Data in Brief (2021)

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Dataset on Emotions using Naturalistic Stimuli (DENS)

Sudhakar Mishra[a,*], Mohammad Asif[a,1], Uma Shankar Tiway[a]

[a] *Indian institute of information technology, Allahabad, India*

ARTICLE INFO

Naturalistic study

self-assessment
normal EEG
Emotion
Multimedia stimuli

ABSTRACT

Emotion is a constructed phenomenon that emerges from the dynamic interaction of multiple components neurologically, physiologically and behaviorally. Such dynamics can not be captured by static and controlled experiments. Hence, the study of emotion with a naturalistic paradigm is needed. In this dataset, multimedia naturalistic stimuli are used to acquire the emotional dynamics using EEG, ECG, EMG and behavioural scales. The stimuli are multimedia videos collected from youtube for 372 affective words, analyzed with multimedia content analysis to filter out non-emotional stimuli and then validated with university students. The validated stimuli had the least variance in subjective ratings on self-assessment scales. The stimuli are then used to acquire neurological dynamics along with peripheral channels and subjective ratings-valence, arousal, dominance, liking, familiarity, relevance and emotion category. Both the raw data and pre-processed data is provided along with the pre-processing pipeline. This data can be utilized to study dynamic activation and connectivity in the whole brain source localization study, understand the mutual interaction between the central and autonomic nervous system, understand temporal hierarchy using multiresolution tools, and perform machine learning-based classification and complex networks analysis. The data is accessible at 10.18112/openneuro.ds003751.v1.0.0

http://dx.doi.org/10.1016/j.dib.xxxx.xx.xxx

2352-3409/ © 2021 The Author(s). Published by Elsevier Inc.

Specifications Table

Subject	Neuroscience: Behavioral
Specific subject area	Emotion processing captured by neural dynamics and self-assessment scales with ecologically valid naturalistic paradigm
Type of data How data were acquired	Electroencephalography (EEG) data Data is acquired using 128 Channel Geodesic EEG System 400. The software used for RAW data acquisition is Net Station. Along with 128 EEG channels, one ECG, two EMG signals are also recorded. The sampling frequency is 250 Hz.
Data format	Raw (.set and .fdt) files and pre-processed (.mat) EEG data
Parameters for data collection	Data was acquired in a quiet lab environment with maintained room temperature. Participants were seated on an immovable chair with an armrest. Participants are instructed and trained for the Emotion task. They were shown emotional multimedia stimuli on the screen and responded with a keyboard and mouse.
Description of data collection	40 participants (with a low score on axis-I) participated in our study of emotions. Resting-state data for these participants is recorded. To study dynamic processing of emotions, participants are shown naturalistic emotional stimuli, and their response is dynamically collected with the precaution that the cognitive load of the task is minimum.
Data source location	Indian Institute of Information Technology Allahabad Devghat, Jhalwa, Prayagraj-211015, U. P. INDIA
Data accessibility	
Related research article	A study of spatio-temporal dynamics of emotion processing using DENS dataset.

Value of the Data

- Emotion is a situated conceptualization. Hence, emotion processing should be studied with a naturalistic paradigm to understand the emotional dynamics with its multi-component constituentsgm. We collected data in an ecologically valid experimental setting using naturalistic emotional stimuli.
- Your second bullet point must explain who can benefit from these data? This data could be utilized by research who are trying to understand emotion dynamics, but their effort is hindered by data collected using a controlled paradigm with artificial stimuli.
- Complex networks analysis and dynamic tracking of causality can be done on this data. In addition, machine learning-based classification can determine the characteristic statistical signature of emotions and their processing.
- Using the data new understanding of emotion processing dynamics can be developed. With the understanding of emotion processing with normal healthy subjects as a reference, brain stimulation studies can be done to cross-verify the results for the healthy subject. With the understanding of emotion mechanisms, emotion disorder conditions using brain stimulation can be created and studied.

