A dataset of continuous affect annotations and physiological signals for emotion analysis

Karan Sharma[1*], Claudio Castellini[1], Egon L. van den Broek[2],
Alin Albu-Schaeffer[1] & Friedhelm Schwenker[3]

December 10, 2018

1. Institute of Robotics and Mechatronics, DLR – German Aerospace Center, Wessling, Germany.
2. Department of Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands.
3. Institute of Neural Information Processing, Ulm University, Ulm, Germany.
*corresponding author(s): Karan Sharma (karan.sharma@dlr.de)

Abstract

From a computational viewpoint, emotions continue to be intriguingly hard to understand. In research, direct, real-time inspection in realistic settings is not possible. Discrete, indirect, post-hoc recordings are therefore the norm. As a result, proper emotion assessment remains a problematic issue. The Continuously Annotated Signals of Emotion (CASE) dataset provides a solution as it focusses on real-time continuous annotation of emotions, as experienced by the participants, while watching various videos. For this purpose, a novel, intuitive joystick-based annotation interface was developed, that allowed for simultaneous reporting of valence and arousal, that are instead often annotated independently. In parallel, eight high quality, synchronized physiological recordings (1000Hz, 16-bit ADC) were made of ECG, BVP, EMG (3x), GSR (or EDA), respiration and skin temperature. The dataset consists of the physiological and annotation data from 30 participants, 15 male and 15 female, who watched several validated video-stimuli. The validity of the emotion induction, as exemplified by the annotation and physiological data, is also presented.

Background & Summary

The field of *Artificial Intelligence* (AI) has rapidly advanced in the last decade and is on the cusp of transforming several aspects of our daily existence. For example, services like customer support and patient care, that were till recently only accessible through human–human interaction, can nowadays be offered through AI-enabled conversational chatbots [1] and robotic daily assistants [2],

respectively. These advancements in interpreting explicit human intent, while highly commendable, often overlook implicit aspects of human–human interactions, especially the role emotions play in the same. Addressing this shortcoming is the aim of the interdisciplinary field of *Affective Computing* (AC, also known as *Emotional AI*), that focuses on developing machines capable of recognising, interpreting and adapting to human emotions [3, 4].

A major hurdle in developing these *affective* machines is the internal nature of emotions that makes them inaccessible to external systems [5]. To overcome this limitation, the standard AC processing pipeline [6] involves: (i) acquiring measurable indicators of human emotions, (ii) acquiring subjective annotations of internal emotions, and (iii) modelling the relation between these indicators and annotations to make predictions about the emotional state of the user. For undertaking steps (i) and (ii) several different strategies are used. For example, during step (i) different modalities like, physiological signals [5, 7, 8], speech [9] and facial-expressions [10, 11] can be acquired. Similarly, the approaches to step (ii) usually vary along the following two main aspects. First, on the kind of annotation scale employed, i.e., either discrete or continuous. Second, on the basis of the emotion-model used, i.e., either discrete emotion categories (e.g., joy, anger, etc.) or dimensional models (e.g., the *Circumplex model* [12]). Traditionally, approaches based on discrete emotional categories were commonly used. However, these approaches were insufficient for defining the strength [8, 6, 11] and accounting for the time-varying nature [13] of emotional experiences. Therefore, nowadays continuous annotation based on dimensional models is preferred and several annotation tools for undertaking the same have been developed [14, 15, 16]. Notwithstanding these efforts at improving the annotation process, a major impediment in the AC pipeline is that both, steps (i) and (ii), require direct human involvement in form of subjects from whom these indicators and annotations are acquired [8, 17]. This makes undertaking these steps a fairly time-consuming and expensive exercise.

To address this issue, several (uni- and multi-modal) datasets that incorporate continuous annotation have been developed. Principal among these are the DEAP [18], SEMAINE [19], RECOLA [20], DECAF [21] and SEWA [22]. The annotation strategy used in these datasets have the following common aspects. First, the two dimensions of the Circumplex model (i.e., *valence* and *arousal*) were annotated separately. Second, in all datasets except SEWA, that uses a joystick, mouse-based annotation tools were used. In recent years, both these aspects have been reported to have major drawbacks [23, 24, 22]. These being, that separate annotation of valence and arousal doesn't account for the inherent relationship between these dimensions [23, 25], and that mouse-based annotation tools are generally less ergonomic than joysticks [23, 24, 26, 22]. To address these drawbacks, we developed a novel *Joystick-based Emotion Reporting Interface* (JERI) that facilitates simultaneous annotation of valence and arousal [27, 28, 16, 29]. A testament to the efficacy of JERI is that in recent years, several similar annotation setups have been presented [25, 30]. However, currently there are no openly available datasets that feature JERI-like setups.

To address this gap, we developed the presented *Continuously Annotated*

Signals of Emotion (CASE) dataset (Data Citation 1). It contains data from several physiological sensors and continuous annotations of emotion. This data was acquired from 30 subjects while they watched several video-stimuli and simultaneously reported their emotional experience using JERI. The physiological measures included in the dataset are from Electrocardiograph (ECG), Blood Volume Pulse (BVP), Galvanic Skin Response (GSR), Respiration (RSP), Skin Temperature (SKT) and Electromyography (EMG) sensors. The annotation data has been previously used for several publications aimed at introducing, analysing and validating this approach to emotion annotation [28, 16, 29]. However, it has not been previously released. To the best of our knowledge, this is the first dataset that features continuous and simultaneous annotation of valence and arousal, and as such can be useful to the wider Psychology and AC communities.

