



Accessibility first: detecting frustration in web browsing for visually impaired and sighted smartphone users

Yasmin Abdelaal¹ · Dena Al-Thani¹

Accepted: 4 October 2023
© The Author(s) 2023

Abstract

Smartphone adoption has become increasingly prevalent in modern society, reflecting the widespread integration of these devices into various aspects of daily life, yet accessibility and usability problems persist. Smartphones pose numerous challenges to individuals with visual impairments (VI). This is due to smaller screen sizes, lack of physical buttons, and prevalent problems. These challenges frustrate users as they are unable to accomplish their goals. In this paper, the physiological response of VI and sighted individuals are compared and examined. The participant's frustration is measured through physiological signals. The GSR is the signal captured to detect the user's frustration caused by the superimposed accessibility and usability web-related issues. GSR, being less obtrusive, is an effective way to measure frustration. There were 13 VI and 16 sighted participants. Using a within-subject and between-subject design, participants completed four tasks, each was completed under frustrating and non-frustrating conditions while wearing an Empatica E4 wristband to collect GSR data. Challenges were both group-specific and mutual. A slowed internet connection and constant page refreshes are frustrating tasks for sighted participants. VI participants faced an unsearchable drop-down list and an inaccessible menu. Pop-up ads and session timeout are common tasks. Among VI participants, there was no significant difference between frustrating and non-frustrating tasks. When VI participants experienced a session timeout, they significantly showed a higher level of arousal. When comparing the physiological responses of sighted and VI participants in the page refresh and session timeout tasks, the results show that the arousal level is significantly higher in VI participants, primarily due to accessibility issues.

Keywords Vision impairment · GSR · Physiological response · Frustration · Accessibility · Usability

1 Introduction

Smartphones have become integral to our daily lives, offering a wide range of services and being used by 92% of active internet users in 2021 [1]. Despite their widespread use, ensuring comprehensive consideration of end users' needs in mobile application development remains a challenge that can result in various issues. End users range in their ability, and web dominance is not limited to a single demographic. This results in accessibility and inclusivity considerations falling through the cracks as web developers target a specific group while overlooking critical accessibility specifications.

Web accessibility, defined as the design and development of web-based content that allows people with disabilities to

access and navigate it effectively, is essential for a website's success in serving individuals with disabilities [2]. Additionally, usability, defined as the effectiveness and satisfaction of achieving goals, is equally crucial for a website's success. Literature has shed light on the importance of addressing accessibility and usability problems by developing a variety of methods to address users' evaluation and identifying common problems. For instance, visual clutter, such as moving images and illegible text, is a common usability problem [3]. In addition, issues such as unnecessary pop-ups for content presentation (e.g., advertisements and forms) and inadequate navigation through webpages were often observed [4]. Inadequate feedback and system issues with assistive technology are common usability problems for visually impaired (VI) users [5]. Studies have provided empirical evidence that these problems can have a negative impact on users [6]. Apart from making the task difficult to complete, when users encounter usability issues, they become frustrated as they are unable to achieve their desired goal.

✉ Dena Al-Thani
dalthani@hbku.edu.qa

¹ College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar

Frustration is a common emotion that arises when individuals are unable to achieve their desired goals [7]. Users may experience frustration when encountering obstacles while trying to complete tasks, such as filling out a form [8]. This frustration can lead to reduced accuracy, slower task completion, and other intangible consequences, including loss of motivation and a decline in the quality of the user experience [8]. Demographic groups with distinct needs, such as the elderly [9] and VI individuals, face even greater challenges when accessibility and usability issues arise. In the context of contemporary and nuanced web and mobile applications, accessibility issues pose significant hurdles that are difficult to overcome. These issues place an additional burden on users, negatively impacting their performance and causing stress and frustration.

When navigating the digital landscape, individuals with vision impairments can encounter a spectrum of accessibility challenges. While certain websites adhere closely to WCAG guidelines ensuring ease of use for screen reader users, others may present more pronounced obstacles. Examples include inadequate feedback, missing alternative text, lack of button titles, difficult navigation, and cluttered pages with much information [5]. Guidelines to enhance web accessibility, such as the Web Content Accessibility Guidelines (WCAG 2.1) have long been developed and adapted. These guidelines have been mandated in a number of countries [10]. Yet, with the rapid evolution of online technologies, accessibility remains an open issue. Accessibility issues faced by people with vision impairment can be difficult and frustrating [11]. For example, in a study where VI users experienced accessibility issues, they reported negative emotions such as tension and irritability [12].

The study's objective is to examine the physiological responses of VI and sighted users to accessibility and usability issues when browsing the web through smartphones. Moreover, to compare the VI and sighted individuals' physiological responses to common usability problems. The paper concludes with a set of recommendations for web designers to reduce frustration and increase inclusiveness.