Data Description

Dataset_description file described the metadata for the dataset. Participants related details are described in participants.json and participants.tsv files. Each subject directory contains a

*Corresponding author: Tel.: +91-7318317901;
e-mail: ust@iiita.ac.in (Uma Shankar Tiway)

self-assessment-feedback.tsv file having entries about the feedbacks given by subject on self-assessment scales-valence, arousal, dominance, liking, familiarity, relevance and emotion category. In addition, it contains information about the time-stamp of mouse click and other details. The EEG directory inside subject directory contains the raw EEG data in .set & .fdt format along with the information about task events in _task-emotion_events.tsv file. The stimuli directory contains stimuli that were used during the experiment. In addition, feedback excel sheet participant_details.xlsx filled by participants is also added. The code directory contains the python code for data collection, python code for data validation and Matlab file for pre-processing the raw data.

Experimental Design, Materials and Methods

1. Stimuli Collection, Selection and Validation

1.1. Stimuli Collection

We have collected 15000 raw multimedia videos using google YouTube API in python. These videos were collected for affective words taken from Affect Net[1, 2][For block diagram see fig-1]. Collectively, 372 affect words from affect-net were used as a search query. One hundred fifty videos were crawled over youtube for each search query, and 150 comments per video were fetched. These comments of the searched videoID's are analyzed using python libraries [2]. All the comments were given some score with positive or negative signs based on the polarity of the sentence. The videos with a high polarity based on the comment analysis were selected to download. These downloaded videos were processed for multimedia content-based analysis (described in the next subsection) following the generation of scene affective vector and calculation of valence and arousal scores for one-minute excerpts from the video. Three high-scoring excerpts on valence and arousal space were selected from each video for the validation study. These selected videos were shown to 661 students (undergraduate and postgraduate students) from IIIT-Allahabad for validation in the next stage (Lab setting for validation, fig-1).

1.2. Extracting stimulus features using multimedia content analysis and performing stimulus selection

Valence and arousal are psycho-physiological dimensions used in literature to represent emotions. Valence gives information about unpleasant to pleasant feelings, whereas arousal informs about the intensity of these emotional feelings [3]. Apart from subjective feedback, the valence and arousal dimension is also used for autonomic rating of stimuli using multimedia content analysis[4]. In the following section, we will describe the extracted audio-visual features associated with valence and arousal. Each video is divided into several one-minute excerpts, and the multimedia content analysis is done for each of these excerpts

1.2.1. Visual Features and their relation to valence and arousal

We have calculated the visual feature vector for each frame which includes motion vector calculated using lukas-kanade point tracking algorithm [5], average motion magnitude [6], shot detection [7], rhythm component [6], brightness of each frame [8, 9, 10] and shot rate [10, 11].

For motion analysis, we are calculating motion vectors using the Lukas-Kanade point tracking algorithm [5]. Motion in scenes is a highly effective stimulating feature to evoke strong emotional responses and can modulate arousal in viewers as reported by [12]. In contrast, valence was reported to be intact with the introduction of motion in the scene. The average of

