# PosEmo – An automated system for measuring user interest and attitude in real time

Mikolaj Buchwald
Poznan Supercomputing and Networking Center
Network Services Department
Polish Academy of Sciences
Poznan, Poland
mikolaj.buchwald@man.poznan.pl

Szymon Kupinski
Poznan Supercomputing and Networking Center
Network Services Department
Polish Academy of Sciences
Poznan, Poland
szymonk@man.poznan.pl

Jan Nowak
Poznan Supercomputing and Networking Center
Network Services Department
Polish Academy of Sciences
Poznan, Poland
jnowak@man.poznan.pl

Magdalena Biadala
Poznan Supercomputing and Networking Center
Network Services Department
Polish Academy of Sciences
Poznan, Poland
mbiadala@man.poznan.pl

Maciej Behnke
Adam Mickiewicz University in Poznan
Department of Psychology and Cognitive Science
Poznan, Poland
maciej.behnke@amu.edu.pl

## Abstract

*Given the increasing prevalence of digital services across various aspects of life, it has become crucial to understand and recognize the mental states of individuals interacting with artificial systems. To address this concern, we aimed to develop the PosEmo – an automated application that can assess individuals' affective states using a video web camera. While studying affective states, we focused on two kinds of emotional behavior: approach/avoidance behavior and behavioral freezing/activation. To measure these behaviors, we use computer vision techniques to track the movement of the participant's head in video recordings, as well as in real-time video streaming. This method offered the seated research participant convenience, replicability, and non-intrusiveness. Drawing from established theoretical frameworks and supported by initial empirical findings, we developed the software and validated it in the online experiment. We found that PosEmo recognized whether people watched negative, neutral, or positive videos. Thus, our innovative approach enables us to accurately estimate people's affective states. In sum, by adopting a human-centered approach, we combined artificial intelligence methodologies to create an innovative system supporting human-computer interaction. Our system's potential research applications span various domains, such as psychology, cognitive science, usability studies, psychotherapy sessions, content quality assessment, and education.*

**Keywords:** image processing, artificial intelligence, affective computing, psychology, emotions

## 1. Introduction

With the recent technological advancements in hardware and software – including sophisticated artificial intelligence (AI) methods – coupled with the availability of open data collected from hundreds and thousands of participants (see, e.g., Behnke, Buchwald, et al., 2022), there is an unprecedented opportunity to gain a deeper understanding of human affective states. There is a vivid interest in systems enabling affective states' recognition in human-centered research fields, such as usability studies (Froschauer et al., 2021) and digital content quality assessment (Huang et al., 2022). However, existing methodologies like AB testing (in usability studies) or web analytics fail to provide comprehensive insights into the cognition and emotions of individuals interacting with digital stimuli, e.g., application interfaces, various dashboards, or video games. There is a lack of a system providing an easy and intuitive means of quantitatively estimating human affective states (e.g., being overwhelmed or interested) available on a large scale in real-time. This especially applies to scenarios where people (users) interact

URI: https://hdl.handle.net/10125/107314
978-0-9981331-7-1

with digital content, such as stimuli in psychological experiments, interfaces in electric vehicles, marketing materials, or remote school classes. To address these limitations, we introduce the PosEmo – a computer vision-based solution that utilizes behavioral methods to measure human affective states. The PosEmo aims to provide software tools that support researchers in various human-related study domains.

## 2. Theoretical foundations of PosEmo

One way individuals display affective states is through behavior. However, the research on affective behavior has been dominated by a focus on facial expressions (Barrett et al., 2019). There is now a proliferation of algorithms and applications for emotion recognition based on facial expressions (facial expression analysis, FEA/FX/FXA), e.g.: AFFDEX (Bishay et al., 2022) – implemented in iMotions software (see https://imotions.com/biosensor/fea-facial-expression-analysis), MorphCast (https://www.morphcast.com), or FaceReader by Noldus (https://www.noldus.com/facereader). Although the idea that there are only a few types of discrete basic emotions is very tempting, summaries of psychological research (Barrett et al., 2019), as well as our experiences with applying such algorithms to real-life, industry scenarios, show three major limitations of detecting discrete emotions from facial expressions.

First, if there are, for instance, seven basic emotions (such as happiness, sadness, fear, etc.; Ekman, 1992) that can be recognized with the facial expressions analysis software, it means that the researcher is provided with seven different measures, which is quite a lot for offline statistical post-processing, and, evidently, too much to handle in real-time emotional feedback applications Johanssen et al., 2019. Second, during a standard psychological experiment or usability research session, when participants are presented with some task, such as watching stimuli or performing some manual task, people tend to show little to no facial expressions different from the ”neutral” facial expression. It means that the researcher can get at most 2-3 meaningful expressions from the, e.g., 45-minute session, thus not differentiating participants' mental processes during the whole session. The third and final issue is that for use-cases other than studying basic facial expressions themselves, researchers are not really interested in the facial expressions *per se*, but they would rather know what is the more general mental state that related to these expressions (e.g., a smile is usually associated with positive valence, see Coles et al., 2022). In other words, one may want to know the general affective state of a participant, i.e., whether the stimuli/the content being watched induces a positive affect (e.g., positive valence, engagement, and interest) or negative affect (e.g., negative valence, boredom, and distraction).

