

## ARTICLE TEMPLATE

# Physiological Data for User Experience and Quality of Experience: A Systematic Review (2018-2022)

Aleph Campos da Silveira<sup>a</sup>, Mariane Lima de Souza<sup>a</sup>, Gheorghita Ghinea<sup>b</sup> and Celso Alberto Saibel Santos<sup>a</sup>

<sup>a</sup>Federal University of Espirito Santo, Brazil; <sup>b</sup>Brunel University London, UK

## ARTICLE HISTORY

Compiled March 11, 2024

## ABSTRACT

The evaluation of human responses in multimedia experiences using physiological data has a well-established presence in the academic literature. However, this field is currently undergoing transformative changes, driven by the accessibility of diverse and cost-effective devices, innovative software analysis methods, and the emergence of novel application domains such as Virtual and Augmented Reality and mulse-media. To address the imperative of contextualizing these evolving trends in a contemporary context, this paper presents a systematic review with the objective of delineating the array of physiological data utilized in assessing Quality of Experience (QoE) and User Experience (UX) in multimedia studies. It also examines the devices employed for data collection and the analytical techniques applied to interpret the acquired data. While our review exposes both constraints and promising discoveries in these domains, it also emphasizes the escalating significance and practicality of leveraging physiological data in user assessments, especially as the boundaries between the physical and digital domains continue to blur.

## KEYWORDS

Physiological Data; Biosignals; User Experience; Quality of Experience

## 1. Introduction

The concept of multimedia has undergone significant evolution. In today's context, it encompasses more than just the combination of different content delivery methods; it also involves the integration of new information technologies and the content they require. Furthermore, multimedia now encompasses the effective communication of ideas and messages through a cohesive array of vectors. In recent years, virtual reality (VR) and augmented reality (AR) have gained substantial interest from investors, the general public, and academic research communities (Cipresso, Giglioli, Raya, & Riva, 2018). While initially limited to stimulating the senses of sight and hearing, efforts have been made to expand the scope of multimedia to incorporate other sensory stimuli through the emerging concept of "mulsemedia" – multiple sensorial media (Ghinea, Andres, & Gulliver, 2011; A. C. Silveira & Santos, 2022).

In view of this ongoing evolution, it becomes necessary to develop new methods for evaluating User Experience (UX) and Quality of Experience (QoE) that can effectively

encompass the diverse modalities, platforms, and technologies offered by multimedia. Although the terms UX and QoE are often used interchangeably in the literature concerning user evaluations of multimedia, there are nuanced distinctions between the two. QoE has primarily emerged from telecommunication research, while UX has its foundations in the field of Human-Computer Interaction (Wechsung & De Moor, 2014). While QoE typically maintains a more technology-centered focus, UX places a greater emphasis on the human aspect. Furthermore, although both fields predominantly adopt quantitative research approaches, QoE tends to lean towards empirical-positivist methods, whereas UX tends to embrace interpretative and constructivist perspectives (Wechsung & De Moor, 2014).

In spite of their differences, both UX and QoE research are increasingly embracing the notion that experience data can be acquired through associative methods, which automatically capture occurrences of stimuli in the multimedia environment. These associative methods are designed to capture unconscious, efficient, involuntary, and uncontrollable bodily reactions, which, in turn, generate evaluations (Corneille & Mertens, 2020). This perspective provides an alternative to traditional evaluation methods that heavily rely on subjective information voluntarily provided by users through surveys, ratings, questionnaires, and similar means. The use of such standard methods can pose challenges, as discussed by Certický et al. (2019), including potential misinterpretations, false data, and time delays in longer experiments. Moreover, when dealing with immersive experiences like VR, questionnaires and surveys can disrupt user immersion, leading to systematic biases as users transition between the virtual and physical worlds to respond to experience-related inquiries (Putze et al., 2020; Safikhani, Holly, Kainz, & Pirker, 2021).

Objective metrics, in contrast to subjective metrics based on questionnaires and surveys, can be obtained through physiological data gathering. Physiological data encompasses any biosignal (e.g., electrodermal activity, gait patterns, blood volume pulse, heart rate, etc.) in living beings that can be measured and monitored. Physiological data is utilized in a wide range of fields such as health and welfare (Beşkardeşler & Arslan, 2019; McKee, 2008; Min, Moon, Cho, & Kim, 2020), entertainment (Dhaouadi & Ben Khelifa, 2020; Grassini, Laumann, Thorp, & Topranin, 2021; Hossain & Kulshreshth, 2019; M. Pinto, Melo, & Bessa, 2018; Voigt-Antons et al., 2021), urban planning (Pei, Guo, & Lo, 2021), architecture (Ergan, Radwan, Zou, Tseng, & Han, 2019), user evaluations (Grassini et al., 2021; Holman & Adebessin, 2019), and education (Alqahtani, Katsigiannis, & Ramzan, 2019; Ciolacu & Svasta, 2021; Giannakos, Sharma, Pappas, Kostakos, & Velloso, 2019; Gruenewald et al., 2018; Korn & Rees, 2019; Lee-Cultura, Sharma, Papavlasopoulou, & Giannakos, 2020).

It is already established that physiological data can enhance the accuracy of user evaluations (Certický et al., 2019; Dey, Phoon, Saha, Dobbins, & Billingham, 2020b; Giannakakis et al., 2019; Keighrey, Flynn, Murray, & Murray, 2021; Liapis et al., 2020; Liapis, Katsanos, & Xenos, 2018). Additionally, authors like Certický et al. (2019); Hussain et al. (2018); Huynh, Kim, Ko, Balan, and Lee (2018) have proposed models and architectures for evaluating QoE and UX using physiological measures to address the issue of imprecision with voluntary and self-report methods. For example, *EngageMon* Huynh et al. (2018) utilizes a combination of sensors from a smartphone, a wristband, and an external camera to accurately determine the engagement level in mobile games.

Other studies also present formalized conceptual proposals for the objective evaluation of QoE. An example is the work by Che, Ma, Yu, and Yue (2019), which aims to explore and measure both the physical and psychological responses and perceptions of

the user’s states (both internal and physical) when interacting with a system. Geared towards learning and education, Ciolacu and Svasta (2021) introduced a model that employs physiological data to measure and regulate learning processes during user interaction with educational content. The authors argue that the teaching-learning process should not solely be evaluated at the end of an exam using subjective methods, but also in real-time throughout the learning experience. Pursuing a similar objective, Gruenewald et al. (2018) proposed a platform for real-time collection of physiological data capable of identifying relevant emotions in a VR-based learning context (such as joy, boredom, and frustration). Lastly, Hussain et al. (2018) presented a platform based on ”lean UX” that integrates product development through continuous measurement of the so-called ”learning loop” (build-measure-learn). The primary focus of this platform is to implicitly and explicitly measure and learn from the subject’s behavior and emotional responses.

To gain a more profound understanding of the application of physiological data for evaluating UX and QoE in digital multimedia content, we conducted an extensive Systematic Literature Review (SLR). The central objective of this review was to address the following Research Question (RQ): *What and how are physiological data being used for user evaluations of digital multimedia content?*

As such, the organization of this article is as follows: Section 2 details the array of physiological signals applicable to the assessment of digital multimedia content, while Section 3 elucidates the databases employed in our review, the keywords utilized for data collection, the selection criteria, and the data filtering process. Subsequently, Section 4 outlines our approach to categorizing the accepted articles in this paper, based on the content (2D, 3D/360, VR/AR) as well as the number and kinds of media used (single media, multimedia, mulsemmedia). In Section 5, we present our analyses and considerations derived from the collected outcomes, whilst challenges identified as a result are described in Section 6. Lastly, Section 7 serves as the platform for our conclusions, in addition to outlining the strengths and limitations of this systematic review.

## **2. Physiological Signals for UX and QoE Evaluation**

### ***2.1. Biofeedback, Biosignals, Physiological Data and Signals***

As highlighted by McKee (2008), biofeedback encompasses both a process and the instruments employed within that process. The process involves utilizing monitored physiological information and subsequently employing biofeedback instruments. On the other hand, the instruments themselves are designed to monitor one or more physiological processes, measure the collected data, and translate those measurements into easily understandable information, such as visual representations or auditory cues. This enables the direct and immediate presentation of monitored and measured physiological data in a simplified manner.

Biofeedback data is closely intertwined with biosignals. According to Giannakakis et al. (2019), biosignals encompass various measurements of human body processes, which can be categorized into two primary types: physical signals and physiological signals. Physical signals encompass indicators of body tension resulting from muscle activity, including pupil size, eye movements, blinking, head and body movements, posture, breathing patterns, facial expressions, and voice. It is important to note that not all physical signals are solely regulated by the Autonomic Nervous System (ANS), and

therefore, their measurement may not be entirely objective. In contrast, physiological signals are directly associated with the ANS, such as cardiac activity, brain function, exocrine activity, and certain muscle excitability measured through electromyography. These physiological signals are closely linked to the ANS and are considered objective metrics for analysis.

In summary, measurements of physiological signals provide insights into the activity of the ANS, offering valuable data on user reflex responses, including stress levels and arousal (Aqajari et al., 2020). These varied biosignals, which have the potential for assessing UX and QoE, will now be discussed in detail.

Throughout this article, the term “Physiological Data” will be used to encompass all the collected biosignals and biofeedback employed for evaluating and enhancing QoE and UX in digital environments. Now, let’s delve into a more comprehensive exploration of these physiological data sources.

## ***2.2. Galvanic Skin Response and Electrodermal Activity***

Galvanic Skin Response (GSR)/Electrodermal Activity (EDA) is used to measure various psychological states, including arousal, attention, and stress (Aqajari et al., 2020). Because GSR/EDA signals can be collected with relatively low-cost devices boasting unobtrusive form factors, as well as the sensitivity of GSR/EDA to psychological processes, it is one of the most popular response systems in psychophysiology (Babaei, Tag, Dingler, & Velloso, 2021). Compared to other physiological signals such as eye tracking, facial expressions, thermal imaging, and EEG, which require a complicated setup, GSR/EDA measurement can be inexpensive and easy to understand. More recent studies have also revealed that GSR/EDA is able to capture the state of arousal (Babaei et al., 2021), as well as both stress and eustress<sup>1</sup> (Tateyama, Ueda, & Nakao, 2019).

According to Aqajari et al. (2020), the two main systems that respond to stress are the Hypothalamic-Pituitary-Adrenal (HPA) axis and the ANS. As we respond to stress by activating or deactivating our sweat glands, changes in sweat gland activity lead to variation in skin conductance, allowing emotional state detection. Most modern GSR/EDA electrodes have an *AgCl* (silver-chloride) contact point with the skin, usually hand and fingers. Data are acquired with sampling rates between 1 – 10 Hz and is measured in units of  $\mu$ Siemens. Since these devices are usually wearable, they are considered to be low intrusive.

## ***2.3. Heart-related Signals***

Already commonplace in e-health applications, Heart Rate (HR), Electrocardiographic (ECG), and other heart-related signals can also be used to gauge the user’s mental state, as the cardiovascular system offers a variety of measurement options that have been correlated with user arousal (Drachen, Nacke, Yannakakis, & Pedersen, 2010; Houzangbe, Christmann, Gorisse, & Richir, 2020; Oliveira, Noriega, Carvalhais, Rebelo, & Lameira, 2020). In this respect, HR is one of the most used signals to collect physiological response (Peruzzini, Grandi, Pellicciari, & Campanella, 2018). The heart is innervated by the two branches of the autonomic nervous system whose activity is associated with cognitive and emotional processes. For that, HR is being used as a

---

<sup>1</sup>A positive form of stress having a beneficial effect on health, motivation, performance, and emotional well-being.

physiological signal for user evaluations and can be an indicator of a user’s physiological state (Peruzzini et al., 2018).

Various signals can be measured using a multitude of tools, the most common of which are HR, Heart Rate Variability (HRV), and Low Frequency/High Frequency (LF/HF) measured through Photoplethysmogram (PPG). PPG is an optically obtained plethysmograph that can detect changes in blood volume within the microvascular tissue bed (Drachen et al., 2010). Often found in smart bands, PPG is typically captured using a pulse oximeter, which shines light on the skin and gauges alterations in light absorption. PPG holds potential for non-invasive stress detection as it relies on passive sensors to furnish pulse data. This data can be exploited to analyze heart-beat intervals and has been demonstrated to be a reliable alternative for assessing HR values in beats per minute for healthy individuals (Houzangbe et al., 2020; Y. Wang, Fischer, & Bry, 2019).

While HR data is generally acquired through wearable devices, some studies have uncovered less intrusive methods for collecting it (G. Du, Long, & Yuan, 2020). This implies that the level of intrusiveness for monitoring HR signals depends on the specific setup. However, as we will demonstrate, distinguishing between stress and eustress states with a single heart monitor can be challenging. For this reason, HR data is frequently utilized in combination with information provided by other devices for physiological data collection.

#### **2.4. *Electroencephalography***

Electroencephalography (EEG) is a method of recording an electrogram of electrical activity in the scalp that has been shown to represent the macroscopic activity of the surface layer of the brain. This measure can be used to explore physiological information about the user and can be a useful tool for UX as EEGs can be taken as an indicator to assess user perception when using products without interruption (Ding, Cao, Qu, & Duffy, 2020). According to Baka, Stavroulia, Magnenat-Thalmann, and Lanitis (2018); Salgado et al. (2018), EEG capture is a difficult task considering the complexity of the human brain and is susceptible to significant noise or unwanted information. However, with the increasing adoption of this method for QoE assessment, there are standard guidelines used in a clinical setting to properly collect and analyze EEG signals (Elor et al., 2020; Li et al., 2021; Sandeep Vara Sankar & Ko, 2020). Nonetheless, some of the practices used to measure EEG can affect the UX and QoE. For example, in order to use a wet/gel electrode, the participant’s scalp needs to be properly prepared, which can be time-consuming and can exhaust the user even before a test or experiment begins. In addition, the user’s freedom of movement is reduced as s/he has to avoid body movement, noise and inaccurate data. For this, EEGs are highly intrusive, although some commercial products, such as those from EMOTIV<sup>2</sup>, seek to make EEG more user-friendly.

#### **2.5. *Other Physiological Signals***

Other physiological signals such as Electromyography (EMG), Electrooculography (EOG), and Breathing Rhythm capture the user’s musculature, eye movement, and breathing in search of patterns that can be also used in QoE assessments. However, in the context of Computer Science, EMG and EOG are mainly used as middleware to

---

<sup>2</sup><http://www.emotiv.com>

allow input of physical actions into a computer as a form of human-computer interaction. However, both EMG and EOG are usually an "intentional" physiological signal, and not part of the ANS. This means that they follow user intent and are not an involuntary signal from the ANS; therefore they should not be treated as an objective signal (Giannakakis et al., 2019).

EOG and Eye Tracking are techniques that can be used to assess the usability of new technologies by shedding light on decision-making through the examination of eye movement patterns, which has the potential to improve usability evaluations by providing ocular data (Cho et al., 2019). While not purely involuntary, except in the case of saccade<sup>3</sup> detection, recognizing where the user is looking and what they are feeling can be an important tool in usability assessments.

Devices to monitor eye movement patterns are widely applied and range from discrete, unobtrusive tracking devices to implantable sensors that capture muscular activity (Cho et al., 2019; Dhaouadi & Ben Khelifa, 2020; Greinacher et al., 2020; Kaluarachchi et al., 2021). In this respect, there is, as will be shown, a wide range of tools with different objectives. Finally, the least commonly used physiological signal was temperature, used only by Liapis et al. (2018) and Rahman, Gedeon, Caldwell, Jones, and Jin (2021).

### 3. Methodology

Systematic literature reviews (SLRs) are the act of identifying, evaluating, and interpreting scientific results relevant to a particular research problem or phenomenon of interest (Carrera-Rivera, Ochoa, Larrinaga, & Lasa, 2022; Kitchenham & Charters, 2007). To perform the SLR underpinning this article, we applied the guidelines proposed in these two references.

The descriptors adopted were:

**(“physiological” AND “data”) OR (“biosignals” AND “data”) OR (“biofeedback” AND “data”) AND (“User” AND “Evaluation”) OR (“QoE” OR “Quality of Experience”) OR ((“User” AND “Evaluation”) OR (“QoE” OR “Quality of Experience”))**

In Boolean search in Titles, Abstracts, and Keywords, for publications starting in 2018 to March 2022. All the collection was conducted between January to March of 2022.

As Section 2.1 has highlighted, whilst biofeedback may be the umbrella term for capturing physiological data from biosignals and returning it for another purpose, the terms biofeedback, biosignals and physiological signals are inextricably linked and is the reason why they were all used when searching datasets.

The papers' database used in this research was the CAPES/Brazil Portal of Periodicals<sup>4</sup> that offers access to the complete texts selected in approximately 21,500 international journals. The selected databases were Scopus, Web of Science, IEEE Xplore and ACM Digital Library.

As depicted in Fig. 1, a total of 337 articles were identified using these descriptors in Scopus. Among these, 83 were selected based on their title and abstracts. In the

---

<sup>3</sup>A rapid and simultaneous movement of both eyes between two or more phases of fixation in the same direction.

<sup>4</sup><https://www-periodicos-capes-gov-br.ez1.periodicos.capes.gov.br/>

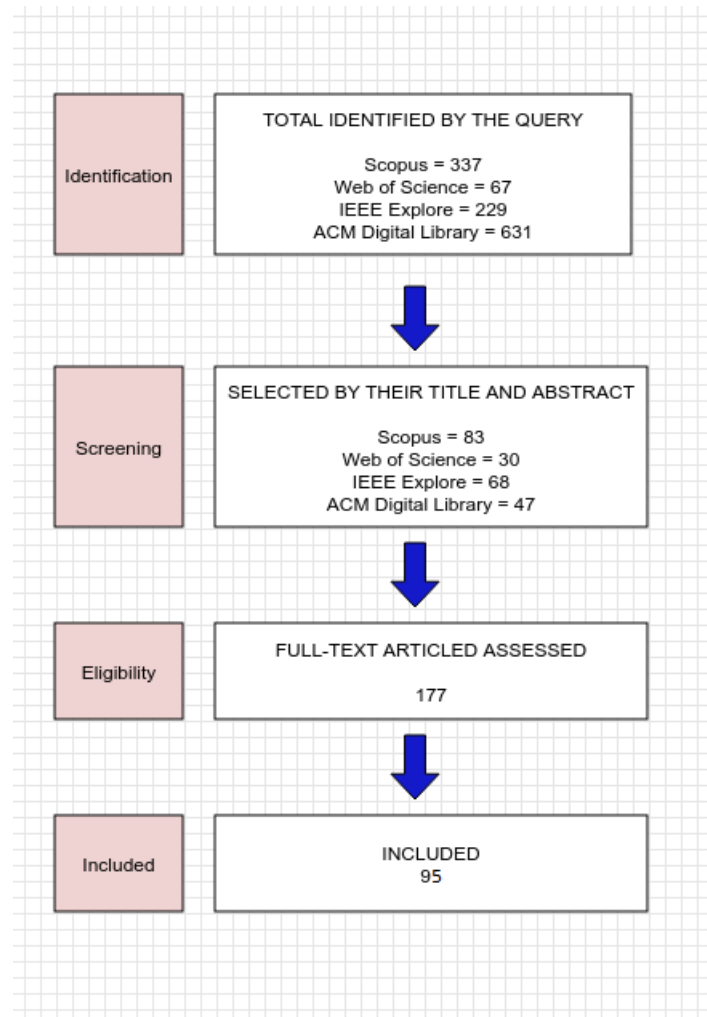


Figure 1.: Articles selection flowchart

Web of Science, the descriptors yielded 67 articles, out of which 30 were chosen. A total of 229 articles were discovered in IEEE Explore, and 68 were selected. In the ACM Digital Library, 631 articles were identified, and 47 were chosen following the initial screening. Subsequently, 33 duplicates were identified by their DOI or Name and subsequently removed. Additionally, 5 articles lacking a DOI identifier and 18 articles that we were unable to retrieve from the database were removed. Consequently, 177 relevant journals and proceedings were identified for this research and subjected to comprehensive reading. Following the thorough screening process, 95 articles met our criteria and were selected for inclusion in this review, following the application of the exclusion criteria. The study participants' inclusion (IC) and exclusion (EC) criteria are as follows:

- IC1 – it is an evaluation of QoE or UX using physiological data.
- EC1 – it is not an evaluation of QoE or UX using physiological data.
- EC2 – it is a short paper, white paper, thesis, and dissertation report.
- EC3 – it is a survey or systematic review paper.
- EC4 – it is not in English.

In the process of conducting this SLR, a rigorous screening and selection process was carried out to identify articles meeting the predefined criteria. This involved the assessment of articles from the databases to ensure the relevance and quality of the included studies. Three authors played roles in screening and selecting articles based on their titles, abstracts, and content. Furthermore, the fourth author proofread the sections dealing with the findings and discussions of the results of the process, ensuring its accuracy and clarity.

