Tellegen and Gilbert Atkinson. 1974. Openness to absorbing and self-altering experiences (" absorption"), a trait related to hypnotic susceptibility. *Journal of abnormal psychology* 83, 3 (1974), 268.
[76] Hee Lin Wang and Loong-Fah Cheong. 2006. Affective understanding in film. *IEEE Transactions on Circuits and Systems for Video Technology* 16, 6 (June 2006), 689–704.
[77] Min Xu, L-T Chia, and Jesse Jin. 2005. Affective content analysis in comedy and horror videos by audio emotional event detection. In *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*. IEEE, 4–pp.

[78] Huan Yang, Baoyuan Wang, Stephen Lin, David Wipf, Minyi Guo, and Baining Guo. 2015. Unsupervised extraction of video highlights via robust recurrent auto-encoders. In *Proceedings of the IEEE International Conference on Computer Vision*. 4633–4641.

[79] Ting Yao, Tao Mei, and Yong Rui. 2016. Highlight detection with pairwise deep ranking for first-person video summarization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 982–990.

[80] Feng Zhou, Xingda Qu, Jianxin Roger Jiao, and Martin G Helander. 2014. Emotion prediction from physiological signals: A comparison study between visual and auditory elicitors. *Interacting with computers* 26, 3 (2014), 285–302.

Received May 2017; revised September 2017; accepted December 2017