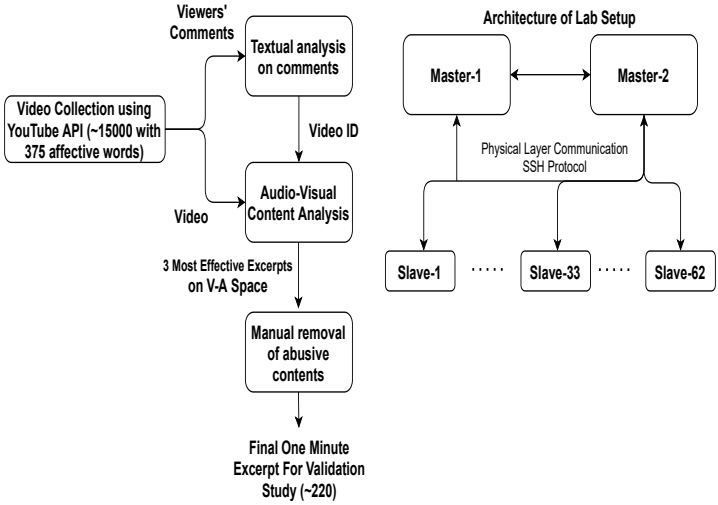


Fig. 1: Left figure represents the procedure which is used to collect the stimuli from internet and then performing textual and multimedia content analysis to finally select around 220 stimuli. The mulitmedia feature which are calculated on these stimuli are described in table-2. Right figure is representing lab configuration where stimuli validation is carried out. All the system had linux mint operating system installed in it. A master-slave architecture based on ssh protocol is implemented. All the stimuli are are divided into 22 blocks randomly into the master system and then distributed to all the slaves dynamically while students are being informed about the experiment by a slide show presentation (supplementary presentation). Once, students are clear about instructions they are assigned one of the slave system. To be noted, the distance between participants were around 2 meters so that they are more focused on their own screen (they are also requested for the same). Lights were off and windows were properly shielded from light so that an immersive environment can be created. Three volunteers were monitoring the activities and were also available for any help to the participants. Once the experiment is done, in master system a script run by the experimenter which collected all the responses from the slave system to the master system.

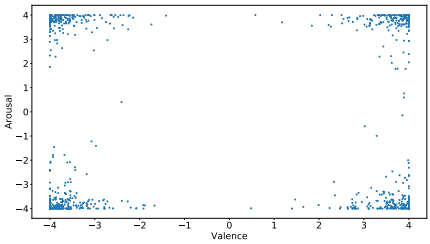


Fig. 2: **Multimedia content based valence-arousal plot:** The valence and arousal is calculated using multimedia content analysis and only those one-minute excerpts are selected and plotted here which are on the periphery. The magnitude of vector is calculated using the formula $e_i = \sqrt{a_i^2 + v_i^2}$. Smaller e_i is closer to the neutral state. Excerpts with a higher e_i score are selected for the validation study.

motion magnitude representing motion for one frame is calculated, and likewise, the calculation for all the subsequent frames [6] is done. For shot detection difference of histogram is calculated. Based on the histogram, a frame is designated as shot boundary if it is more than a particular threshold [7]. A high shot rate could introduce stress and accented moments, whereas a low shot rate could be used to de-accentuate an action [13]. In this way, the shot rate can be used to introduce arousal in the scene. To calculate rhythm component difference between two subsequent shot frames' index is taken into account [6]. Brightness feature of every frame is calculated using value element of HSV image space representation [8, 9, 10].

1.2.2. Audio Features and their relation to valence and arousal

For the audio analysis, we have used the pyAudioAnalysis python library [14]. Using this library, 34 features have been calculated. It includes Zero Crossing Rate(ZCR); short-term energy; short-term entropy of energy; spectral centroid; spectral spread; spectral entropy; spectral flux; spectral roll-off; 13 features of MFCC; 12 features of chroma vector; and chroma deviation. These features have been used to create an audio scene effect vector. [11] have also used these vectors for music classification in 7 emotion categories. Following the concept of using short-duration clip segments [11] to detect better the effects change in scene, we have used a segment of 200ms for calculating the above-mentioned audio affect vector.

Features in this audio scene affect vector have been reported to be associated with valence and arousal. In the literature [15, 16] following association of audio features with valence and arousal have been reported. ZCR [8, 9, 17]; short-time energy [8, 9, 10, 11]; short-term entropy of energy [11]; spectral entropy [11]; spectral flux [8, 17, 11]; spectral rolloff [8, 17, 11]; MFCC [17, 11] and chroma deviation [11] reported to be associated with arousal. On the other hand, spectral centroid; spectral spread; MFCC [11] and chroma vector [11] reported in association with valence.

Using the audio-visual feature vectors as described in this section, we have calculated the valence and arousal and plotted them as shown in fig-2). Using the valence and arousal vector, the emotional highlight score e_i for each one minute excerpt is calculated using the formula $e_i = \sqrt{a_i^2 + v_i^2}$. Smaller e_i is closer to the neutral state. Excerpts with a higher e_i score are selected for the validation study. The valence and arousal are plotted in V-A space for each excerpt extracted from the video (as shown in fig-2).