## 4. Findings

To categorize our findings, initially, three main categories were established, depending on the number and type of media, and on the environment type as follows:

- (1) QoE/UX studies using physiological data for single media:
  - (a) Studies with 2D content.
- (2) QoE/UX studies using physiological data for multimedia:
  - (a) Studies with 2D content.
  - (b) Studies with 3D/360 content.
  - (c) Studies with VR/AR content.
- (3) QoE/UX studies using physiological data for mulsemmedia:
  - (a) Studies with 2D content.

Between 2018 and 2022, the majority of studies were categorized under traditional 2D content, multimedia content, and VR environments. It is important to highlight the proliferation of HMDs is extending to the academic realm, matching the number of studies focused on traditional multimedia. Furthermore, only a limited number of studies employing mulsemmedia content were identified, which involves the utilization of multiple sensory devices synchronized with traditional audiovisual content. This underscores an unexplored research area that can be pursued with the assistance of physiological data.

### *4.1. QoE/UX studies using physiological data for single media*

#### *4.1.1. 2D Content*

Most studies in this category used physiological data to measure QoE of commercial or educational applications. Accordingly, Beşkardeşler and Arslan (2019) used GSR/EDA and HR to evaluate a Web Application that provided visual stress tests and basic settings for them. Results show that the proposed system is promising as a tool in remotely monitoring the stress level of the subjects. In a similar study, Bruun (2018) used GSR/EDA to capture physiological data while users interacted with a web application. The main objective of this study was to show that identifying and reading GSR/EDA can be accurate even when done by non-specialists; furthermore, it was found that non-specialists analyzing GSR/EDA sensor data were able to detect close to 100% of all subjective events marked by the participants.

In their work, Isiaka, Adamu, and Adamu (2021) harnessed GSR/EDA in conjunction with skin temperature, eye tracking, and user activity attributes to develop a window-based dynamic control system. This system incorporates biosensors, offering a novel approach to comprehending and managing user emotions and behavior while

they interact with a user interface. The biosensors collect secondary data attributes from these synchronized physiological readings and employ them for two primary purposes: detecting optimal emotional responses and assessing users' stress levels. The authors successfully captured stress-induced spikes in their sensors. These spikes, in turn, trigger an integrated interface within users' affective states on the same web pages. In a related study that also seeks to understand how a user responds while interacting with an interface, Katada, Okada, Hirano, and Komatani (2020) employed GSR/EDA and HR data to gather mental states while users engaged with a dialogue system. Their findings suggest that physiological signals can aid in detecting the implicit aspects of negative sentiments, which are acoustically/visually indistinguishable. Furthermore, Rahman, Hossain, and Gedeon (2020) utilized GSR/EDA to capture human emotions during interactions with emotionally "charged" videos. Their aim was to recognize emotions across up to seven distinct emotional categories. To validate the accuracy of their results, they employed a leave-one-observer-out approach and a neural network classifier. These methods yielded high performance, allowing their classifiers to achieve high accuracy in successfully identifying the seven emotional categories within this task.

In their study, Courtemanche et al. (2018) leveraged both an eye tracker, as well as GSR/EDA and EEG measurements to gain deeper insights into users' experiences within a 2D virtual environment. In this experimental setup, participants were presented with image stimuli comprising standardized pictures sourced from the International Affective Picture System (IAPS), displayed individually on a gray background. The Tobii eye tracker was employed to record parameters such as eye movement, pupil diameter, and EEG activity, facilitated by the Netstation acquisition software and EGI amplifiers. GSR/EDA measurements were captured using a wireless MP-150 Biopac amplifier with two electrodes placed on the palm of the non-dominant hand. This innovative physiological tool effectively utilized these biosignals to identify regions within the images where users were undergoing different emotional and cognitive states at a higher frequency. The results of the study demonstrated that physiological heatmaps outperformed standard gaze heatmaps in pinpointing emotionally significant areas within the user interface. In a similar setup, Cho et al. (2019) employed eye-tracking technology to augment the Think-Aloud methodology to effectively detect crucial usability concerns within a proposed mobile application. This setup enabled the researchers to pinpoint 19 critical errors across four distinct activities, with two of these activities yielding task completion rates of less than 78%. In order to gain a more comprehensive comprehension of these usability issues, an exhaustive analysis of participants' concurrent eye movements and verbal commentaries was conducted through an in-depth problem analysis, highlighting that eye trackers and Think-Aloud methods can be merged for better user evaluation understanding.

In contrast to setups necessitating dedicated hardware for eye tracking, Kaluarachchi et al. (2021) introduced *EyeKnowYou*, a self-contained DIY toolkit employing a webcam mounted on the head that focuses on the user's eye. This approach allowed for the determination of increased cognitive load based on physiological attributes and simultaneous identification of whether the user is genuinely looking at a screen, achieved through reflections on the eye's corneal surface. Also solely using an infrared camera, S. Chen and Epps (2020) captured head movement and eye tracking to gain a deeper insight into users' mental load during the evaluation of simulated tasks. The results revealed a correlation between tasks with high loads and associated head and eye movements. A similar research approach was employed by H. He et al. (2018) in their study, the authors introduced a methodology aimed at identifying human emo-

tions and cognitive states through the analysis of eye-gaze behavior, head motion, and verbal input. Within this experimental framework, the researchers gathered user eye-gaze data during interactions with a 2D screen. This data was then merged with information on object movement and object characteristics using machine learning (ML) algorithms, resulting in the creation of an eye-gaze-based object attention prediction module. Additionally, an affective computing module, based on eye movement and head pose, was employed to infer the user's emotional state from non-verbal cues. This emotional evaluation was subsequently combined with the emotional content extracted from verbal input, leading to a comprehensive analysis of the user's overall emotional and cognitive state. Notably, the authors achieved accuracies ranging from 86.3% to 92.1%, highlighting the high accuracy and practical utility of the attention prediction module in predicting human attention states and preferences.

Peruzzini et al. (2018) utilized a mixed environment to engage users in interactions with virtual and digital items. User performance was captured and digitized, emulating human-machine interaction. Simultaneously, physiological data collection was employed to objectify users' physical and mental workload during task execution. This system was subsequently applied to an industrial case study centered around agricultural machinery driving and control. The aim was to assist in defining a new cabin and its control board, encompassing considerations such as seat features, positioning and grouping of controls, and placement of supplementary devices. By utilizing EEG, Li et al. (2021) adopted a quantitative approach to assess fluctuations in QoE as users encountered varying levels of video buffering. Their approach introduced a more objective measurement methodology, effectively addressing the challenges associated with the inconsistency of subjective data quality. By comparing Buffering-Time-EEG (BT-EEG) scores with Mean Opinion Score (MOS) and DR Score, smaller variances indicated that the evaluation scores, based on an equivalent number of evaluation samples, closely aligned with the overall population average in a statistical context. Consequently, the authors asserted that the proposed BT-EEG score not only demonstrates enhanced evaluation capabilities but also aligns more closely with the genuine standard score for QoE.

With an emphasis on educational advancements, Ciolacu and Svasta (2021) introduce a pedagogical model tailored for Education 4.0, employing principles from diverse domains encompassing electronic technology, artificial intelligence (AI), healthcare, multimodal learning analytics, and education. The authors contend that harnessing multimodal data derived from online courses and wearable devices can offer invaluable insights into the specific learning context and enrich students' subjective experiences, including emotional states, stress perception, and well-being. In their research, they employed wearable devices equipped with integrated sensors to examine physiological signals such as HR, photoplethysmography, and blood oxygen levels. The authors expressed optimism about the results, underscoring the need for further exploration in the context of an adaptive framework for self-regulated learning. Through biofeedback, both participating students and researchers gained awareness of the body's physiological responses, further underscoring the potential for enhancing the learning process.

In a related study, Holman and Adebessin (2019) employed EEG technology to evaluate users' emotional responses during their interaction with electronic articles. The experiment was designed to record the participants' brain activity using EEG patterns, enabling the identification of their emotional states. Each participant was tasked with reading three distinct electronic news articles, each strategically crafted to elicit varying emotions. The results demonstrated the effectiveness of EEG in capturing the users' emotional experiences. Nevertheless, it is important to note that the au-

thors made an assumption that the spikes detected by the software indicated stronger psycho-physiological responses from the users. However, it is worth considering that these spikes could also be associated with head motion. Further discussion regarding artifacts will be elaborated upon in subsequent sections.

Audio content has also been subject to evaluation through the analysis of physiological data, as exemplified in the study by R. Du et al. (2022). Similar in nature to the research conducted by Kim, Lee, and Kim (2021), this study utilized EEG to effectively classify emotions while users listened to music. The authors introduced a hybrid model comprising a one-dimensional convolutional neural network and both bidirectional and unidirectional long short-term memory components (1D-CNN-BiLSTM) for the purpose of classifying user emotions based on EEG data. The experimental outcomes showcased the superiority of the proposed method over traditional LSTM and 1D-CNN-LSTM models, with accuracy rates of 94.85%, 98.41%, and 99.27% achieved in the valence classification task of music-induced emotion, respectively. Similarly, the accuracy for the arousal classification task was recorded at 93.40%, 98.23%, and 99.20%, respectively.

In a study conducted by Tateyama et al. (2019), GSR/EDA was employed to categorize human emotions elicited by music stimulation. The study introduced a concept aimed at differentiating distress from eustress. The results indicated that changes in the GSR/EDA exhibited similarities during both distress-inducing and eustress-inducing tasks, suggesting that distinguishing between distress and eustress based solely on skin measurements during the tasks may be challenging. However, a notable distinction emerged when assessing GSR/EDA levels after the tasks. Specifically, the GSR/EDA remained higher at the conclusion of listening to negative mood induction music following a distress-inducing task as compared to that following a eustress-inducing task. This utilization of GSR/EDA provided insights into the enduring nature of distress in contrast to eustress.

Just as in the study of Rahman et al. (2020) and with similar objectives, Rahman et al. (2021) employed EDA, HR, Eye tracking and other physiological signals to capture and classify human emotions while listening to different music genres. Employing neural networks, Rahman et al. (2021) were able to classify emotions with high accuracy, with the highest accuracy of 99.2% across all different classification methods.

In their research, de Santana and Otani (2021) utilized a GSR/EDA sensor and conducted two case studies. The first case study involved a cloud-based support application designed to assist speakers during presentations by identifying stress events. The second case study involved a quantitative analysis during game-play of a mobile game, where arousal events during winning and losing matches were compared. The results demonstrated that GSR/EDA could effectively assess QoE in both scenarios, albeit with certain caveats related to noise and false data. In a comparable study, Lee, Ryu, Park, and Yun (2020) utilized physiological data to scrutinize the user experience during conversational agent-assisted interactions while watching television. To facilitate this investigation, wearable devices were employed to gather GSR/EDA and HR data from the users. The collected physiological data significantly enhanced the quality of user responses. The authors emphasized the value of physiological measurements in collecting objective data through the analysis of physiological recordings during interactions and in identifying shifts in user emotions.

In an attempt to enhance traditional evaluation methods, Liapis et al. (2018) conducted an investigation into how the Think Aloud protocol could be improved through real-time monitoring of physiological data. To accomplish this, the authors observed users as they interacted with a web application and completed a series of tasks. The

participants' physiological signals, including GSR/EDA, skin temperature, and blood volume pulse, were recorded. Analysis of the collected data revealed that participants using the Think Aloud mode with physiological monitoring reported significantly more usability issues. This suggests that physiological measurements uncovered hidden data that would not have been identified through a standard subjective assessment. With a similar goal in mind, Hussain et al. (2018) proposed and evaluated a framework (Lean UX) for the evaluation of user interfaces. This framework captures and measures multiple signals in real-time, such as eye tracking for visual attention and EEG for rapidly detecting emotions, motivations, engagement (arousal), cognitive workload, and frustration levels.

In another study with a similar aim of enhancing traditional evaluation methods, Liapis, Katsanos, Karousos, Xenos, and Orphanoudakis (2021) introduced a method that utilized GSR/EDA and HR measurements to enhance Retrospective Think Aloud (RTA) procedures. In RTA, participants verbalize their interaction experiences after completing a task or a set of tasks on a 2D display. While the authors did not specify which sensor was more effective for their ultimate research objective, their overall findings suggested that incorporating physiology into RTA interventions appears to be the most suitable approach for evaluators interested in detecting a greater number of usability issues compared to the conventional strictly oriented RTA approach.

## ***4.2. QoE/UX studies using physiological data for multimedia***

### *4.2.1. 2D Content*

Due to the nature of multimedia environment, only a handful studies could be classified in this category. Some of the explored more than one category, such as in the work of Tsai, Pan, and Hu (2021). The objective of this study conducted by Tsai et al. (2021) was to investigate the affective state in VR while users watched videos displayed in different formats: 2D video, panoramic video, and 360 video. To conduct this study, the authors developed an experimental procedure and an emotion annotation and monitoring system. This system not only assessed users' affective states but also collected physiological signals and observed behavioral movements while users watched videos in different formats. The signals captured in this study included EEG, HR, GSR/EDA, and eye gaze patterns. The researchers achieved success, noting that users perceived panoramic and 360 videos more favorably.

In their research, Markova, Ganchev, and Kalinkov (2018) introduce a wearable system designed for continuous monitoring and recording of negative emotions, high emotional arousal levels, and high-arousal-negative-valence states using physiological signals such as skin conductivity and ECG. This system employs a client-server architecture, utilizing wireless data acquisition devices (Shimmer3 GSR+ and Shimmer3 ECG) to collect physiological data. These signals are then transmitted via Bluetooth to a mobile phone, which serves both as the user interface and manages data aggregation and transmission to the server. Signal processing and classification tasks are conducted on the server side using specialized software, while the client side handles user interface and data communication. During data acquisition, the study utilized audio-visual stimuli, specifically seventeen musical clips, with sixteen categorized within the four quadrants of the valence-arousal emotion model. An additional video clip was used to establish a baseline relaxation condition at the start of each recording session. The experimental results suggest that the proposed approach has the potential to automate the detection of emotional arousal, negative emotional states, and high-arousal-

negative-valence conditions from physiological signals. Reported detection accuracy for a single-stimuli recording ranged between 75% to 85%.

Frey, Ostrin, Grabli, and Cauchard (2020) instead of using physiological data to evaluate and measure QoE, used it to dynamically change a storytelling environment. *Physiologically Driven Storytelling* is a concept that combines two perspectives on the use of digital tools: interactive narratives as a way to relate to fictive or real-life events and physiological measurements as a way to assess and affect people. Using GSR/EDA and eye tracking, the authors conducted two studies, one to uncover story characteristics that can elicit engagement and increase perceived similarity, and a second to assess what states can be measured when reading a story and which physiological features are the most useful to do so. The authors found that constructs linked to cognitive processes resulted in higher classification accuracy than constructs related to affect.

Towards a similar objective, the study reported in (Gurita et al., 2019) employed GSR/EDA to identify states of mind while a user interacted with a system. The demo developed by the authors explored the opportunity of how dynamic visualization stimuli and bio-feedback sensors might provide insight into mental states that the user is not aware of. The system design is the result of a project undertaken by the authors to employ the potential of this type of sensor in the context of therapies. Early results suggest a correlation between the visualization and the emotions of users, as well as users' interest and engagement with the system. Further research can be beneficial for creating a greater therapy experience, as well as generating new user testing methods and game interactions that would make use of those variables. Ajenaghughrure, Sousa, Kosunen, and Lamas (2019) also used EEG to evaluate the level of trust. This study aimed to show that it is possible to detect users' trust levels in AI technologies and provides details on what physiological signals are suitable for real-time trust detection, as well as details on the predictive ML model used.

#### 4.2.2. 3D/360 Content

The primary emphasis within multimedia environments was on commercial and serious games. In alignment with this, Daylamani-Zad, Spyridonis, and Al-Khafaaji (2022) utilized GSR/EDA and HR to evaluate the capacity of serious games to induce stress in firefighters. Their research findings underscored the effective elicitation of stress in the study participants. This was substantiated through both self-assessment and physiological measurements, and the outcomes demonstrated statistical significance. Consequently, with both physiological apparatuses, the authors successfully identified and verified the accomplishment of the game's primary objective, namely, the induction of stress. In another gaming study, Alves, Gama, and Melo (2018) developed a first-person shooter video game with a unique feature - dynamic adjustments of in-game difficulty and environmental settings. This adaptation relied on real-time monitoring of the player's mental state during gameplay, assessed through HR, GSR/EDA, and EEG devices. If the player exhibited excessive stress, the game would automatically fine-tune its difficulty level. The results, however, were mixed. While there was empirical evidence suggesting that users experienced a heightened flow state and achieved higher scores when playing a game that adapted to their mental state, their results revealed a different trend. It became evident that players actually achieved a more elevated flow state and obtained higher scores when engaging with a game that adjusted itself based on their performance. A more in-depth statistical analysis further confirmed that players attained a superior flow state and higher scores in a performance-based adaptable game, contradicting the notion that a mental state-based adaptable game

led to the same outcomes.

Another gaming session evaluation was conducted by Certický et al. (2019). They aimed to create a reliable model for predicting players' enjoyment during gameplay by capturing various physiological signals, including HR, GSR/EDA, respiratory activity, and EEG. Multiple psychophysiological measures, such as heart rate, electrodermal activity, and respiratory activity, were combined with self-reported data to create training sets for ML algorithms. Subsequently, the authors trained and compared the results of four different ML models, with the best one achieving an accuracy of approximately 96%, further suggesting that psychophysiological measures can indeed be employed to assess the enjoyment of digital entertainment consumers. In their player performance study, Darzi, Wondra, McCrea, and Novak (2019) compared classification accuracies across psychological dimensions and game difficulty preferences. They utilized three data modalities: physiology, performance, and personality traits, with 30 participants playing a computer game at nine difficulty levels. This involved recording various physiological measures, including GSR/EDA, HR, eye tracking, respiration, EEG, and performance variables. Participants also completed a short questionnaire about perceived difficulty, enjoyment, valence, arousal, and desired difficulty adjustments. Combining these data modalities, the authors used four classifier types to classify questionnaire dimensions into two, three, or multiple classes. Notably, accuracy varied across dimensions, with the highest being 97.6% for two-class and 84.1% for three-class classification. The study highlighted the significance of normalized physiological data while acknowledging the influence of game difficulty, personality, and performance.

Huynh et al. (2018) used HR and GSR/EDA to gauge the sense of engagement in mobile games, asserting that measuring and predicting a gamer's engagement can serve as an effective indicator for determining the success and QoE of a mobile game. The results indicated that their system, with the utilization of sensing channels, could classify three levels of engagement with an average accuracy of up to 85% in cross-sample evaluation and 77% in cross-subject evaluation. In a separate study, Klarkowski, Johnson, Wyeth, Phillips, and Smith (2018) employed GSR/EDA to assess players' interactions with various game difficulty setups. The results demonstrated decreased GSR/EDA in the low-challenge video game condition compared to the medium and high challenge conditions, with consistent patterns found between the medium and high challenge conditions. Overall, these findings suggested that GSR/EDA response varies with challenge. In a related study, conducted by Satti et al. (2021), the objective was to evaluate the influence of gaming and identify heightened stress levels using GSR/EDA signals. Employing ML methodologies, the authors quantified stress levels during user engagement with a video game. The outcomes of this investigation yielded moderately average results, with accuracy rates hovering around 60%, and precision in identifying stress and relaxation levels approximately at 50%.

Also in a game environment and using a range of devices, including Tobii eye tracking for capturing gaze, an Empatica E4 wristband, and a Kinect Skeleton, Lee-Cultura et al. (2020) investigated users' feelings while interacting with Motion-Based Educational Games. This allowed them to identify features that lead to rapid and highly accurate predictions of players' academic performance. The study also provided real-time proactive feedback to players to support them throughout their educational gaming experience. Yang, Rifqi, Marsala, and Pinna (2018) used HR, GSR/EDA, EMG, respiration, and body movement to evaluate QoE in gaming platforms. Statistical analysis revealed a relationship between game events and associated emotions with a positive correlation identified between players' perception of immersion and amusement in the

game experience evaluation. The authors also established that physiological signals are effective modalities for predicting players' game experiences. They demonstrated that fusing features from behavioral and meta-information modalities significantly enhances prediction performance. In another study, Manjunatha et al. (2020) used EEG and eye tracking to analyze the playing styles of different users. The study involved 20 participants with prior experience in playing real-time strategy games, and the data obtained enabled the authors to detect three distinct play styles between the players.

In addition, a study by Dhaouadi and Ben Khelifa (2020) employed HR, GSR/EDA, and EMG to develop and train several ML models using data collected from gamers during their gameplay sessions. Their primary objective was to propose an automated stress monitoring system tailored for young gamers, including children and adolescents. Wearable sensors were employed to capture data, which was subsequently analyzed and processed using deep learning algorithms, including Long Short-Term Memory (LSTM) and Deep Neural Networks (DNN). These models enable the real-time quantitative estimation of gamers' stress levels. Notably, the study yielded two highly accurate models for stress detection, underscoring the potential of psychophysiological data in assessing user experience and quality of experience. With a similar aim, the study conducted by Sandeep Vara Sankar and Ko (2020) harnessed ML techniques to effectively identify user fatigue during video game sessions. The authors devised a game-based, multi-objective experimental setup to assess the performance of BCI with an EEG setup and investigate cognitive states such as fatigue and attention deficits during extended BCI tasks. Their proposed 2D-CNN model excelled in classifying target/non-target objects, achieving an impressive 86.12% accuracy, while also successfully discerning fatigue-induced attention deficits in users. Notably, the classification results from their model showed significant promise compared to conventional methods like Bayesian Linear Discriminant Analysis (BLDA) at 68.48%, Support Vector Machine (SVM) at 70.23%, and bagging tree algorithms at 76.54%, also highlighting that this setup based in EEG can be successfully used as a mean of measurement of UX and QoE.

