Movie Genre Classification by Exploiting MEG Brain Signals

Pouya Ghaemmaghami[1], Mojtaba Khomam Abadi[1,4], Seyed Mostafa Kia[1,2,3], Paolo Avesani[1,2,3], and Nicu Sebe[1]

[1] Dept. of Information Engineering and Computer Science, University of Trento, 38123 Trento, Italy
[2] NeuroInformatics Laboratory (NILab), Bruno Kessler Foundation, Trento, Italy
[3] Centro Interdipartimentale Mente e Cervello (CIMeC), University of Trento, Italy
[4] Semantic, Knowledge and Innovation Lab (SKIL), Telecom Italia
p.ghaemmaghami@unitn.it,khomamiabadi@disi.unitn.it,seyedmostafa.kia@unitn.it,avesani@fbk.eu,sebe@disi.unitn.it

Abstract. Genre classification is an essential part of multimedia content recommender systems. In this study, we provide experimental evidence for the possibility of performing genre classification based on brain recorded signals. The brain decoding paradigm is employed to classify magnetoencephalography (MEG) data presented in [1] to four genre classes: Comedy, Romantic, Drama, and Horror. Our results show that: 1) there is a significant correlation between audio-visual features of movies and corresponding brain signals specially in the visual and temporal lobes; 2) the genre of movie clips can be classified with an accuracy significantly over the chance level using the MEG signal. On top of that we show that the combination of multimedia features and MEG-based features achieves the best accuracy. Our study provides a primary step towards user-centric media content retrieval using brain signals.

Keywords: multimedia content retrieval, MEG, genre classification, brain decoding, signal processing.

1 Introduction

Movies are one of the most important sources for entertaining people. Nowadays, thanks to advances in technology, people have access to a large number of movies from various sources. From this fact has emerged the need for automatic movie recommendation and automatically detecting the genre of a movie is an important ingredient of a good recommender system [17, 26]. So far, various automatic genre classification methods have been proposed based on audio-visual features such as average shot length, color variance, saturation, brightness, grayness, motion, and visual excitement [14, 20, 17, 2, 8, 26]. However, analyzing only the audio-visual features may not be sufficient for this task as these features may fail to capture the personal preferences of the human viewer. On the other hand, a recommendation system that can access the viewer's perception of the movie (e.g. via psycho-physiological data), might be able to perform better.

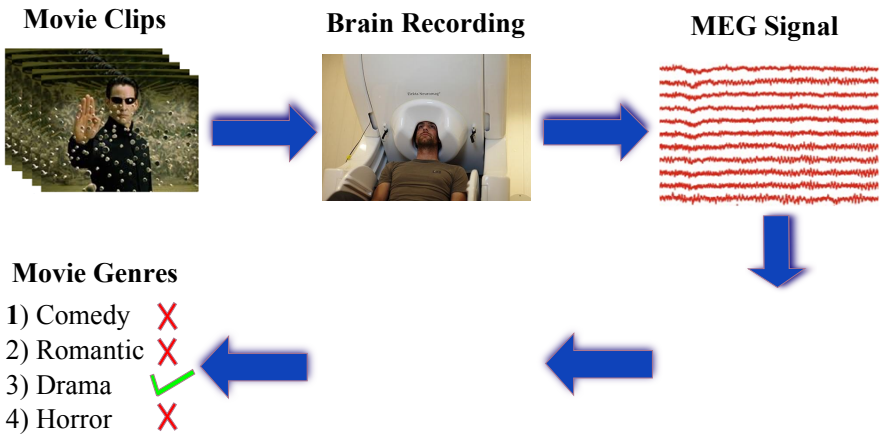


Fig. 1: The framework used in this study for movie genre classification using MEG signals.

Prior works on brain studies have shown that, in the human brain, low-level visual features are encoded in the early visual cortex and high-level visual features are encoded in inferior temporal cortex [15, 22, 21]. Recent papers on brain signal decoding demonstrate that some of these low-level visual features such as orientation, direction of motion and color of visual stimulus can be successfully decoded from brain signals [9, 10, 7]. In a typical brain decoding paradigm, different categories of stimuli are presented to the participants participating in the experiment, while their brain signal is recorded simultaneously. Then a classifier is employed to classify the recorded data into the target stimulus classes. If the classifier performs above chance on the test set, it can be concluded that the stimuli related activities are encoded in the brain signal [6, 4, 3].