Methods

Participants

Thirty volunteers (15 males, age 28.6±4.8 years and 15 females, age 25.7±3.1 years; range of age 22-37 years) from different cultural backgrounds participated in the data collection experiment. These participants were recruited from an organisation-wide call for volunteers sent at the Institute of Robotics and Mechatronics, DLR. Upon registering for the experiment, an email containing general information and instructions for the experiment was sent to the participants. In this email, they were asked to wear loose clothing and men were asked to preferably shave facial hair, to facilitate the placement of sensors. All participants had a working proficiency in English and were communicated to in the same. More information on the sex, age-group, etc., of the participants is available in the metadata to the dataset (Data Citation 1).

Ethics Statement

This experiment is compliant with the World Medical Association's Declaration of Helsinki, that pertains to the ethical principles for medical research involving human subjects, last version, as approved at the 59th WMA General Assembly, Seoul, October 2008. Data collection from participants was approved by the institutional board for protection of data privacy and by the work council of the German Aerospace Center. A physician is part of the council that approved the experiment.

Experiment Design

The experiment was setup with a *within subjects design*. Accordingly, *repeated measures* were made and all participants watched and annotated the different video-stimuli used for the experiment. To avoid carry-over effects, the order of the videos in a viewing session was modified in a pseudo-random fashion, such

Figure 1: The typical experiment setup (left) shows, a participant watching a video and annotating using JERI (joystick circled). The video-playback window (center) with the embedded annotation interface, that has the Self-Assessment Manikin (SAM) added to the valence and arousal axes (right).

that the resulting video sequence was different for every participant. To isolate the emotional response elicited by the different videos, they were interleaved by a two-minute long blue screen. This two-minute period also allowed the participants to rest in between annotating the videos. More information on the video-sequences is available in the dataset (Data Citation 1).

Experiment Protocol

On the day of the experiment, the participants were provided an oral and a written description of the experiment. After all questions regarding the experiment were addressed, the participants were asked to sign the informed consent form. Then, a brief introduction to the 2D circumplex model was provided and any doubts about the same were clarified. Following this, physiological sensors were attached and the participant was seated facing a 42" flat-panel TV (see Figure 1, left). Detailed information was then provided on the annotation procedure. It was emphasised to the participants that they should annotate their emotional experience resulting from the videos, and not the emotional content of the videos. To accustom the participants to the annotation interface, they were asked to watch and annotate five short practice videos. During this practice session, the experiment supervisor intervened whenever the participant asked for help and if required, provided suggestions at the end of every video. This session was also used to inspect the sensor measurements and if required, make appropriate corrections. After the practice session, the experiment was initiated and lasted for approximately 40 minutes. At the end of the experiment, feedback on the annotation system was acquired using the SUS questionnaire [16, 31]. Then, the sensors were removed and refreshments were offered. The participants were also encouraged to share any further insights they had on the experiment.

Annotation Interface

Figure 1 (right) shows the design of the annotation interface. It is based on the 2D circumplex model that has been supplemented with the Self-Assessment-Manikin (SAM) [32] on its co-ordinate axes. These manikin depict different
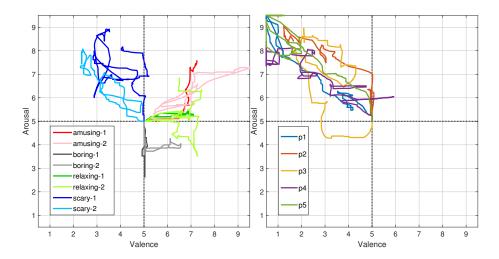
Figure 2: The plot on the left shows the annotations from one participant for the different videos (see Table 1) in the experiment. The annotations for the 'scary-2' video by the first five participants (labelled as p1–p5) can be seen in the plot on the right.

valence (on X–axis) and arousal (on Y–axis) levels, thereby serving as a non-verbal guide to the participants during annotation. The red pointer in the figure shows the resting/neutral position. The participants were instructed to annotate their emotional experience by moving/holding the red pointer in the appropriate region of the interface. Since the annotation was done over the entire length of a video, it results in a continuous 2D trace of the participant's emotional experience (see Figure 2). The position of the annotation interface inside the video-playback window is shown in Figure 1 (center). This position can be easily changed, but since none of the participants requested that, it was retained as shown for all participants. The annotation interface was developed in the National Instruments (NI) LabVIEW programming environment.

Videos

In this experiment, the aim was to elicit multiple emotional states, namely *amusing, boring, relaxing* and *scary*, through video-stimuli. To this end, 20 videos previously used by other studies were shortlisted [33, 34, 35]. The emotional content of these videos was then verified in a pre-study, where 12 participants (no overlap with the participants of this study) viewed and rated these videos remotely using a web-based interface. Based on the results of this pre-study and further internal reviews, eight videos were selected for the main experiment, such that there were two videos each for every emotional state that we wanted to elicit. Additionally, three other videos were also used in the experiment, i.e., the start-video, the end-video and the interleaving blue-screen videos. More

information on the videos is available in Table 1, in the Usage Notes section and in the dataset (Data Citation 1).