In related studies, Schmidt, Uhrig, and Reuschel (2020) aimed to gain a better understanding of QoE in Cloud Gaming. They employed GSR/EDA and HR to collect data that could shed light on the impact of screen size and network conditions on QoE. While no significant differences were observed with respect to screen size (all screen sizes used in the research were larger than 5 inches) and immersion, the authors did report notable differences in overall gaming QoE and immersion resulting from a reduction in video quality. Furthermore, among all the extracted physiological features, only HRV was found to be significantly affected by network conditions. Similarly, Ye, Ning, Backlund, and Ding (2021) used GSR/EDA and HR to gauge the level of flow experienced during a gaming session. Their findings indicated that different game challenge settings evoke distinct cognitive or emotional experiences in users. The study showed that the flow experience of users can be discerned through specific features in GSR/EDA and HR signals.

Darzi, McCrea, and Novak (2021) used EEG, HR, and GSR/EDA for biofeedback control of difficulty in games. However, this study revealed no observed differences between the dynamically and static game difficulty. The authors stated that no clear benefit of using biofeedback controllers was detected and results did not support the use of physiological measures for effective exercise gaming in practical settings. In one rare instance, the authors' conclusion is that physiological sensors are expensive, time-consuming, and inconvenient; thus, they should not be used unless they show a clear benefit. In related work, Ninaus, Tsarava, and Moeller (2019) developed an adaptive

game that would trigger events if low arousal was detected by low heart rate. The adaptive mode did not affect the type of events that happened during the game, but the frequency with which they occurred and their difficulty: if in a specific scenario, there were 4 events in total, but the participant’s heart rate value dropped at least 5 bpm below baseline for at least 10 seconds, then the scenario would adapt, triggering more challenging events. On the other hand, if participants’ heart rates increased by at least 5 bpm above baseline for at least 10s, the scenario adapted to the player’s high arousal and stress, triggering a helping action. The results evidenced that participants, unaware of the true nature of the pilot study, experienced the most challenging and fascinating adaptive version of the game. In contrast, the non-adaptive game was rated as easier. However, no differences were found in the players’ perceived flow, which can be attributed to the limited sample size. More importantly, though, the difficulty of the non-adaptive version was rated as being significantly higher, whereas this was not the case for the adaptive version of the game. This is in line with the authors’ argument that the implemented biofeedback as controllers worked as expected, making the game neither too easy nor too difficult. In their study, Houzangbe et al. (2020) employed a desktop PC along with a Mio LINK HR wristband, which has the capability to estimate heart rate using photoplethysmography. The VR setup consisted of an HTC Vive VR System and headphones. Utilizing a game inspired by *Portal*, participants were instructed to “voluntarily control their heart rate”. In each trial, participants could view their current heart rate on the screen, along with a target heart rate to achieve. As they reached the designated heart rate zone, the mechanisms within different rooms progressively activated. Participants had to maintain their heart rate within the specified zone for a predetermined duration to fully activate the mechanisms and the teleporter, ultimately completing the trial. The purpose of this experiment was to gain insights into player engagement during a game controlled by physiological parameters.

Hossain and Kulshreshth (2019) employed EEG and HR to assess stereoscopic 3D viewing on a 2D display. Initial findings indicated that stereoscopic 3D does offer advantages in tasks where depth information is relevant for the game task at hand. Additionally, participants in the 3D group exhibited lower stress levels and higher heart rates, suggesting a greater sense of engagement and presence in stereoscopic 3D conditions. Similarly and utilizing EEG, Riaz, Majid, and Mir (2021) conducted a study to explore the impact of four audio-visual High Dynamic Range (HDR) videos on users. They extracted and selected time-domain features from EEG signals recorded across 5 channels, employing an SVM classifier to classify emotions into two states: arousal and valence. Their classification outcomes demonstrated accuracy rates of 80.55% for arousal and 70.37% for valence, highlighting the potential of HDR videos as potent stimuli for emotion recognition and the use of EEG to capture mental state. The measurement of flow was also explored by Maier, Elsner, Marouane, Zehnle, and Fuchs (2019). They collected physiological data including GSR/EDA, HR, and body temperature, which were then input into a deep neural network. This approach achieved an accuracy of 67.50% in distinguishing between high-flow and low-flow states.

In this category, a unique approach to collecting physiological data was identified in the research by G. Du et al. (2020). In an effort to circumvent the use of wearable and intrusive equipment for physiological data capture, they employed an innovative contactless method involving Facial Emotions and HR to evaluate user reactions to four distinct types of games. Consequently, the authors were able to automatically detect and measure four emotions. While this system does have high requirements, such as the need for excellent lighting conditions, a non-contact setup is appreciated

as it reduces the preparation time and discomfort for the user.

To explore QoE with 360 content, physiological data were used in two cases, the first of which is the study of Elor et al. (2020), who used physiological data to compare how two setups - CAVE and VR - were perceived by users. EEG, GSR/EDA, and HR were collected at run time while users played Project Star Catcher. The order of which the system was played was counterbalanced (half of the users experienced the VR setup first, while CAVE was the first setup experienced by the other half) to prevent bias and order effects. Results showed that VR surpassed CAVEs in the feeling of presence with the exercise game, supporting the hypothesis that a modern VR such as the HTC Vive is more engaging and produces better results than the more expensive room-scale CAVE setup. The second is the work of L. He et al. (2018) who employed EEG and motion capture techniques in order to better understand the potential of VR to improve QoE in the context of theatrical performance. Their results show that VR applications were able to replicate live performances of a theater scenario by providing a higher sense of presence, higher engagement levels, and a stronger desire to see live performances. After studying three replaying conditions, including the middle camera video, an edited video, and the panoramic video, the authors concluded that the viewer response to the edited video is more comparable to that of the panoramic video. Subsequently, in a comparative study, they used the panoramic video and the edited conventional video of a theater performance to evaluate the viewer response and it was found that the engagement level evoked by the panoramic video was higher than that of the conventional video. However, the result of the information recall evaluation shows that the viewers of conventional video had a better performance on the narrative-based questions. Most users were able to memorize the narrative plots, while the viewers of VR sometimes missed some key elements of the storyline. This means that while VR evoked a higher sense of presence, it also made it more difficult for the user to concentrate on the play.

In a study akin to Ciolacu and Svasta (2021), Giannakos et al. (2019) used diverse data sources, including click-stream data, eye tracking, EEG, video, and physiological data, during a learning game experiment. Their goal was to better comprehend the learning experience through multimodal data. Their findings emphasize the effectiveness of multimodal data in predicting learning performance, highlighting its superiority over traditional data streams like keystrokes. This research suggests that integrating advanced technologies can enhance the understanding of unscripted tasks in learning environments. It contributes by identifying critical features for skill development, evaluating data stream predictive power, and guiding future learning technology design and experiments. Also in this sense, in order to better understand mental workload, in a similar study to Wulvik, Dybvik, and Steinert (2020), Korn and Rees (2019) used physiological data to investigate mental workload and learning. Using facial expressions and GSR/EDA with standard performance measures, the study confirmed known correlations like increased speed and increased error rate. In addition, the analysis of the physiological data provided evidence for two major affective effects: the gamification of work and learning tasks incites significantly more positive emotions and increases emotions altogether.

Also in order to improve the quality of work life with serious games, Wulvik et al. (2020) investigated the workload and effect of a virtual environment simulating a ship control deck in two different scenarios. In these experiments, physiological data were collected through HR and GSR/EDA signals. The IMotions software platform was used to present stimuli and collect and synchronize both subjective and physiological data. The HR sensor was fixed with five leads to the participants' chests according to

the instructions provided by Shimmer, with lead VX<sup>5</sup> in position six. The GSR/EDA sensor was attached to the middle part of the index and middle fingers of the left hand. After fixing the sensors, participants were seated in front of a computer screen in the simulator environment. Participants were instructed on how the experiment would proceed and instructed to follow instructions given on the screen or via audio. The results showed significant changes in mental and physiological states between the two experimental scenarios. Moreover, there was an overall increase in excitement, discontent, stress and workload in the scenario where participants sailed a large ship in a narrow harbor compared to the scenario where they sailed in the open sea. Also, Mijović et al. (2019) replicating a work environment, the authors used EEG to understand the mental state of a worker in a job environment. Such a system would detect a drop in mental and physical performance so that appropriate action (eg, a break or a change in task) could be taken. In addition, the authors sought a system that could prevent the occurrence of operational errors and improve the worker’s experience.

#### 4.2.3. VR/AR Content

A considerable amount of studies yielded by this review are in this category. For instance, the work of Elor et al. (2020) showed, using physiological, that VR was better perceived by users than CAVE. In a similar vein, other studies also used physiological data to better understand how VR can improve QoE, how it differs from concrete reality, and how they promote a better experience than traditional 2D displays: Alqahtani et al. (2019) used HR and GSR/EDA to evaluate a system in multiple environments: AR, VR and tablet based. The results of the study indicated comparatively higher levels of QoE for users of the augmented reality and tablet platforms. Hynes, Flynn, Lee, and Murray (2019) used GSR/EDA to acquire objective measurements to investigate how different users felt in a VR environment compared to concrete reality. The objective metrics showed that VR yields efficiency and productivity gains over paper-based instruction. Physiological data also showed that the non-VR users experienced higher stress, with increases in skin temperature, HR, and significantly higher increases in GSR/EDA than their VR-based counterparts. Collins et al. (2019), using GSR/EDA and HR, explored with a new methodology how VR improves QoE. The authors implemented the methodology in a VR learning context to demonstrate its implementation and possible outcomes. By sensing the users’ physiological responses, they were able to predict with high accuracy the degree to which the VS and environmental stimuli impacted the subjects’ emotional states. Through HR and GSR/EDA, they were able to measure users’ cognitive load and better understand how users were dealing with the virtual environment. C. Chen et al. (2021) aimed to determine the impact of a 9D-VR Seat on user immersion, by analyzing their physiological responses. 20 volunteers experienced the VR roller coaster and VR big pendulum. Through analysis of both HR as well as visualization and statistical data, the authors found that there is a significant effect on HR when users experience VR, and that the excitement points of the VR content were related to changes in HR. In related work, Che et al. (2019) used HR signals of users captured during an experiment across several VR environments, such as ‘Slingshot’ and ‘Longbow’ developed by the game company VALVE. During the experiment, the experience data of 80 testers were recorded and four types of interaction events were defined and correlated with the HR change, such as high HR while grabbing or touching in a VR environment. With a similar setup Keighrey

---

<sup>5</sup>V<sub>x</sub>-WCT is the ECG vector signal measured from the Wilson’s Central Terminal (WCT) voltage to the V<sub>x</sub> position

et al. (2021) investigate AR/VR environments with GSR/EDA and HR. The results however were different from those of Collins et al. (2019); Dey, Phoon, Saha, Dobbins, and Billinghamurst (2020a); Souza, Nedel, Kopper, Maciel, and Tagliaro (2018), in that users in the AR and tablet groups experienced a higher enjoyment than VR, and although the capture of GSR/EDA revealed an increase in physiological arousal for the VR group, which could be attributed to the technology.

The measurement of presence in VR has also been explored via physiological data: Souza et al. (2018) used GSR/EDA, HR, and Behavior Observation (BO) to investigate and improve the level of presence in a VR environment. With physiological data, the authors were able to create a Dynamic Stimulus (DS). A significant difference was found in respect of the monitored physiological measures between the setup with no stimulus and DS condition, showing that it was possible to dynamically modulate the VR experience using real-time HR and GSR/EDA to increase the effectiveness of VR experiences. Dey et al. (2020a), using EEG, HR and GSR/EDA recorders, measured the sense of presence in a VR environment. The authors identified differences in physiological activities at different levels of presence. It was concluded that higher presence in calm virtual environments could be characterized by increased HR, elevated beta and theta activities in the frontal region, and increased alpha activity in the parietal region of the brain. Athif et al. (2020) also investigated the concept of presence in VR. Accordingly, EEG, HR and GSR/EDA signals were captured while experiencing custom designed VR scenarios with factors contributing to presence being suppressed and unsuppressed. The authors were able to detect significant variation with the suppression of the realism factor. The variations of activity in the temporal region lead to the assumption of insular cortex activation which may be related to the sense of presence. Baka et al. (2018) used EEG to capture participants' brain activity and the sense of presence during their exposure to different virtual and real environments. Both groups reported a high level of presence, although the real environments claimed higher levels and brain activity. Also, since the environment did not require any direct interaction of the participant, the authors noticed that interaction does not always play a significant role in user presence and immersion. Another related study is that of Dey et al. (2020b) who, using a combination of EEG, GSR/EDA, and HR, evaluated the sense of presence in a VR platform, reporting that there is a difference in brain activity with differing levels of presence in Virtual Environments (VEs). HR was also found to be higher in the high presence, with higher visual fidelity, VR setup than in the low presence environment, with low graphics and minimalist interaction. The high presence setup provided an embodied hand with human-like skin color, while the low presence only had a white detached palm, which represented the hand. Also, in high presence, animals could also be "touched" and would respond by running away, while in low presence the animals did not respond to user interactions. Another study in this context is that of A. Islam, Ma, Gedeon, Hossain, and Liu (2019), which used GSR/EDA to analyze drivers' responses to various driving simulator setups (single monitor, multiple monitors, and VR). They studied 23 participants engaged in virtual driving across different configurations. After data preprocessing, a neural network classifier achieved up to 90% accuracy in predicting simulator performance, also highlighting the potential of using physiological data to better understand users' interactions with a computer system.

In yet another study targeting presence, students were recruited among volunteers from the student population at a Norwegian university (Grassini et al., 2021). The VR scene presented a first-person roller-coaster ride. The VR scenario was chosen for the ability to induce, even after short use, a high sense of presence. For two minutes before

the start of the experiment and for the entirety of the virtual roller-coaster ride session, task-irrelevant tones were played via two speakers situated 45 cm behind the subject's position at 20 cm to each other. Participants were asked not to react to the tones but were told that the tones were external to the VR environments and were playing as part of the experiment. EEG results revealed central brain areas that were associated with the sense of presence. Voigt-Antons et al. (2021) using a VR environment based in a horror game to induce the state of fear measured by HR variation. In this study, it was also proposed the possibility of such data being used as a control factor to improve gaming experience instead of using physiological data to evaluation purposes only. Petrescu et al. (2020) used a VR acrophobia game in an environment with HR, GSR/EDA, and EEG collectors to better understand how subjects reacted to height stimuli. After processing the data, the authors were able to detect the users' anxiety state with high accuracy, that is, with a collection of physiological tools, it is possible to assess the user's mental state when interacting with a system. While not a strictly QoE-related study as such, its results can still be used for this purpose.

Cybersickness, a common issue in virtual environments (VEs), can be examined through physiological data and this can be used for UX and QoE. In a study by Katsigiannis, Willis, and Ramzan (2019), HR and GSR/EDA were employed to detect cybersickness within a Smart-Exercise-Bike Virtual Reality System. The analysis revealed that texture quality and frame rate significantly influenced perceived visual quality in the VR application, though not simulator sickness scores. Interestingly, the quality setting had no impact on participants' physiological responses in terms of HR and GSR/EDA. However, a weak correlation was observed between these physiological parameters and simulator sickness scores. It's noteworthy that engaging in strenuous activities, such as cycling, led to increased sweating and cardiac activity, potentially confounding GSR/EDA and HR signals related to simulator sickness, underscoring the impact of physical activity on physiological data.

In a related vein, Martin, Mathieu, Pallamin, Ragot, and Diverrez (2020) utilized HR and GSR/EDA data to detect VR cybersickness in 103 participants during VR video game sessions. Employing ML techniques, they trained models to predict cybersickness intensity, achieving up to 75% explained variance in a regression approach and up to 91% accuracy in a classification approach. These outcomes suggest the potential for real-time cybersickness detection, facilitating the timely application of countermeasures. Similarly, R. Islam, Ang, and Quarles (2021) opted for physiological data over subjective measures for real-time cybersickness detection. Their study primarily focused on HR and GSR/EDA, omitting EEG due to its complexity and extensive pre-processing requirements. Lastly, in a study by Al-Ashwal et al. (2021), GSR/EDA signals were used to detect cybersickness. They observed an increase in GSR/EDA among users experiencing cybersickness; however, this phenomenon was not consistent across all participants.

The measurement of stress in VR is another user-centric aspect that has been explored and is the subject of a study undertaken by Barathi, Proulx, O'Neill, and Lutteroth (2020). They used a GSR/EDA and HR detector to detect users' mental states in order to personalize and optimize their experience. They conducted two experiments, the first of which compared the impact of physical exertion and gamification on physiological measurements during rest, conventional exercise, VR exercise gaming, and sedentary VR gaming. The second experiment compared underwhelming, overwhelming, and optimal VR exercise gaming scenarios. The authors identified gaze fixations, eye blinks, pupil diameter, and skin conductivity as physiological measures suitable for affect recognition in VR exercise gaming and analyzed their utility

in determining affective valence and arousal. In a similar study, Chauhan, Reithinger, and Mackey (2018) used HR to assess stress levels as users interacted with a VR-based gaming environment for chronic pain relief. The proposed system achieved an accuracy of 93% in detecting stress levels during the sessions. A key-point of this study is that, while most systems are invasive, impractical, or intermittent, their proposed method could be used to avoid time-consuming setups that could stress or annoy users.

The identification of mental states via physiological data is also a topic for ML studies. Kalatzis, Stanley, and Prabhu (2021) used HR and Respiration Rhythm signals as objective data to extract HRV and Respiration Rate Variability features from 25 participants during two VR scenarios. This was done in order to create a validated multimodal dataset for affective state classification. However, Gupta et al. (2020) did not use physiological data to measure QoE. Instead, the authors used EEG, GSR/EDA, and HR to measure the level of trust of a user interacting with a Virtual Assistant in VR. The authors concluded that with this novel method, it is possible to quantify human trust using physiological data, suggesting new ways of using physiological data to explore UX and QoE.

Pei et al. (2021) used GSR/EDA and HR signals for emotion detection and proposed a method that can be used for dynamic and repeatable evaluation of design before the project is completed. Simultaneously, this method can help designers measure users' emotional perceptions and thus provide an objective basis for design optimization. In a similar study M. Pinto et al. (2018) used physiological data not to analyze user QoE but to improve it in a video game setting. Using GSR/EDA, users had their gaming experience modified by their mental state. This is a new method of using physiological information as it allows a dynamic way to improve QoE. In the study of Zou and Ergan (2021) three pairs of virtual environments were developed, each pair corresponding to one human experience, respectively stress, anxiety, and aesthetic pleasure. Specifically, for each pair of VEs, one VE had design features configured to invoke a positive experience, whereas the other VE had the same set of design features but configured to invoke a negative experience. This study successfully employed the use of EEG, a GSR/EDA, and PPG to better understand how the user experienced multiple virtual environment scenarios.

Based on the premise that questionnaires break user immersion and can hinder VR QoE evaluations, Putze et al. (2020) used GSR/EDA, respiration, and blood volume pressure (BVP) to understand how users interact with questionnaires within and outside of a VR environment. The purpose of this study was to show that, while subjective measurements can be useful, they can also break user immersion when used to evaluate VR setups. This is because while questionnaires are widespread measurement instruments to assess subjective responses to a particular experience in VR user studies, study results show evidence that switching between VR and physical reality leads to a break in presence.

Spatial sound has also been evaluated with physiological data: Moraes, Flynn, Hines, and Murray (2020) investigated how users dealt with spatial auditory localization in VR. Results from the experiments showed that users were able to understand the task of identifying spatial sound and learn how to use a system built to treat Central Auditory Processing Disorder (CAPD), also reporting that the two interaction methods were evaluated, eye gaze pointing and wand pointing, are quite comparable. This application can also be expanded to the health domain as a tool to improve assessment methodologies for people diagnosed with auditory processing disorders. In another study, Moraes, Flynn, Hines, and Murray (2021) used physiological data to understand how users reacted to spatial sound in a VR environment. Using an E4

Empatica wristband, GSR/EDA, HR, temperature, blood volume pulse, and acceleration were gathered. Results show that physiological data were successful in capturing users' feelings and that spatial sound improved QoE.