Audio Features		
Features	**Description**	**V/A**
MFCC	Mel frequency cepstral coefficients from a cepstral representation where the frequency bands are not linear but distributed according to the mel-scale.	V,A
Spectral entropy	Entropy of the normalized spectral energies for a set of sub-frames	A
ZCR	The rate of sign-changes of the signal during the duration of a particular frame	A
Short-term energy		
Short-term entropy of energy	The entropy of sub-frames normalized energies. It can be interpreted as a measure of abrupt changes	
Spectral centroid	Spectral centroid is the ”center of mass” of the spectrum and represents the brightness of a sound	V
Spectral spread	The second central moment of the spectrum	V
Spectral flux	The squared difference between the normalized magnitudes of the spectra of the two successive frames	A
Spectral roll-off	The frequency below which 90% of the magnitude distribution of the spectrum is concentrated	A
Chroma vector	A 12-element representation of the spectral energy where the bins represent the 12 equal tempered pitch classes of western type music(semitone spacing)	A
Chroma deviation	The standard deviation of the 12 chroma coefficients	A
Video Features		
Motion Vector	Motion vector using lukas-kanade algo	A
Corner		
Average of motion magnitude	Intensity difference between two successive frames	A
shot length		V,A
Rhythm component		V
Brightness		V
Shot rate		A

Table 2: **Audio-Visual Features** Audio-video multimedia feature which has been used for multimedia context analysis.

1.3. Stimulus Validation

Participants Details: All the participants are recruited from the Indian Institute of Information Technology, Allahabad, India. The mean age of the participants is 21(±3). The gender ratio of participation is skewed with more participation by male students. In the validation study, 548 males and 113 females had participated. Only a negligible number of participants were left-handed(36 out of 661). Mostly, participants are engineering undergrads fig-3a. As it is known that India is a multilingual and multicultural country, we asked participants about their mother tongue in the response form. Participation came from a wide variety of native language speakers(29 languages). Although, Hindi was dominant, followed by Telugu, Punjabi, Marathi, and Bengali fig-3c. Hence, these stimuli can be used in multilingual and multicultural conditions also. In addition, participants rated their mood overall during the experiment (6.73 ± 1.97), which were overall towards the pleasant side. Concerning participant characteristics, gender distribution was 85% males and 15% females. The average age of participants was 23.3 ± 1.4 years and the average mood rating was 6.73 ± 1.97 on 1-10 scale. Distribution of cognitive load during the experiment

Since the number of candidate stimuli came after multimedia content analysis, for validation were large (230 stimuli), we needed enough volunteers, in proportion, for the validation study. As mentioned above, six hundred sixty-one volunteers volunteered in the study. The statistics about volunteers are provided in fig-3a, 3c, 3b, and summary statistics about rating is plotted in fig-3d and also attached with the submission online.

We did our validation study in a lab setting described in the (fig-1). All the participants were placed approximately 2 meters apart, and an immersive environment was created for the experiment (by switching off the lights and creating a quiet space). The validation study took three months to complete starting from data from the first batch of participants. As mentioned, 230 stimuli were divided into blocks. Each block contains only ten stimuli which created 23 blocks. Before the arrival of the new set of participants, the blocks are recreated by re-assigning videos randomly and automatically so that any experimenter bias can be avoided (e.g., more negative videos in a block than others). Periodically, we removed those stimuli which had higher variation in their ratings. In this case, heuristically, we set 3 (value of standard deviation) as a threshold value for valence, arousal, and dominance scales. This procedure allowed us to get more samples for the stimuli, which caused less variation (i.e., more consistent) in valence, arousal, and dominance responses by subjects. Fig-3d shows a scatter plot for the mean valence and mean arousal calculated for each video stimuli across the participants (for 69 stimuli). The fig-3d clearly shows that even after the fully automatic procedure, the stimuli for the fourth quadrant were not rated by the participants. Hence, we are not getting a perfect 'U' shape. However, the data is not linear and is inclined towards a 'U' shape. These results are in sync with the previous reports on valence-arousal space, where they are non-linearly related with the 'U' shape[3].