In this paper, we address the specific problem of genre classification of movie clips using magnetoencephalography (MEG) data. Our contribution is two-fold: first, we use the correlation analysis to show that genre related information is present in the visual and temporal areas of the brain; second, we illustrate that these genre related brain signals can be decoded to target genre classes using the brain decoding paradigm. We tested our hypotheses on DECAF dataset [1] which contains 30 subjects who watched 36 movie clips. Figure 1 illustrates the overall framework used in our study.

The rest of this paper is organized as follows. In section 2 we briefly review the state of the art on genre classification in the multimedia content analysis context. Then, in section 3 we present the dataset, data preprocessing and feature extraction. Furthermore, we discuss the method used for movie genre annotation and present our correlation and classification analyses. Section 4 presents our experimental results and a brief discussion. Finally, section 5 concludes this paper by stating the future directions.

2 State of the Art

In the literature various content-based genre classification approaches have been proposed based on audio-visual features [14, 20, 17, 2, 8, 26]. Rasheed, et al. [17] used four low-level visual features (average shot length, color variance, motion content and lighting key) to classify over a hundred movie previews into four broad genre categories (Comedy, Action, Drama and Horror). In a similar study [8], the authors used the same low-level visual features and slow and fast moving effects to classify movie previews into three different genres. Elsewhere, Zhou, et al. [26] represented over one thousand movie trailers using a bag-of-visual-words model with shot classes as vocabularies. Then they mapped these bag-of-visual-words models to high-level movie genres.

Apart from audio-visual features, movies can be classified into different genres based on their emotional content. In other words, movie genre can also be described via the affective content of the movie. This emotional content induces an emotional experience in the viewer [23]. In fact, the emotions that are elicited in response to a video clip contain useful information regarding the genre of the video clip [19]. For example, in case of Horror and Action movies, it has been shown that movie segments with high emotion intensity cover the major part of the movie highlights [24].

The common approach for predicting multimedia affect is a content-centric approach, in which audio-visual features of the movie are used for affect prediction. Many researchers have investigated the affective contents of the video clips. Xu, et al. [23], analyzed the affective content of comedy and horror movies by detecting emotional segments. Soleymani, et al. [19] showed that a Bayesian classification approach can tag movie scenes into three affective classes (calm, positive excited and negative excited). They used content-based features extracted from each shot of 21 full length movies. In another study [24], a hierarchical model for analyzing movie affective contents was proposed. The proposed model, firstly, detects the emotional intensity level of the movie using fuzzy clustering on arousal features. Secondly, emotion types (Anger, Sad, Fear, Happy and Neutral) are detected using valence related features. Finally, Hidden Markov Models (HMMs) are applied to capture the context information. A similar hierarchical approach using conditional random fields (CRFs) was proposed in [25].

The second approach regarding multimedia affect prediction is a user-centric approach that aims to capture the emotion of the viewer. In [18], 64 movie scenes were shown to participants while their physiological responses were recorded. Their results showed a significant correlation between self-assessed emotional responses of the participants and the predicted affect from physiological responses. Koelstra, et al. [11] employed electroencephalogram (EEG) to record brain activity of 32 participants watching 40 music video excerpts. A similar study was done in [1] with magnetoencephalography (MEG) data, in which 30 subjects were watching 36 movie clips. These two studies suggest that EEG and MEG data effectively encode emotional information. In this study, we use the MEG data presented in [1], and we show that the extra spatial resolution and the more

user-friendly environment provided by the MEG device provide positive factors in the direction of capturing the emotional response of subjects.

Our brief review of state of the art reveals that movie genre classification has been achieved so far only with content-based multimedia features. On the other hand, user-centered affect recognition can be accomplished using neuroimaging techniques. However, up to now, the efficacy of the brain decoding approaches on movie genre classification has not been investigated. Therefore, our study aims at exploring the possibility of classifying movie genres using MEG data.