Source	Video-Label	Video-ID	Intended Attributes		Dur. [s]
			Valence	Arousal	
Hangover	amusing-1	1	med/high	med/high	185
When Harry Met Sally	amusing-2	2	med/high	med/high	173
European Travel Skills	boring-1	3	low	low	119
Matcha: The way of Tea	boring-2	4	low	low	160
Relaxing Music with Beach	relaxing-1	5	med/high	low	145
Natural World: Zambezi	relaxing-2	6	med/high	low	147
Shutter	scary-1	7	low	high	197
Mama	scary-2	8	low	high	144
Great Barrier Reef	startVid	10	-	-	101
Blue screen with end credits	endVid	12	-	-	120
Blue screen	bluVid	11	-	-	120

Table 1: The source, label, ID used, intended valence-arousal attributes and the duration of the videos used for the dataset.

Sensors & Instruments

The physiological sensors used for the experiment were selected based on their prevalence in AC datasets and applications [3, 6, 18, 10]. Other sensors and instruments were chosen based on either the recommendations of the sensor manufacturer or on how interfaceable they were with the data acquisition setup. More details on these sensors and instruments are provided in this subsection and Table 2.

ECG sensor. The electrical signal generated by the heart muscles during contraction can be detected using an ECG sensor. The procedure used involves placement of three electrodes in a triangular configuration on chest of the participant. Two electrodes are placed on the right and left *coracoid processes* and the third on the *xiphoid process* [36]. This sensor also pre-amplifies and filters the detected electric signal.

BVP sensor. Also known as a Photoplethysmography (PPG) sensor, it emits light into the tissue and measures the reflected light. The amount of observed reflected light varies according to the blood flowing through the vessels, thus serving as a measure for cardiac activity. This sensor was placed on the middle finger of the non-dominant hand [36].

GSR sensor. Also known as Electrodermal Activity (EDA) sensor, it measures the variation in electrical conductance resulting from sweat released by the glands on the skin. The two electrodes emanating from this sensor were placed on the index and ring fingers of the non-dominant hand [36].

Respiration sensor. The expansion and contraction of the chest cavity can be measured using a Hall effect sensor placed around the *pectoralis major* muscle [36]. Thus, this sensor measures the respiration rate of the participant.

Skin temperature sensor. Small variations in skin temperature were measured and converted to electrical signals using an epoxy rod thermistor. This sensor was placed on the pinky finger of the non-dominant hand [36].

EMG sensors. The surface voltage associated with muscle contractions can be measured using a surface-Electromyography (sEMG, simply EMG) sensor. Previous research in AC has generally focused on three muscles. These are the *corrugator supercilii* and *zygomaticus major* muscle groups on the face, and the *trapezius* muscle on the upper-back. Accordingly, a total of three EMG sensors (one each for the aforementioned muscles) were used for the experiment. These sensors also pre-amplify and perform an analog Root Mean Square (RMS) conversion on the measured raw EMG signal [36].

Sensor isolators. The sensor manufacturer recommends using a 'sensor isolator' to ensure electrical isolation between the participants and the powered sensors. Accordingly, the physiological sensors were indirectly connected to the data acquisition module, through these sensor isolators (see Figure 3).

Data acquisition modules. A 32-channel (16-channel differential) Analog-to-Digital Conversion (ADC) module with 16-bit resolution was used to acquire the output voltages from the sensor isolators (indirectly, the sensors). This module is connected to a Data Acquisition (DAQ) system that transfers the data to the acquisition PC.

Joystick. The joystick is the only instrument in the experiment that is directly controlled by the participants. The used joystick is a generic digital gaming peripheral that features a return spring. This provides the user proprioceptive feedback about the location of the pointer in the interface, thereby helping to mitigate the cognitive load associated with simultaneous tasks of watching the video and annotating her emotional experience [16, 25].

Sensor/Instrument	No.	Manufacturer	Model	Conversions	
				Equations	Units
ECG sensor	1	Thought Technology	SA9306	$V_{out} = (V_{in} - 2.8)/50 \cdot 10^3$	mV
BVP sensor	1	Thought Technology	SA9308M	$BVP\% = 58.962 V_{in} - 115.09$	$\%$
GSR sensor	1	Thought Technology	SA9309M	$G = 24 V_{in} - 49.2$	μS
Respiration sensor	1	Thought Technology	SA9311M	$R\% = 58.923 V_{in} - 115.01$	$\%$
Skin temp. sensor	1	Thought Technology	SA9310M	$T = 21.341 V_{in} - 32.085$	$^{\circ}C$
EMG sensor	3	Thought Technology	SA9401M-50	$V_{out}(RMS) = (V_{in} - 2)/4000 \cdot 10^6$	μV
Sensor Isolator	2	Thought Technology	SE9405AM	-	-
ADC module	1	National Instruments	NI 9205	-	-
DAQ system	1	National Instruments	cDAQ-9181	-	-
Joytick	1	Thrustmaster	T.16000M	$pos_{out} = 0.5 + 9 \cdot (pos_{in} + 26225)/52450$	-

Table 2: The type, number (No.), manufacturer and model of different sensors and instruments used in the experiment. Wherever applicable, the conversion equations used to transform the logged input values to the desired output units/scales (see last column) are also presented.