2. Experiment

As mentioned above, we adapted the naturalistic paradigm with complex and naturalistic stimuli for our probe on emotion processing. The experiment is designed to fill the gap between emotion processing observed in a controlled lab setting and an ecologically valid naturalistic environment. The presented single-trial experiment localizes the emotional event felt while the participant is immersed in the context presented through naturalistic stimuli. At the same time, it addresses pitfalls of the naturalistic paradigm reported earlier (e.g., continuous tracking of

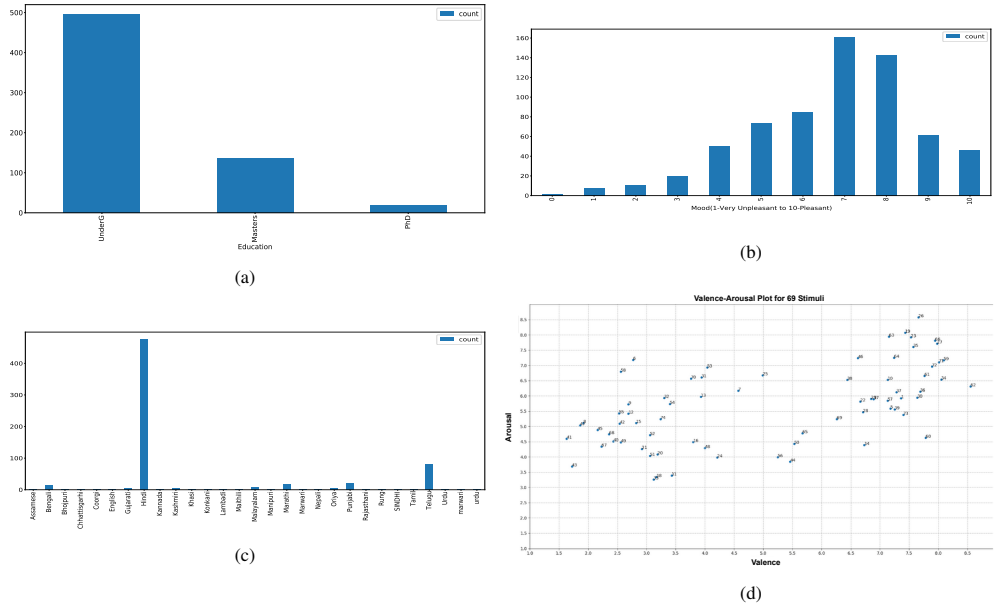


Fig. 3: (a)Time duration before click when participants consciously responded to their feeling of any emotion. (b) Education, (c) Mood, (d) Language. (e)The dots in the scatter plot is representing mean value of valence and arousal for each stimuli. The annotated number is the serial number against the video as per the video_id given in excel file (summary_data_frame).

feedback of emotional state [18]). A total of 40 participants participated in this experiment with mean age (23.35 ± 1.34) and three female participants. All were enrolled in the institute in the bachelor and master courses.

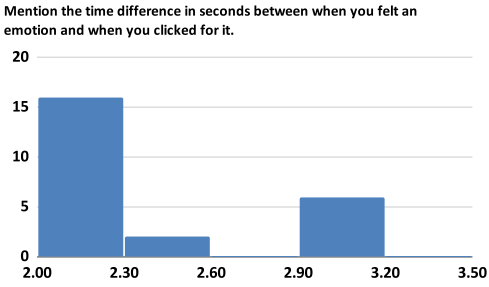
2.1. Experiment Setup

Laboratory Environment

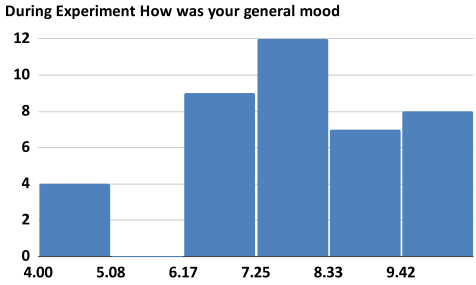
The experiment was performed in a single laboratory environment. During the experiment, all the lights of the room were switched off. The chair with armrest was immovable, and the screen was kept at a 1-meter distance from the participant. The stimuli were presented on a 15.6-inch screen display (1024x768 resolution) using a dedicated computer system of Intel core i5 3rd generation, 3.2GHz. However, the resolution of the video stimuli was reduced to 800x600. For the audio, Sennheiser CX 180 Street II in-Ear Headphone was used. Participants were allowed to adjust the audio volume to their comfort level during training. For response, a mouse and a keyboard were also placed.