2 Background

2.1 Accessibility

Addressing the realm of accessibility challenges, particularly among individuals who are blind or visually impaired, remains an ongoing concern, especially within the context of mobile touchscreen devices. A study that investigated whether these problems are adequately addressed in the recent release of the Web Content Accessibility Guidelines

(WCAG 2.1) discussed that a multitude of critical accessibility problems are often overlooked during the development of mobile apps and websites [13]. The study revealed a total of 34 major problems and highlighted the need for better adherence to guidelines and improvements in WCAG 2.1 to effectively address the challenges faced by blind and VI users on mobile devices. Some of the critical accessibility problems identified includes dynamic content, carousel, inaccessible captcha, difficult navigation, and labelling [13].

Venturing beyond the scope of mobile devices, the spectrum of accessibility complications expands its reach to encompass social media platforms as well [14]. A significant issue identified was the lack of access to embedded content in images. An evaluation of Facebook posts revealed that almost half of the images contained inaccessible text for VI users. Despite social media companies' efforts, text within images remains inaccessible. The authors highlight the inequality resulting from partial accessibility and emphasizes the need for inclusive design and development practices. Accessibility problems can also arise during software updates and website redesigns, leading to user frustration [15]. This can occur when organizations lack formal processes to ensure that changes comply with accessibility requirements. The negative consequences of such accessibility issues were evident in user surveys, highlighting the need for better consideration of accessibility during the update process.

Instances of accessibility-related service shortcomings on retailer websites trigger adverse consumer responses among those with visual impairments [16]. These reactions include negative word-of-mouth, sharing negative experiences on social media, and even the avoidance of the retailer's other accessible sales channels. The study [16] emphasized the lasting consequences of accessibility problems, leading to a decline in customer satisfaction and engagement. The prevalence of content accessibility challenges on popular websites hinders effective information access for users with disabilities [17]. In fact, significant majority of websites fail to meet accessibility standards, hindering users with disabilities from accessing information effectively [17]. Continuous reporting of content accessibility, as opposed to a dichotomous approach, has been suggested to improve accessibility levels. This highlights the needs for comprehensive accessibility practices in ensuring a more inclusive and frustration-free browsing experience.

Frustrations can arise when users encounter difficulties in locating specific links due to their similarity in starting words or inadequate keyword usage [18]. Clear and intuitive navigation systems are crucial to prevent such frustrations and ensure users can access desired information efficiently. Overall, the impact of accessibility problems on user

frustration is evident across various studies [13–18]. These problems hinder individuals with disabilities from fully engaging with digital platforms. In a world of continually advancing internet technologies, it becomes crucial to prioritize comprehensive accessibility strategies to enhance user experiences, promote inclusivity, and alleviate frustrations and other undesired experiences associated with inaccessible content.

2.2 Understanding frustration

Frustration is a form of physiological arousal that yields a physiological response. In the Pleasure-Arousal-Dominance (PAD) model, physiological arousal quantifies the strength of the emotion [19]. Physiological arousal is used to measure strong emotions such as the stress [20], frustration [21], attention [22], and anxiety [23]. Identifying the user's arousal can help find areas that frustrate the user, making frustration an important factor in the user experience. For example, when analysing a user's gaming experience, low arousal can indicate that the user is bored; therefore, developers should create games that are exciting and engaging. [24]. Furthermore, detecting arousal in the learning domain can assist teachers in determining whether the learning material is engaging [25]. Arousal can also be used as a proxy for attention detection [26]. In addition, frustration can occur because of accessibility and usability issues, when the user is unable to achieve their goal due to an inaccessible website. Individuals with disabilities, such as vision impairment, face numerous challenges that result in high arousal that hinders them from completing a simple task. Timed responses while filling out a form, hamburger menus, inaccessible CAPTCHA, dynamic content, and a variety of other challenges exist [13].

Frustration can have serious implications. Excessive frustration, for example, can lead to physical and psychological illnesses such as depression, aggressive behavior, insomnia, nightmares, and anger [27]. Furthermore, excessive frustration can lead to a loss of confidence and self-esteem due to constantly being unable to achieve the desired goal [27]. Increased blood pressure, raised body temperature, constricted blood vessels, and other reflexes are examples of frustration [28].