Data Acquisition

Figure 3 shows the experiment and the data acquisition setup. The video-playback, the annotation interface and the data acquisition components were
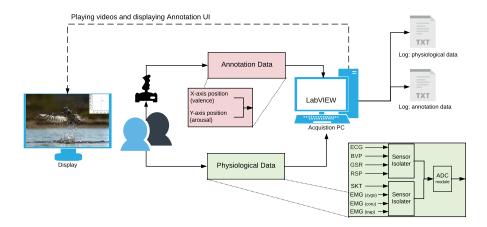
Figure 3: The schematic shows the various aspects of the experiment and the data acquisition setup. The arrows indicate the direction of the data-flow. The solid and the dotted lines indicate the primary and secondary tasks of the acquisition process, respectively.

all directly managed through LabVIEW. This allows for a seamless integration of all these different components. The open-source VLC media player was used for video-playback. The joystick was directly connected to the acquisition PC over a USB port. The physiological data was acquired over Ethernet using the DAQ system. The acquisition rate for the annotation and the physiological data was 20 Hz and 1000 Hz, respectively. The acquired data was augmented with the timestamp provided by the acquisition PC and logged in two different text files, i.e., one each for the physiological and the annotation data. The same process was repeated for all participants resulting in 60 (30×2) log files.

Data Preprocessing

The procedure used for transforming the raw log files into the presented dataset is summarized in this subsection. In the following, step 1 was performed once and steps 2–5 were iteratively applied to log files for each participant.

1. *Duration of the videos:* using the *ffprobe* tool from the *FFmpeg* multimedia framework [37], the exact duration (in milliseconds) of the videos was determined and has been made available in the dataset (Data Citation 1).

2. *Transforming raw data:* the sensor input received by the sensor isolators gets modified before being transferred to the DAQ system. To rectify the effects of this modification, the logged voltages need to be transformed. This was achieved by applying the equations presented in Table 2 to yield the desired output with specific units/scales (Table 2, last column). Similarly, the logged annotation data, which was in the integer

interval $[-26225 .. 26225]$, was also rescaled to the annotation interface interval $[0.5 .. 9.5]$ by using the equations presented in Table 2.

3. *Data interpolation:* a common problem in data acquisition and logging is the latencies that can be introduced during any of these processes. This was also evident in our data, where, e.g., the time between the subsequent samples of the annotation data was occasionally more than the expected 50 ms. To address this issue, linear interpolation was performed on the physiological and the annotation data. For undertaking the same, first, two time-vectors with sampling intervals of 1 ms (for the physiological data) and 50 ms (for the annotation data) were generated based on the time-duration of the logged data. These vectors serve as the query points for the interpolater that determines the value at these points by fitting a line between the corresponding discrete samples in the logged data. As a result of the interpolation process, the resulting sampling intervals for the physiological and the annotation data were 1 ms and 50 ms, respectively. For the eventuality that other researchers might prefer to use either the non-interpolated data or different interpolation methods, the original non-interpolated data is also available in the dataset (Data Citation 1).

4. *Adding the video-IDs:* the log files contain timestamps, but do not have information identifying the duration and the order of the videos. Hence, the extracted video-durations and a lookup table of the video-sequences were used to identify the data segments pertaining to each video. Then, this information was added as an extra column to the log files, containing the different video-IDs (see Table 1). This process was also undertaken for the non-interpolated data.

5. *Saving the data:* the resulting data from the aforementioned steps was saved into two different *comma-separated values* (csv) files, i.e., one each for the physiological and the annotation data. The csv format was chosen as it is natively accessible by different programming and scientific computing frameworks.

Code availability

The LabVIEW-based graphical code for the experiment and data acquisition is highly specific to the sensors and equipment used in our experiment. It has therefore not been made available with this dataset (Data Citation 1). Nevertheless, readers who wish to replicate the experiment can contact the corresponding author for further assistance. For readers who want to reproduce the experiment, we hope the detailed description provided in this article will suffice.

The data preprocessing steps outlined in the previous subsection were implemented in MATLAB 2014b. The linear interpolation was performed using the *interp1* function. The raw log files, that the data preprocessing code acts upon, are not a part of the released dataset because they do not contain the video-IDs. Hence, the preprocessing code has also not been released. Nevertheless, both

the raw data and the preprocessing code, are available to interested researchers upon request.

Data Records

The presented CASE dataset comprises of the processed data resulting from the aforementioned experiment. This dataset is hosted as a single archive file on the *figshare* data repository (Data Citation 1) and is organised into three main directories. These are, (i) the *interpolated*, (ii) the *non-interpolated* and (iii) the *metadata* directories. At the root and the subsequent sub-directories, detailed *README* files explaining the contents of these directories have been provided.

metadata

This directory contains auxiliary information about the experiment organised into the following three files:

1. *participants.xlsx:* this Excel file contains the participant-ID, the sex, the age-group and the ID of the video-sequence used, for all participants in the experiment.