Fortunately, affective behavior is not limited to facial expressions and includes body sway or whole-body movements, including leaning forward, reclining, or freezing (Behnke et al., 2021). Body sway is a component of affective behavior that is organized by its direction and intensity (Bradley and Lang, 2007). It is proposed that affective states are organized around two motivational systems: approach-oriented and avoidance-oriented (Bradley and Lang, 2007). For instance, pictures of mutilated bodies (Hagenaars et al., 2012), dangerous animals (Hillman et al., 2004), and guns (Eerland et al., 2012) cause individuals to display defensive behavior. In contrast, pictures of smiling individuals (Gea et al., 2014), attractive individuals, and delicious food cause people to display approach behavior. For affective behavior intensity, individuals decrease their body mobility facing negatively evaluated stimuli—response to avoid threats (Azevedo et al., 2005) – a behavioral freezing (Lang and Bradley, 2013). Negative stimuli, including social threat (Roelofs et al., 2010), mutilation images (Azevedo et al., 2005), anticipating electrical shock (Hashemi et al., 2019), and recalling sad memories (Fawver et al., 2014), cause individuals to reduce body movement. In contrast, pleasant stimuli cause individuals to move more (Ciria et al., 2017; Hagenaars et al., 2014; Naugle et al., 2011; Stins et al., 2015).

However, existing methods for measuring affective behavior often rely on additional hardware, such as force platforms (Hagenaars et al., 2012), wired psychophysiological apparatus (Buchwald et al., 2019), eye-trackers (Bykowski and Kupiński, 2018), or high-resolution/multispectral cameras (Veshneva and Singatulin, 2021). Furthermore, software available for preparing and conducting affective research (e.g., iMotions https://imotions.com; Kulke et al., 2020) is primarily desktop-based and assumes the participant's physical presence in a laboratory setting. These solutions typically require the involvement of trained technicians, leading to increased costs and expertise necessary to carry out such studies. Finally, a sophisticated statistical post-processing must occur to obtain meaningful and usable results for a researcher or client.

PosEmo aims to tackle these challenges and offer a comprehensive solution that advances the measurement of user experience and the study of human affect. The primary research question of the work described in the current article is whether it is possible to develop a reliable algorithm for real-time detection of affective states in online scenarios – i.e., outside of the laboratory settings, with conditions closely reminiscent of the natural way people interact with technology.

## 3. Contribution

Since the previously published research in the domain of video-based, behavior-related affective states detection, the following contributions were made: (1) the algorithm for real-time affective states detection was created; hence, the need to perform the whole experimental procedure first, including calibration, was elevated; (2) proof-of-concepts for two application scenarios were developed and tested: the utilization of emotional states classification in education via the videoconferencing tool, and an online web application utility for presenting the material and gathering data – for basic research, usability studies and/or marketing materials assessment. To delve into details, as compared to, e.g., Behnke et al., 2021 – the results from the previous study were strengthened, utilizing a novel dataset that was collected outside of the laboratory, in a more ecologically-valid (natural) settings (i.e., the participants were viewing the materials on their own computers, in their homes). Moreover, apart from the offline post-factum analysis being performed, a machine learning model was created, which allows the decoding of the emotional states in real-time, on shorter periods of time (up to 60 seconds). In the study, the open-source modules, such as Google's MediaPipe (face/head detection) and Django framework (web application), were utilized; however, the ultimate goal was to leverage the functionalities of these base building blocks in order to come up with a novel utility for detection of affective states of the participants in real-time, based on their behavior in the front of the video camera. To that extent, we believe such a goal was achieved, as explained in the following sections.

## 4. PosEmo description

PosEmo is software that infers the affective states of a participant (a user) by evaluating one's affective behavior – focusing on head sway. To retrieve the position of the head of the participant, we utilize Google's mediapipe algorithms (https://google.github.io/mediapipe; see also Zhang et al., 2020) and their implementations in Python and JavaScript frameworks, depending on the use-case that is being considered.

The PosEmo approach to affective behavior can be applied outside of the laboratory using an off-shelf web camera without additional specifications for background or light. Limited resource consumption allows, e.g., implementing this as an Internet of Things (IoT) – thus increasing the safety of car drivers and industrial machine operators or increasing enjoyment in end-user experience, such as gamers' excitement while testing new games.

