TELKOMNIKA, Vol. 13, No. 4, December 2015, pp. 1343~1351
ISSN: 1693-6930, accredited A by DIKTI, Decree No: 58/DIKTI/Kep/2013
DOI: 10.12928/TELKOMNIKA.v13i4.2735

Recognition of Emotions in Video Clips: The Self-Assessment Manikin Validation

Dini Handayani*[1], Abdul Wahab[2], Hamwira Yaacob[3]
Computer Science Department, Kulliyyah of Information and Communication Technology
International Islamic University Malaysia
*Corresponding author, email: dini.handayani@gmail.com[1], abdulwahab@iium.edu.my[2], hyaacob@iium.edu.my[3]

Abstract

Many research domains use video contents as stimuli for study on human emotions.A video content within a particular genre or a specific the meevokes dynamic emotions that are highly useful in many research fields. The present study proposed a set of video-clip stimuli that embody four emotions under specific genres of movies, namely happiness, calmness, sadness and fear. Two experiments (a preliminary and a validation) were conducted in order to validate the video clips. Self-Assessment Manikin was utilized to rate the videos. All the video clips were rated with respect to valence and arousaljudgment. In the preliminary experiment, the video clipswere rated in terms of how clearly the expected emotions wereevoked. The validation experiment was conducted to confirm the results from preliminary experiment, and only video clips with high recognition rates were included into data set.

***Keywords**: SAM, stimuli, video emotion, valence, arousal*

1. Introduction

Emotional responses to a video content may well be one of the most complex tasks that humans can accomplish. There has been a research trend towardsthe affective computing community todevelop a stimuli repositories and recognize human emotions stimulated by watching video clips, as shown in Table 1. When watching a video, aperson experiences emotion based on his/her cognitive perception and appraisal of the situation depicted in the video [1]. For this reason, it is necessary to understand a human cognitive perception of a given situation and its relation to his/her emotions [2].

Although there is an increasing interest in the recognition of emotions using video stimuli, many questions remain; how do the videosevoke emotions, and to what extent can they do so?To answer these questions, for a start, a set of video stimuli needs to be established.The aim of this study is to provide such stimuli set. Here, four categories of emotionare used; 'happy', 'calm', 'sad', and 'fear'. They are defined on the dimensions of valence and arousal. Valence ranges from positive (pleasant) to negative (unpleasant) while arousal ranges from excited (active) to calm (passive). As presented in Figure 1, the corresponding dimensions of valence and arousal are depicted as horizontal and vertical axes, respectively, on a Cartesian coordinate space. The video stimuli set have to consist of 2^2 = 4 videos of expressions corresponding to the combinations of {*pleasant*, *unpleasant*} ∪ {*active*, *passive*} for each of the emotions.

Two experiments were conducted in order to validatethe video-clip stimuli. In the preliminary experiment, the participants rated the video clips based on the valence and arousal judgement. The aim was to determine which video clips that can be clearly identified (in terms of emotional response) within an optimal duration of time. In the validation experiment, these video clips were rated to findthe ones with highest accuracy that would form the dataset. The rest of the paper is organized as follows: Related works are reviewed in Section 2. Section 3 present and described the material and method. Section 4 presents development of stimuli. Current open issues, future work, and conclusions are covered in Section 5.

Received September 2, 2015; Revised October 29, 2015; Accepted November 20, 2015

2. Related Works

Emphasizing on development of the stimuli set, reviews on seven selected scientific literatures were done based on several categories including the database name,stimuli set size as well as affect representation as shown in Table 1.