2. *video_sequences.xlsx:* the video-stimuli were shown in a unique sequence to every participant. The columns of this Excel file contain video-IDs that indicate the ordering of the video-stimuli in these sequences.

3. *videos.xlsx:* in addition to the attributes already presented in Table 1, this Excel file contains further information on the used video-stimuli. This includes, the videos' durations in milliseconds, links to the IMDb/YouTube entries for the videos' sources, URLs to the videos and the time-window for the videos at these URLs. More information on how to acquire these videos is presented in the Usage Notes section.

interpolated and non-interpolated

Both of these directories share similar structure and filenaming conventions, with the only difference being the process used to generate the files contained in them (see the Subsection on Data Processing). Hence, unless stated otherwise, the description provided here is applicable to the files in both of these directories. The data contained in these directories is organised into two sub-directories. These being, (i) the *annotations* and (ii) the *physiological* directories containing the participant-wise annotation and physiological data, respectively. An overview of the data records in these sub-directories is provided below.

annotations/sub_XX.csv

This directory contains 30 annotation files titled `sub_XX.csv`, where `XX` are natural numbers in the set $\{1, 2, \ldots, 30\}$ denoting the IDs of the participants.

The column-name and the content for the four columns in each csv file are as follows:

- *Column 1: jstime.* Time in milliseconds from the beginning of the video-viewing session to the end.

- *Column 2: valence.* The scaled X-axis value of the joystick position in the interface (see Table 2).

- *Column 3: arousal.* The scaled Y-axis value of the joystick position in the interface (see Table 2).

- *Column 4: video.* Contains the sequence of video-IDs that indicates the ordering and duration of the different video-stimuli for the given participant.

physiological/sub_XX.csv

This directory contains 30 physiological data files titled `sub_XX.csv`, where `XX` are natural numbers in the set $\{1, 2, \ldots, 30\}$ denoting the IDs of the participants. The column-name and the content for the 10 columns in each csv file are as follows:

- *Column 1: daqtime.* Time in milliseconds from the beginning of the video-viewing session to the end.

- *Columns 2–9: ecg, bvp, gsr, rsp, skt, emg_ zygo, emg_ coru and emg_ trap.* The transformed sensor output values for each of 8 physiological sensors used in the experiment. More information on the sensors, the transformations applied, and the outputs units for these values, is available in Table 2 and the README files in these directories.

- *Column 10: video.* Contains the sequence of video-IDs that indicates the ordering and duration of the different video-stimuli for the given participant.

Note. The *jstime* and the *daqtime* columns in the above mentioned files contain timestamps provided by a common clock on the logging computer. They have been named differently due to the different sampling intervals used for logging these files, i.e., 50 ms and 1 ms, respectively.

Technical Validation

Annotation Data

The quality and the reliability of the annotation data has been thoroughly addressed in our previous works [27, 28, 16, 29]. A summary of the relevant highlights from these works is presented below.

In [27, 28] several different exploratory data analyses were presented. These analyses provided an initial intuition into the annotation patterns for the different video-stimuli. For example, the annotations for the two scary videos had in general low valence and high arousal values. They were thus different from annotations for the amusing videos which had relatively high valence and medium arousal. These differences can also be seen in the annotations presented in Figure 2 (left). The initial exploratory results presented in [27, 28] were then formally validated in [16], where Multivariate ANOVA (MANOVA) was used to quantify the statistical significance of the differences in the annotations for these videos. The 'usability' of our annotation approach was validated using the *System Usability Scale* (SUS) questionnaire. According to the ratings received on the same, the annotation setup had 'excellent' usability as the participants found it to be consistent, intuitive and simple to use [16]. In [16, 29], several different methods for analysing the annotation patterns in continuous 2D annotations were presented. The results of these continuous methods were concurrent to the results of the MANOVA. Also in [16, 29] several different methods for extracting additional information from these continuous annotations have been presented. For example, the *Change Point Analysis* method in [16] automatically detects the major change-points in the annotation data that can be used to segment the annotations into several salient segments. For comparison with the physiological data, some results for the annotation data are presented in the next subsection.

Sensor	Extracted Features
ECG	Heart Rate (HR)
	Inter-Beat Interval (IBI)
	Standard Deviation (SD) of NN-intervals (SDNN)
BVP	Heart Rate (HR)
	Inter-Beat Interval (IBI)
	Standard Deviation (SD) of NN-intervals (SDNN)
GSR	Skin Conductance Level (SCL)
	Skin Conductance Response (SCR)
Respiration	Respiration Rate (RR)
	Interval of Respiration peaks
Skin Temperature	Temperature
	SD of Temperature (SDT)
EMG–zygomaticus	Amplitude of the signal
EMG–corrugator	Amplitude of the signal
EMG–trapezius	Amplitude of the signal

Table 3: The sensors and the various features extracted from the sensor signals.

Sensor	Feature Selected
ECG	mean HR
BVP	Standard Deviation (SD) of NN-intervals (SDNN)
GSR	mean SCR
Respiration	mean RR
Skin Temperature	SD of Temperature (SDT)
EMG–zygomaticus	mean amplitude (mean Zygo)
EMG–corrugator	mean amplitude (mean Corr)
EMG–trapezius	mean amplitude (mean Trap)

Table 4: The sensors and the features selected from each sensor.