The measurement of workload was also the subject of the work by Tremmel et al. (2019), who utilized EEG to estimate cognitive workload within VR UX. They hypothesized that by recording EEG while using a VR device, user interactions in the virtual environment could be adapted in real-time based on cognitive state. The authors stated that they were successful in discriminating among three levels of workload. In another similar study, Gruenewald et al. (2018) used GSR/EDA, EEG, HR, and Eye Tracking to develop a physiological data acquisition and processing approach to recognize emotions relevant in a learning context: happiness, boredom, and frustration. However, the results showed that the ML models had difficulties in obtaining convincing recognition rates for the least represented classes.

Ergan et al. (2019) provided a method that fused VR and noninvasive body sensor networks to quantify human experience in architectural spaces. Using a set of biometric sensors, physiological metrics such as GSR/EDA, EEG, and HR were captured and examined while subjects were navigating and performing tasks in a VE. The integrated platform was used to quantify the sense of stress and anxiety through structured user experiments in a visualization laboratory using alternate VEs configured by varying the related set of architectural design features. However, capturing eye movement in VR environments can be tricky, since the eye is usually covered by the HMD. However, Moinnereau, Oliveira, and Falk (2020) were able to validate a low-cost "do-it-yourself" solution based on EXG sensors<sup>6</sup>, as opposed to expensive camera-based solutions.

Min et al. (2020) employed GSR/EDA and HR to evaluate how VR could help users regulate blood pressure and stress. The experiment found a significant effect of the content type to the changes in these measures and confirmed that the "calm" content was helpful for one to self-regulate to a lower heart rate and blood pressure, stable GSR/EDA, and that "disturbing" content worked in the opposite way. The authors applied this result to calm down and stabilize the vital signs of patients during actual coronary angiography and catheterization operations, eliciting positive comments from both patients and the operating team.

### ***4.3. QoE/UX studies using physiological data for Mulsemmedia***

#### *4.3.1. 2D Content*

As the field of mulsemmedia is expanding, it was expected to find some studies using physiological data to better understand how mulsemmedia and multisensory interaction affect QoE. Accordingly, Mesfin et al. (2020) used eye tracking devices and HR monitor wristbands to capture users' physiological data whilst they were experiencing mulsemmedia. After each video clip, users were asked to complete an on-screen questionnaire with a set of questions related to smell, sound, and haptic effects targeting their enjoyment and perception of the experiment. Results showed that the eye gaze and HR results showed a significant influence of the cross-modally mapped multi-sensorial effects on the users' QoE and highlighted that when the olfactory content is crossmodally congruent with the visual content, the visual attention of the users seems shifted towards the correspondent visual feature. In a following study of the same author with a similar setup, Mesfin et al. (2021) during their investigation on

---

<sup>6</sup>ExG sensors are used to measure electrical activity between a pair of sensors (differential)

how intensity and valence of introducing odor to multimedia content is perceived by the user, showed that while eye gaze patterns were unaffected by the experimental conditions, valence has a significant impact upon user heart rates.

In related work, Mesfin, Hussain, Covaci, and Ghinea (2019) used eye tracking and HR to evaluate a mulsemmedia environment with different audio setups. The experiment revealed that high/low-pitched audio, associated haptic effect, induced a significantly higher heart rate in experimental group participants compared to those from the control group. Results showed that the average level of agreement of the experimental group of participants to self-reported questions is generally lower than that of the control group. This implies that respondents enjoyed viewing the videos with original audio more than viewing them with high-/low-pitched audio. This again shows that more research needs to be undertaken to better understand the use of cross-modal correspondences in mulsemmedia. Also, the higher heart rates and different eye-gaze patterns of users who experienced cross-modally mapped mulsemmedia allude to increased levels of stress and excitement, which have to be investigated further.

In an unusual use of physiological signals, Greinacher et al. (2020) investigated respiratory rhythm signals to instigate the impact of visual and haptic stimuli in a VR exercise game to assess and improve breath-movement synchronicity (BMS). Results, however, were mixed regarding the ability of visual stimuli to improve BMS, as the authors stated that although visual stimuli were rated higher than haptic stimuli, the users preferred the latter.

#### 4.4. Data Visualization

##### 4.4.1. Media Types in Different Domains

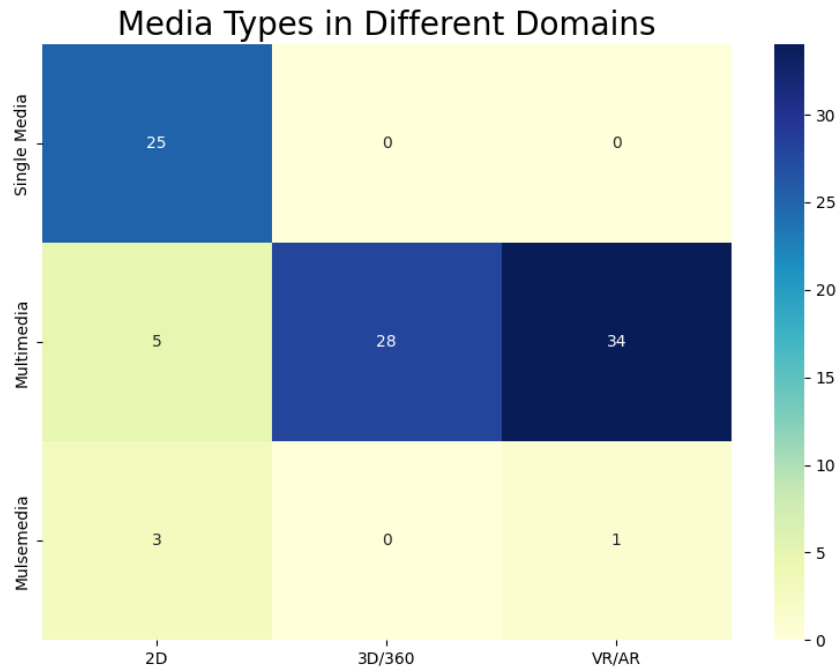


Figure 2.: Distribution of the publications in different domains

This heatmap chart (Figure 2) visualization offers a comprehensive perspective on

the distribution of media types within different domains. Notably, “Single Media” exhibits a dominant presence in the “2D” content category. In contrast, “Multimedia” shows a more balanced distribution across all three categories, with “VR/AR” demonstrating the highest prevalence. This indicates that physiological signals are being increasingly being used in this setup, on par with the standard multimedia content.

#### 4.4.2. Physiological Signals in Different Media Studies

The results presented in the radar chart (Figure 3) showcase the variations in physiological signal measurements across different media types. Multimedia consistently stands out with the highest measurements in parameters such as HR and EEG activity. Single media closely follows multimedia in most measurements. Mulsemedia and Other studies show lower usage of physiological devices overall, with “HR” and “EEG” being the most frequently used in these categories.

Physiological Signals in Different Media Studies (Log Scale)

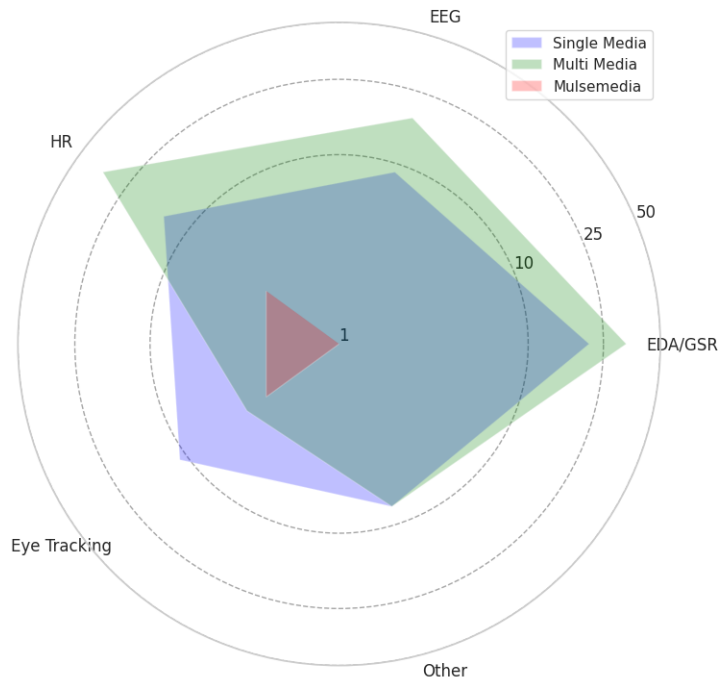


Figure 3.: Distribution of the publications.

It can be observed that “Multimedia” studies use “HR” (Heart Rate) devices significantly more than the other types. “Single Media” studies also make extensive use of “HR” devices but have comparatively higher usage of “EDA/GSR” and “EEG” devices. “Mulsemedia” and “Other” studies show lower usage of physiological devices overall, with “HR” and “EEG” being the most frequently used in these categories.

## 5. Discussion

In this section, we discuss in detail how physiological signals are being collected, processed and analyzed, and towards what end – all, in order to better understand the purpose of using physiological data in UX and QoE studies.

### 5.1. Subjective and Objective Evaluations

With the advent of new interactive technologies, such as VR/AR and mulsemmedia, the best way to evaluate them is also being sought. Existing methods for evaluating new environments include subjective methods, objective evaluation methods, and the combination of both (Che et al., 2019).

Subjective evaluation methods are mainly realized through users' self-assessment when the direct experience of several factors of the product can be evaluated through oral evaluation or through questionnaires. In contrast, objective assessment methods primarily assess the user's physiological signals such as brain waves, HR, eye movements, etc. (Che et al., 2019). The vital difference between these methods is that the former is done under the user's direct control over their responses, while the latter captures implicit biofeedback data to measure how the user is engaging with a product or system.

As stated in Section 4, when employed for user evaluations, physiological data are mostly utilized to measure presence or the "feeling of being there" (Athif et al., 2020; Baka et al., 2018; Dey et al., 2020a; Grassini et al., 2021; Gupta et al., 2020; L. He et al., 2018; Keighrey et al., 2021; Magalie, Sameer, & Philippe, 2018; Tremmel et al., 2019) in VR environments<sup>7</sup>. Presence, as well as other related metrics such as QoE, are commonly measured subjectively through questionnaires, which can suffer from participant biases, dishonest answers, recency effects, and fatigue, making post-experience questionnaire somewhat unreliable (Certický et al., 2019; Dey et al., 2020a, 2020b; Holman & Adebessin, 2019; Hynes et al., 2019; X. Wang & Lin, 2020) and making the "real" UX hidden (Peruzzini et al., 2018). It can also break user immersion and make it difficult for subjects to accurately remember their feelings of presence after leaving the VR experience (Putze et al., 2020) and, under certain conditions, requires a large number of participants and a substantial time to analyze the subjective data (Peruzzini et al., 2018).

These challenges have led to efforts to measure presence using physiological signals using objective metrics such as EEG, HR, and GSR/EDA/EDA. These signals can be recorded continuously, without any conscious effort on the part of the participant, and hold the potential of yielding more objective data than subjective methods that require participants to introspect their experience and obtain descriptions on predefined psychometric scales (Schmidt et al., 2020). For instance, in the study of Dey et al. (2020a), EEG combined with HR and GSR/EDA revealed significant differences in pleasure levels in virtual universes that user questionnaires were not able to show. In the same study, biofeedback was able to predict a higher level of satisfaction, with users having higher HR and higher brain activity in more detailed universes, revealing data that subjective assessment alone was not able to.

As shown in Table 1, there is a growing emphasis on employing physiological data in studies regarding the use of VR/AR, with 32.67% of surveyed papers falling in this category, closely followed at 32.63% by studies in the more traditional single 2D display setup. This is in keeping with the popularization of VR glasses and their adoption in academic environments for purposes other than commercial ones, such as teaching and training. Moreover, with the popularization of VR/AR environments and mulsemmedia<sup>8</sup>, the premise that the near future will see a shift towards the increasing adoption of physiological data to better understand multisensory environments seems

---

<sup>7</sup>There are even proposals of integrated systems to detect users' trust level in AI technologies, and of biofeedback sensors with VR such as EmteqVR (Mavridou et al., 2019).

<sup>8</sup><https://www.ericsson.com/en/6g/internet-of-senses>

reasonable. We now turn our attention to the signals underlying physiological data and how they are collected, measured, processed, and classified.

### 5.2. Models for Emotion Classification

For emotion classification with biofeedback, some studies, such as (Tsai et al., 2021) and (J. Pinto, Fred, & da Silva, 2019), have adopted the *Circumplex Model of Affect* proposed by (Posner, Russell, & Peterson, 2005) and depicted in Figure 4a. Others, such as (H. He et al., 2018) and (Ma, Chen, & Lin, 2018), have utilized the *PAD Emotional State Model*.

The *Circumplex model* describes emotions using the two dimensions of valence and arousal, while the *PAD model* uses three dimensions to represent emotions: pleasure, excitement and dominance. Valence and arousal are present in both models and the former is used to contrast states of pleasure (e.g. happiness) and displeasure (e.g. anger), while the latter is used to contrast states of low arousal (e.g. calm) and high arousal (e.g. excited). However, the *PAD* has the addition of a Dominance-Submission Scale that represents control and domination vs. the controlled or submissive feels. The addition of this dimension can be used to differentiate, for example, fear and anger. While fear and anger are both unpleasant, anger is a dominant emotion; on the other hand, fear is a submissive emotion. Although Circumplex and PAD are the most common models adopted, other proposals were also identified, such as the work of J. Pinto et al. (2019) which focuses on classifying emotions with two valence classes and two arousal classes rather than just a single one like the Circumplex.

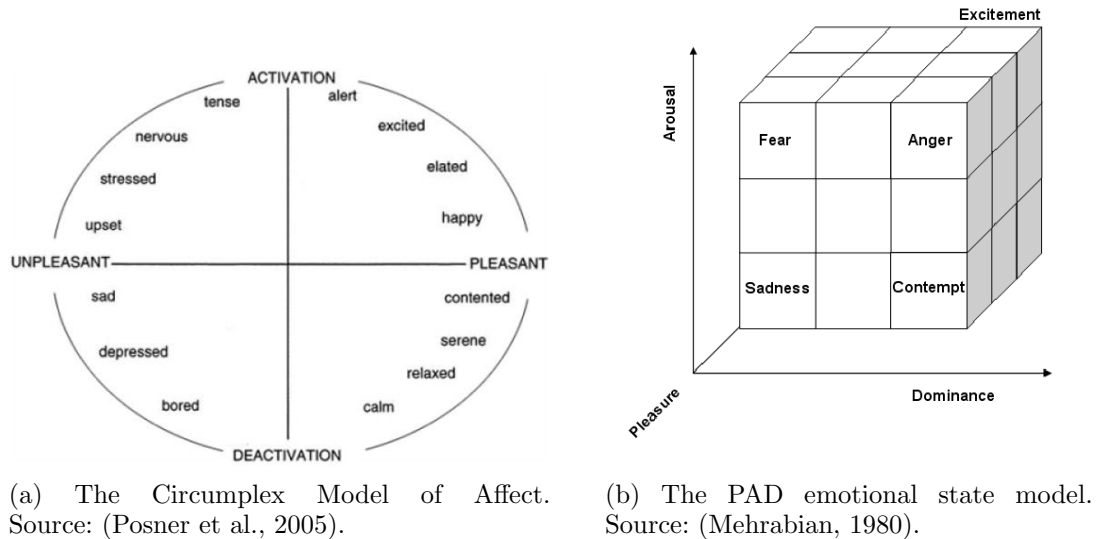


Figure 4.: Emotion classification models most used in the screening process

Table 1.: Where physiological data were collected.

<b>Single Media Studies</b>	
2D	Beşkardeşler and Arslan (2019); Bruun (2018); S. Chen and Epps (2020); Cho et al. (2019); Ciolacu and Svasta (2021); Courtemanche et al. (2018); de Santana and Otani (2021); R. Du et al. (2022); H. He et al. (2018); Holman and Adebessin (2019); Hussain et al. (2018); Isiaka et al. (2021); Kaluarachchi et al. (2021); Katada et al. (2020); Kim et al. (2021); Lee et al. (2020); Li et al. (2021); Liapis et al. (2020, 2021, 2018); Peruzzini et al. (2018); Rahman et al. (2021, 2020); Tateyama et al. (2019); Y. Wang et al. (2019)
3D/360°	–
VR/AR	–
<b>Multimedia Studies</b>	
2D	Ajenaghughrure et al. (2019); Frey et al. (2020); Gurita et al. (2019); Markova et al. (2018); Tsai et al. (2021)
3D/360°	Alves et al. (2018); Certický et al. (2019); Darzi et al. (2021, 2019); Daylamani-Zad et al. (2022); de Santana and Otani (2021); Dhaouadi and Ben Khelifa (2020); G. Du et al. (2020); Elor et al. (2020); Giannakos et al. (2019); L. He et al. (2018); Hossain and Kulshreshth (2019); Houzangbe et al. (2020); Huynh et al. (2018); Klarkowski et al. (2018); Korn and Rees (2019); Lee-Cultura et al. (2020); Maier et al. (2019); Manjunatha et al. (2020); Mijović et al. (2019); Ninaus et al. (2019); Riaz et al. (2021); Sandeep Vara Sankar and Ko (2020); Satti et al. (2021); Schmidt et al. (2020); Wulvik et al. (2020); Yang et al. (2018); Ye et al. (2021)
VR/AR	Al-Ashwal et al. (2021); Alqahtani et al. (2019); Athif et al. (2020); Baka et al. (2018); Barathi et al. (2020); Chauhan et al. (2018); Che et al. (2019); C. Chen et al. (2021); Collins et al. (2019); Dey et al. (2020a, 2020b); Ergan et al. (2019); Grassini et al. (2021); Gupta et al. (2020); Hynes et al. (2019); A. Islam et al. (2019); R. Islam et al. (2021); Kalatzis et al. (2021); Katsigiannis et al. (2019); Keighrey et al. (2021); Martin et al. (2020); Min et al. (2020); Moinnereau et al. (2020); Moraes et al. (2020, 2021); Pei et al. (2021); Petrescu et al. (2020); J. Pinto et al. (2019); M. Pinto et al. (2018); Putze et al. (2020); Souza et al. (2018); Tremmel et al. (2019); Voigt-Antons et al. (2021); Zou and Ergan (2021)
<b>Mulsemedia Studies</b>	
2D	Mesfin et al. (2019, 2020, 2021)
3D/360°	–
VR/AR	Greinacher et al. (2020)

### ***5.3. Collecting and Measuring Physiological Signals in UX and QoE studies***

As Table 2 highlights, GSR/EDA is the most used physiological signal for evaluations, with 61.05% of surveyed papers employing this signal, with HR being the second most common physiological signal used in UX/QoE studies (55.78%), due to this particular signal being easiest to capture and read (Hynes et al., 2019; Klarkowski et al., 2018; Ye et al., 2021). The use of EEG signals, whilst not as frequently employed as GSR/EDA and HR, is garnering increasing attention, being present in 33% of studies, mainly due to the popularization of commercial devices to capture this data (e.g. EMOTIV). Each such signal will now be taken in turn and explored in more detail.

#### *5.3.1. Galvanic Skin Response/ Electrodermal Activity*

GSR/EDA, being the most common physiological data used in UX/QoE studies, can be captured for a multitude of purposes. For instance, it has been used to measure flow in video games (Hynes et al., 2019; Klarkowski et al., 2018; Ye et al., 2021), to improve the accuracy of questionnaires (Moraes et al., 2020), to detect anxiety in VR (Petrescu et al., 2020), as well as cybersickness in the same environment (Al-Ashwal et al., 2021; Martin et al., 2020).

However, GSR/EDA data has limitations that are often overlooked. In the context of emotion research, for instance, GSR/EDA data can offer insights into the intensity of an emotion (arousal), but not whether it is positive or negative (valence). For example, de Santana and Otani (2021) created a low-cost GSR/EDA sensor that can be applied in various contexts, such as healthcare (like in patient monitoring or stroke detection), UX evaluation (in frustration detection or adaptive systems), content recovery, security (in the detection of incidents) or even in autonomous cars (in the management of interruptions). However, as the authors themselves stated, a crucial aspect is how to differentiate between arousal and stressful events. This aspect depends on the context and an HCI specialist was the stakeholder needed to design the experiment and define a means to support data triangulation. The same difficulties in detecting the right emotion were identified by de Santana and Otani (2021) and Satti et al. (2021). Satti et al. (2021) stated that stress classification with GSR/EDA alone was also difficult and that a classification accuracy of more than 63.3% in right stress detection could not be achieved. Similar results were obtained by Collins et al. (2019), with an average accuracy of only 48.36% for cognitive load with GSR/EDA. One of the identified reasons for the low accuracy was that the delayed nature of GSR/EDA makes it difficult to isolate emotional responses against frequent cognitive events (Collins et al., 2019). Due to this, the GSR/EDA takes longer to stabilize, making it difficult to make the correct decision for classification algorithms.

In order to overcome the lack of reliability of GSR/EDA to correctly differentiate stress from eustress, Tateyama et al. (2019) proposed a new method, but the results were mixed. However, Markova et al. (2018) claimed they were able to continuously monitor and identify negative emotions, high levels of emotional arousal, and high arousal-negative valence states from a combination of ECG and EDA with accuracy for a single stimulus recording between 75% and 85%. Another attempt to overcome this obstacle is the work of Tateyama et al. (2019), who developed a system capable of distinguishing the specific form of stress (distress vs. eustress) and were able to identify that the response during a state of distress leads to specific skin conductance responses that are not the same as psychological eustress (beneficial stress). Solutions

like these could overcome the weaknesses of using GSR/EDA for emotion recognition.