Table 1. Mood and Emotion Stimuli Repositories

No	Source	Name	Size	Affect Representation
1	Koelstra et al., 2012 [2]	Database for Emotion Analysis using Physiological Signal (DEAP)	40 music videos; a minuteeach.	Valence, arousal, and dominance.
2	Sandra Carvalho, Jorge Leite, Santiago Galdo-Alvarez, 2012 [3]	Emotional Movie Database (EMDB)	50 film clips; 40 seconds each.	Valence, arousal, and dominance.
3	Douglas-Cowie, Cowie, & Sneddon, 2007 [4]	HUMAINE Database	50 clips; 5 to 180 seconds each.	Intensity, arousal, valence, dominance, and predictability.
4	Schaefer, Nils, Sanchez, & Philippot, 2010 [5]	FilmStim	70 film clips;1 to 7minutes each.	Six emotionsdiscreet and 15 mixed feeling scores.
5	M. Soleymani, Lichtenauer, Pun, & Pantic, 2012 [6]	MAHNOB-HCI	20 film clips; 35 to 117 seconds each.	Arousal, valence, dominance, and predictability.
6	Schedl et al., 2014 [7]	VIOLENT SCENE DATABASE	25 full movies.	Not reported.
7	Baveye, Dellandrea, Chamaret, & Chen, 2015 [8]	LIRIS-ACCEDE	9,800 film clips;8 to 12 seconds each.	Valence and arousal.

With regard to the affect representation, one study used discrete approach to describe emotion, while some others represented emotions in either 2D valence-arousal space or 3D valence-arousal-dominance, as suggested by psychologists.

Although there are many sets of video stimuli as mentioned above, most of them are protected by copyrights and thus not freely available. For video stimuli that were freely available online, some of them no longer do. This prompts the need of a freely available dataset thatis suitable for research on human emotions.

Figure 1. Circumplex Model of Affect from [9] with the emotional state colors [10], wherebythe *x*-axis is for valence scaled and *y*-axis is for arousal scaled

3. Materials and Method

3.1. Emotional Model

The most straightforward way to express emotions is by using categorical approach or discrete labels, such as 'anger', 'contempt', 'disgust', 'fear', 'sad', 'surprise', and 'happy'. On the other hand, psychologists often express emotions in an *n*-dimensional space. Russell [9] proposed a two-dimensional affective space model for measuring emotions known as circumplex model of affect. It is composed of valence and arousal.

Bialoskorski et al., labeled emotional states with colours [10], as illustrated in Figure 1. Happy emotional state, indicated in orange, is defined as having positive valence and high degree of arousal. Calm emotional state, indicated in green, is defined as having positive valence but low degree of arousal. Sad emotional state, indicated in blue, is defined as having negative valence and low degree of arousal. Fear, indicated in red, is defined as having negative valence but high degree of arousal.

3.2. Self-Assessment Manikin (SAM)

3.2.1. Representation

The commonly used technique to validate the emotion stimuli is SAM [11]. SAM is a self-reporting affective state measurement, using cartoon like manikin (see Figure 2) to plot basic emotions on the affective space. A nine-point pictorial scale was utilized for the purpose of this study. In the following, two sets of manikin were used. The first set is the scoring for valence, the range is from nine (happy) to one (sad). The second set is the scoring for arousal, the range is from nine (active) to one (passive).

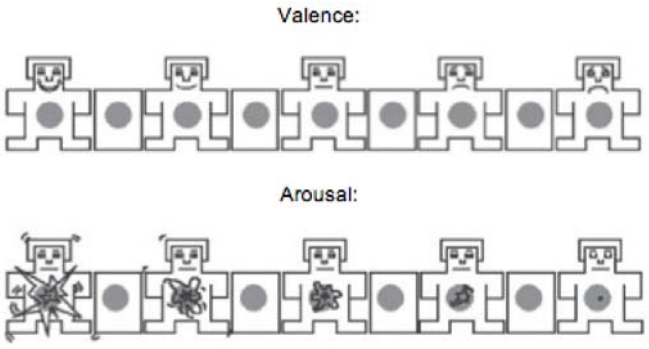


Figure 2. SAM

In order to measure the agreement between emotion labels from video clips and that of SAM, the SAM arousal and valence scores were translated into four emotional states: 'happy', 'calm', 'sad', and 'fear' as described in Section 3.1. Based on the two axes in Figure 1, each participant had to select one of the SAMs. A SAMis defined as 'happy' when the levels of valence and arousalare above 5:

$$(valence \geq 5) \cap (arousal \geq 5) \quad (1)$$

'Calm' is when the levels of valenceis above 5 and arousal below 5:

$$(valence \geq 5) \cap (arousal < 5) \quad (2)$$

'Sad' is when the levels of valence is below 5 and arousal below 5:

$$(valence < 5) \cap (arousal < 5) \quad (3)$$

'Fear' is when the levels of valence is below 5 and arousal above 5:

$$(valence < 5) \cap (arousal \geq 5) \quad (4)$$

3.3 Experiment Protocol and Setup

The participants were briefed about the experiment through a consent form and a verbal introduction. Participants were also instructed on how to fill in their SAM forms. The approximate time interval between the start of a trial and the end of the self-reporting phase was three minutes and ten seconds. Eights video clips were played from the proposed dataset in random order for each participant. The entire protocol took 30 minutes on average, in addition of five minutes of setup time (see Figure 3).

The proposed video set consisted of video clips selected from Asian movie scenes, commercial advertisements, and online resources. The selection criteria for the video clips were as follows: (i) the video clips should be understandable without explanation, (ii) the duration of the video clips should be relatively short, and (iii) the video clips should evoke only one emotion (rather than multiple emotions) from the participants.

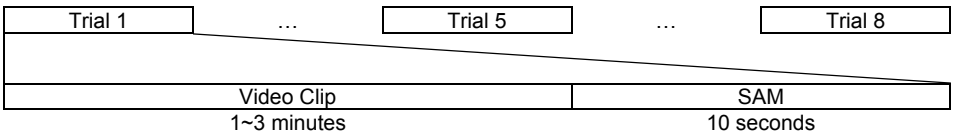


Figure 3. There were eight trials in each experimental session.Each trial was conducted with a video clip. The self-reporting phase was done at the end of each trial

After watching each video clip, the participants were given an SAM form each and asked to provide the following information: (i) valence score, (ii) arousal score, and (iii) the confirmation if they watched the clip prior to the experiment. The experiment was performed in a classroom environment with controlled temperature and illumination.

4. Development of Stimuli

4.1. Description of the Video Set

In this proposed set of video-clip stimuli, four categories of emotion ('happy', 'calm', 'sad', and 'fear') were set. The clips were taken from different films and shows of various genres to express these emotions. In a study done by Ekman, happiness, sadness, and fear were considered as basic emotions[12]. Calmness was notconsidered as a basic emotion in Ekman's study, butitis in this study as the opposite of fear, mirroring the fact that happiness is the opposite of sadness.

'Happy', 'calm', 'sad', and 'fear'are defined as regions along valence and arousal axes as illustrated in Figure 1, together with their explanations in Section 3.1.

'Happy' videos are considered as 'arousing' and 'pleasant'. 'Calm' videos are considered as 'low arousing' and 'pleasant'. 'Sad'video is considered as 'low arousing' and 'unpleasant'. 'Fear' videos are considered as 'arousing' and 'unpleasant'.

To calculate the participants' perception rate (in percentage) in recognizing an emotion in a happy video, V_h, the following formula is used:

$$V_h = \frac{100}{n} * \sum_{i=1}^{n} Ihi \quad (5)$$

where n is the group size of the participants,$lh_i$is the emotional intensity level of each participant when they watch a happy video. Likewise, equation (5) can be re-written for Calmness, Sadness and Fearness intensity level as V_c, V_s and V_f.

4.2. Preliminary Study

4.2.1. Participants

Forty-eightyoung and healthy participants (26 women and 22 men) of different races (Malay and non-Malay) and educational backgrounds volunteered to participate in the preliminary experiment at International Islamic University of Malaysia (IIUM). Their ages varied between 19 to 39 years old, with mean (M) of 21.56 years old and standard deviation (SD) of 5.11 years. They had different educational backgrounds from undergraduate to postgraduate students with different English proficiency from intermediate to native speakers.