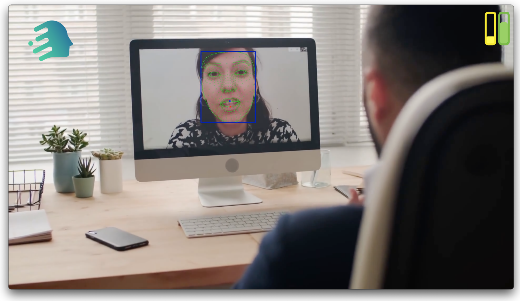

**Figure 1. PosEmo functionalities overview, an augmented reality vision.**

The presented Algorithm 1 is a simplified version of the implementation. It is only to help you understand how our algorithm works. The input data is an array of vectors $V\{x_{00}, x_{01}, x_{0M}, ..., x_{N0}, x_{N1}, x_{NM}\}$ of size $M \times 3 \times N$ where x is vector $\{xCenter, yCenter, bboxarea\}$ per frame, N is common length of all vectors, and M is the number of all videos and the ground truth - *y_vector* is generated. Each vector V contains a sequence of x vectors, where each x contains the parameters: xCenter (x coordinate of the center of the bounding box), yCenter (y coordinate of the center of the bounding box), and bboxarea (area of the bounding box). The sequence of vectors x in vector V corresponds to the order of recorded video frames in time. The *y_vector* contains only information about the stimuli type of the video (0 - Negative, 1 - Neutral, 2 - Positive). Since the X data is sorted by the type of video stimulus, that is [0, 0, 0, 1, 1, 1, 2, 2, 2], the *y_vector* is generated based on this order. After loading and initializing the data, the vector V is reduced to a dimension with two parameters for each video, so the *reduced_vector* is $2 \times M$ size. Then, the K-Means algorithm is initialized, which is initially supposed to find 15 video groups. This allows for a better fit of the model. Then, the number of groups is reduced to three based on determining each group's dominant video stimuli type. In this way, each of the 15 groups is assigned one of the three video groups. On this basis, when making predictions for each video, a group from 1 to 15 is determined, which corresponds to a group of one of the three video types (Negative, Neutral, or Positive). The returned vector *kgroup_pred* contains information about the type of video stimulus in the movie.

### 4.1. PosEmo measures and application

We use the stream from the participant/user camera (the camera directed at the respondent's upper body, including the head) to calculate, based on the body sway, the Interest and Attitude metrics for each respondent (see Fig. 1 and 2 for visualizations). The interest measure indicates the lean of the head towards the camera (approach and avoidance behavior), with higher numbers indicating being closer (displaying greater interest or approach behavior). The attitude measure indicates the intensity of head sway (behavioral freezing/activation), with higher numbers indicating more movement (displaying greater activity or activation behavior). We used Interest and Attitude names, rather than the approach/avoidance and activation/freezing, to simplify PosEmo usage and understanding for the end users who do not have expertise in psychology and might be confused by the original terms. Interest and attitude metrics are measured during the trial for each participant, and they can be aggregated (averaged) across participants, providing a value of each metric for each interaction (a stimulus or a trial). The interest and attitude metrics are calculated with in-house algorithms and scripts, implemented mostly in the Python programming language. We provide API access to our servers on demand, where the detected head or face bounding boxes can be analyzed to retrieve the measures of the participant's affective states.

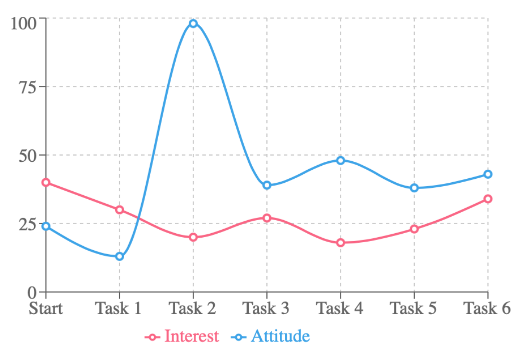


**Figure 2. PosEmo measures: interest (approach vs. avoidance), and attitude (activity vs. freezing).**

As a proof-of-concept in educational scenarios, the PosEmo measures have been integrated with an open-source videoconferencing tool – eduMeet (https://github.com/edumeet, https://edumeet.org, see also Montanari et al., 2020), and the resulting software system (see Fig. 3) has been presented at the GÉANT

---

**Algorithm 1:** Dimension reduction with t-SNE and data clustering with K-Means

**Input** : Input array of vectors contain xCenter, yCenter, bbox area for each frame $V\,\{x_{00}, x_{01}, x_{0M}, ..., x_{N0}, x_{N1}, x_{NM}\}$ of size $M \times 3 \times N$ where x is vector $\{xCenter, yCenter, bboxarea\}$ per frame, N is common length of all vectors and M is the number of all videos

**Initialize:** $y\_vector \leftarrow generate\_y\_as\_stimuli\_type(V)$

$reduced\_vector \leftarrow TSNE(V)$ t-SNE vector reduction from $M \times 3 \times N$ dimension to $2 \times M$ dimension

$kmeans \leftarrow KMeans(n\_clusters = 15)$

$kmeans.fit(reduced\_vector)$

$kgroups \leftarrow get\_kgroups(kmeans, reduced\_vector)$ From 15 cluster get 3 cluster where is most common video type

**for** $i \leftarrow 0$ *to* $M$ **do**

  $kgroup\_pred \leftarrow fit\_to\_group([reduced\_vector[i]])$ Fit video to one of the 3 video type groups

**end**

**Output** : kgroup_pred represents each video fitted to one of the 3 groups

---

European National Research and Education Networks (NRENs) annual meeting TNC22 (https://www.posemo.io/tnc22). The presented scenario considered utilizing PosEmo as a way to improve the feedback information during remote class sessions, in which non-verbal communication, eye contact, and facial expressions are limited. We also released a plugin for web applications that can be used to visualize PosEmo measures in real time – it is available as an open-source solution at our GitHub repository: https://github.com/PosEmo/posemo-meter.

## 5. PosEmo validation

### 5.1. Study design

We ran an online experiment to evaluate the algorithms used in the PosEmo. In our study, participants watched nine videos, namely three positive, three negative, and three neutral video clips. Building upon the theoretical models (Bradley and Lang, 2007) and supporting empirical evidence (Behnke et al., 2021), we expected that people should behave differently when watching videos of different valence, and the PosEmo should recognize this difference.