3 MEG Signal Analysis

In this section, we describe the employed MEG dataset, the analysis procedure, and our experimental setup.

3.1 Dataset

In this study we used the MEG dataset reported in [1]. This dataset contains 36 movie clips cropped from Hollywood movies (see table 2 for their titles). The MEG brain signals of 30 volunteers were recorded while they were watching the movie clips. All movie clips were shown with 20 frames/second at a screen refresh rate of 60 Hz, and they were projected onto a screen placed about a meter in front of the subject inside the MEG acquisition room. The MEG data were recorded with 1KHz sampling rate in a magnetically shielded room with controlled illumination using a Electa Neuromag system, which has 306 channels via 102 triple-sensors, i.e., 102 magnetometers and 204 planar gradiometers (see [1] for more details).

3.2 MEG Feature Extraction

The MEG data preprocessing has been handled using the MATLAB Fieldtrip toolbox [16]. Following [1], after extracting the MEG trials, we preprocessed the MEG recordings as follows:

Band-pass filtering: Upon downsampling the MEG signal to 300 Hz, in order to remove the noise generated by external perturbations such as moving vehicles or muscle activity, high-pass and low-pass filtering with cut-off frequencies of 1 Hz and 95 Hz are performed, respectively.

Time-Frequency analysis: The spectral power of MEG data between 3 and 45 Hz was estimated using the Welch method with a window size of 300 samples. Following [1], (i) we discarded the magnetometer sensors because they are generally prone to noise and (ii) we used a standard Fieldtrip function to combine the spectral power of planar gradiometers to obtain 102 combined-gradiometer spectral power for each trial.

MEG Features extraction: MEG features are computed by averaging the spectral power over four major frequency bands: theta (3:7 Hz), alpha (8:15 Hz), beta (16:31 Hz), and gamma (32:45 Hz). For each trial of a given subject, the output of this procedure is a 3-dimensional matrix with the following dimensions: 102 (number of the MEG combined-gradiometer sensors) $\times$ 4 (major frequency bands)$\times L$, where L is the length of a video clip in seconds.

3.3 Multimedia Content Feature Extraction

Following [1], the low-level audio-visual features (listed in Table 1) are extracted for each second of the movie clips. The extracted multimedia content analysis (MCA) features include 49 video features and 56 audio features. Hence, for each video, we have 105 (low-level multimedia features) $\times$ L features.

Table 1: Extracted audio-visual features from each movie clip (the number of features is listed in the parenthesis).

Audio features	**Description**
MFCC features (39)	MFCC coefficients [13], derivative of MFCC, MFCC Autocorrelation (AMFCC)
Energy (1) and Pitch (1)	Average energy of audio signal [13] and first pitch frequency
Formants (4)	Formants up to 4400Hz
Time frequency (8)	mean and std of: MSpectrum flux, Spectral centroid, Delta spectrum magnitude, Band energy ratio [13]
Zero crossing rate (1)	Average zero crossing rate of audio signal [13]
Silence ratio (2)	Mean and std of proportion of silence in a time window [13]
Video features	**Description**
Brightness (6)	Mean of: Lighting key, shadow proportion, visual details, grayness, median of Lightness for frames, mean of median saturation for frames
Color Features (41)	Color variance, 20-bin histograms for hue and lightness in HSV space
Motion (1)	Mean inter-frame motion [1]
VisualExcitement (1)	Features as defined in [1]

3.4 Annotating Movie Genres

In order to annotate movie genres, three human observers were asked to watch the movie clips and classify each movie into four genres: Comedy, Romantic, Drama, Horror. The movie genres were picked based on the majority voting between the observers. To evaluate the consistency of the genres across subjects, we measured the agreement between annotators' labeling using the Cohen's Kappa measurement. The average κ across observers is $77\% \pm 2\%$ ($p - value < 0.001$) that suggests a *substantial agreement* [12] between the annotators. Furthermore,

we employed the Cohen's kappa to evaluate the agreement between the movie genres obtained from the majority voting, with the genres obtained from the Internet Movie Database (IMDB)[5]. The average κ across the two labels is 72% ($p-value < 0.001$) that shows a *substantial agreement* between our picked labels (sign majority voting) and the labels obtained from the IMDB. The lack of *full agreement* between these two labels is mainly due to the fact that the employed movie clips in [1] are not necessarily representing the whole movie theme. The genre labels provided by this study augment the dataset proposed in [1]. From here on we refer to the majority voting labels resulting from the annotation process as the ground-truth (see table 2 for the obtained ground-truth labels).