4.2.2. Material

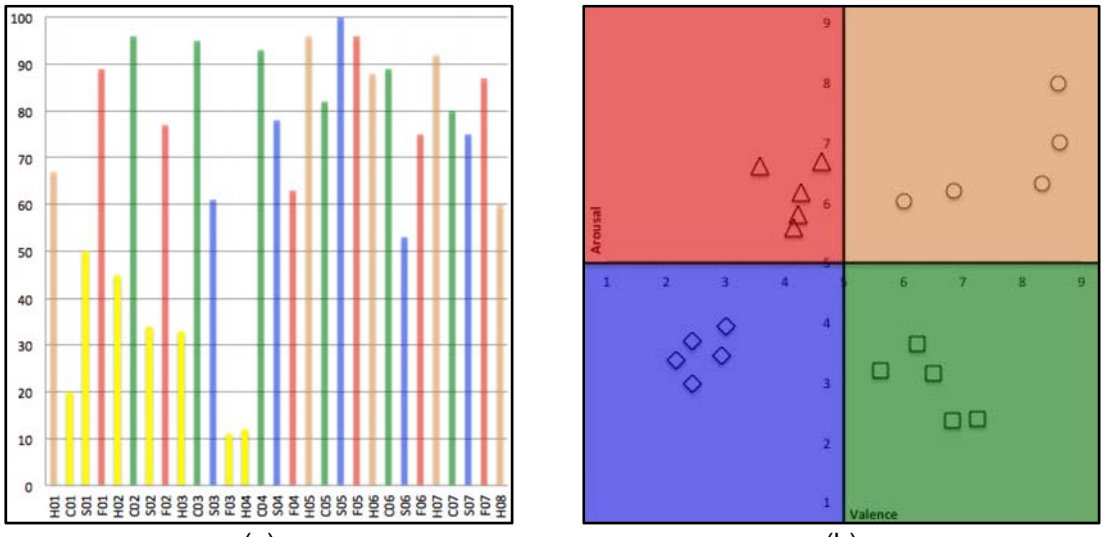
The set of 29 video clips consisting of four categories of emotion were used for the preliminary study. Some of these clipscame with English subtitles. The experiment also examined the role of language in the study of emotions for multiracial participants.The duration of the clipsvaried between 20 seconds to 177 seconds, with M of 87.10 seconds and SD of 37.28 seconds.

4.2.3. Procedure

Each participant wasasked to fill in the SAM form after watching a video clip. Eight video clips were rated by each participant. They were also asked to confirm if they have viewed the clips before; the aim was to obtain their genuine emotional responses.

4.2.4. Results

Table 2 lists the video clips used in preliminary and validation experiments. They are presented and organized by categories of emotion ('H' for 'happy', 'C' for 'calm', 'F' for 'fear', and 'S' for 'sad'). The percentage of the best recognition, duration, and result from the preliminary experiment are shown for each video clip.The result of preliminary experiment is shows in Figure 4. 20 out of 29 video clips are included for validation experiment; consist of five happy videos, five calm videos, five sad videos and five fear videos.



(a) (b)

Figure 4. (a) Percentage rating for each video clip. Yellow bar indicated the excluded video clips for validation study. (b) Mean rating on a 9-points scale obtained for each video clip on valence and arousal. Each symbol represents one video clip.

Each imagewas rated 13 times in average. For each video clip, the recognition result was based on the user perception percentage on the video clips with certain expected emotions.

4.2.5. Discussion

In general, video clips'that express sadness and happiness were slightly better recognisedcompared to those expressing calmness and fear. Psychologists suggested videos to be 1 to 10 minutes of lengthto evoke a particular emotion [13]. For this reason, Video 8 was excluded from the data set eventhough the accuracy was high, due to its short length. However, video 26 also excluded from the data set, due to the long duration. The results showed that the participants recognized multiple emotions in video clips that have durationsof more than two minutes.