EEG Setup

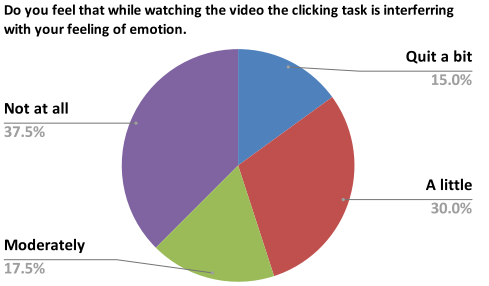
For recording EEG, 128 Channel Geodesic EEG System 400 was used. A dedicated computer system, iMac, was used for recording and storing these signals. 128 EEG channels cap (HCGSN v.1.0) and three peripheral signals (ECG and MEG electrode pairs with Physio16 Package) were correctly placed. Physiological signals were recorded at a sampling rate of 250 Hz using net station software.



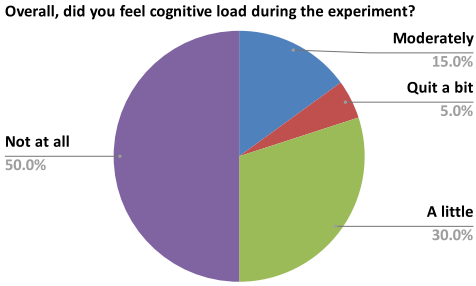
(a) Time Difference (seconds) between click and awareness of feeling



(b) General Mood of participants during experiment



(c) Click is Interfering with emotion feeling



(d) Cognitive Load during experiment

Fig. 4: **Participants Questionnaire** (a) Time difference between the moment when participants got aware about emotional feelings and the clicking, (b) overall general mood of the participants while performing the experiment, (c) how much a non-emotional clicking task was interfering with the feeling of emotions, (d) how much cognitive load participants felt during the experiment.

2.2. EEG Experiment

2.2.1. Experiment Overview

In the experiment room, first of all, participants filled the consent form. Then, the participant was thoroughly informed about the experiment and protocols using a presentation (experiment presentation [online]) by the experimenter.

Next, the participant was led to the experiment room. The EEG cap was placed correctly on the participant's head. Electrodes are adjusted and refilled with the saline solution for conduction (if needed). The impedance is checked and kept under 30Kohm. In addition, peripheral channels ECG & EMG are placed correctly at their desired spot on the body. The right and left ECG electrodes are placed directly below the clavicle near the right and left shoulder. The EMG electrodes are placed on the backside of the neck, left and right lateral gastrocnemius, and on the participant's hand for click. Then, a prototype experiment (i.e., Training) was done by the participant. Up to this phase, the participant was free to ask any doubt.

When the whole experimental protocols are cleared, and the participant is familiar with the whole system and instructions, the experimenter checks and confirms that the impedance is below 30KOhm, confirms that all the room lights are off, leaves the room, and start recording the signals. Participants were allowed to call the experimenter in case of any discomfort.

The experimenter checks the impedance after each trial (on the recording screen) while the participant responds on the response windows. If the impedance goes above the threshold, the experimenter pauses the experiment, applies the saline solution to assure that the impedance is below the threshold, asks the participant for water. Then the experimenter leaves the room with lights off, and the recording resumes again.

After performing the experiment successfully, participants filled the google form asking for the demographic details. We also asked for the time difference between when they got aware of felt emotion and clicked. Six participants did not respond. 30 out of 34 remainings reported that click was done between 2-3 seconds after awareness for the feeling of emotion (fig-4a). In addition, we asked participants to rate their mood (fig-4b), cognitive load during the experiment (fig-4d), and how much they felt interfered by the clicking while watching the stimuli (fig-4c). In the end, participants are asked for tea. After completing everything, they are asked to leave the room.

2.2.2. Experiment Paradigm

The paradigm is comprised of bio-calibration, training, and testing modules which are described in detail below.

Bio-calibration: The experiment in the laboratory starts with bio-calibration. Participants were to perform a specific task when the instructions for it comes on the monitor screen in front of them. Twenty tasks were performed as instructed to the participants. The tasks include Look right, Look left, Lookup, Look down, Wiggle your left toe, Hold your breath, Take a deep breath, Breath-in & breath-out, Grit teeth, Make the click event, Move hand slowly between keyboard and mouse, Press right arrow key in the keyboard, Imagine the legs movement on a beat, Imagine the finger movement on the beat, Look straight ahead, Move the eyeball left, right, up, down, Tongue movement, Chewing, Swallowing, Close the eyes. These signals are recorded to create a template and recognize them as noise in the EEG signal (if considered part of the analysis).