Table 2.: Physiological signals used in the screened articles.

<b>Physiological Signals in Single Media Studies</b>	
GSR/EDA	Beşkardeşler and Arslan (2019); Bruun (2018); Courtemanche et al. (2018); de Santana and Otani (2021); Giannakos et al. (2019); Huynh et al. (2018); Isiaka et al. (2021); Katada et al. (2020); Klarkowski et al. (2018); Lee et al. (2020); Lee-Cultura et al. (2020); Liapis et al. (2021, 2018); Peruzzini et al. (2018); Rahman et al. (2021, 2020); Satti et al. (2021); Tateyama et al. (2019); Wulvik et al. (2020); Yang et al. (2018)
EEG	Courtemanche et al. (2018); R. Du et al. (2022); Giannakos et al. (2019); Holman and Adebessin (2019); Hossain and Kulshreshth (2019); Hussain et al. (2018); Li et al. (2021); Manjunatha et al. (2020)
HR	Beşkardeşler and Arslan (2019); Ciolacu and Svasta (2021); Giannakos et al. (2019); Hossain and Kulshreshth (2019); Huynh et al. (2018); Katada et al. (2020); Lee et al. (2020); Lee-Cultura et al. (2020); Liapis et al. (2021, 2018); Peruzzini et al. (2018); Rahman et al. (2021); Y. Wang et al. (2019); Wulvik et al. (2020); Yang et al. (2018)
Eye Tracking	S. Chen and Epps (2020); Cho et al. (2019); Courtemanche et al. (2018); Giannakos et al. (2019); H. He et al. (2018); Hussain et al. (2018); Isiaka et al. (2021); Kaluarachchi et al. (2021); Manjunatha et al. (2020); Peruzzini et al. (2018); Rahman et al. (2021)
Other	Ciolacu and Svasta (2021); Hussain et al. (2018); Isiaka et al. (2021); Lee-Cultura et al. (2020); Liapis et al. (2018); Peruzzini et al. (2018); Rahman et al. (2021); Yang et al. (2018)
<b>Physiological Signals in Multimedia Studies</b>	
GSR/EDA	Al-Ashwal et al. (2021); Alqahtani et al. (2019); Athif et al. (2020); Barathi et al. (2020); Certický et al. (2019); Collins et al. (2019); Darzi et al. (2019); Daylamani-Zad et al. (2022); Dey et al. (2020a, 2020b); Dhaouadi and Ben Khelifa (2020); Elor et al. (2020); Ergan et al. (2019); Gruenewald et al. (2018); Gupta et al. (2020); Hynes et al. (2019); A. Islam et al. (2019); R. Islam et al. (2021); Katsigiannis et al. (2019); Keighrey et al. (2021); Maier et al. (2019); Markova et al. (2018); Martin et al. (2020); Moraes et al. (2021); Pei et al. (2021); Petrescu et al. (2020); M. Pinto et al. (2018); Putze et al. (2020); Schmidt et al. (2020); Souza et al. (2018); Tsai et al. (2021); Ye et al. (2021); Zou and Ergan (2021)
EEG	Alves et al. (2018); Athif et al. (2020); Baka et al. (2018); Certický et al. (2019); Darzi et al. (2019); Dey et al. (2020a, 2020b); Elor et al. (2020); Ergan et al. (2019); Grassini et al. (2021); Gruenewald et al. (2018); Gupta et al. (2020); L. He et al. (2018); Petrescu et al. (2020); Sandeep Vara Sankar and Ko (2020); Tremmel et al. (2019); Tsai et al. (2021); Zou and Ergan (2021)

Continued on next page

Table 2 – continued from previous page

<b>Physiological Signals in Multimedia Studies</b>	
HR	Alqahtani et al. (2019); Alves et al. (2018); Athif et al. (2020); Barathi et al. (2020); Certický et al. (2019); Chauhan et al. (2018); Che et al. (2019); C. Chen et al. (2021); Collins et al. (2019); Darzi et al. (2019); Daylamani-Zad et al. (2022); Dey et al. (2020a, 2020b); Dhaouadi and Ben Khelifa (2020); Elor et al. (2020); Ergan et al. (2019); Gruenewald et al. (2018); Gupta et al. (2020); R. Islam et al. (2021); Kalatzis et al. (2021); Katsigiannis et al. (2019); Keighrey et al. (2021); Maier et al. (2019); Markova et al. (2018); Martin et al. (2020); Moraes et al. (2021); Pei et al. (2021); Petrescu et al. (2020); Putze et al. (2020); Schmidt et al. (2020); Souza et al. (2018); Tsai et al. (2021); Voigt-Antons et al. (2021); Ye et al. (2021); Zou and Ergan (2021)
Eye Tracking	Darzi et al. (2019); Gruenewald et al. (2018); Moinnereau et al. (2020); Moraes et al. (2020); Tsai et al. (2021)
Other	Certický et al. (2019); Darzi et al. (2019); Dhaouadi and Ben Khelifa (2020); G. Du et al. (2020); Maier et al. (2019); Moraes et al. (2021); Putze et al. (2020); Souza et al. (2018)
<b>Physiological Signals in Mulsemedia Studies</b>	
GSR/EDA	–
EEG	–
HR	Mesfin et al. (2019, 2020, 2021)
Eye Tracking	Mesfin et al. (2019, 2020, 2021)
Other	Greinacher et al. (2020)

On top of the lack of information about valence and despite its simplicity, recording and analyzing GSR/EDA signals may also be delicate and complicated, involving knowledge from psychology, physiology, and signal processing. The underlying psycho-physiological mechanism of GSR/EDA is complex and not well-understood, leaving various factors unnoticed that potentially affect the signal (Babaei et al., 2021). Houzangbe, Christmann, Gorisse, and Richir (2019) in their multimodal experiment also failed to detect cognitive load with GSR/EDA, although, with post-processing and using Fast Fourier Transform (FFT), they were able to detect statistical significance between low and high cognitive load. The authors suggest that the amplitude of the GSR/EDA signal may reflect a person’s level of cognitive load, therefore, it could be used as a key resource for ML algorithms in detecting cognitive load. Alves et al. (2018) initially used GSR/EDA (and HR) for their evaluation but were unable to effectively differentiate between mental states. Moreover, GSR/EDA produced too many missing values in this research and was subsequently discarded. Zou and Ergan (2021) and Dey et al. (2020b) endorse these claims: when the GSR/EDA data were analyzed, the generic signal characteristics for these sensors were not effective in determining user experiences or presence.

GSR/EDA is also subject to changes in ambient temperature. For instance, hot climates stimulate sweating and can alter the skin’s conductivity index while cold environments can make measurements difficult. Conclusions like these were presented in

the work of Suni-Lopez, Condori-Fernandez, and Catala (2020) who stated that, although they could use GSR/EDA data for user evaluation, authors also recommended that for the identification of emotional triggers, a context analyzer capable of processing environmental and physiological data is needed to avoid data corruption. This was corroborated in another experiment conducted by Georges, Courtemanche, Fredette, and Doyon-Poulin (2020) where it was noted that, as GSR/EDA is very sensitive to the environment, it may not be able to capture accurate signals outside of a controlled laboratory environment.

Table 3 compiles a comprehensive list of GSR/EDA/EDA collector devices, revealing Shimmer GSR/EDA and E4 Empatica as the foremost choices, featured in 11 and 7 studies, respectively. Yet, the table also underscores the diversity in available GSR/EDA/EDA options, encompassing devices like Bitalino with pluggable GSR/EDA/EDA, Neulog GSR/EDA sensor, and Grove - GSR/EDA Sensor, among others.

While commercial devices hold sway in usage, it is noteworthy that open-source and DIY solutions, such as DIY GSR/EDA, maintain a presence, signaling the inclination of some researchers towards tailor-made solutions that align precisely with their distinctive research requisites.

### *5.3.2. Heart Related Physiological Data*

Alongside GSR/EDA, HR is one of the most popular physiological signals captured in UX/QoE experiments. Being able to be collected by commercial wearable devices, they are usually cheap and easy to read, as shown in the work of Santamaria-Granados, Mendoza-Moreno, Chantre-Astaiza, Munoz-Organero, and Ramirez-Gonzalez (2021) who, using only a commercial wearable (Xiaomi Mi Band), proposed an architecture capable of recommending tourist attractions based on the emotional states of users in different contexts. Despite being only 44% accurate in detecting the user’s emotional state so the system could recommend activities, it is worth noting that the study used a low-precision commercial wearable sensor. In addition, the study was carried out in the “real world”, outside of a controlled laboratory environment, which can hinder results.

HR is usually measured by photoplethysmography, and being a skin sensor is also closely related to GSR/EDA. It is therefore unsurprising that both signals are used together in studies, such as the work of Kim et al. (2021), who proposed a musical classification system through emotion recognition using induced signals of GSR/EDA and HR. The system predicted user emotions by applying data acquired through wearable GSR/EDA and HR sensors. Similar setups were found in the studies of Pei et al. (2021), Y. Wang et al. (2019), and Martin et al. (2020).

Although HR is easy to collect and analyze, HR can only reflect arousal but not valence, that is, one cannot judge the type of emotion underlying the HR. For instance, HR alone was not accurate in measuring emotions in the study of Ergan et al. (2019), when the results were combined with other signals, they helped to measure stress levels more clearly. Zou and Ergan (2021) also claimed that only the frequency domain features from the HR sensor were useful in measuring user experience. However, this question may be resolved using subjective questionnaires for supplementary research (Pei et al., 2021). In a similar study, Chauhan et al. (2018) proposed a prototype that employs an ear-clip HR sensor, an Arduino microcontroller, and a supervised learning algorithm. To acquire a training data set, authors ran stress induction experiments to track HRV metrics and Discrete Wavelet Transform (DWT) coefficients and then

<b>GSR/EDA Device</b>	
Bitalino with pluggable GSR/EDA	Beşkardeşler and Arslan (2019); J. Pinto et al. (2019); Souza et al. (2018)
Neulog GSR/EDA sensor NUL-217	Elor et al. (2020)
Pip biosensor	Keighrey et al. (2021)
Laxtha PolyG-A	Min et al. (2020)
Mindfield eSense	Ye et al. (2021)
TRIGbox	Schmidt et al. (2020)
Grove - GSR/EDA Sensor	Certický et al. (2019); Daylamani-Zad et al. (2022); Gurita et al. (2019)
biosignalsPlux Explorer	Tateyama et al. (2019)
NeXus-10 physiological platform	Liapis et al. (2020, 2018)
Mindplace Thoughtstream GSR/EDA	Bruun (2018)
Biopac MP-150	Courtemanche et al. (2018); Yang et al. (2018)
Biopac MP-160	Kalatzis et al. (2021)
GSR/EDAsensor2	Darzi et al. (2019)
E4 Empatica	Collins et al. (2019); Giannakos et al. (2019); Huynh et al. (2018); Hynes et al. (2019); A. Islam et al. (2019); R. Islam et al. (2021); Katada et al. (2020); Lee-Cultura et al. (2020); Maier et al. (2019); Moraes et al. (2020); Rahman et al. (2021, 2020)
Shimmer3 GSR/EDA+	Barathi et al. (2020); Dey et al. (2020a, 2020b); Ergan et al. (2019); Gupta et al. (2020); Korn and Rees (2019); Markova et al. (2018); Martin et al. (2020); Petrescu et al. (2020); Satti et al. (2021); Tsai et al. (2021); Wulvik et al. (2020)
Biopac EDA amplifier	Klarkowski et al. (2018)
Equivital EQ02	Al-Ashwal et al. (2021)
Nexus10	Liapis et al. (2021); M. Pinto et al. (2018); Putze et al. (2020)
DIY GSR/EDA	Athif et al. (2020); de Santana and Otani (2021); Katsigiannis et al. (2019)

Table 3.: GSR/EDA devices

trained an AdaBoost ensemble classifier to 93% 4-fold cross-validation accuracy and 93% precision.

Much in the same way that GSR/EDA was used in the work of Moraes et al. (2020), studies have used HR to improve results obtained via questionnaires from an application that monitors HR and, through it, measures the level of fun, as authors hypothesized that heart rate variability could be due to joy (Che et al., 2019; Zou & Ergan, 2021). HR was also used to improve the accuracy of EEG signals (Alves et al., 2018), with authors stating that results showed that the measure that better complements the EEG is the HR.

The work by Peruzzini et al. (2018) presents a methodology and proposes a configuration to monitor performance and evaluate the UX in a truck movement scenario. In particular, the proposed configuration integrated different methods for acquiring physical and physiological parameters that measure HR. However, the authors did not explicitly state how the objective of the experiment was achieved, although they did claim that the proposed setup was capable of capturing and measuring mental states. In another HR-only experiment, Alqahtani et al. (2019) were able to accurately detect with a 61.22% accuracy the level of difficulty in a learning platform, implying that this biosensor alone can be used for emotion recognition and dynamic learning.

However, results with HR are mixed: in a two-scenario experiment based on ship navigation, Wulvik et al. (2020) investigated workload with HR and GSR/EDA in a simulated ship control room environment. The authors observed an increase in arousal, displeasure, stress, and workload in the scenario where participants navigated a large ship in a narrow harbor compared to the scenario where they navigated on an open sea. Thus, the authors stated that GSR/EDA was able to register clearer results than HR. However, in another study (Dey et al., 2020a) with the same biofeedback setup, only the HR was able to provide significant results in a QoE experiment. This study thus claims that HR can be useful for future studies on mental immersion, complementing more traditional subjective measures in a multi-method evaluation approach, while GSR/EDA should be studied further to be conclusive for future research. The previously cited work in subsection 5.3.1 by Collins et al. (2019) also found that, while GSR/EDA was unable to predict user stress due to its delayed nature, HR was a better predictor because it does not suffer from delays and is one of the fastest responses to the sympathetic nervous system.

Table 4 presents all the identified HR collector devices. Shimmer devices, such as the Shimmer3 ECG device and the SHIMMER v2.0 wireless ECG sensor, are the most commonly used HR sensors. Although Shimmer devices dominate, there is a noticeable trend in using wristbands for heart rate measurement. Devices like Mio Link, Fitbit, and Polar, such as Polar H10 and Polar OH1 Sensor, showcase the popularity of wearable wristband-based solutions for HR data collection.

In summary, while Shimmer devices are widely used for HR data collection, the choice of HR collector devices varies, with a growing trend in wristband-based solutions and a wide range of options available to researchers.

### 5.3.3. *Electroencephalography Signals*

EEG, previously used exclusively for medical purposes, is gaining popularity due to new devices being introduced to the market and commercial use. General results show that EEG sensors are more effective when compared to GSR/EDA and HR (Zou & Ergan, 2021) as EEG alone is able to reach high level of accuracy in emotion recognition (Elor et al., 2020). Using EEG alone, Riaz et al. (2021) were able to classify emotions

<b>HR Device</b>	
Bitalino with pluggable ECG	Alves et al. (2018); Beşkardeşler and Arslan (2019); Souza et al. (2018)
Zephyr Bioharness 3	Certický et al. (2019); Peruzzini et al. (2018)
TRIGbox	Schmidt et al. (2020)
Grove - Finger-clip Heart Rate Sensor	Daylamani-Zad et al. (2022)
Vivoactive watch	Hossain and Kulshreshth (2019)
Reflectance PPG Sensor	Chauhan et al. (2018)
NeXus-10	Liapis et al. (2020, 2018); Putze et al. (2020)
Shimmer3 ECG device	Dey et al. (2020a, 2020b); Ergan et al. (2019); Gupta et al. (2020); Markova et al. (2018); Martin et al. (2020); Petrescu et al. (2020); Wulvik et al. (2020)
SHIMMER v2.0 wireless ECG sensor	Alqahtani et al. (2019)
Mindfield eSense	Ye et al. (2021)
g.USBamp	Darzi et al. (2021, 2019)
Polar H10	Dhaouadi and Ben Khelifa (2020); Voigt-Antons et al. (2021)
Polar OH1 Sensor	Elor et al. (2020)
Mio Link	Houzangbe et al. (2020); Mesfin et al. (2019, 2020, 2021)
Fitbit Surge	Certický et al. (2019)
Fitbit PurePulse	Y. Wang et al. (2019)
PSL-iECG2	Min et al. (2020)
Omron M7 Intelli IT	Min et al. (2020)
Fitbit Charge HR	Keighrey et al. (2021)
DIY HR collector	Athif et al. (2020)

Table 4.: HR devices

with an accuracy of 80.55% for arousal and 70.37% for valence. Holman and Adebessin (2019), using a commercial product (EMOTIV), were able to collect EEG data from participants while reading online news articles. Fluctuations in the different channels of participants' EEG data allowed the authors to confirm that participants experienced negative emotions when reading news articles online. Also, due to the nature of online news, nearly all ten participants experienced the emotion of anger. Using the same setup, the work of Hossain and Kulshreshth (2019) showed that participants had lower levels of stress and higher levels of engagement and arousal in the stereoscopic 3D group.

Reading EEG, however, is not a trivial task: as mentioned before, EEG capture is difficult not only because of the complexity of the human brain and its susceptibility to noise but also in respect of interpreting its data Salgado et al. (2018). There was no standard found in the papers here presented, however, some general guidelines were identified in the works of Elor et al. (2020), Li et al. (2021), and Sandeep Vara Sankar and Ko (2020). In general, EEG data collection begins with the setup of an EEG capture device. In clinical experiments, according to Wan et al. (2019), 10-20 International electrode standards are followed to position the electrodes from nasion toinion and other measurements from one ear to the end of the other ear, usually from pre auricular or mastoid, for all different subjects. However, with the advent of commercial EEG capture devices, this task is becoming more trivial. After setup capture data, they are pre-processed and used as presented in the section 5.4.

Five bands can be captured by EEG: alpha, beta, delta, theta, and gamma (Elor et al., 2020). Elor et al. (2020) stated that the alpha band occurs at frequencies between 7 and 12 Hz and is usually associated with stress-related neural activity and relaxation, although Baka et al. (2018) stated that alpha frequencies are also a source of human emotions. The beta band occurs at frequencies between 12 and 30 Hz and is usually associated with focus as well as active cognition such as arousal, anxiety, and concentration. Increases in beta waves were also correlated with active, busy, or anxious thinking and concentration (Baka et al., 2018). The delta band occurs at frequencies between 0.5 and 4 Hz and is correlated with consciousness and sleep. Delta waves have been found to have more activity during deep sleep, where the deeper the sleep, the greater the activity. The theta band occurs at frequencies between 4 and 7 Hz and is associated with motor processing, although some works imply that they are also related to fatigue (Sandeep Vara Sankar & Ko, 2020) and human emotions (Li et al., 2021). This includes spikes in theta activity for planning motor behavior, path spatialization, memory, and learning. The gamma band occurs at frequencies between 30 and 100 Hz and has been correlated with thought, awareness and meditation. Also, according to Elor et al. (2020), there are conflicting opinions on whether gamma bands are reliable due to biological artifacts such as eye movement and jaw clenching.

Normally EEG signals for emotion recognition are captured in frontal channels. However, Zou and Ergan (2021) states that looking only at data collected in the frontal area may not be as revealing as data collected in the back of the brain. Thus, to measure the human experience in designed controlled spaces, one must examine the signal from the rear and front areas. Furthermore, to monitor human vigilance and stress levels, frontal asymmetries were observed as strong indicators of stress level separation.

In addition to emotions, Manjunatha et al. (2020) used EEG to reveal different tactics in users playing a *real-time strategy* game, where each user performed a search and rescue mission in the presence of opposing teams. In this scenario, three sets of tactics were revealed by EEG measurements: offensive, cautious, and target discovery.

This is unique compared to the studies mentioned so far, as the authors aimed to find playing styles with EEG rather than looking for emotions and imply a potential of EEG to capture patterns from different scenarios.

Even though increasingly popular, the use of EEG together with VR displays may be problematic due to interference between the systems. Although some studies claim that the interference can be minimal and EEG data quality is not affected (Grassini et al., 2021), there are still challenges such as costs, accuracy of sensors, data transfer errors or inconsistency, and ease of use for devices (Elor et al., 2020; Ye et al., 2021). Giannakos et al. (2019) also recognized that the cost of collecting EEG data may shy away authors from using it, as well as adopting it as an everyday tool outside the lab environment.

Furthermore, besides requiring an exhaustive setup of specialists to acquire good quality signals (Ergan et al., 2019; Salgado et al., 2018), EEG suffers from variability between subjects (Roy et al., 2019). According to Roy et al. (2019), this is due to physiological differences between individuals, which severely affect the performance of models that are meant to generalize across subjects, preventing the transferability of model parameters between sessions/subjects. This means that, even after a model is trained with physiological signals collected from a user, it could lead to errors and inaccuracies when used to classify another batch of physiological data because of the subjective differences in physiological structure and psychological states between users, leading to signals from different sessions/subjects on the same task show different features and distribution.