Additionally, some of the video clipsthat had no subtitle failed to be recognized by participants. It shows that language plays an important role in recognizing emotions.

The participants could still correctly recognize the videos' emotions eventhough it was theirfirst time watching them. A total of 20 video clips were included for validation experiment.

4.3. Validation Study

4.3.1. Participants

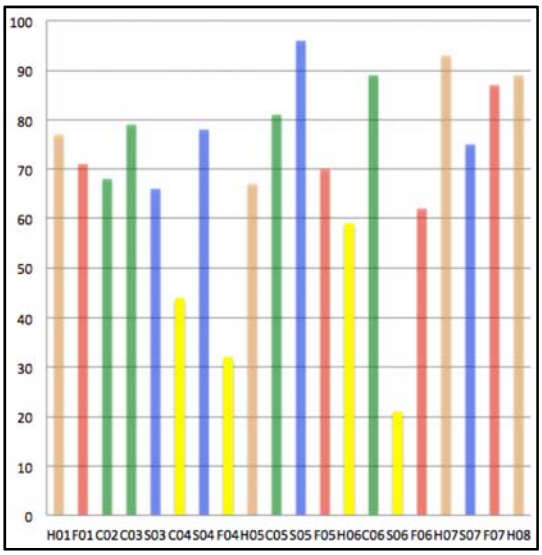
A different group of participants: 113 young and healthy participants (54 women and 59 men) of different races (Malay and non-Malay) and educational backgrounds volunteered to participate in the validation experiment at IIUM. They were undergraduate students with different English proficiency from intermediate to native speakers. In addition, their ages varied between 19 to 21 years old, with M of 19.99 years old and SD of 0.81 years.

4.3.2. Material

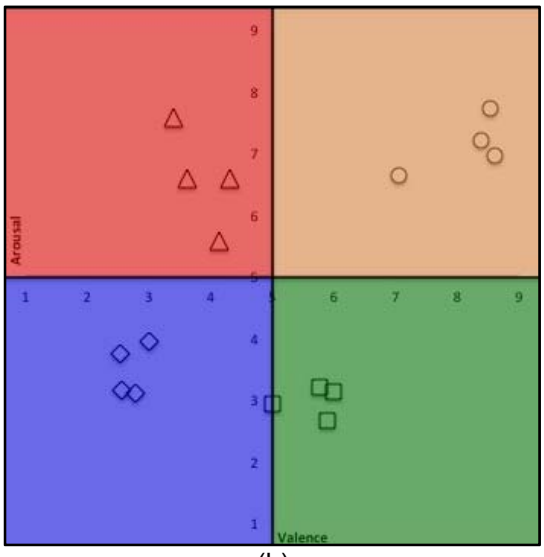
The set of 20 video clips consisting of fourcategories of emotion from the preliminary study was used. Due to language barrier, English subtitles were added to every video clips to avoid failure of recognition. The video clips' durations varied between 60 seconds to 103 seconds, with M of 76.6 seconds and SD of 16.20 seconds.

4.3.3. Procedure

Similar with the preliminary study, participants were asked to fill in their SAM forms after watching a video clip. Eight video clips were rated by each participant. The participants were also asked if they had seen the video clips before.



(a)



(b)

Figure 5. (a) Percentage rating for each video clip. Yellow bar indicated the excluded video clips for a final data set. (b) Mean rating on a 9-points scale obtained for each video clip on valence and arousal. Each symbol represents one video clip.

4.3.4. Results

The percentage of the best recognition, duration, and result from the validation experiment are shown for each video clip, as listed in Table 2.The result of validation experiment is shows in Figure 5.

In the validation experiment, each video was rated four times in average. For each video clip, the recognition result was based on the user perception percentage on the video clips with certain expected emotions.Finally, 16 out of 20 video clips are included for data set; consist of four happy videos, four calm videos, four sad videos and four fear videos.