Physiological Data

In the Background & Summary section, the typical AC processing pipeline was presented. The final objective of this pipeline is to develop machine learning models that can infer the emotional state of humans from (in the given case) physiological signals. To achieve the same, it is critical that the physiological responses to the different video-stimuli are discernible from each other and are ideally correlated to annotation data. If indeed these patterns exist, they would validate the quality and the value of this data. To determine the same, we extracted several features from the physiological data and performed *Principal Components Analysis* (PCA) on these features. The details and results of this analysis is presented as follows.

Feature Extraction

The feature extraction was performed iteratively over the physiological data files for each participant. First, the data for a given participant was segmented into chunks for the different video-stimuli. Then, from the sensor data pertaining to each of these video-chunks, several features were extracted (see Table 3).

For the technical validation presented here, one predominantly used feature for each sensor was selected and where applicable, the mean of this feature across the given video-chunk was calculated. Similarly, the mean valence and arousal values across each video-chunk were calculated. The selected physiological features are presented in Table 4. As a result, for the 30 participants who each watched eight emotional video-stimuli, we have 240 (30×8) values for each of these selected features. Due to inter-personal differences, the participants have a different baseline value for each of these extracted features. These differences can be detrimental to the comparison of these features across all participants and were therefore removed using *Z-score* standardisation across each participant. The same was also done for the annotation data. The violin-plots in Figure 4 show the distributions of the selected features and annotation data, across the 4 different video-labels (see Table 1). From the figure, it is apparent that some of the physiological features (consider the top eight panels) charac-
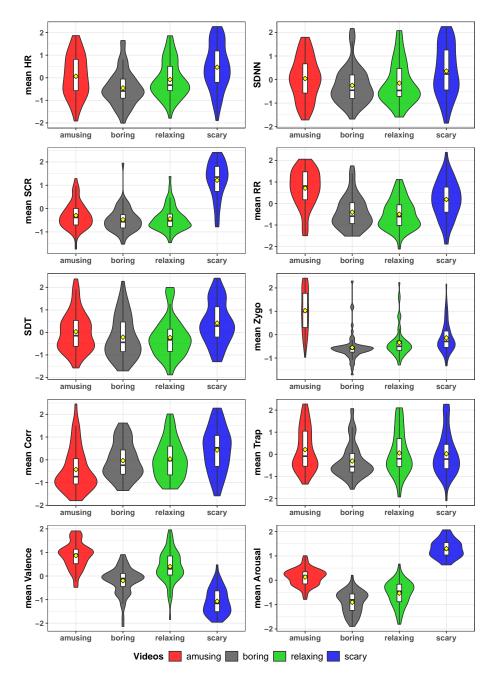
Figure 4: "Violin" plots of the distribution of the selected features and the mean annotation (valence & arousal) values across different types of videos. The box plots embedded in each violin plot show the Interquartile Range (IQR) for each considered variable, while a yellow diamond marks the mean of the distribution.

terise specific types of videos. For instance, scary videos result in high values of SCR and elevated HR, while amusing videos elicit accelerated respiration rates and activity of the *zygomaticus* muscles. Boring and relaxing videos, as expected, elicit similar values of all features. These results are in line with previous research [5, 6], where, e.g., HR and SCR were determined to be positively correlated to arousal. This effect can also been seen in our data, where the reported arousal levels (see bottom-left panel) for scary videos are higher than for the other videos. Similarly, *zygomaticus* activity which has been reported to be positively correlated to valence (see bottom-left panel), also exhibits similar patterns in our data.
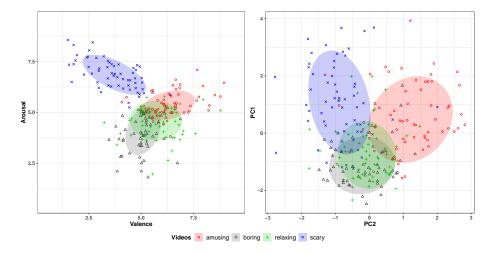

Figure 5: Scatter plots of the mean annotation data and the first two principal components of the physiological data, labelled according to the types of videos. Ellipses denote one standard deviation.

PCA

PCA is a commonly used dimensionality reduction technique [38, 39]. This allows for the visualization of the given data in a lower dimensional space, where e.g., spatial distributions of the data can be analysed. The result of PCA for the selected *Z-scored* features is shown in Figure 5 (right), where the scores on the first two principal components are shown. The scatter plot on the left in Figure 5 shows the mean valence and arousal values across the different video-labels. The data ellipses show the standard deviation of data pertaining to these video-labels. As is evident from this figure, the physiological and the annotation data form concurrent clusters. These two figures validate the data in an even more prominent way than Figure 4. Valence and arousal values (left panel) of scary videos are concentrated in the upper-left quadrant, those for the amusing videos are in the upper-right, and the others lie in the middle with low arousal values,

as one would expect. This is confirmed by the right panel, in which the four types of videos are represented analogously on the plane obtained using the first two principal components of the physiological features. This seems to indicate that the physiological features somehow "match" the joystick annotations. Of course, this serves only an initial investigation and a more rigorous analysis is required to fully exploit the potential of the database. Nevertheless, the results provided here show that the presented dataset has several viable characteristics that would make it of interest to our research community.