Table 5 presents all the identified EEG collector devices. Similar to the trends observed with GSR/EDA and HR devices, there is a notable shift towards using commercial EEG devices. This suggests that researchers are increasingly relying on well-established commercial EEG solutions like EMOTIV for their data collection needs.

Notwithstanding the prevalence of Emotiv, Table 5 also highlights the diversity in EEG device selection, including options like g.GAMMAcap2 EEG, OpenBCI EEG cap, SMARTIN EEG, and others.

#### 5.3.4. *Electrooculography and Eye Tracking*

Similar to EMG, EOG may not be directly linked to the ANS, but it can still serve as a valuable tool for assessments, such as measuring pupil dilation or saccadic eye movement. In this regard, researchers such as S. Chen and Epps (2020) and Kaluarachchi et al. (2021) employed cameras to capture eye tracking data, speech, and head movements to gauge cognitive load and establish associations between mental workload and task-related events. Stress was also the focus of investigation in the work of Isiaka et al. (2021), who utilized eye tracking and GSR/EDA to measure stress and mental overload during web navigation sessions.

Apart from mental workload measurement, eye tracking has also been used towards the recognition of human emotions, such as in the work of Gruenewald et al. (2018) and H. He et al. (2018) who used eye tracking among other sensors to feed an ML algorithm for recognition of human emotions. Cho et al. (2019), in a study related to the work of Courtemanche et al. (2018) and Isiaka et al. (2021), used eye tracking in conjunction with GSR/EDA and EEG to create heatmaps of human emotions in order to create a tool that answered "*Where in the picture do people tend to feel something?*" or "*Where in the interface do people tend to experience the most cognitive load?*" This study, however, in contrast with the conclusion of Zou and Ergan (2021) and Elor et al. (2020), showed no correlation between EEG heatmaps and arousal scores, and

<b>EEG Device</b>	
Emotiv Epoc+ EEG headset	Dey et al. (2020a, 2020b); R. Du et al. (2022); Ergan et al. (2019); L. He et al. (2018); Holman and Adebesin (2019); Hos-sain and Kulshreshth (2019)
g.GAMMAcap2 EEG	Athif et al. (2020)
OpenBCI EEG cap	Gupta et al. (2020)
SMARTIN EEG	Mijović et al. (2019)
LADYBIRD, Guger Technologies	Tremmel et al. (2019)
Looxid Link	Tsai et al. (2021)
BIOSEMI Active Two	Baka et al. (2018)
InteraXon Muse 2—Brain Sensing Head-band	Elor et al. (2020)
ANTNeuroEEG	Grassini et al. (2021)
G-Tech EGG	Ajenaghughrure et al. (2019)
Neurosky Mindwave Mobile	Certický et al. (2019)
Electrical Geodesics	Courtemanche et al. (2018)
Brain Productis BrainAmps	Li et al. (2021)
BAlert X24 EEG headset	Manjunatha et al. (2020)
ENOBIO 20 EEG	Giannakos et al. (2019)
32-channel Neuroscan	Sandeep Vara Sankar and Ko (2020)

Table 5.: EEG collection devices

low correlation of GSR/EDA heatmaps while pupillary dilation remained significantly correlated with arousal scores. Nonetheless, Cho et al. (2019) also remarked that the heatmaps presented were created using only one physiological signal at a time, implying that using multiple signals at once may achieve better results.

As stated in section 4, capturing eye movement in VR can be fraught with obstacles. However, one success story was found: Moinnereau et al. (2020) used ExG nodes to measure saccadic eye movement and, with the help of ML and SVM, were able to recognize saccadic eye movement with an accuracy of 76%.

In the study of Giannakos et al. (2019), the authors claimed that eye tracking was especially good for the task of predicting learning in educational games, although they stated that merging multiple sources of signals had the potential to further increase the accuracy of this prediction. Similar to the findings of Cho et al. (2019). Peruzzini et al. (2018) also made use of eye tracking to simulate and learn truck driving. However, the authors stated that was unclear which use of biofeedback was most relevant.

For spatial sound localization, the studies of Moraes et al. (2020, 2021) also used eye tracking and GSR/EDA to assess the role of physiological signals in a VR-based sound localization task. The pupillary response is correlated with the level of cognitive load, visual attention and memory. Therefore, increasing values suggest a higher level of cognitive load for the task, indicating that the task is more challenging for the user. Although the GSR/EDA did not show differences between the groups studied, the authors suggested that the increasing values of the GSR/EDA and pupillary response over time suggested that the task was challenging and mentally demanding.

---

<b>Eye Tracking</b>	
Tobii X-60 (Tobii Technology AB)	Courtemanche et al. (2018)
Tobii Pro Glasses 2	Peruzzini et al. (2018)
Tobii 1750 eye tracker	Isiaka et al. (2021)
Tobii Pro Nano	Manjunatha et al. (2020)
Tobii X3-120	Giannakos et al. (2019)
Tobii X2-30	Cho et al. (2019)
HMD with Tobii	Al-Ashwal et al. (2021); Moraes et al. (2021)
Kinect2.0	G. Du et al. (2020)
Gazepoint GP3	Darzi et al. (2021, 2019)
Eye Tribe	Mesfin et al. (2019, 2020, 2021); Rahman et al. (2021)
Emotient FACET	Korn and Rees (2019)
DIY Camera or Device	Barathi et al. (2020); S. Chen and Epps (2020); Kaluarachchi et al. (2021); Moinnereau et al. (2020)

Table 6.: Eye Tracking and Face Recorders

Table 6 presents all the identified Eye Tracker and Face collector devices. Tobii commercial products, such as Tobii X-60, Tobii Pro Glasses 2, Tobii 1750 eye tracker, Tobii Pro Nano, Tobii X3-120, and Tobii X2-30, were the most commonly used eye trackers in the studies. Some DIY devices were also identified, mainly intended for use alongside Head-Mounted Displays (HMDs). This indicates a need for customized

solutions in situations where off-the-shelf products may not fully meet the research requirements. Furthermore, the presence of DIY cameras or device solutions suggests that researchers are exploring open-source or cost-effective alternatives to tailor their setups to specific research needs while minimizing costs.

### 5.3.5. Electromyography Signals and Muscle Activity

As remarked in section 2.5, EMG signals are not usually the subject of the ANS, and thus cannot be used as an involuntary and objective physiological signal. Even so, some muscle signals captured by electromyography can still be considered involuntary: in some studies, emotion recognition and user evaluation were measured from involuntary facial muscle movements (Dhaouadi & Ben Khelifa, 2020). For example, facial EMG is being used to measure emotional states (eg, arousal and valence) during gaming for positive improvement. However, facial EMG requires an adequate laboratory environment and technical knowledge to deal with artifacts, generating intrusion and intimacy problems.

Eye tracking is also an enabler of observational methods for capturing the behavior of users by capturing their faces, like the work of Hussain et al. (2018), which used cameras to detect facial expressions and eye movement, alongside other physiological signals. In a similar study, Yang et al. (2018) used physiological signals such as facial monitoring (alongside HR, GSR/EDA and breathing rhythm) to explore the gaming experience in video games. While the authors did not explicitly state which biofeedback was most useful, facial coding was responsive to game-play features, showing greater variance than HR.

Table 7 presents all the identified Electromyography Signals and Muscle Activity collector devices.

Devices		
Kinect Skeleton	Lee-Cultura et al. (2020)	A body tracking device that can capture body movement
Actigraph wGT3X-BT	Dhaouadi and Ben Khelifa (2020)	Sleep/Awake body Accelerometer
NoitomTM motion capture sensor	L. He et al. (2018)	Accelerometer for head movement capture
SweetZpot FLOW	Greinacher et al. (2020)	A device for capturing Breathing Rhythm

Table 7.: Devices for Miscellaneous Physiological Signal Capture

### 5.3.6. Other Physiological Data and Devices

Besides the aforementioned physiological signals, other physiological information that can be used for user evaluations, such as respiratory rhythm and temperature, was also identified. In this subsection, we present devices that are too specific to be categorized individually. As mentioned in subsection 5.3.2, Kalatzis et al. (2021) were able to classify emotion with ECG and respiratory rhythm, with the latter revealing being the main component to demonstrate emotion. Greinacher et al. (2020) used respiratory rhythm to develop respiratory synchrony. Breathing rhythm is easy to measure and furnishes data that provides valuable insights into correct movement execution,

especially when analyzed in conjunction with additional real-time movement data. The authors compared visual breathing indication with haptic guidance for athletes to maintain correct, efficient, and healthy Breath-Motion Synchronicity (BMS) during indoor simulation. The results show a positive significant impact of purely verbal instructions and purely tactile feedback on the BMS and no significant impact of visual feedback.

Temperature has also been used to recognize human emotions in the studies of Barathi et al. (2020); Isiaka et al. (2021); Liapis et al. (2021); Maier et al. (2019); Moraes et al. (2021) and Rahman et al. (2021). This means that, besides the more traditional GSR/EDA and HR, other physiological signals can be used for evaluations, although how they are associated with emotions and stress measurement is not always clear (Babaei et al., 2021; Peruzzini et al., 2018).

#### ***5.4. Data Processing: Machine Learning and other methods***

Regardless of the signal collected, the process of capturing physiological data typically involves five steps (Cao, 2020): (a) setting up the data capture device; (b) collecting physiological data; (c) preprocessing the data to eliminate noise and artifacts; (d) generating samples; and finally, (e) identifying patterns. This can be accomplished in various ways, such as classifying emotions, fatigue, and attention, which can be achieved through supervised or unsupervised ML methods (Cao, 2020; Haratian & Timotijevic, 2018; Rahman et al., 2021, 2020; Sandeep Vara Sankar & Ko, 2020).

The preprocessing methods vary among studies. To reduce noise in EEG data, Darzi et al. (2019) applied a band-pass filter with a range of 2–60 Hz, whereas Baka et al. (2018) and Ergan et al. (2019) used a range of 1-50 Hz. Typically, the signal is segmented into time series data (Baka et al., 2018), and Fast Fourier Transforms (FFT) (Ergan et al., 2019; Ye et al., 2021) were applied to each segment for each electrode, allowing for analysis in the frequency domain. The size of each segment varies across studies, ranging from 1 second (Ergan et al., 2019) to 120 seconds (Baka et al., 2018; Martin et al., 2020). Additionally, Max-Min normalization was employed in the work of Rahman et al. (2021) to ensure that all EDA, HR, and other values fell within the range of 0 to 1. This work also introduced a new visualization technique known as Gingerbread Animation, which employs a 2D representation of a human body to visually depict the time series of physiological signals propagating on that 2D surface.

In a study focused on detecting human emotions, G. Du et al. (2020) employed a bidirectional long-term and short-term memory network to learn HR classifications, while a convolutional neural network (CNN) was utilized to learn facial expressions. Experimental results demonstrated that this model achieved high accuracy and low computation time in identifying four emotions. Similar outcomes were reported by Darzi et al. (2019) and Sandeep Vara Sankar and Ko (2020), who used a configuration combining EEG, HR, Eye Tracker, and GSR/EDA with subjective measures. They were able to accurately classify various psychological dimensions with a remarkable 97.6% accuracy, employing regression analysis and a Support Vector Machine (SVM). The authors concluded that, in comparison to other classification methods, physiological measures provided more informative results than performance and personality data. They emphasized that despite the relatively high cost and measurement challenges, physiological data should be collected when the goal is to classify psychological dimensions. However, not every application successfully recognized human emotions

with ML-based models; Gruenewald et al. (2018) failed to do so and hypothesized that it was due to the small size of their dataset and the substantial data requirements for properly training deep learning models.

In addition to the detection of user emotion, ML is also being used to detect and improve user well-being from physiological signals. To this end, Martin et al. (2020) used ML fed with EEG data to detect cybersickness in VR, achieving a high accuracy for this purpose. Similar results were obtained by Martin et al. (2020). For this, three ML algorithms were evaluated: SVM, Gradient Boosting and Random Forest. These were able to predict the intensity of cybersickness and showed an accuracy of up to 75% in a regression approach and an accuracy of up to 91% in a classification approach. Furthermore, with trained models, they were able to create real-time recognition.

Several classification models were also constructed by Ye et al. (2021) and Maier et al. (2019) to detect "flow" in video games. Using multimodal data, Dhaouadi and Ben Khelifa (2020) presented another related study aiming at multimodal stress recognition. Accordingly, by employing ECG, EMG and GSR/EDA to feed a deep learning network and a Long Short-Term Memory (LSTM) model, the authors were able to predict emotions with 95% accuracy using LSTM, while DNN obtained the best accuracy rate of 65%.

Having discussed how biophysical signals are analyzed and classified, the next section will discuss how they are collected and measured.

## 6. Challenges

Most authors agree that monitoring and collecting physiological data is not a trivial task. Problems regarding sensor sensitivity, external environments, difficulty in capturing, and the presence of noise are all noted in the literature. Obstacles should be highlighted: latency and synchronization issues. This is even greater when it comes to multimodal collection, as different physiological data have different durations and latency for a given stimulus. For instance, HR measures change more rapidly than GSR/EDA values, but more slowly than pupil size (Courtemanche et al., 2018).

### *6.1. Standardized Protocols for Biofeedback-enhanced Evaluations*

Concerns also revolve around the absence of standardized protocols for conducting Biofeedback-enhanced evaluations. While we have laid the groundwork with a fundamental framework for such assessments, as detailed in the five-step approach outlined in Section 5.4, there has been limited effort invested in establishing a universally recognized procedure. In response to this critical issue, we have conducted an in-depth exploration, documented in a recently published article (A. Silveira et al., 2023). These guidelines have been crafted to ensure the continual engagement and motivation of users throughout biofeedback sessions, especially in long-term applications. This is particularly crucial because user discomfort and resistance can pose significant challenges, particularly in situations involving high-cost devices, invasive sensors, or scenarios where users may feel self-conscious about the data collection process. It is imperative to explore strategies that enhance user acceptance and comfort, as these factors play a pivotal role in achieving the desired outcomes. Although initial guidelines were proposed in this article, ongoing research is aimed at refining and enhancing their practical implementation.

### ***6.2. Challenges in Biofeedback Data Collection***

Furthermore, depending on the setup, biofeedback demands not only high-cost devices but also necessitates meticulous experiment preparation, an uncomfortable experience for users (Ye et al., 2021), and labor-intensive post-processing of data. Another obstacle to consider is that, when sensors are employed in isolation, certain forms of biofeedback may not yield the anticipated results. For instance, in a study conducted by Ergan et al. (2019), various physiological data were utilized to measure stress levels in a variety of VR architectural environments. While HR data alone did not achieve the desired outcomes, the inclusion of GSR/EDA and EEG data proved successful for the authors. A similar conclusion was drawn by Lee-Cultura et al. (2020), who found that combining multimodal features derived from eye tracking and other physiological data provided the most accurate predictions in a learning context.

### ***6.3. Variability in Physiological Measurements***

Not all physiological measurements are universally applicable. In a study by Tateyama et al. (2019), it was observed that GSR/EDA failed to distinguish between various states, such as stress and eustress. Additionally, due to its gradual and slow variation (typically changing within 1.5–6.5 seconds after the stimulus, as noted by Ye et al. (2021)), GSR/EDA may not be suitable for detecting states requiring rapid detection or characterized by high variations in induced mental states, as seen in fast-paced games or action movies.

### ***6.4. Contextual Influence***

The contextual variant, specifically the individual’s surrounding environment, can significantly influence the outcomes of GSR/EDA signals. This issue is frequently overlooked in studies, potentially undermining the practical utility of GSR/EDA for emotion detection purposes if this issue is not properly addressed. However, it is worth noting that GSR/EDA remains the most straightforward and cost-effective signal to capture or measure and can effectively detect the emotional state of users in tightly controlled environments.

### ***6.5. Heart Rate as a Physiological Signal***

Compared to other physiological signals, HR may be more successful in achieving better results depending on the setup and environment as they are now nearly ubiquitous in wearable gadgets, thereby providing for cheaper and faster ways to collect, process and classify its data. However, it suffers from similar problems to GSR/EDA: the increase in HR, for example, is present in both fear and arousal. Differentiating one from the other is still an ongoing obstacle.

### ***6.6. EEG for Emotion Detection***

For emotion detection, although expensive, EEG is the most suitable. While problematic to set up and collect data, with enough pre-processing and post-processing, researchers were able to detect emotions more accurately than any other tool. However, it is still restricted to laboratory research and far from practical applications,

although commercial EEG applications such as EMOTIV can leverage real application scenarios. Another issue is also that, while traditional assessment methods can be applied “anywhere and anytime”, physiological measurements still rely on controlled laboratory settings. Aiming to solve this problem, Georges et al. (2020) proposed the use of GPS and physiological data to create emotional geographical heatmaps, which delineate areas where users experience specific emotional states in outdoor environments that can be used to infer information for future biofeedback applications.

### ***6.7. Physiological Data in Physical Activities***

Physiological data can become particularly susceptible to interference when integrated into applications involving physical movement, such as exercise, exercise-based gaming, or outdoor activities, as they are more prone to disruptions caused by respiration and motion. Moreover, high-intensity exercises are physically demanding and lead to increased sweating, elevated user temperature, changes in breathing patterns, and alterations in movement—all of which can significantly impact physiological measurements, such as GSR/EDA (Barathi et al., 2020).

### ***6.8. Challenges in Real-world Environments***

Adapting data collection in real-world environments poses a significant challenge due to the unpredictable nature of environmental factors and user behaviors. Additionally, it is essential to consider Cross-Cultural Factors, as cultural influences can impact how individuals respond to biofeedback interventions or stimuli. Similarly, addressing User Variability is crucial, as individuals exhibit diverse physiological responses to stimuli, making the establishment of universal benchmarks or standards for biofeedback challenging. This also poses a threat to long-term monitoring, as conducting biofeedback evaluations over extended periods can be demanding due to concerns such as sensor fatigue, data drift, and shifts in user behavior or physiological responses as time progresses.

### ***6.9. Latency in Biofeedback***

Few articles, even those studying real-time capture and identification of biosignals, have cited latency as a challenge. However, Petrescu et al. (2020) declare that response latency was an issue: the effects of sensor latency, network transmission latency, and screen refresh latency can cause a delay between a user’s bodily reaction and changes in displayed visual feedback. This obstacle needs to be carefully considered when designing interfaces for biofeedback training. For instance, Courtemanche et al. (2018) used third-party systems to synchronize their measurements: Syncbox was used to send TTL (Transistor-Transistor Logic) signals every 10s to each recording device (GSR/EDA, Eye Tracking and EEG) so the multimodal data could be sorted.

In real-time environments, the latency issue is even greater, although Martin et al. (2020) were able to train models that could detect cybersickness in “real-time”. Based on the processing chain and the trained models, a Proof-of-Concept was implemented by the authors that allowed estimating VR sickness level every second. Preliminary evaluation showed that it is suitable for real-time applications as detection was provided in less than 30 milliseconds.

Petrescu et al. (2020), addressing the problem of latency and other challenges, pro-

posed a protocol for monitoring and processing biophysical signals in VR applications. The elements of this protocol are designed to allow real-time estimates with low latencies. Therefore, although the proposed protocol largely follows the structure of most protocols present in the literature, the need to make estimates in real time led to the use of a smaller time window for feature extraction and a regression model to estimate the level of anxiety. Its results showed that its protocol reached average accuracy close to other studies, being, therefore, adequate.

### **6.10. Tools for Multi-Physiological Signal Measurement**

Lastly, some tools were found for the simultaneous measurement of multiple physiological signals. For example, in addition to the aforementioned *Synbox* (Courtemanche et al., 2018), we also identified *PhysiOBS* (Liapis et al., 2021), an observational analysis tool that aims to support researchers and professionals in the demanding task of post-processing of biofeedback analysis for Data UX. However, *PhysiOBS* can only help researchers identify stress-related periods through its integrated stress assessment model that uses video footage and EDA as input data.

### **6.11. Ethical Considerations**

Additional challenges arise when gathering physiological data, particularly within real-time or immersive settings, giving rise to ethical dilemmas concerning user consent, safeguarding data privacy, and the potential for sensitive information to be misused. Consequently, the imperative of securing this collected physiological data cannot be overstated, as any breaches or unauthorized access could result in severe repercussions. Implementing robust data encryption and protective measures becomes essential in addressing these concerns.

## **7. Conclusion**

Returning to our research question – *What and how are physiological data being used for user evaluations of digital multimedia content?* – this systematic review has unveiled a diverse array of physiological signals being employed for evaluations, with an emerging emphasis on the utilization of EEG in this context. Additionally, we have observed a growing trend in the application of physiological data to investigate VR environments. However, paradoxically, considering the inherent nature of physiological data, its usage in multisensory and mulsemmedia-based environments remains scarce. Consequently, we hypothesize that future studies will address these untapped opportunities and focus on bridging these gaps.