The research was carried out using a representative Polish sample (n = 103; 50 women; age range = from 19 to 60, $M$=43, $SD$=11.05). The participants were recruited using Ariadna, i.e., the Polish research panel, Poland's biggest independent nationwide research panel

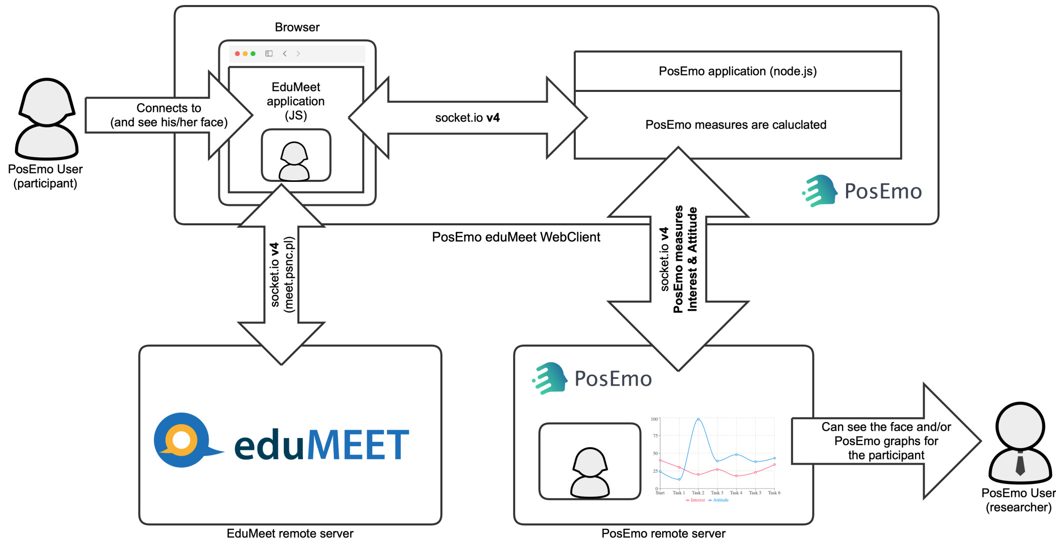


**Figure 3. PosEmo integration with the eduMeet – open-source conferencing application.**

(over 110,000 active panel members). Appreciating their involvement, participants received points that could be redeemed for rewards from a diverse selection of several hundred products offered by the panel organizers. Before their participation, all participants provided informed consent, and the study was conducted anonymously. The Bioethical Committee of Poznan University of Medical Sciences (802/22) approved the study protocols.

Each participant ran the study on her/his computer, which lasted approximately 30 minutes. In this way, we received data of diverse origins, which made it possible to exclude the influence of hardware on the common features of the collected data. The experiment started with completing consent questionnaires. Next, participants played an audio/video clip using the test system, which will serve as a reference measure for the volume level of sound on the participants' computer. The main part of the study consisted of a series of 9 experimental trials. At the beginning of each trial, a 20-second washout video clip – which is a combination of colored rectangles – was displayed to serve as a baseline for the actual stimulus. The participants then watched a 60-second video of varying valence – positive, negative, and neutral clips. The video clips were presented in counter-balanced order. After each video, participants reported their valence, arousal, interest, and motivation on the 9-item scale (Cowen and Keltner, 2017).

We used stimuli from an emotion-eliciting video clip database with prior evidence of reliability and validity (Behnke et al., 2020; Hussain et al., 2017; Kaczmarek et al., 2021; Schaefer et al., 2010). To elicit positive affective states, we used validated video clips: (a) Summer Olympic Games (Athletes performing successfully and showing their joyful reactions); (b) Delicious Food (presentation of delicious food and deserts); (c) Carlton Draught "Beer Chase" (A police pursuit of thieves in which neither side wants to dodge a drop of beer held in their hands). To elicit neutral affective states, we used validated video clips: (a) "Blue" (A man organizes the drawers in his desk, or a woman walks down an alley); (b) "The Lover" (The character walks around town); (City life scenes); (c) "Twin Peaks: Fire Walk with Me" (the character sweeps the floor in the bar). To elicit negative affective states, we used validated video clips: (a) "Schindler's list" (In a concentration camp, countless lifeless bodies are being handled and stacked by fellow prisoners. Amidst this bleak scene, Schindler's attention is abruptly drawn to a deceased young girl, clad in a vivid red jacket); (b) "The Blair Witch Project" (In a house filled with piercing screams, the characters walk around, when suddenly, one of them starts screaming); (c) "Trainspotting" (A drug addict is suffering from withdrawal symptoms and violent diarrhea in an extremely dirty public restroom).

## 5.2. Analysis and Results

We used data from the participants who recorded their videos with a minimum frame rate at an average level of 10 FPS for each stimulus video. In total, information for 915086 frames was obtained for 927 stimuli (103 participants × 9 videos), which gives an average vector length of 987.15 per video, and average FPS of 16.45. Because the number of frames per second varies depending on computer performance and network bandwidth, resulting in different vector lengths of videos, the videos were reduced to 1 FPS by combining the values of the reduced data using the median. Subsequently, the vectors were then aligned to the most frequent vector length (because the time of the stimuli varied minimally – in a scale of milliseconds). The obtained data allowed us to compute 927 vectors (i.e., samples) with 177 parameters (features). Of 126 participants who completed the study, 23 did not meet the technical criteria described above, resulting in 103 analyzed participants.