4.3.5. Discussion

Ultimately, 16 videos were chosen from the validation experiment. Only videos with the recognition accuracy of more than 60% were included for the stimuli dataset. In addition, with regard to the duration, the video clips were kept as short as possible to avoid multiple emotional recognition.

5. General Discussion and Conclusion

In this study, a set of video stimuli had been proposed. Some important issues were discussed; specifically, the duration of the video clips, the authenticity of participants' emotions while watching the video clips, and finally the use of subtitles for multiracial participants. The use of SAM had also been shown to be an effective tool to recognize emotions from valence and arousal dimensions.

Table 2. The video clips used in experimental study

No	Code	Source	Duration (minutes)	Best Recognition Preliminary Experiment (%)	Preliminary Experiment Result	Best Recognition Validation Experiment (%)	Validation Result
1	H01	Maxis Hari Raya 2013 TVC (Eng.)	1	67	Included for validation.	77	Included for dataset.
2	C01	Incredible India	1.58	20	Excluded.		
3	S01	Touching Thai Advertisement, Shows How A Single Act of Kindness Could Change Your Life	2.57	50	Excluded.		
4	F01	The Grudge Movie, Scariest Horror Scene	1.43	89	Included for validation.	71	Included for dataset.
5	H02	PETRONAS Jahit 60s TVC	1	45	Excluded.		
6	C02	Beach DVD-Wave-With Relaxing Beaches and Sea Sounds	1.03	96	Included for validation.	68	Included for dataset.
7	S02	Raya TVC PTS Media Group - 'Ibu, Al-Fatihah Tu Apa'	2.57	34	Excluded.		
8	F02	Missing Our Deals Will Haunt You - Little Girl TV Advert	.20	77	Excluded.		
9	H03	[Thai TVC] 'Mae Toi' - Thai Life Insurance	1.57	33	Excluded.		
10	C03	'Havasupai Indian Waterfall Relaxation' The Classic Video by David Huting	1	95	Included for validation.	79	Included for dataset.
11	S03	A Blind Father and His Daughter-Short Sad Story	1	61	Included for validation.	66	Included for dataset.
12	F03	Proton Advertisement, Seat Belt	1	11	Excluded.		
13	H04	CNY Commercials 2013 - BERNAS - 'Ka Fan'	1.3	12	Excluded.		
14	C04	Cuia de Viagem-Langkawi	1.43	93	Included for validation.	44	Excluded.
15	S04	BERNAS-Chinese New Year Commercial-Family Reunion Dinner 'Sek Fan'	1	78	Included for validation.	78	Included for dataset.
16	F04	The Grudge 3-scariest scene	1.20	63	Included for validation.	32	Excluded.
17	H05	Nido Milk-You're My Number One 2014 TVC Sharon Cuneta & Barbie Almalbis	1.30	96	Included for validation.	67	Included for dataset.

18	C05	Relaxing DVD-Mangrove Journey-Tropical Waterfalls with Nature Sounds	1	82	Included for validation.	81	Included for dataset.
19	S05	The Saddest Commercial Ever	1.30	100	Included for validation.	96	Included for dataset.
20	F05	The Ring-best scene as a horror movie	1.38	96	Included for validation.	70	Included for dataset.
21	H06	Dtac TriNet-Happiness	1.21	88	Included for validation.	59	Excluded.
22	C06	Robin Bird Chirping and Singing - Song of Robin Red Breast Birds - Robins	1.12	89	Included for validation.	89	Included for dataset.
23	S06	'Crash' Saddest scene	1.36	53	Included for validation.	21	Excluded.
24	F06	The scariest scene ever-The Eye-Horror movie	1.17	75	Included for validation.	62	Included for dataset.
25	H07	Baby Laughing Hysterically at Ripping Paper	1	92	Included for validation.	93	Included for dataset.
26	C07	Heartwarming Thai Commercial - Thai Good Stories	2.55	80	Excluded.		
27	S07	Line TVC-Closer	1.30	75	Included for validation.	75	Included for dataset.
28	F07	The most scary scene on roof	1.09	87	Included for validation.	87	Included for dataset.
29	H08	Tourism Australia's new ad	1	60	Included for validation.	89	Included for dataset.