This review provides insights into the diverse array of physiological signals used for evaluations and highlights the increasing emphasis on the utilization of EEG and VR environments. It acknowledges the limitations and challenges associated with the use of physiological data in evaluations and proposes areas of future research to address these limitations. Furthermore, this review discussed the importance of standardization, interdisciplinary collaboration, ethical considerations, longitudinal studies, and the integration of physiological and subjective measures. Overall, it provides a comprehensive exploration of how physiological data are used in user evaluations of digital multimedia content.

However, it should be noted that evaluations incorporating physiological data are not without their limitations. In general, the challenges associated with the use of such data stem from several factors. These include the low detection accuracy of certain sensors, the reliance on professional-grade devices that hinder the application of these methods outside controlled laboratory environments, and latency issues that affect the real-time capture and identification of physiological data. Consequently, it is reasonable to assume that future research endeavors should prioritize addressing and mitigating these shortcomings.

Through our analysis, we have successfully identified a diverse range of devices utilized for the collection of physiological data. Given the multitude of options available, this article can serve as a valuable resource for stakeholders, aiding them in making informed decisions regarding the selection of appropriate devices, as well as guiding them on the processing, interpretation, and analysis of the collected data. However, despite the identification of various methods and ML models, we have not been able to determine an optimal approach for processing and interpreting physiological signals. This finding underscores the need for future endeavors to explore and establish more effective methodologies in this domain.

In addition to the previous conclusions, further insights can be drawn from the analysis conducted in this systematic review. Firstly, standardization and the development of guidelines are imperative in this field. Given the wide range of physiological signals, devices, and methodologies employed in user evaluations, establishing common protocols and best practices is essential. This standardization will not only improve the comparability and reproducibility of studies but also facilitate collaboration and knowledge sharing among researchers and practitioners.

Interdisciplinary collaboration emerges as a key factor in leveraging the potential of physiological data for user evaluations. Collaboration between experts from various fields such as psychology, human-computer interaction, neuroscience, and data science can significantly contribute to advancing our understanding of the relationship between physiological responses and user experiences. By combining expertise, insights, and methodologies from different disciplines, researchers can gain deeper insights into the underlying mechanisms of user experiences with digital multimedia content.

Ethical considerations should also be at the forefront of research endeavors. As the use of physiological data becomes more prevalent, researchers and practitioners need to address important ethical aspects. This includes ensuring data privacy, obtaining informed consent from participants, and employing responsible data handling practices. Adhering to ethical guidelines is crucial to safeguard the rights and well-being of individuals involved in these studies.

While the review primarily focused on cross-sectional studies, the potential of longitudinal investigations should not be overlooked. Conducting longitudinal studies that capture physiological responses over an extended period can provide valuable insights into the stability and variability of physiological measures. This longitudinal perspective can shed light on how user experiences evolve and change over time, enabling a deeper understanding of the dynamic nature of these experiences.

Moreover, the integration of physiological data with subjective measures such as self-reports and surveys can contribute to a more comprehensive comprehension of user experiences. By combining objective physiological indicators with subjective feedback provided by users, researchers can achieve a well-rounded viewpoint that encompasses both physiological responses and subjective perceptions. This unified approach enables a more precise and nuanced evaluation of user experiences with digital multimedia content.

The findings of this SLR highlight the increasing attractiveness and operational feasibility of using physiological data as an objective lens to evaluate user experiences. To further advance this field, it is crucial to focus on standardization, interdisciplinary collaboration, ethical considerations, longitudinal studies, and the integration of physiological and subjective measures. By addressing these aspects, researchers and practitioners can harness the full potential of physiological data to enhance the design, development, and assessment of interactive digital multimedia experiences.

In conclusion, the utilization of physiological data presents a compelling and practical alternative to subjective, self-reported measures, as it enables an objective examination of UX and QoE. It is noteworthy that the use of physiological data for user evaluations of digital multimedia content encounters similar considerations regarding preparation time when compared to subjective approaches. However, the escalating adoption of cyberphysical and AR/XR systems, combined with the widespread availability and accessibility of physiological data collection devices, serve as significant catalysts for leveraging physiological data in the digital realm. These trends underscore the growing appeal and operational feasibility of employing physiological data in assessing user experiences within the digital landscape.

## Acknowledgement(s)

This study was financed partly by the Coordination for the Improvement of Higher Education Personnel (CAPES, Brazil) – Finance Codes 88887.570688/2020-00 and 88881.689984/2022-01, National Council for Scientific and Technological Development (CNPQ, Brazil) – Finance Code 307718/2020-4, and Fundação de Amparo à Pesquisa e Inovação do Espírito Santo (FAPES, Brazil) – Finance Code 2021-GL60J.

## References

- Ajenaghughrure, I. B., Sousa, S. C., Kosunen, I. J., & Lamas, D. (2019). Predictive model to assess user trust: A psycho-physiological approach. In *10th Indian conference on human-computer interaction*. New York, NY, USA: ACM. Retrieved from <https://doi.org/10.1145/3364183.3364195>
- Al-Ashwal, W., Asadi, H., Mohamed, S., Alsanwy, S., Kooijman, L., Nahavandi, D., ... Nahavandi, S. (2021). Cybersickness measurement and evaluation during flying a helicopter in different weather conditions in virtual reality. In *2021 IEEE international conference on systems, man, and cybernetics (SMC)* (p. 2152-2157).
- Alqahtani, F., Katsigiannis, S., & Ramzan, N. (2019). Ecg-based affective computing for difficulty level prediction in intelligent tutoring systems. In *2019 UK/China emerging technologies (UCET)* (p. 1-4).
- Alves, T., Gama, S., & Melo, F. S. (2018). Flow adaptation in serious games for health. In *2018 IEEE 6th international conference on serious games and applications for health (SeGAH)* (p. 1-8).
- Aqajari, S. A. H., Naeini, E. K., Mehrabadi, M. A., Labbaf, S., Rahmani, A. M., & Dutt, N. (2020). *GSR analysis for stress: Development and validation of an open source tool for noisy naturalistic GSR data*. arXiv. Retrieved from <https://arxiv.org/abs/2005.01834>
- Athif, M., Rathnayake, B. L. K., Nagahapitiya, S. M. D. B. S., Samarasinghe, S. A. D. A. K., Samarasinghe, P. S., Peiris, R. L., & De Silva, A. C. (2020). Using biosignals for objective measurement of presence in virtual reality environments. In *2020 42nd annual international conference of the IEEE engineering in medicine biology society (EMBC)* (p. 3035-3039).

- Babaei, E., Tag, B., Dingler, T., & Velloso, E. (2021). A critique of electrodermal activity practices at CHI. In *2021 CHI conference on human factors in computing systems*. New York, NY, USA: ACM. Retrieved from <https://doi.org/10.1145/3411764.3445370>
- Baka, E., Stavroulia, K. E., Magnenat-Thalmann, N., & Lanitis, A. (2018). An eeg-based evaluation for comparing the sense of presence between virtual and physical environments. In *Proceedings of computer graphics international 2018* (p. 107–116). New York, NY, USA: ACM. Retrieved from <https://doi.org/10.1145/3208159.3208179>
- Barathi, S. C., Proulx, M., O’Neill, E., & Lutteroth, C. (2020). Affect recognition using psychophysiological correlates in high intensity vr exergaming. In *2020 CHI conference on human factors in computing systems* (p. 1–15). New York, NY, USA: ACM. Retrieved from <https://doi.org/10.1145/3313831.3376596>
- Beşkardeşler, K., & Arslan, R. B. (2019). Web-based evaluation of user stress using stroop test and wearable hardware. In *2019 11th international conference on electrical and electronics engineering (ELECO)* (p. 959-962).
- Bruun, A. (2018). It’s not complicated: A study of non-specialists analyzing gsr sensor data to detect ux related events. In *10th Nordic conference on human-computer interaction* (p. 170–183). New York, NY, USA: ACM. Retrieved from <https://doi.org/10.1145/3240167.3240183>
- Cao, Z. (2020). A review of artificial intelligence for EEG-based brain-computer interfaces and applications. *Brain Science Advances*, *6*(3), 162-170. Retrieved from <https://doi.org/10.26599/BSA.2020.9050017>
- Carrera-Rivera, A., Ochoa, W., Larrinaga, F., & Lasa, G. (2022). How-to conduct a systematic literature review: A quick guide for computer science research. *MethodsX*, *9*, 101895. Retrieved from <https://www.sciencedirect.com/science/article/pii/S2215016122002746>
- Certický, M., Čertický, M., Sinčák, P., Magyar, G., Vaščák, J., & Cavallo, F. (2019). Psychophysiological indicators for modeling user experience in interactive digital entertainment. *Sensors*, *19*(5). Retrieved from <https://www.mdpi.com/1424-8220/19/5/989>
- Chauhan, U., Reithinger, N., & Mackey, J. R. (2018). Real-time stress assessment through ppg sensor for vr biofeedback. In *20th international conference on multimodal interaction: Adjunct*. New York, NY, USA: ACM. Retrieved from <https://doi.org/10.1145/3281151.3281156>
- Che, X., Ma, S., Yu, Q., & Yue, C. (2019). An event-based user experience evaluation method for virtual reality applications. In *2019 IEEE smartworld, ubiquitous intelligence computing, advanced trusted computing, scalable computing communications, cloud big data computing, internet of people and smart city innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCOM/IOP/SCI)* (p. 767-774).
- Chen, C., Yao, L., Li, G., Wang, L., Liu, K., Wang, G., & Chen, X. (2021). Research on the relationship between virtual reality content and physiological response. *J. Phys. Conf. Ser.*, *2026*(1), 012014.
- Chen, S., & Epps, J. (2020). Multimodal coordination measures to understand users and tasks. *ACM Trans. Comput.-Hum. Interact.*, *27*(6). Retrieved from <https://doi.org/10.1145/3412365>
- Cho, H., Powell, D., Pichon, A., Kuhns, L. M., Garofalo, R., & Schnall, R. (2019). Eye-tracking retrospective think-aloud as a novel approach for a usability evaluation. *International Journal of Medical Informatics*, *129*, 366-373. Retrieved from <https://www.sciencedirect.com/science/article/pii/S1386505619302291>
- Ciolacu, M. I., & Svasta, P. (2021). Education 4.0: Ai empowers smart blended learning process with biofeedback. In *2021 IEEE global engineering education conference (EDUCON)* (p. 1443-1448).
- Cipresso, P., Giglioli, I. A. C., Raya, M. A., & Riva, G. (2018). The past, present, and future of virtual and augmented reality research: A network and cluster analysis of the literature. *Frontiers in Psychology*, *9*. Retrieved from <https://doi.org/10.3389/fpsyg.2018.02086>

- Collins, J., Regenbrecht, H., Langlotz, T., Said Can, Y., Ersoy, C., & Butson, R. (2019). Measuring cognitive load and insight: A methodology exemplified in a virtual reality learning context. In *2019 IEEE international symposium on mixed and augmented reality (ISMAR)* (p. 351-362).
- Corneille, O., & Mertens, G. (2020). Behavioral and physiological evidence challenges the automatic acquisition of evaluations. *Current Directions in Psychological Science*, *29*(6), 569-574. Retrieved from <https://doi.org/10.1177/0963721420964111>
- Courtemanche, F., Léger, P.-M., Dufresne, A., Fredette, M., Labonté-Lemoyne, E., & Sénécal, S. (2018). Physiological heatmaps: A tool for visualizing users' emotional reactions. *Multimedia Tools Appl.*, *77*(9), 11547-11574. Retrieved from <https://doi.org/10.1007/s11042-017-5091-1>
- Darzi, A., McCrea, S. M., & Novak, D. (2021). User experience with dynamic difficulty adjustment methods for an affective exergame: Comparative laboratory-based study. *JMIR Serious Games*, *9*(2), e25771. Retrieved from <https://doi.org/10.2196/25771>
- Darzi, A., Wondra, T., McCrea, S., & Novak, D. (2019). Classification of multiple psychological dimensions in computer game players using physiology, performance, and personality characteristics. *Frontiers in Neuroscience*, *13*. Retrieved from <https://www.frontiersin.org/article/10.3389/fnins.2019.01278>
- Daylamani-Zad, D., Spyridonis, F., & Al-Khafaaji, K. (2022). A framework and serious game for decision making in stressful situations; a fire evacuation scenario. *International Journal of Human-Computer Studies*, *162*, 102790. Retrieved from <https://www.sciencedirect.com/science/article/pii/S1071581922000192>
- de Santana, V. F., & Otani, L. M. F. (2021). Measuring quantitative situated user experience with a mobile galvanic skin response sensor. In *XX Brazilian symposium on human factors in computing systems*. New York, NY, USA: ACM. Retrieved from <https://doi.org/10.1145/3472301.3484339>
- Dey, A., Phoon, J., Saha, S., Dobbins, C., & Billinghamurst, M. (2020a). A neurophysiological approach for measuring presence in immersive virtual environments. In *2020 IEEE international symposium on mixed and augmented reality (ismar)* (p. 474-485).
- Dey, A., Phoon, J., Saha, S., Dobbins, C., & Billinghamurst, M. (2020b). Neurophysiological effects of presence in calm virtual environments. In *2020 IEEE conference on virtual reality and 3d user interfaces abstracts and workshops (VRW)* (p. 744-745).
- Dhaouadi, S., & Ben Khelifa, M. M. (2020). A multimodal physiological-based stress recognition: Deep learning models' evaluation in gamers' monitoring application. In *2020 5th international conference on advanced technologies for signal and image processing (atsip)* (p. 1-6).
- Ding, Y., Cao, Y., Qu, Q., & Duffy, V. G. (2020). An exploratory study using electroencephalography (EEG) to measure the smartphone user experience in the short term. *International Journal of Human-Computer Interaction*, *36*(11), 1008-1021. Retrieved from <https://doi.org/10.1080/10447318.2019.1709330>
- Drachen, A., Nacke, L. E., Yannakakis, G., & Pedersen, A. L. (2010). Correlation between heart rate, electrodermal activity and player experience in first-person shooter games. In *5th ACM siggraph symposium on video games* (p. 49-54). New York, NY, USA: ACM. Retrieved from <https://doi.org/10.1145/1836135.1836143>
- Du, G., Long, S., & Yuan, H. (2020). Non-contact emotion recognition combining heart rate and facial expression for interactive gaming environments. *IEEE Access*, *8*, 11896-11906.
- Du, R., Zhu, S., Ni, H., Mao, T., Li, J., & Wei, R. (2022). Valence-arousal classification of emotion evoked by chinese ancient-style music using 1D-CNN-BiLSTM model on EEG signals for college students. *Multimedia Tools and Applications*. Retrieved from <https://doi.org/10.1007/s11042-022-14011-7>
- Elor, A., Powell, M., Mahmoodi, E., Hawthorne, N., Teodorescu, M., & Kurniawan, S. (2020). On shooting stars: Comparing CAVE and HMD immersive virtual reality exergaming for adults with mixed ability. *ACM Trans. Comput. Healthcare*, *1*(4). Retrieved from <https://doi.org/10.1145/3396249>

- Ergan, S., Radwan, A., Zou, Z., Tseng, H.-A., & Han, X. (2019). Quantifying human experience in architectural spaces with integrated virtual reality and body sensor networks. *J. Comput. Civ. Eng.*, *33*(2), 04018062.
- Frey, J., Ostrin, G., Grabli, M., & Cauchard, J. R. (2020). Physiologically driven storytelling: Concept and software tool. In *2020 CHI conference on human factors in computing systems* (p. 1–13). New York, NY, USA: ACM. Retrieved from <https://doi.org/10.1145/3313831.3376643>
- Georges, V., Courtemanche, F., Fredette, M., & Doyon-Poulin, P. (2020). Emotional maps for user experience research in the wild. In *Extended abstracts of the 2020 chi conference on human factors in computing systems* (p. 1–8). New York, NY, USA: ACM. Retrieved from <https://doi.org/10.1145/3334480.3383042>
- Ghinea, G., Andres, F., & Gulliver, S. (2011). *Multiple sensorial media advances and applications: new developments in mulsemedia*. IGI Global.
- Giannakakis, G., Grigoriadis, D., Giannakaki, K., Simantiraki, O., Roniotis, A., & Tsiknakis, M. (2019). Review on psychological stress detection using biosignals. *IEEE Transactions on Affective Computing*, 1-1.
- Giannakos, M. N., Sharma, K., Pappas, I. O., Kostakos, V., & Velloso, E. (2019). Multimodal data as a means to understand the learning experience. *International Journal of Information Management*, *48*, 108-119. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0268401218312751>
- Grassini, S., Laumann, K., Thorp, S., & Topranin, V. d. M. (2021). Using electrophysiological measures to evaluate the sense of presence in immersive virtual environments: An event-related potential study. *Brain and Behavior*, *11*(8), e2269. Retrieved from <https://onlinelibrary.wiley.com/doi/abs/10.1002/brb3.2269>
- Greinacher, R., Kojić, T., Meier, L., Parameshappa, R. G., Möller, S., & Voigt-Antons, J.-N. (2020). Impact of tactile and visual feedback on breathing rhythm and user experience in VR exergaming. In *2020 twelfth international conference on quality of multimedia experience (QoMEX)* (p. 1-6).
- Gruenewald, A., Kroenert, D., Poehler, J., Brueck, R., Li, F., Littau, J., ... Niehaves, B. (2018). [regular paper] biomedical data acquisition and processing to recognize emotions for affective learning. In *IEEE 18th international conference on bioinformatics and bioengineering (bibe 2018)* (p. 126-132).
- Gupta, K., Hajika, R., Pai, Y. S., Duenser, A., Lochner, M., & Billinghamurst, M. (2020). Measuring human trust in a virtual assistant using physiological sensing in virtual reality. In *2020 IEEE conference on virtual reality and 3d user interfaces (VR)* (p. 756-765).
- Gurita, A. E., Soliman, M., Neetson-Lemkes, W., Chung, V., Hansen, N. B., Korte, G., & Vosmeer, M. (2019). Exploring states of mind: Emotion visualization with bio-feedback sensors in a 3d environment. In *Companion publication of the 2019 on designing interactive systems conference 2019 companion* (p. 33–36). New York, NY, USA: ACM. Retrieved from <https://doi.org/10.1145/3301019.3325150>
- Haratian, R., & Timotijevic, T. (2018). On-body sensing and signal analysis for user experience recognition in human-machine interaction. In *2018 4th international conference on frontiers of signal processing (ICFSP)* (p. 50-55).
- He, H., She, Y., Xiahou, J., Yao, J., Li, J., Hong, Q., & Ji, Y. (2018). Real-time eye-gaze based interaction for human intention prediction and emotion analysis. In *Proceedings of computer graphics international 2018* (p. 185–194). New York, NY, USA: ACM. Retrieved from <https://doi.org/10.1145/3208159.3208180>
- He, L., Li, H., Xue, T., Sun, D., Zhu, S., & Ding, G. (2018). Am i in the theater? usability study of live performance based virtual reality. In *24th ACM symposium on virtual reality software and technology*. New York, NY, USA: ACM. Retrieved from <https://doi.org/10.1145/3281505.3281508>
- Holman, M., & Adebesin, F. (2019). Taking the subjectivity out of ux evaluation with emotiv epoc+. In *South African institute of computer scientists and information technologists 2019*. New York, NY, USA: ACM. Retrieved from