2.3 Arousal detection

Arousal can be detected using both subjective and objective measures. Questionnaires such as the Self-Assessment Manikin (SAM), in which participants answer questions about their emotions after completing an experiment, are examples of subjective measures. Subjective measures, while simple to use, are not very accurate because participants can alter their

responses because of social pressure. Objective measures, on the other hand, detect emotions through physiological signals expressed by the human nervous system and thus cannot be controlled by the individual. Several studies have used sensors to detect physiological signals such as heart rate and galvanic skin response to gain real-time insights into participants' emotional states. Although used for emotion detection, physiological signals are primarily used for biofeedback. GSR, for example, can be used for biofeedback training, in which the individual is trained to gain voluntary control over their autonomic response; this can be used in an epilepsy therapy [29]. Similarly, ECG is primarily used to diagnose and monitor heart conditions.

When used for emotion detection, physiological signals can possess some limitations as they could be easily distorted. For example, the EEG has sensitive electrodes that are attached to the participant's head, yet the participant must remain still as any small body or head movements may accidentally detach the electrodes from the scalp [30]. Similarly, noise can be associated with the ECG due to the high sensitivity of the data [31]. The ET, although proven to be effective in detecting frustration [21], cannot be used since its limited to the sighted demographic and cannot be used with VI users. The GSR, is sensitive to temperature [32], motion [33], and has minor latency issues [34]. Yet, these constraints can be alleviated by controlling the temperature, limiting motion, and calculating latency that is 1–5 s [35]. The main advantages of GSR over other physiological sensors are its ease of use, low cost of sensors, and simple visual interpretation of the signals. People with vision impairment can also use the GSR signal, unlike other physiological signals such as ET. Thus, for this study, the GSR is the signal used for arousal detection.

GSR is the change in the electrical conductivity of the skin because of sweat production. The electrodes can be positioned in many locations depending on the sensor; for example, they can be placed in the pointer, middle, and index finger (e.g., A tool called Shimmer31). It can also be used as a wristband (e.g., Empatica E4). The GSR signal is composed of phasic and tonic signals. There are two types of inferences drawn from the phasic signal: the non-specific skin conductance responses (NS-SCRs) and event-related skin conductance responses (ER-SCRs). The NS-SCRs are responses that are not associated with an event or stimuli, while the ER-SCR is a result of interacting with a triggering stimulus [36]. In this study, the ER-SCRs are analyzed, and details about the methodology are explained in Sect. 3.

2.4 Frustration detection in HCI

In healthcare, researchers, doctors, and healthcare professionals are intrigued by the prospect of using physiological

signals to monitor patients and train medical students. GSR and HRV, for example, have been used to monitor patient stress and prevent addictive behavior [37]. A previous study aimed to validate the feasibility of using physiological signals to continuously monitor stress and prevent addictive behaviors to prevent alcohol relapse [37], while two physiological signals were monitored, GSR and HRV. The results revealed that the GSR data was primarily clean (87.86%) and was consistent with previous research regarding peaks per minute.

Physiological signals are applied in the learning field to detect frustration in students. This is a field of interest since adolescents spend most of their time in a school setting and are exposed to higher social stressors than children and adults during this period [38, 39]. Students' emotional states [40] and engagement with the learning material [41–43] were explored in a class setting. Exploring an ecologically valid environment such as a school, rather than a controlled environment, was addressed [44]. Similarly, researchers are working on machine learning models to detect frustration in real-world situations, such as when using a computer to complete a search query or book a flight.

In 2019, Matthews et al. examined the impact of usability issues on participant frustration using eye tracking [21]. Pupillary response and gaze behavior were used to examine the relationship between arousal and usability problems [21]. The pupillary response was used to track changes in arousal, whereas gaze behavior was used to determine focal attention. The stimulus consists of common usability problems that caused frustration. The participants were required to complete a series of web-based tasks, each in a normal and disruptive mode. Normal tasks were completed without any webpage alterations while disruptive mode included a usability problem. Pop-ups, mouse malfunction, blue screen of death, and session timeout were all examples of problems in the disruptive mode. The results demonstrated that this method could distinguish between normal and disruptive tasks. When users completed a disruptive task with a session timeout, their arousal level was significantly higher. The increase in arousal is due to completing the task in double the time required for the other tasks, thus affecting the accumulated arousal. When the rest of the tasks were compared, it was discovered that there was no significant difference between the usability problems.

3 Method

The user study took place in a controlled setting, and participants were recruited by contacting potential participants and snowballing. It included 29 individuals divided into two

groups: VI ($n=13$) and sighted ($n=16$). The study consisted of qualitative data collection methods (e.g., questionnaire and interview) as well as quantitative methods (e.g., measuring GSR signal). The quantitative data was acquired through the Empatica e43 wristband to record the electrodermal activity of the participants.