Training: The next two modules, namely training, and testing were similar in steps. The only differences are in terms of the number of trials and duration of the resting state recording. The

training and testing modules are further divided into four sub-modules as follows.
1. Quiz: This sub-module was performed to assess the participants' understanding of the Valence and Arousal scales. A set of multiple-choice questions (stating an imaginary emotional scenario in English textual form on the screen; experiment presentation [online]) were displayed to the participants one by one (total 11 questions; experiment presentation [online]). After each question, participants had to select one option from the four given choices. The four choices were a) High Valence High Arousal (HVHA), b) High Valence Low Arousal (HVLA), c) Low Valence High Arousal (LVHA), d) Low Valence Low Arousal (LVLA). The experiment goes to the next sub-module once the participants answered at least 10 out of 11 questions correctly.
2. Resting-State Recording: Participants were instructed to sit in the relaxing posture, no physical movement, neither click nor use mouse for the duration of the fixation mark appearing on the screen (eyes opened). The duration of this baseline was 10 seconds for training.
3. Stimuli Presentation (Trials) & responses: In this sub-module, participants have presented a 1-minute long stimulus video. Participants were instructed to click anywhere on the screen with the help of the mouse during the stimulus video presentation only when they felt any emotion. The number of clicks was left intended upon the participants with no lower or upper limit. Next, after the stimulus window, the response window appears.

4. Participants' Self Assessment: In the set of response windows, after each stimulus video, participants had to do self-assessments on continuous(valence, arousal, dominance, and relevance), discrete(liking and familiarity) for each trial. First, participants were required to rate for the whole stimulus on the scales mentioned above. After that, each click event has to be categorized in a particular emotion category selected from a list of presented emotions for each quadrant of valence-arousal space.

There was no time limit during the self-assessment. Once the participants performed the self-assessment on the response window and provided feedback, a trial is completed. Between each trial, fixation marks appear. Simultaneously, a straightforward arithmetic calculation is presented to the participants to reduce the effect of the last presented stimuli. It is optional to attend this calculation. The participant can take as long as he/she wants before going to the successive stimuli by performing the left mouse click. After clicking, the subsequent trial begins, and the same process repeats. For valence, arousal, and dominance, participants had to choose a point on a continuous scale ranging from 1 to 9 with the help of the mouse. Manikins were also visible to assist the participants (as shown in the experiment presentation [online]). The valence, arousal, and dominance scales range from unpleasant to pleasant, inactive to active, and submissive to dominant, respectively. For liking and familiarity, participants had to choose a point on a discrete scale of range 1 to 5. Both the scales range from Least to Much. For Relevance, participants had to choose a point on a continuous scale ranging from 1(Not related) to 5(Completely related) spectrum. For each click-event, a new window (click window; for reading convenience) appears after rating the scales mentioned above. In each click window, three consecutive images (the image frames at 20 frames before click, at click, 20 frames after click, from left to right) are shown on a single screen to aid the participants in recalling the context and the emotion they felt. In the bottom panel of the same window, the participants chose an emotion they felt while clicking. To choose an emotion, they first had to select a category among four categories- HVHA, HVLA, LVHA, LVLA and then selected an emotion from the presented list of emotions.

Only three trials were given in the training module. So that participants can get used to the whole paradigm & response behaviour. After the training, the experimenter entered the room, asked for any doubt, refilled the saline solution on scalp electrodes, ensured the impedance was

below 30 Kohm, and informed the participant that the next module would be the testing module.

Testing: The same paradigm is followed for the testing module as it was for the training module, except the quiz is not presented this time, the duration of resting-state is increased to 80 seconds, and the total number of trials is increased to 11 compared to 3 in the training module. After the experiment, the general questionnaire is presented using a google form which participants were asked to fill.

Overall subjective feedback: At the end of the experiment, subjects are asked to fill the google form asking for name, age, gender, and other details (Detail of Google Form [online]). Participants are also asked if they felt cognitive load during the experiment. The rating distribution is as follows: not at all:20, a little:12, moderately:0, quit a bit:2, very much:0. Participants are asked, "Do you feel that while watching the video, the clicking task is interfering with your feeling of emotion. For example, you were more concerned about clicking than getting involved with the video and feel emotions." They responded as follows: not at all:15, a little:12, moderately:7, quit a bit:6, very much:0.