- <https://doi.org/10.1145/3351108.3351139>
- Hossain, E., & Kulshreshth, A. K. (2019). Exploring the effects of stereoscopic 3d on gaming experience using physiological sensors. In *Symposium on spatial user interaction*. New York, NY, USA: ACM. Retrieved from <https://doi.org/10.1145/3357251.3358751>
- Houzangbe, S., Christmann, O., Gorisse, G., & Richir, S. (2019). Effects of voluntary heart rate control on user engagement in virtual reality. In *2019 IEEE conference on virtual reality and 3d user interfaces (VR)* (p. 982-983).
- Houzangbe, S., Christmann, O., Gorisse, G., & Richir, S. (2020, 01). Effects of voluntary heart rate control on user engagement and agency in a virtual reality game. *Virtual Reality*, *24*(4), 665-681. Retrieved from <https://doi.org/10.1007/s10055-020-00429-7>
- Hussain, J., Khan, W. A., Hur, T., Bilal, H. S. M., Bang, J., Hassan, A. U., ... Lee, S. (2018). A multimodal deep log-based user experience (UX) platform for UX evaluation. *Sensors*, *18*(5). Retrieved from <https://www.mdpi.com/1424-8220/18/5/1622>
- Huynh, S., Kim, S., Ko, J., Balan, R. K., & Lee, Y. (2018). Engagemon: Multi-modal engagement sensing for mobile games. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, *2*(1). Retrieved from <https://doi.org/10.1145/3191745>
- Hynes, E., Flynn, R., Lee, B., & Murray, N. (2019). A quality of experience evaluation comparing augmented reality and paper based instruction for complex task assistance. In *2019 IEEE 21st international workshop on multimedia signal processing (MMSP)* (p. 1-6).
- Isiaka, F. M., Adamu, A., & Adamu, Z. (2021). Integration of biosensor to a window-based control system for user emotion detection to static and dynamic visual contents of webpages. *Int. J. Crowd Sci.*, *5*(3), 257-270.
- Islam, A., Ma, J., Gedeon, T., Hossain, M. Z., & Liu, Y.-H. (2019). Measuring user responses to driving simulators: A galvanic skin response based study. In *2019 IEEE international conference on artificial intelligence and virtual reality (AIVR)* (p. 33-337).
- Islam, R., Ang, S., & Quarles, J. (2021). Cybersense: A closed-loop framework to detect cybersickness severity and adaptively apply reduction techniques. In *2021 IEEE conference on virtual reality and 3d user interfaces abstracts and workshops (VRW)* (p. 148-155).
- Kalatzis, A., Stanley, L., & Prabhu, V. G. (2021). Affective state classification in virtual reality environments using electrocardiogram and respiration signals. In *2021 IEEE international conference on artificial intelligence and virtual reality (AIVR)*. IEEE.
- Kaluarachchi, T. I., Sapkota, S., Taradel, J., Thevenon, A., Matthies, D. J., & Nanayakkara, S. (2021). Eyeknowyou: A DIY toolkit to support monitoring cognitive load and actual screen time using a head-mounted webcam. In *Adjunct publication of the 23rd international conference on mobile human-computer interaction*. New York, NY, USA: ACM. Retrieved from <https://doi.org/10.1145/3447527.3474850>
- Katada, S., Okada, S., Hirano, Y., & Komatani, K. (2020). Is she truly enjoying the conversation? analysis of physiological signals toward adaptive dialogue systems. In *2020 international conference on multimodal interaction* (p. 315-323). New York, NY, USA: ACM. Retrieved from <https://doi.org/10.1145/3382507.3418844>
- Katsigiannis, S., Willis, R., & Ramzan, N. (2019). A QoE and simulator sickness evaluation of a smart-exercise-bike virtual reality system via user feedback and physiological signals. *IEEE Transactions on Consumer Electronics*, *65*(1), 119-127.
- Keighrey, C., Flynn, R., Murray, S., & Murray, N. (2021). A physiology-based QoE comparison of interactive augmented reality, virtual reality and tablet-based applications. *IEEE Transactions on Multimedia*, *23*, 333-341.
- Kim, H.-G., Lee, G. Y., & Kim, M.-S. (2021). Dual-function integrated emotion-based music classification system using features from physiological signals. *IEEE Transactions on Consumer Electronics*, *67*(4), 341-349.
- Kitchenham, B. A., & Charters, S. (2007). *Guidelines for performing systematic literature reviews in software engineering* (Tech. Rep. No. EBSE 2007-001). Keele University and Durham University Joint Report.
- Klarkowski, M., Johnson, D., Wyeth, P., Phillips, C., & Smith, S. (2018). Don't sweat the small stuff: The effect of challenge-skill manipulation on electrodermal activity. In *2018*

- annual symposium on computer-human interaction in play* (p. 231–242). New York, NY, USA: ACM. Retrieved from <https://doi.org/10.1145/3242671.3242714>
- Korn, O., & Rees, A. (2019). Affective effects of gamification: Using biosignals to measure the effects on working and learning users. In *12th ACM international conference on pervasive technologies related to assistive environments* (p. 1–10). New York, NY, USA: ACM. Retrieved from <https://doi.org/10.1145/3316782.3316783>
- Lee, S., Ryu, H., Park, B., & Yun, M. H. (2020). Using physiological recordings for studying user experience: Case of conversational agent-equipped tv. *International Journal of Human-Computer Interaction*, *36*(9), 815–827. Retrieved from <https://doi.org/10.1080/10447318.2019.1693166>
- Lee-Cultura, S., Sharma, K., Papavaslopoulou, S., & Giannakos, M. (2020). Motion-based educational games: Using multi-modal data to predict player’s performance. In *2020 IEEE conference on games (cog)* (p. 17–24).
- Li, Z., Geng, B., Tao, X., Duan, Y., Gao, D., & Hu, S. (2021). Video quality measurement for buffering time based on eeg frequency feature. In *2021 IEEE international conference on image processing (ICIP)* (p. 1424–1428).
- Liapis, A., Katsanos, C., Karousos, N., Sotiropoulos, D., Xenos, M., & Orphanoudakis, T. (2020). Stress heatmaps: A fuzzy-based approach that uses physiological signals. In A. Marcus & E. Rosenzweig (Eds.), *Design, user experience, and usability. case studies in public and personal interactive systems* (pp. 268–277). Cham: Springer International Publishing.
- Liapis, A., Katsanos, C., Karousos, N., Xenos, M., & Orphanoudakis, T. (2021). User experience evaluation: A validation study of a tool-based approach for automatic stress detection using physiological signals. *Int. J. Hum. Comput. Interact.*, *37*(5), 470–483.
- Liapis, A., Katsanos, C., & Xenos, M. (2018). Don’t leave me alone: Retrospective think aloud supported by real-time monitoring of participant’s physiology. In M. Kurosu (Ed.), *Human-computer interaction. theories, methods, and human issues* (pp. 148–158). Cham: Springer International Publishing.
- Ma, J., Chen, C.-C., & Lin, Y.-C. (2018). Emotional and cognitive assessment of use of functional animation. In *International conference on machine vision and applications* (p. 61–65). New York, NY, USA: ACM. Retrieved from <https://doi.org/10.1145/3220511.3220516>
- Magalie, O., Sameer, J., & Philippe, B. (2018). Toward an automatic prediction of the sense of presence in virtual reality environment. In *6th international conference on human-agent interaction* (p. 161–166). New York, NY, USA: ACM. Retrieved from <https://doi.org/10.1145/3284432.3284452>
- Maier, M., Elsner, D., Marouane, C., Zehnle, M., & Fuchs, C. (2019). DeepFlow: Detecting optimal user experience from physiological data using deep neural networks. In *Twenty-eighth international joint conference on artificial intelligence*. International Joint Conferences on Artificial Intelligence Organization. Retrieved from <https://doi.org/10.24963/ijcai.2019/196>
- Manjunatha, H., Distefano, J. P., Jani, A., Ghassemi, P., Chowdhury, S., Dantu, K., ... Esfahani, E. T. (2020). Using physiological measurements to analyze the tactical decisions in human swarm teams. In *2020 IEEE international conference on systems, man, and cybernetics (SMC)* (p. 256–261).
- Markova, V., Ganchev, T., & Kalinkov, K. (2018). Detection of negative emotions and high-arousal negative-valence states on the move. In *2018 advances in wireless and optical communications (RTUWO)* (p. 61–65).
- Martin, N., Mathieu, N., Pallamin, N., Ragot, M., & Diverrez, J.-M. (2020). Virtual reality sickness detection: an approach based on physiological signals and machine learning. In *2020 IEEE international symposium on mixed and augmented reality (ismar)* (p. 387–399).
- Mavridou, I., Seiss, E., Kostoulas, T., Hamedi, M., Balaguer-Ballester, E., & Nduka, C. (2019). Introducing the EmteqVR interface for affect detection in virtual reality. In *2019 8th international conference on affective computing and intelligent interaction workshops and demos (ACIIW)*. IEEE.

- McKee, M. G. (2008). Biofeedback: an overview in the context of heart-brain medicine. *Cleveland Clinic Journal of Medicine*, 75(Suppl.2), S31–S34.
- Mehrabian, A. (1980). *Basic dimensions for a general psychological theory: Implications for personality, social, environmental, and developmental studies*. Cambridge: Oelgeschlager, Gunn & Hain.
- Mesfin, G., Hussain, N., Covaci, A., & Ghinea, G. (2019). Using eye tracking and heart-rate activity to examine crossmodal correspondences QoE in mulsemmedia. *ACM Trans. Multimedia Comput. Commun. Appl.*, 15(2). Retrieved from <https://doi.org/10.1145/3303080>
- Mesfin, G., Hussain, N., Kani-Zabihi, E., Covaci, A., Saleme, E. B., & Ghinea, G. (2020). QoE of cross-modally mapped mulsemmedia: an assessment using eye gaze and heart rate. *Multimedia Tools and Applications*, 79(11-12), 7987–8009. Retrieved from <https://doi.org/10.1007/s11042-019-08473-5>
- Mesfin, G., Saleme, E. B., Ademoye, O. A., Kani-Zabihi, E., Santos, C. A., & Ghinea, G. (2021). Less is (just as good as) more - an investigation of odor intensity and hedonic valence in mulsemmedia qoe using heart rate and eye tracking. *IEEE Transactions on Multimedia*, 23, 1095-1105.
- Mijović, P., Milovanović, M., Ković, V., Mijović, B., Gligorijević, I., Minović, M., & Živanović Mačužić, I. (2019). Communicating the user state: Introducing cognition-aware computing in industrial settings. *Safety Science*, 119, 375-384. Retrieved from <https://www.sciencedirect.com/science/article/pii/S092575351732043X>
- Min, S., Moon, J.-g., Cho, C.-H., & Kim, G. J. (2020). Effects of immersive virtual reality content type to mindfulness and physiological parameters. In *26th ACM symposium on virtual reality software and technology*. New York, NY, USA: ACM. Retrieved from <https://doi.org/10.1145/3385956.3418942>
- Moinnereau, M.-A., Oliveira, A., & Falk, T. H. (2020). Saccadic eye movement classification using exg sensors embedded into a virtual reality headset. In *2020 IEEE international conference on systems, man, and cybernetics (SMC)* (p. 3494-3498).
- Moraes, A. N., Flynn, R., Hines, A., & Murray, N. (2020). Evaluating the user in a sound localisation task in a virtual reality application. In *2020 twelfth international conference on quality of multimedia experience (QoMEX)* (p. 1-6).
- Moraes, A. N., Flynn, R., Hines, A., & Murray, N. (2021). The role of physiological responses in a VR-based sound localization task. *IEEE Access*, 9, 122082–122091.
- Ninaus, M., Tsarava, K., & Moeller, K. (2019). A pilot study on the feasibility of dynamic difficulty adjustment in game-based learning using heart-rate. In A. Liapis, G. N. Yannakakis, M. Gentile, & M. Ninaus (Eds.), *Games and learning alliance* (pp. 117–128). Cham: Springer International Publishing.
- Oliveira, T., Noriega, P., Carvalhais, J., Rebelo, F., & Lameira, V. (2020). How deep is a virtual reality experience? virtual environments, emotions and physiological measures. In F. Rebelo & M. M. Soares (Eds.), *Advances in ergonomics in design* (pp. 462–471). Cham: Springer International Publishing.
- Pei, W., Guo, X., & Lo, T. (2021). Biofeedback in the dynamic vr environments: A method to evaluate the influence of architectural elements on human spatial perception. In *2021 workshop on algorithm and big data* (p. 27–33). New York, NY, USA: ACM. Retrieved from <https://doi.org/10.1145/3456389.3456400>
- Peruzzini, M., Grandi, F., Pellicciari, M., & Campanella, C. E. (2018). User experience analysis based on physiological data monitoring and mixed prototyping to support human-centre product design. In *Advances in ergonomics in design* (pp. 401–412). Springer International Publishing.
- Petrescu, L., Petrescu, C., Mitruț, O., Moise, G., Moldoveanu, A., Moldoveanu, F., & Leordeanu, M. (2020). Integrating biosignals measurement in virtual reality environments for anxiety detection. *Sensors*, 20(24). Retrieved from <https://www.mdpi.com/1424-8220/20/24/7088>
- Pinto, J., Fred, A., & da Silva, H. P. (2019). Biosignal-based multimodal emotion recognition in a valence-arousal affective framework applied to immersive video visualization. In *2019*

- 41st annual international conference of the IEEE engineering in medicine and biology society (EMBC) (p. 3577-3583).
- Pinto, M., Melo, M., & Bessa, M. (2018). Use of the physiological response to improve the gaming experience. In *2018 international conference on graphics and interaction (ICGI)* (p. 1-7).
- Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology*, *17*(03). Retrieved from <https://doi.org/10.1017/s0954579405050340>
- Putze, S., Alexandrovsky, D., Putze, F., Höffner, S., Smeddinck, J. D., & Malaka, R. (2020). Breaking the experience: Effects of questionnaires in vr user studies. In *2020 CHI conference on human factors in computing systems* (p. 1–15). New York, NY, USA: ACM. Retrieved from <https://doi.org/10.1145/3313831.3376144>
- Rahman, J. S., Gedeon, T., Caldwell, S., Jones, R., & Jin, Z. (2021). Towards effective music therapy for mental health care using machine learning tools: Human affective reasoning and music genres. *Journal of Artificial Intelligence and Soft Computing Research*, *11*(1), 5–20. Retrieved from <https://doi.org/10.2478/jaiscr-2021-0001>
- Rahman, J. S., Hossain, M. Z., & Gedeon, T. (2020). Measuring observers' EDA responses to emotional videos. In *31st australian conference on human-computer-interaction* (p. 457–461). New York, NY, USA: ACM. Retrieved from <https://doi.org/10.1145/3369457.3369516>
- Riaz, M., Majid, M., & Mir, J. (2021). Emotion recognition using electroencephalography in response to high dynamic range videos. In *2021 international conference on information technology (ICIT)* (p. 565-570).
- Roy, Y., Banville, H., Albuquerque, I., Gramfort, A., Falk, T. H., & Faubert, J. (2019). Deep learning-based electroencephalography analysis: a systematic review. *J. Neural Eng.*, *16*(5), 051001.
- Safikhani, S., Holly, M., Kainz, A., & Pirker, J. (2021). The influence of in-vr questionnaire design on the user experience. In *27th ACM symposium on virtual reality software and technology*. New York, NY, USA: ACM. Retrieved from <https://doi.org/10.1145/3489849.3489884>
- Salgado, D. P., Martins, F. R., Rodrigues, T. B., Keighrey, C., Flynn, R., Naves, E. L. M., & Murray, N. (2018). A QoE assessment method based on EDA, heart rate and EEG of a virtual reality assistive technology system. In *9th ACM multimedia systems conference* (p. 517–520). New York, NY, USA: ACM. Retrieved from <https://doi.org/10.1145/3204949.3208118>
- Sandeep Vara Sankar, D., & Ko, L.-W. (2020). Evaluation of fatigue and attention levels in multi-target scenario using cnn. In *2020 international computer symposium (ICS)* (p. 247-251).
- Santamaria-Granados, L., Mendoza-Moreno, J. F., Chantre-Astaiza, A., Munoz-Organero, M., & Ramirez-Gonzalez, G. (2021). Tourist experiences recommender system based on emotion recognition with wearable data. *Sensors*, *21*(23). Retrieved from <https://www.mdpi.com/1424-8220/21/23/7854>
- Satti, F. A., Hussain, M., Hussain, J., Kim, T.-S., Lee, S., & Chung, T. (2021). User stress modeling through galvanic skin response. In *2021 15th international conference on ubiquitous information management and communication (IMCOM)* (p. 1-6).
- Schmidt, S., Uhrig, S., & Reuschel, D. (2020). Investigating the relationship of mental immersion and physiological measures during cloud gaming. In *2020 twelfth international conference on quality of multimedia experience (QoMEX)* (p. 1-6).
- Silveira, A., Raisamo, R., Spyridonis, F., Covaci, A., Ghinea, G., & Santos, C. A. S. (2023). Guidelines for conducting biofeedback-enhanced qoe studies in mulsemmedia-enhanced virtual reality. In *Brazilian symposium on multimedia and the web* (p. 32–40). New York, NY, USA: ACM. Retrieved from <https://doi.org/10.1145/3617023.3617029>
- Silveira, A. C., & Santos, C. A. S. (2022). Ongoing challenges of evaluating mulsemmedia QoE.

- In *2nd workshop on multisensory experiences (SensoryX'22)*. Porto Alegre, RS, Brasil: SBC. Retrieved from <https://sol.sbc.org.br/index.php/sensoryx/article/view/20005>
- Souza, V. C. d., Nedel, L., Kopper, R., Maciel, A., & Tagliaro, L. (2018). The effects of physiologically-adaptive virtual environment on user's sense of presence. In *2018 20th symposium on virtual and augmented reality (SVR)* (p. 133-142).
- Suni-Lopez, F., Condori-Fernandez, N., & Catala, A. (2020). Understanding implicit user feedback from multisensorial and physiological data: A case study. In *IEEE/ACM 42nd international conference on software engineering workshops* (p. 563-569). New York, NY, USA: ACM. Retrieved from <https://doi.org/10.1145/3387940.3391466>
- Tateyama, N., Ueda, K., & Nakao, M. (2019). Development of an active sensing system for distress detection using skin conductance response. In *2019 8th international conference on affective computing and intelligent interaction (ACII)*. IEEE.
- Tremmel, C., Herff, C., Sato, T., Rechowicz, K., Yamani, Y., & Krusienski, D. J. (2019). Estimating cognitive workload in an interactive virtual reality environment using EEG. *Frontiers in Human Neuroscience*, *13*. Retrieved from <https://www.frontiersin.org/article/10.3389/fnhum.2019.00401>
- Tsai, T.-C., Pan, T.-Y., & Hu, M.-C. (2021). Investigating the affective state of vr hmd user when watching videos displayed in different formats. In *2021 IEEE international conference on artificial intelligence and virtual reality (AIVR)* (p. 242-246).
- Voigt-Antons, J.-N., Spang, R., Kojić, T., Meier, L., Vergari, M., & Möller, S. (2021). Don't worry be happy - using virtual environments to induce emotional states measured by subjective scales and heart rate parameters. In *2021 IEEE virtual reality and 3d user interfaces (VR)* (p. 679-686).
- Wan, X., Zhang, K., Ramkumar, S., Deny, J., Emayavaramban, G., Siva Ramkumar, M., & Hussein, A. F. (2019). A review on electroencephalogram based brain computer interface for elderly disabled. *IEEE Access*, *7*, 36380-36387.
- Wang, X., & Lin, L. (2020). Research on the construction of product material perceptual evaluation model based on EEG physiological cognition. *E3S Web Conf.*, *179*, 01016.
- Wang, Y., Fischer, N., & Bry, F. (2019). Pervasive persuasion for stress self-regulation. In *2019 IEEE international conference on pervasive computing and communications workshops (percom workshops)* (p. 724-730).
- Wechsung, I., & De Moor, K. (2014). Quality of experience versus user experience. In S. Möller & A. Raake (Eds.), *Quality of experience: Advanced concepts, applications and methods* (pp. 35-54). Cham: Springer International Publishing.
- Wulvik, A. S., Dybvik, H., & Steinert, M. (2020). Investigating the relationship between mental state (workload and affect) and physiology in a control room setting (ship bridge simulator). *Cognition, Technology & Work*, *22*(1), 95-108.
- Yang, W., Rifqi, M., Marsala, C., & Pinna, A. (2018). Towards better understanding of player's game experience. In *2018 ACM international conference on multimedia retrieval* (p. 442-449). New York, NY, USA: ACM. Retrieved from <https://doi.org/10.1145/3206025.3206072>
- Ye, X., Ning, H., Backlund, P., & Ding, J. (2021). Flow experience detection and analysis for game users by wearable-devices-based physiological responses capture. *IEEE Internet of Things Journal*, *8*(3), 1373-1387.
- Zou, Z., & Ergan, S. (2021). Evaluating the effectiveness of biometric sensors and their signal features for classifying human experience in virtual environments. *Advanced Engineering Informatics*, *49*, 101358. Retrieved from <https://www.sciencedirect.com/science/article/pii/S1474034621001117>

## Author Biographies



**Aleph Campos da Silveira** is a PhD student in Computer Science from the Federal University of Espirito Santo (UFES), Brazil. He also holds a BSc degree in Computer Science and Information Systems and an MSc in Education, both from the Federal University of Lavras (UFLA), Brazil. His research interests include Usability Evaluation, Multisensory Human-Computer interaction, Biofeedback, and Games.



**Mariane Lima de Souza** is an Associate Professor at UFES, Brazil, with a Ph.D. in Developmental Psychology from Federal University of Rio Grande do Sul (UFRGS), Brazil. Her research focuses on Development and Cognition, particularly cognitive neuroscience and experimental phenomenology. She has contributed to various research projects and has extensive teaching experience at both undergraduate and postgraduate levels. She has also supervised numerous students in their academic pursuits.



**Gheorghita Ghinea** is a Professor of Mulsemmedia Computing at Brunel University. He received his B.Sc. and B.Sc. (Hons.) degrees in Computer Science and Mathematics and his M.Sc. degree in Computer Science from the University of the Witwatersrand, Johannesburg, South Africa. He also received his Ph.D. degree in Computer Science from the University of Reading, U.K. His research interests include the confluence of computer science, media, and psychology, focusing on building adaptable cross-layer end-to-end communication systems incorporating user multisensorial and perceptual requirements.



**Celso Alberto Saibel Santos** is a Professor in the Department of Informatics at the Federal University of Espirito Santo (UFES). He received his Ph.D. in Informatics from Université Paul Sabatier de Toulouse III, France, his M.Sc. in Electrical Engineering from the University of São Paulo (USP), Brazil, and his B.Sc. in Electrical Engineering from UFES, Brazil. His research interests include Multimedia Systems, Computing Systems Engineering, and Multisensory Applications.