Usage Notes

Videos

Due to copyright issues, we cannot directly share the videos as a part of this dataset (Data Citation 1). Nonetheless, to help users in ascertaining the emotional content of these videos in more detail and if required, to replicate the experiment, we have provided links to websites where these videos are currently hosted (see /metadata/videos.xlsx). We are aware that these links might become unusable in the future and that this can cause inconvenience to the users. In such an eventuality, we encourage the users to contact us, so that we can assist them in acquiring/editing the videos.

Feature Extraction and Downstream Analysis

The code used for the technical validation of the dataset was developed in MAT-LAB 2014b and R-language (version 3.3.3). The feature extraction was done in MATLAB using open-source toolboxes/code like TEAP [40] and an implementation of the Pan Tompkins QRS detector [41]. The PCA analysis was performed in R using the *prcomp* function from the *stats* package. This code is available to interested researchers upon request. Users of the dataset (Data Citation 1) interested in leveraging the continuous nature of the provided annotations are advised to check our previous works [16, 29]. The analysis presented in these works was primarily undertaken in R-language and can be easily reproduced.

Data Citations

1. Sharma, K., Castellini, C., van den Broek, E., Albu-Schaeffer, A. & Schwenker, F. **CASE Dataset** (2018). Available at: `https://rmc.dlr.de/download/CASE_dataset/CASE_dataset.zip`

Acknowledgements

data acquisition software and assisting in data collection, and Dr. Freek Stulp for his help and support in developing this dataset.

Author Contributions

KS co-developed the data acquisition software, collected the data, developed the dataset, contributed to the technical validation and composed the manuscript.

CC co-designed the experiment, helped with participant enrolment, supervised the data collection, contributed to the technical validation and the composition of the manuscript.

EL assisted in the design and the development of the annotation interface, supervised the data processing and edited the manuscript.

AA verified the dataset and edited the manuscript.

FS co-designed the experiment, supervised the technical validation and edited the manuscript.

Competing financial interests

The author(s) declare no competing financial interests.

References

[1] van den Oord, A. *et al.* Wavenet: A generative model for raw audio. *CoRR* **abs/1609.03499** (2016). URL http://arxiv.org/abs/1609.03499. 1609.03499.

[2] Hagengruber, A., Leidner, D. & Vogel, J. Edan: Emg-controlled daily assistant. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '17, 409–409 (ACM, New York, NY, USA, 2017).

[3] Picard, R. W. *Affective Computing* (MIT Press, Cambridge, MA, USA, 1997).

[4] McStay, A. *Emotional AI: The Rise of Empathic Media* (SAGE, 2018).

[5] van den Broek, E. L. *et al.* Affective man-machine interface: Unveiling human emotions through biosignals. In Fred, A., Filipe, J. & Gamboa, H. (eds.) *Biomedical Engineering Systems and Technologies*, 21–47 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2010).

[6] van den Broek, E. L. *Affective Signal Processing (ASP): Unraveling the mystery of emotions*. Ph.D. thesis, University of Twente, Enschede (2011).

[7] Hanke, M. *et al.* A studyforrest extension, simultaneous fmri and eye gaze recordings during prolonged natural stimulation. *Scientific Data* **3**, 160092 EP – (2016). URL `http://dx.doi.org/10.1038/sdata.2016.92`. Data Descriptor.

[8] Gatti, E., Calzolari, E., Maggioni, E. & Obrist, M. Emotional ratings and skin conductance response to visual, auditory and haptic stimuli. *Scientific Data* **5**, 180120 EP – (2018). Data Descriptor.

[9] Schwenker, F. & Scherer, S. (eds.) *Multimodal Pattern Recognition of Social Signals in Human-Computer-Interaction - 4th IAPR TC 9 Workshop, MPRSS 2016, Cancun, Mexico, December 4, 2016, Revised Selected Papers*, vol. 10183 of *Lecture Notes in Computer Science* (Springer, 2017).

[10] Taamneh, S. *et al.* A multimodal dataset for various forms of distracted driving. *Scientific Data* **4**, 170110 EP – (2017). URL `http://dx.doi.org/10.1038/sdata.2017.110`. Data Descriptor.

[11] Soleymani, M., Koelstra, S., Patras, I. & Pun, T. Continuous emotion detection in response to music videos. In *Face and Gesture 2011*, 803–808 (2011).

[12] Russell, J. A. Core affect and the psychological construction of emotion. *Psychological review* **110**, 145 (2003).

[13] Soleymani, M., Asghari Esfeden, S., Fu, Y. & Pantic, M. Analysis of EEG signals and facial expressions for continuous emotion detection. *IEEE Transactions on Affective Computing* (2015).

[14] Cowie, R. *et al.* FEELTRACE: An instrument for recording perceived emotion in real time. In *ISCA tutorial and research workshop (ITRW) on speech and emotion*, 19–24 (2000).