Acknowledgments

We wish to extend our sincere thanks to Prof. Narayanan Srinivasan, CBCS, University of Allahabad. He kindly gave the SM some very useful pieces of advice related to the experiment. Dr Sonia baloni ray, IIIT-Delhi, has introduced SM to the EEG technology and taught him how to use it. She demonstrated it to him by letting him participate in her attention study. Then, we would like to thank Mr Amit Tiwary (then M.tech student), Mr Pravin Srivastav (then PhD candidate), and Ms Anandpreet Kaur (then PhD candidate) for helping us in conducting the validation study.

References

[1] A. Valitutti, C. Strapparava, O. Stock, Developing affective lexical resources., PsychNology Journal 2 (2004) 61–83.
[2] F. Å. Nielsen, A new anew: Evaluation of a word list for sentiment analysis in microblogs, arXiv preprint arXiv:1103.2903 (2011).
[3] J. A. Russell, A circumplex model of affect., Journal of personality and social psychology 39 (1980) 1161.
[4] A. Yazdani, E. Skodras, N. Fakotakis, T. Ebrahimi, Multimedia content analysis for emotional characterization of music video clips, EURASIP Journal on Image and Video Processing 2013 (2013) 1–10.
[5] C. Tomasi, T. Kanade, Detection and tracking of point, Technical Report, features. Technical Report CMU-CS-91-132, Carnegie, Mellon University, 1991.
[6] A. Hanjalic, L.-Q. Xu, Affective video content representation and modeling, IEEE transactions on multimedia 7 (2005) 143–154.
[7] A. Pardo, Simple and robust hard cut detection using interframe differences, in: Iberoamerican Congress on Pattern Recognition, Springer, 2005, pp. 409–419.
[8] Y. Cui, S. Luo, Q. Tian, S. Zhang, Y. Peng, L. Jiang, J. S. Jin, Mutual information-based emotion recognition, in: The Era of Interactive Media, Springer, 2013, pp. 471–479.
[9] S. Zhang, Q. Tian, Q. Huang, W. Gao, S. Li, Utilizing affective analysis for efficient movie browsing, in: 2009 16th IEEE International Conference on Image Processing (ICIP), IEEE, 2009, pp. 1853–1856.
[10] G. Irie, T. Satou, A. Kojima, T. Yamasaki, K. Aizawa, Affective audio-visual words and latent topic driving model for realizing movie affective scene classification, IEEE Transactions on Multimedia 12 (2010) 523–535.
[11] H. L. Wang, L.-F. Cheong, Affective understanding in film, IEEE Transactions on circuits and systems for video technology 16 (2006) 689–704.
[12] R. F. Simons, B. H. Detenber, T. M. Roedema, J. E. Reiss, Emotion processing in three systems: The medium and the message, Psychophysiology 36 (1999) 619–627.

[13] B. Adams, C. Dorai, S. Venkatesh, Novel approach to determining tempo and dramatic story sections in motion pictures, in: Proceedings 2000 International Conference on Image Processing (Cat. No. 00CH37101), volume 2, IEEE, 2000, pp. 283–286.

[14] T. Giannakopoulos, pyaudioanalysis: An open-source python library for audio signal analysis, PloS one 10 (2015) e0144610.

[15] S. Wang, Q. Ji, Video affective content analysis: a survey of state-of-the-art methods, IEEE Transactions on Affective Computing 6 (2015) 410–430.

[16] S. Wu, T. H. Falk, W.-Y. Chan, Automatic speech emotion recognition using modulation spectral features, Speech communication 53 (2011) 768–785.

[17] S. Arifin, P. Y. Cheung, Affective level video segmentation by utilizing the pleasure-arousal-dominance information, IEEE Transactions on Multimedia 10 (2008) 1325–1341.

[18] M. Andric, S. Goldin-Meadow, S. L. Small, U. Hasson, Repeated movie viewings produce similar local activity patterns but different network configurations, NeuroImage 142 (2016) 613–627.