3.5 Correlation Analysis

We calculate the Pearson correlation between the 102 combined MEG gradiometers in each frequency band (θ , α, β, and γ) and audio-visual features extracted from movie clips. The obtained p-values were first fused over all clips and then over all subjects using the Fisher's method [5]. To correct our results for multiple comparisons, we performed the Boferroni correction before reporting significant results. The results of the correlation analysis are discussed in section 4.

3.6 Classification Procedure

In the classification experiments we used a naive Bayes classifier to decode the MEG and MCA features into four genre classes. To do this we performed the following analysis:

MEG-based and MCA-based descriptors: MEG/MCA descriptor of each trial is calculated by averaging the MEG/MCA features over time. Hence, the length of each MEG/MCA descriptor is 408 (4 bands $\times$ 102 triple-sensors) and 105, respectively.

MEG-based user-centric classification: The classification of the MEG-descriptors is repeated 30 times (corresponding to the 30 users) independently. For each user, the 36 MEG-descriptors of 36 movie clips are used as samples.

MCA-based video classification: For the video-centric genre classification, the 36 MCA-descriptors are used as samples.

MEG+MCA: The MEG descriptors of each subject are concatenated to the MCA descriptors and a feature vector of 408+105=513 features is used for genre classification.

Single-subject classification scenario: We used the Naive Bayes classifier under the leave-one-clip-out cross-validation schema for each subject separately.

[5] `http://www.imdb.com`

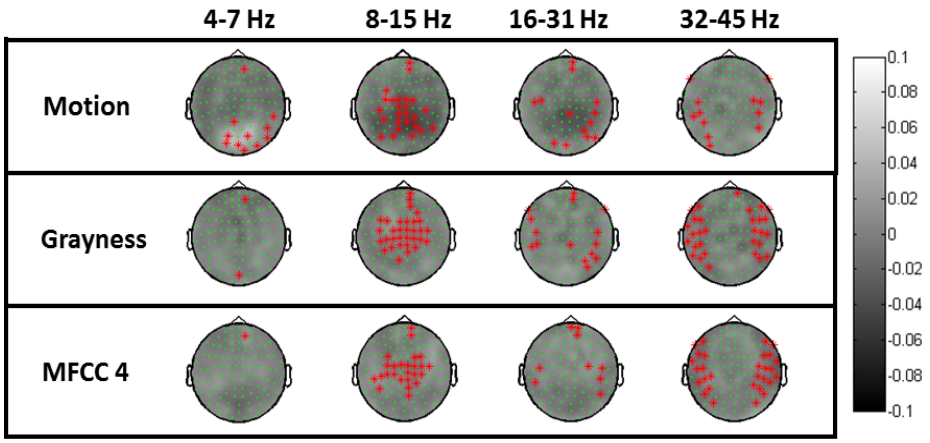


Fig. 2: Pearson correlation analysis between the MEG responses and audio-visual features. Correlation over each channel is denoted by the gray level, and significant correlations are marked with red $\star$.

The ground-truth labels are used as the target labels in the classification procedure.

Population analysis: To evaluate the efficacy of MEG/MCA descriptors at the population level, for each video clip, we computed the majority vote over predictions of the single-subject classification across 30 subjects.