To visualize and find similarities in the collected data, the 177 resulting features were reduced to two dimensions (a 2D representation in a principal component analysis manner). To reduce the number of extracted features, we utilized the t-SNE algorithm (Cai and Ma, 2022) because of the outliers' presence (Li et al., 2017). The transformation made it possible to visualize and verify whether the data grouped according to the viewed stimulus video type (see Fig. 4). The data were additionally grouped using the K-Means algorithm (Xu and Wunsch, 2005) to create a background on the graph. The algorithm grouped the affective behavior into three groups that corresponded well to the types of the video stimuli (i.e., negative, neutral, and positive) (see Fig. 4). We found the greatest degree of clustering for negative and positive videos. Neutral videos were less accurately clustered, but it can still be seen that they were grouped together.

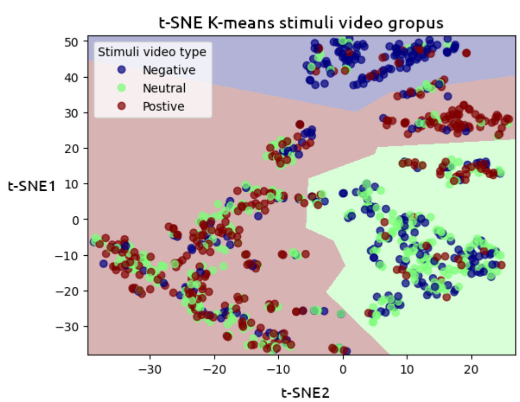


**Figure 4. Video grouped into three groups (Negative, Neutral and Positive) using t-SNE and K-Means algorithms.**

Next, we ran a multivariate analysis of variance (MANOVA) (Warne, 2014) to determine the statistical significance of the differences observed between the groups of results.

We found the difference in means between stimulus film types based on the results of Wilks' lambda variance for the t-SNE1 parameter (see Table 1), and based on the t-SNE2 parameter (see Table 2). The p-value is less than 0.001, and so the null hypothesis, suggesting no differences in behavior while watching different types of affective stimuli, is rejected (see Tables 3 and 4; see also Warne, 2014). This suggests that people displayed different affective behaviors to the different types of affective videos, and the observed difference is too large to be explained solely by chance. Thus, it is possible to recognize from the data recorded during the study whether the stimulus video viewed was negative, neutral, or positive.

| Stimuli_type | n | mean | std |
|---|---|---|---|
| Negative | 309 | 3.419 | 13.326 |
| Neutral | 309 | -2.695 | 17.292 |
| Positive | 309 | -5.020 | 18.835 |

**Table 1. Means and standard errors of groups for t-SNE1.**

| Stimuli_type | n | mean | std |
|---|---|---|---|
| Negative | 309 | 11.957 | 27.605 |
| Neutral | 309 | -5.053 | 18.860 |
| Positive | 309 | 3.163 | 22.749 |

**Table 2. Means and standard errors of groups for t-SNE2.**

We also checked the relationship between the type of stimulus video (negative, neutral, and positive – as defined in the ground truth database) and the respondents' answers in questionnaires associated with every video (i.e., with every sample). The questionnaires included items on emotions induced by the stimuli, where on a scale from 0 to 9, the respondents reported the state level of the experienced valance, arousal, motivation, and interest after watching the video (see the "Study design" section above). We observed that positive videos always had associated with them more positive valence and higher levels of motivation and interest, while negative videos had the lowest levels, with neutral videos ranked in the middle values (see Fig. 5). We also used a one-way ANOVA to compare the differences between the mean responses for each stimuli type (negative, neutral, and positive) within each emotional state (Cardinal and Aitken, 2013). The analysis revealed that for each of these models, the p-values were less than 0.05 (see Table 5), from which we can conclude that there were statistically significant

| Statistic | Value | Num DF | Den DF | F Value | Pr >F |
|---|---|---|---|---|---|
| Wilks' lambda | 0.959 | 2.000 | 924.000 | 19.700 | 0.000 |
| Pillai's trace | 0.041 | 2.000 | 924.000 | 19.700 | 0.000 |
| Hotelling-Lawley trace | 0.043 | 2.000 | 924.000 | 19.700 | 0.000 |
| Roy's greatest root | 0.043 | 2.000 | 924.000 | 19.700 | 0.000 |

**Table 3. Multivariate linear model - Intercept.**

| Statistic | Value | Num DF | Den DF | F Value | Pr >F |
|---|---|---|---|---|---|
| Wilks' lambda | 0.953 | 2.000 | 924.000 | 22.637 | 0.000 |
| Pillai's trace | 0.045 | 2.000 | 924.000 | 22.637 | 0.000 |
| Hotelling-Lawley trace | 0.049 | 2.000 | 924.000 | 22.637 | 0.000 |
| Roy's greatest root | 0.049 | 2.000 | 924.000 | 22.637 | 0.000 |

**Table 4. Multivariate linear model - Stimuli type.**

differences in the averages and that there was an effect of the type of video on the answers provided by the respondents.