With regard to the validation of the stimuli, as a future work, additional experiments to measure emotions are needed with an implicit approach such as electro encephalogram (EEG) to automatically recognize participants' emotions when they watch these video clips. Although there are other measurement tools available, they seem less suitable for recognition of emotionsat first glance.

In conclusion, by creating this dataset, it is hoped that it can resolve the lack of availability of previous data sets and it can be easily shared and used by other researchers in the field of affective computing.

Acknowledgment

This work is supported by Fundamental Research Grant Scheme (FRGS) funded by the Ministry of Higher Education (Grant code: FRGS14-*137-0378).

References

[1] KR Scherer. "What are emotions ? And how can they be measured ?". *Soc. Sci. Inf.* 2005; 44(4): 695–729.

[2] S Koelstra, C Muhl, M Soleymani, JS Lee, A Yazdani, T Ebrahimi, T Pun, A Nijholt and I (Yiannis) Patras. "DEAP : A Database for Emotion Analysis Using Physiological Signals". *IEEE Trans. Affect. Comput.* 2012; 3(1): 18–31.

[3] ÓFG Sandra Carvalho, Jorge Leite, Santiago Galdo-Álvarez. "The Emotional Movie Database (EMDB): A Self-Report and Psychophysiological Study". *Appl. Psychophysiol. Biofeedback*. 2012; 37(4): 279–294.

[4] E Douglas-Cowie, R Cowie and I Sneddon. "The HUMAINE database: addressing the collection and annotation of naturalistic and induced emotional data". *Affect. Comput. Intell. Interact.* 2007: 488–500.

[5] A Schaefer, F Nils, X Sanchez and P Philippot. "Assessing The Effectiveness of a Large Database of Emotion-Eliciting Films: A New Tool for Emotion Researchers". *Cogn. Emot.* 2010: 1–36.
[6] M Soleymani, J Lichtenauer, T Pun, and M Pantic. "A Multimodal Database for Affect Recognition and Implicit Tagging". *IEEE Trans. Affect. Comput.* 2012; 3(1): 42–55.
[7] M Schedl, M Sjoberg, I Mironica, B Ionescu, VL Quang, YG Jiang, and CH Demarty. "VSD2014: A Dataset for Violent Scenes Detection in Hollywood Movies and Web Videos". *Content-Based Multimed. Index. (CBMI), 2015 13th Int. Work.* 2014.
[8] Y Baveye, E Dellandréa, C Chamaret and L Chen. "LIRIS-ACCEDE: A Video Database for Affective Content Analysis". *Affect. Comput. IEEE Trans.* 2015: 1–14.
[9] J Russell. "A circumplex model of affect". *J. Pers. Soc. Psychol.* 1980.
[10] LSS Bialoskorski, JHD Westerink and EL van den Broek. "Mood Swings: An affective Interactive Art System". *ICST Inst. Comput. Sci. Soc. Informatics Telecommun. Eng. 2009*. 2009: 181–186.
[11] M Bradley and PJ Lang. "Measuring Emotion: The Self-Assessment Manikin and The Semantic Differential". *J. Behav. Ther. Exp. Psychiat.* 1994; 25(I).
[12] P Ekman, D Matsumoto and WV Friesen. "Facial Expression in Affective Disorders". *What face Reveal. Basic Appl. Stud. Spontaneous Expr. Using Facial Action Coding Syst.* 1997.
[13] M Soleymani, M Pantic and T Pun. "Multimodal Emotion Recognition in Response to Videos". *IEEE Trans. Affect. Comput.* 2012; 3(2): 211–223.