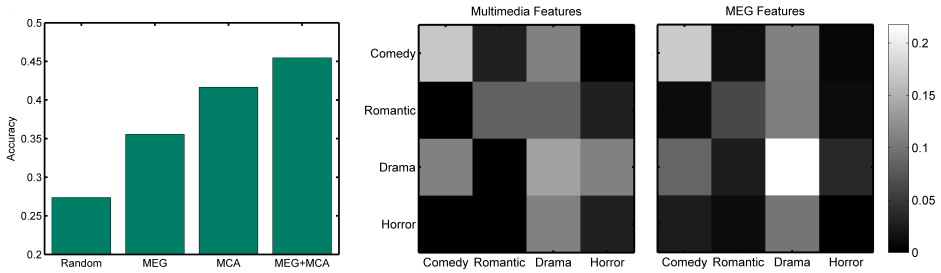
4 Results and Discussion

4.1 Correlation Results

The calculated Pearson correlation between 102 MEG sensors in each frequency band and audio-visual features extracted from movie clips is shown in Figure 2. This figure shows two visual features (motion and grayness) and an audio feature (the forth MFCC coefficient). As one can see, audio-visual features are significantly correlated with MEG sensors in temporal area of the brain in the γ band (32:45 Hz). This part of the brain processes the visual information as well as the audio information. Furthermore in the α band (8-15 Hz), the extracted motion feature is significantly correlated with the MEG sensors located in the posterior part of the brain, confirming previous studies [9, 7].

4.2 Classification Results

Figure 3a summarizes the results of the single-subject classification scenario. It compares the accuracy of four-class classification based on the MEG and MCA features with the chance level (27.4%). The chance level is computed by feeding



(a) (b)

Fig. 3: (a) Comparison between accuracy of MEG and multimedia features with random inputs in the single-subject scenario. (b) Confusion matrix for four-class genre classification using multimedia and MEG features. x and y axes represent predicted and actual labels, respectively.

random numbers with normal distribution into the classification procedure for 100 times. In the MEG case, the mean accuracy of 35.6% is obtained over 30 subjects which is significantly ($p - value < 0.001$) higher than the chance level. This significant difference suggests the existence of genre related information in the recorded brain activity. Furthermore, combining MEG features of each subject with MCA features provides higher accuracy (45.5%) than employing only MCA features (41.7%). This improvement suggests the existence of complementary genre related information in the MEG brain signal.

Figure 3b shows the confusion matrices for four-class genre classification using MCA and MEG features. To facilitate the comparison, the confusion matrices are normalized with respect to the total number of samples (30×36 in the MEG case and 36 in the MCA case). Even though the classification accuracy using MCA features is higher than using MEG features, confusion matrices show significantly similar patterns ($p - value < 2 \times 10^{-5}$). In both cases, the comedy and drama genres are predicted with higher confidence while romantic and horror genres are almost indistinguishable from other categories.

Finally we performed a population analysis on the predicted labels from the MEG signal alone and the combined MEG and MCA features. To do this, the predicted genres at the population level are computed by majority voting over the predicted labels of single-subject predictions. The results are summarized in table 2. The population level accuracy for both MEG and MEG+MCA features is 55.6% which is significantly higher than classification accuracy of only MCA features (41.7%). Despite the same performance in MEG and MEG+MCA scenarios, examining the predicted genres in table 2 shows that the combined features are more successful in predicting the romantic genre. This is mainly because the MEG features are weaker than the MCA features in classifying the romantic genre (compare the 4th and the 5th columns of table 2). This fact con-

firms the existence of complementary genre related information in brain signals and multimedia contents.

Table 2: Movie clip titles, ground-truth labels, and predicted labels.