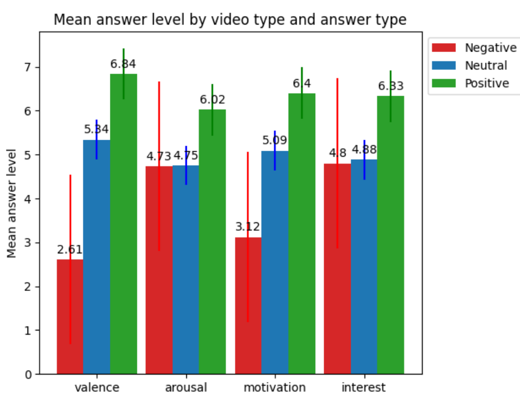


**Figure 5. Mean level of emotional state according to the type of stimulus video grouped by valance, arousal, motivation and interest (error bars represent 95% confidence intervals).**

| Emotional state | F-value | p-value |
|---|---|---|
| pleasure | 2000.641 | 0.0 |
| arousal | 3157.321 | 0.0 |
| motivation | 2117.732 | 0.0 |
| interest | 2468.725 | 0.0 |

**Table 5. One-way ANOVA results for a given emotional state correlated with video type.**

The creation of this algorithm made it possible to examine the dependencies and sense in the collected data, including its performance for particular classes of stimuli. The results presented in Table 6 show that the highest precision - 0.73 - was achieved for Negative films. Classification recall turned out to be the highest for positive films - 0.80. For neutral films,

| | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| Negative | 0.73 | 0.34 | 0.47 | 309 |
| Neutral | 0.42 | 0.33 | 0.37 | 309 |
| Positive | 0.45 | 0.80 | 0.58 | 309 |
| Accuracy | | | 0.49 | 927 |
| Macro avg | 0.53 | 0.49 | 0.47 | 927 |
| Weighted avg | 0.53 | 0.49 | 0.47 | 927 |

**Table 6. Performance metrics for the t-SNE algorithm – classification of psychophysiological states with PosEmo.**

none of the metrics has given outstanding results. This outcome may reflect that it is hard to observe an explicit behavior that would let us distinguish neutral stimuli from other categories among the participants' videos. The confusion matrix (Figure 6) presents the number of films classified to each category.

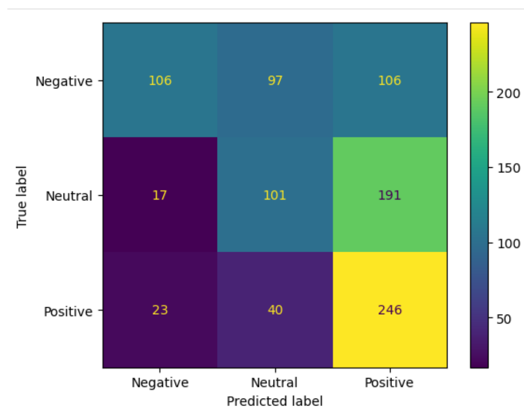


**Figure 6. Confusion matrix of t-SNE dimension reduction and K-Means clustering for 1FPS videos.**

## 6. Discussion

PosEmo is an innovative software application that effectively deduces the emotional states of participants or users by analyzing their head movement patterns. Drawing from established theoretical frameworks (Bradley and Lang, 2007) and supported by initial empirical findings (Behnke et al., 2021), we developed a software that, by retrieving participants' head positions, recognized whether people watched negative, neutral, or positive videos. This innovative approach enables us to estimate people's affective states accurately.

However, our software system primarily functions as a quantitative tool, which excels in scenarios involving the study of multiple participants over an extended duration, typically spanning a few minutes or more. This characteristic highlights one of the significant limitations of our software. Another constraint we encounter is the reluctance of individuals to share their images due to concerns about data privacy and biases associated with artificial intelligence algorithms (Behnke, Saganowski, et al., 2022). To address these concerns, PosEmo fully adheres to the General Data Protection Regulation (GDPR), a comprehensive regulation within European Union (EU) law that safeguards data protection and privacy in the EU and the European Economic Area.

A key aspect of the PosEmo service is that it processes sensitive data entirely on the user's computer in the web-browser. Only anonymized data about head position and movement measures are transmitted to the cloud server architecture. According to GDPR Art. 11 ”Processing which does not require identification” in the PosEmo cloud service the identification of the subject is no longer possible. Despite these privacy measures, it remains challenging to convey to non-technical individuals, including decision-makers and representatives from research and development funding agencies, that our primary objective is to assist individuals while prioritizing data privacy and security.

To ensure security, we leverage stable and established platforms and frameworks such as Ubuntu operating system that is used for our virtual servers, Django for web applications, and PostgreSQL for database technologies. Additionally, the institution where the studies are conveyed and where the data are stored (i.e., the authors' affiliation) is certified and audited with ISO:27001 for implementing robust information security protocols, as well as ISO:9001 for maintaining quality control and reliable procedures.

Our future plans involve continuously developing and refining the concepts underlying PosEmo. We intend to validate our measures in diverse scenarios, starting from fundamental psychological research and extending to their practical applications in commercial settings. The availability and scalability of software solutions like PosEmo have the potential to advance the realm of the Internet of Behavior (IoB).

PosEmo is a compelling example of a solution initially developed and validated within controlled laboratory settings (Behnke et al., 2021; Mitchell, 2012). The promising potential of PosEmo lies in its applicability as a smart Internet-of-Behavior device, capable of leveraging AI-embedding platforms such as NVIDIA Jetson Nano (https://developer.nvidia.com/embedded-computing), or Raspberry Pi (https://www.raspberrypi.org/). By integrating PosEmo into IoT devices, we can enhance the safety of machine operators in industries like manufacturing and agriculture. Furthermore, these IoT implementations have the potential to reduce maintenance costs within various production processes across different sectors (see, e.g., Teixeira and Behrens, 2020, Płóciennik et al., 2021).