[15] Nagel, F., Kopiez, R., Grewe, O. & Altenmueller, E. EMuJoy: Software for continuous measurement of perceived emotions in music. *Behavior Research Methods* **39**, 283–290 (2007).

[16] Sharma, K., Castellini, C., Stulp, F. & den Broek, E. L. V. Continuous, real-time emotion annotation: A novel joystick-based analysis framework. *IEEE Transactions on Affective Computing* (2017). To appear.

[17] Kächele, M., Schels, M. & Schwenker, F. The influence of annotation, corpus design, and evaluation on the outcome of automatic classification of human emotions. *Frontiers in ICT* **3**, 27 (2016). URL `https://www.frontiersin.org/article/10.3389/fict.2016.00027`.

[18] Koelstra, S. *et al.* Deap: A database for emotion analysis ;using physiological signals. *IEEE Transactions on Affective Computing* **3**, 18–31 (2012).

[19] McKeown, G., Valstar, M., Cowie, R., Pantic, M. & Schroder, M. The semaine database: Annotated multimodal records of emotionally colored conversations between a person and a limited agent. *IEEE Transactions on Affective Computing* **3**, 5–17 (2012).

[20] Ringeval, F., Sonderegger, A., Sauer, J. & Lalanne, D. Introducing the recola multimodal corpus of remote collaborative and affective interactions. In *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, 1–8 (2013).

[21] Abadi, M. K. *et al.* DECAF: MEG-based multimodal database for decoding affective physiological responses. *IEEE Transactions on Affective Computing* **6**, 209–222 (2015).

[22] Ringeval, F. *et al.* AVEC2017: Real-life depression, and affect recognition workshop and challenge. In *Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge*, AVEC '17, 3–9 (ACM, New York, NY, USA, 2017). URL `http://doi.acm.org/10.1145/3133944.3133953`.

[23] Metallinou, A. & Narayanan, S. Annotation and processing of continuous emotional attributes: Challenges and opportunities. In *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*, 1–8 (2013).

[24] Baveye, Y., Dellandréa, E., Chamaret, C. & Chen, L. Deep learning vs. kernel methods: Performance for emotion prediction in videos. In *Affective Computing and Intelligent Interaction (ACII), 2015 International Conference on*, 77–83 (IEEE, 2015).

[25] Girard, J. M. & C. Wright, A. G. Darma: Software for dual axis rating and media annotation. *Behavior Research Methods* (2017).

[26] Yannakakis, G. N. & Martinez, H. P. Grounding truth via ordinal annotation. In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, 574–580 (2015).

[27] Antony, J., Sharma, K., van den Broek, E. L., Castellini, C. & Borst, C. Continuous affect state annotation using a joystick-based user interface. In *Proceedings of Measuring Behavior 2014: 9th International Conference on Methods and Techniques in Behavioral Research*, 268–271 (2014).

[28] Sharma, K., Castellini, C. & van den Broek, E. L. Continuous affect state annotation using a joystick-based user interface: Exploratory data analysis. In *Measuring Behavior 2016: 10th International Conference on Methods and Techniques in Behavioral Research*, 500–505 (2016).

[29] Sharma, K. *et al.* A functional data analysis approach for continuous 2-D emotion annotations. *Web Intelligence* (2018). To appear.

[30] Karashima, M. & Nishiguchi, H. Continuous affect rating in cartesian space of pleasure and arousal scale by joystick without visual feedback. In Stephanidis, C. (ed.) *HCI International 2017 – Posters' Extended Abstracts*, 316–323 (Springer International Publishing, Cham, 2017).

[31] Sauro, J. & Lewis, J. R. When designing usability questionnaires, does it hurt to be positive? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, 2215–2224 (ACM, New York, NY, USA, 2011).

[32] Bradley, M. M. & Lang, P. J. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry* **25**, 49–59 (1994).

[33] Gross, J. J. & Levenson, R. W. Emotion elicitation using films. *Cognition & Emotion* **9**, 87–108 (1995).

[34] Hewig, J. *et al.* A revised film set for the induction of basic emotions. *Cognition and Emotion* **19**, 1095–1109 (2005).

[35] Bartolini, E. E. *Eliciting Emotion with Film: Development of a Stimulus Set*. Master's thesis, Wesleyan University (2011).

[36] Thought Technology, 8205 Montreal/Toronto Blvd. Suite 223, Montreal West, QC H4X 1N1 Canada. *Physiology Suite*, 5.1 edn. URL `http://www.thoughttechnology.com/pdf/manuals/SA7971%20V5.1%20Physiology%20Suite..pdf`.

[37] FFmpeg: A complete, cross-platform solution to record, convert and stream audio and video. URL `https://ffmpeg.org/`.

[38] Jolliffe, I. *Principal component analysis* (Wiley Online Library, 2005).

[39] Ringner, M. What is principal component analysis? *Nature biotechnology* **26**, 303–304 (2008).

[40] Soleymani, M., Villaro-Dixon, F., Pun, T. & Chanel, G. Toolbox for emotional feature extraction from physiological signals (TEAP). *Frontiers in ICT* **4**, 1 (2017).

[41] Sedghamiz, H. *Matlab Implementation of Pan Tompkins ECG QRS detector*. (2014).