ID	Titles	Ground-Truth	MCA	MEG	MEG+MCA
1	Ace-Ventura: Pet Detective	COMEDY	COMEDY	COMEDY	COMEDY
2	The Gods Must be Crazy II	COMEDY	DRAMA	COMEDY	COMEDY
3	Liar Liar	COMEDY	ROMANTIC	COMEDY	COMEDY
4	Airplane	COMEDY	COMEDY	COMEDY	COMEDY
5	When Harry Met Sally	COMEDY	COMEDY	COMEDY	COMEDY
6	The Gods Must be Crazy	COMEDY	DRAMA	COMEDY	COMEDY
7	The Hangover	COMEDY	DRAMA	COMEDY	DRAMA
8	Up	COMEDY	COMEDY	DRAMA	COMEDY
9	Hot Shots	COMEDY	DRAMA	COMEDY	DRAMA
10	August Rush	ROMANTIC	DRAMA	DRAMA	DRAMA
11	Truman Show	ROMANTIC	DRAMA	DRAMA	DRAMA
12	Wall-E	ROMANTIC	HORROR	COMEDY	DRAMA
13	Love Actually	ROMANTIC	DRAMA	DRAMA	DRAMA
14	Remember the Titans	DRAMA	HORROR	DRAMA	DRAMA
15	Legally Blonde	COMEDY	COMEDY	DRAMA	DRAMA
16	Life is Beautiful	COMEDY	COMEDY	COMEDY	COMEDY
17	Slumdog Millionaire	ROMANTIC	ROMANTIC	DRAMA	ROMANTIC
18	House of Flying Daggers	ROMANTIC	ROMANTIC	DRAMA	ROMANTIC
19	Gandhi	DRAMA	DRAMA	DRAMA	DRAMA
20	My girl	DRAMA	COMEDY	DRAMA	COMEDY
21	Lagaan	DRAMA	COMEDY	DRAMA	COMEDY
22	Bambi	DRAMA	HORROR	DRAMA	DRAMA
23	My Bodyguard	DRAMA	DRAMA	DRAMA	DRAMA
24	Up	ROMANTIC	ROMANTIC	DRAMA	DRAMA
25	Life is Beautiful	DRAMA	DRAMA	DRAMA	DRAMA
26	Remember the Titans	DRAMA	COMEDY	DRAMA	DRAMA
27	Titanic	DRAMA	HORROR	DRAMA	DRAMA
28	Exorcist	HORROR	HORROR	DRAMA	DRAMA
29	Mulholland Drive	DRAMA	COMEDY	DRAMA	COMEDY
30	The Shining	HORROR	DRAMA	DRAMA	DRAMA
31	Prestige	DRAMA	HORROR	COMEDY	DRAMA
32	Alien	HORROR	DRAMA	DRAMA	DRAMA
33	The untouchables	DRAMA	DRAMA	COMEDY	DRAMA
34	Pink Flamingos	HORROR	DRAMA	COMEDY	DRAMA
35	Crash	DRAMA	DRAMA	DRAMA	DRAMA
36	Black Swan	HORROR	DRAMA	DRAMA	DRAMA
	Accuracy		41.7%	55.6%	55.6%

5 Conclusions

In this paper, we presented an approach for classification of movie clips into four genres (Comedy, Romantic, Drama, Horror) using MEG brain signals. We experimentally illustrated that there exists a significant correlation between audio-visual multimedia features and the MEG signal. This finding opens the door of research toward prediction and reconstruction of multimedia features using brain signals, generally known as mind reading. Furthermore, a naive Bayes classifier has been used to perform the genre class prediction using the features extracted from MEG sensors. Our classification results confirm the possibility of user-centered classification of movies into four broad genres only based on brain signals. In addition, our analysis supports the hypothesis of existence of complementary genre related information in the brain signal and in multimedia contents. To the best of our knowledge, this is the first effort in the direction

of user-centered movie genre classification using brain signals. We plan to extend this work by employing more effective and sophisticated brain decoding approaches in order to improve our classification results. Furthermore, in the future work, we plan to replicate the experiments using the brain signals recorded by portable brain recording devices such as the Emotiv sensors.

Acknowledgments. This work has been supported by the MIUR Cluster project Active Ageing at Home.