### 6.1. Conclusion

PosEmo represents an ambitious endeavor to gain deeper insights into human affect on a larger scale than ever before, utilizing cutting-edge artificial intelligence techniques. In the current study, we presented that video-based classification of affective states is possible in real-time, utilizing material gathered from conditions closely resembling the natural way people interact with technology. Nowadays, nearly everyone possesses a digital camera at their disposal, which is primarily used for capturing photos or streaming videos. However, we believe these same devices present an unprecedented opportunity to – through study – better understand the interaction between human beings and their surrounding (digital) environment.

## 7. Acknowledgements

Part of the dissemination work and conference presence has been financed by the *PosEmo* grant TANGO-5-A/0051/2021 from the Polish National Research and Development Center (https://www.gov.pl/web/ncbr).

The authors would also like to thank prof. Łukasz D. Kaczmarek for his invaluable insights and hard work planning psychophysiological research at the Adam Mickiewicz University in Poznań, Poland, which was one of the inspirations to create the PosEmo system.

## References

Azevedo, T. M., Volchan, E., Imbiriba, L. A., Rodrigues, E. C., Oliveira, J. M., Oliveira, L. F., Lutterbach, L. G., & Vargas, C. D. (2005). A freezing-like posture to pictures of mutilation. *Psychophysiology*, *42*(3), 255–260.

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological science in the public interest*, *20*(1), 1–68.

Behnke, M., Bianchi-Berthouze, N., & Kaczmarek, L. D. (2021). Head movement differs for positive and negative emotions in video recordings of sitting individuals. *Scientific reports*, *11*(1), 1–8.

Behnke, M., Buchwald, M., Bykowski, A., Kupiński, S., & Kaczmarek, L. D. (2022). Psychophysiology of positive and negative emotions, dataset of 1157 cases and 8 biosignals. *Scientific Data*, *9*(1), 1–15.

Behnke, M., Gross, J. J., & Kaczmarek, L. D. (2020). The role of emotions in esports performance. *Emotion*.

Behnke, M., Saganowski, S., Kunc, D., & Kazienko, P. (2022). Ethical considerations and checklist for affective research with wearables. *IEEE Transactions on Affective Computing*.

Bishay, M., Preston, K., Strafuss, M., Page, G., Turcot, J., & Mavadati, M. (2022). Affdex 2.0: A real-time facial expression analysis toolkit. https://doi.org/10.48550/ARXIV.2202.12059

Bradley, M. M., & Lang, P. J. (2007). Emotion and motivation. In J. T. Cacioppo, L. G. Tassinary, & G. Berntson (Eds.), *Handbook of psychophysiology* (3rd ed., pp. 581–607). Cambridge University Press.

Buchwald, M., Kupiński, S., Bykowski, A., Marcinkowska, J., Ratajczyk, D., & Jukiewicz, M. (2019). Electrodermal activity as a measure of cognitive load: A methodological approach. *2019 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*, 175–179. https://doi.org/10.23919/SPA.2019.8936745

Bykowski, A., & Kupiński, S. (2018). Automatic mapping of gaze position coordinates of eye-tracking glasses video on a common static reference image. *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*, 1–3.

Cai, T. T., & Ma, R. (2022). Theoretical foundations of t-sne for visualizing high-dimensional clustered data.

Cardinal, R. N., & Aitken, M. R. (2013). *Anova for the behavioral sciences researcher*. Psychology Press.

Ciria, L., Muñoz, M., Gea, J., Peña, N., Miranda, J., Montoya, P., & Vila, J. (2017). Head movement measurement: An alternative method for posturography studies. *Gait & posture*, *52*, 100–106.

Coles, N. A., March, D. S., Marmolejo-Ramos, F., Larsen, J. T., Arinze, N. C., Ndukaihe, I. L., Willis, M. L., Foroni, F., Reggev, N., Mokady, A., et al. (2022). A multi-lab test of the facial feedback hypothesis by the many smiles collaboration. *Nature human behaviour*, 1–12.

Cowen, A. S., & Keltner, D. (2017). Self-report captures 27 distinct categories of emotion bridged by continuous gradients. *Proceedings of the national academy of sciences*, *114*(38), E7900–E7909.

Eerland, A., Guadalupe, T. M., Franken, I. H., & Zwaan, R. A. (2012). Posture as index for approach-avoidance behavior. *PLoS One*, *7*(2), e31291.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, *6*(3-4), 169–200. https://doi.org/10.1080/02699939208411068

Fawver, B., Hass, C. J., Park, K. D., & Janelle, C. M. (2014). Autobiographically recalled emotional states impact forward gait initiation as a function of motivational direction. *Emotion*, *14*(6), 1125.

Froschauer, R., Kurschl, W., Wolfartsberger, J., Pimminger, S., Lindorfer, R., & Blattner, J. (2021). A human-centered assembly workplace for industry: Challenges and lessons learned. *Procedia Computer Science*, *180*, 290–300.

Gea, J., Muñoz, M. A., Costa, I., Ciria, L. F., Miranda, J. G., & Montoya, P. (2014). Viewing pain and happy faces elicited similar changes in postural body sway. *PLoS One*, *9*(8), e104381.

Hagenaars, M. A., Roelofs, K., & Stins, J. F. (2014). Human freezing in response to affective films. *Anxiety, Stress & Coping*, *27*(1), 27–37.

Hagenaars, M. A., Stins, J. F., & Roelofs, K. (2012). Aversive life events enhance human freezing responses. *Journal of Experimental Psychology: General*, *141*(1), 98.

Hashemi, M. M., Gladwin, T. E., de Valk, N. M., Zhang, W., Kaldewaij, R., van Ast, V., Koch, S. B., Klumpers, F., & Roelofs, K. (2019). Neural dynamics of shooting decisions and the switch from freeze to fight. *Scientific reports*, *9*(1), 1–10.

Hillman, C. H., Rosengren, K. S., & Smith, D. P. (2004). Emotion and motivated behavior: Postural adjustments to affective picture viewing. *Biological psychology*, *66*(1), 51–62.

Huang, C.-C., Lin, M.-C., Cheng, C.-F., Tuan, P.-L., & Chen, Y.-C. (2022). Revisiting effectiveness of social marketing and networking on crowdfunding platform during covid-19. *Transylvanian Review*, *30*(1).

Hussain, Z., Zhang, M., Zhang, X., Ye, K., Thomas, C., Agha, Z., Ong, N., & Kovashka, A. (2017). Automatic understanding of image and video advertisements. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1705–1715.

Johanssen, J. O., Bernius, J. P., & Bruegge, B. (2019). Toward usability problem identification based on user emotions derived from facial expressions. *2019 IEEE/ACM 4th International Workshop on Emotion Awareness in Software Engineering (SEmotion)*, 1–7. https://doi.org/10.1109/SEmotion.2019.00008

Kaczmarek, L. D., Behnke, M., Enko, J., Kosakowski, M., Guzik, P., & Hughes, B. M. (2021). Splitting the affective atom: Divergence of valence and approach-avoidance motivation during a dynamic emotional experience. *Current Psychology*, *40*, 3272–3283.

Kulke, L., Feyerabend, D., & Schacht, A. (2020). A comparison of the affectiva imotions facial expression analysis software with emg for identifying facial expressions of emotion. *Frontiers in Psychology*, *11*, 329.

Lang, P. J., & Bradley, M. M. (2013). Appetitive and defensive motivation: Goal-directed or goal-determined? *Emotion Review*, *5*(3), 230–234.

Li, W., Cerise, J. E., Yang, Y., & Han, H. (2017). Application of t-sne to human genetic data. *Journal of Bioinformatics and Computational Biology*, *15*(04). https://doi.org/10.1142/s0219720017500172

Mitchell, G. (2012). Revisiting truth or triviality: The external validity of research in the psychological laboratory [PMID: 26168439]. *Perspectives on Psychological Science*, *7*(2), 109–117. https://doi.org/10.1177/1745691611432343

Montanari, M., Barth, I., Lariccia, S., Pantazatos, D., De Carnero, F. M., Sansone, N., & Toffoli, G. (2020). Using learning analytics in a next generation digital learning environment to transition from face-to-face to remote learning during the coronavirus crisis. *Proceedings of the 13th international conference of education, research and innovation*, 6257–6265.

Naugle, K. M., Hass, C. J., Joyner, J., Coombes, S. A., & Janelle, C. M. (2011). Emotional state affects the initiation of forward gait. *Emotion*, *11*(2), 267.

Płóciennik, M., Palma, R., Mueller, S., Kupiński, S., Błaszczak, M., Brahma, S., Radziuk, A., & Światkowski, A. (2021). Leveraging agri-food iot solutions to connect apiary owners and farmers. *2021 16th International Conference on Telecommunications (ConTEL)*, 152–157. https://doi.org/10.23919/ConTEL52528.2021.9495980

Roelofs, K., Hagenaars, M. A., & Stins, J. (2010). Facing freeze: Social threat induces bodily freeze in humans. *Psychological science*, *21*(11), 1575–1581.

Schaefer, A., Nils, F., Sanchez, X., & Philippot, P. (2010). Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers. *Cognition and emotion*, *24*(7), 1153–1172.

Stins, J. F., van Gelder, L. M., Oudenhoven, L. M., & Beek, P. J. (2015). Biomechanical organization of gait initiation depends on the timing of affective processing. *Gait & Posture*, *41*(1), 159–163.

Teixeira, I. T., & Behrens, F. H. (2020). Iot-based indicator for industrial accident risks. *Brazilian Technology Symposium*, 148–155.

Veshneva, I., & Singatulin, R. (2021). The status functions application for multispectral data images processing in virtual reality systems. In *Cyber-physical systems* (pp. 225–235). Springer.

Warne, R. T. (2014). A primer on multivariate analysis of variance (manova) for behavioral scientists. *Practical Assessment, Research Evaluation*, *19*. https://doi.org/10.7275/sm63-7h70

Xu, R., & Wunsch, D. (2005). Survey of clustering algorithms. *IEEE Transactions on Neural Networks*, *16*(3), 645–678. https://doi.org/10.1109/TNN.2005.845141

Zhang, F., Bazarevsky, V., Vakunov, A., Tkachenka, A., Sung, G., Chang, C.-L., & Grundmann, M. (2020). Mediapipe hands: On-device real-time hand tracking. *arXiv preprint arXiv:2006.10214*.