References

1. Abadi, M., Subramanian, R., Kia, S., Avesani, P., Patras, I., Sebe, N.: Decaf: Meg-based multimodal database for decoding affective physiological responses. IEEE Transactions on Affective Computing (2015)
2. Brezeale, D., Cook, D.J.: Using closed captions and visual features to classify movies by genre. In: International Workshop on Multimedia Data Mining (2006)
3. Carlson, T.A., Hogendoorn, H., Kanai, R., Mesik, J., Turret, J.: High temporal resolution decoding of object position and category. Journal of vision 11(10) (2011)
4. Cox, D.D., Savoy, R.L.: Functional magnetic resonance imaging (fmri) brain reading: detecting and classifying distributed patterns of fmri activity in human visual cortex. Neuroimage 19(2), 261–270 (2003)
5. Fisher, R.A.: Statistical methods for research workers. Quarterly Journal of the Royal Meteorological Society 82(351), 119–119 (1956)
6. Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., Pietrini, P.: Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293(5539), 2425–2430 (2001)
7. Haynes, J.D., Rees, G.: Decoding mental states from brain activity in humans. Nature Reviews Neuroscience 7(7), 523–534 (2006)
8. Huang, H.Y., Shih, W.S., Hsu, W.H.: A film classifier based on low-level visual features. In: IEEE Workshop on Multimedia Signal Processing. pp. 465–468 (2007)
9. Kamitani, Y., Tong, F.: Decoding motion direction from activity in human visual cortex. Journal of Vision 5(8), 152–152 (2005)
10. Kamitani, Y., Tong, F.: Decoding the visual and subjective contents of the human brain. Nature neuroscience 8(5), 679–685 (2005)
11. Koelstra, S., Muhl, C., Soleymani, M., Lee, J.S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., Patras, I.: Deap: A database for emotion analysis; using physiological signals. IEEE Transactions on Affective Computing 3(1), 18–31 (2012)
12. Landis, J.R., Koch, G.G.: The Measurement of Observer Agreement for Categorical Data. Biometrics 33(1), 159–174 (1977)
13. Li, D., Sethi, I.K., Dimitrova, N., McGee, T.: Classification of general audio data for content-based retrieval. Pattern Recognition Letters 22(5), 533–544 (2001)
14. Nam, J., Alghoniemy, M., Tewfik, A.H.: Audio-visual content-based violent scene characterization. In: International Conference on Image Processing (1998)
15. Obermayer, K., Blasdel, G.G.: Geometry of orientation and ocular dominance columns in monkey striate cortex. The Journal of neuroscience 13(10), 4114–4129 (1993)
16. Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.M.: Fieldtrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Computational intelligence and neuroscience (2010)

17. Rasheed, Z., Sheikh, Y., Shah, M.: On the use of computable features for film classification. IEEE Transactions on Circuits and Systems for Video Technology 15(1), 52–64 (2005)
18. Soleymani, M., Chanel, G., Kierkels, J.J., Pun, T.: Affective characterization of movie scenes based on multimedia content analysis and user's physiological emotional responses. In: IEEE International Symposium on Multimedia (2008)
19. Soleymani, M., Kierkels, J.J., Chanel, G., Pun, T.: A bayesian framework for video affective representation. In: International Conference on Affective Computing and Intelligent Interaction (2009)
20. Sugano, M., Isaksson, R., Nakajima, Y., Yanagihara, H.: Shot genre classification using compressed audio-visual features. In: International Conference on Image Processing (2003)
21. Tanaka, K.: Mechanisms of visual object recognition: monkey and human studies. Current opinion in neurobiology 7(4), 523–529 (1997)
22. Wang, G., Tanaka, K., Tanifuji, M.: Optical imaging of functional organization in the monkey inferotemporal cortex. Science 272(5268), 1665–1668 (1996)
23. Xu, M., Chia, L.T., Jin, J.: Affective content analysis in comedy and horror videos by audio emotional event detection. In: IEEE International Conference on Multimedia and Expo (2005)
24. Xu, M., Jin, J.S., Luo, S., Duan, L.: Hierarchical movie affective content analysis based on arousal and valence features. In: ACM Multimedia (2008)
25. Xu, M., Xu, C., He, X., Jin, J.S., Luo, S., Rui, Y.: Hierarchical affective content analysis in arousal and valence dimensions. Signal Processing 93(8), 2140–2150 (2013)
26. Zhou, H., Hermans, T., Karandikar, A.V., Rehg, J.M.: Movie genre classification via scene categorization. In: ACM Multimedia (2010)