This study adopted the same method from [21] by employing a 4X2 within-subject design in which VI and sighted participants completed four tasks using a mobile phone in a frustrating and non-frustrating interaction. The non-frustrating tasks were not altered and should be completed without facing any difficulties. The frustrating tasks were manually programmed into well-known governmental webpages to induce frustrations through accessibility and usability problems. The frustrations are chosen specifically for each group based on causes of end-user frustrations for the sighted [45] and accessibility issues frustrating the VI users [13]. The sighted individuals' alterations include slow performance and page refresh. While the VI individuals' alterations include a non-searchable drop-down button and a non-responsive menu. The mutual alterations for sighted and VI individuals are pop-ups and session timeout.

The level of arousal is identified by calculating the number of peaks within each task [9, 46, 47]. The most commonly used way of calculating the level of arousal is through the average for a participant [48]. The average of SCR in a task is calculated and used as the level of arousal [49]. The collected data is then examined quantitatively and qualitatively, starting from data cleaning through artefact removal, followed by signal decomposition and peak detection, and statistical analysis of the quantitative data.

3.1 Study hypotheses

Hypothesis 1 (H1) Frustrating tasks cause significantly higher levels of arousal.

The study aims to investigate the physiological reaction of the VI and sighted participants to usability problems while completing the frustrating task. Previous studies have demonstrated that usability problems, such as session timeout, slowed internet connection, and system malfunction can frustrate end users [9, 21, 45].

Hypothesis 2 (H2) VI individuals show significantly higher levels of arousal when encountering usability problems in comparison to sighted individuals.

The study also aims to compare the physiological responses of VI and sighted users when encountering accessibility and usability problems. According to a recent evaluation of the research on mental health outcomes and existing therapies for individuals with VI, anxiety, and stress levels can be higher in those who have visual impairments [50].

Table 1 Participant demographic information

	Sighted group (n = 16)	VI group (n = 13)
<i>Gender</i>		
Female	11 (68.8%)	5 (38.5%)
Male	5 (31.3%)	8 (61.5%)
<i>Age</i>		
21–30	10 (62.5%)	6 (46.2%)
31–40	4 (25%)	6 (46.2%)
>41	1 (6.3%)	0
<i>Mobile phone over computer</i>		
Yes	12 (75%)	13 (100%)
No	4 (25%)	0
<i>Consumption time</i>		
1–2 h	5 (31.6%)	0
2–3 h	0	2 (15.4%)
3–4 h	1 (6.6%)	4 (30.8%)
4–5 h	0	1 (7.7%)
5–6 h	7 (43.8%)	0
6–7 h	0	1 (7.7%)
More than 7	3 (18.8%)	5(38.5%)

Hypothesis 3 (H3) Slow webpage performance yields the highest levels of arousal when compared to other usability problems for the sighted. The flight booking task yields the highest levels of arousal when compared to other usability problems for the VI.

The study aims to identify the usability problems that yield the highest level of arousal. Previous research has indicated that session timeout was proven to yield the highest significant level of arousal when compared to other frustrations [21]. Yet the high increase in the arousal for the session timeout is associated with having more time to complete the task [21]. This hypothesis aims at exploring other usability problems and their physiological impact on participants. According to Google, 53% of webpage visitors are likely to leave a website if it loads for more than 5 s [51].

3.2 Participants

The participants were recruited under a set of conditions: (a) Arabic speakers, (b) iPhone users (c) age range of 18 to 65, with no health conditions related to cardiovascular and respiratory diseases, and no severe psychological disorders (d) must avoid eating, drinking caffeinated beverages, and smoking two hours before the study. The term VI encompasses individuals who are either partially sighted or completely blind. In this work, our primary focus is on individuals who have no vision, i.e. completely blind and rely on speech-based screen readers as their means of interacting with technology. The choice of Arabic language and iPhones as the primary device was

informed by consultation with the technology officer at the Qatar Culture Centre for the Blind, who confirmed that VI users in Qatar predominantly speak Arabic and commonly use iPhones. Demographic information for the participants is summarized in Table 1.

3.3 Procedure

3.3.1 Pre-study questionnaire and procedure

Participants completed a questionnaire regarding their demographics and mobile phone usage habits. They were provided with the Empatica E4 wristband to wear on their non-dominant hand, except for VI participants who wore it on their less-used hand due to smartphone accessibility. VI participants were also asked to adjust Voiceover settings as desired (e.g., speed and language). After wearing the wristband, participants were instructed to relax while listening to soothing music to collect baseline data and allow the wristband to adapt to skin temperature. Subsequently, participants engaged in a series of frustrating and non-frustrating tasks.

Participants were not informed of the time specified to complete each task, to reduce the cognitive load on the participants. Participants were instructed to continue attempting to complete the task until directed to do so. Participants had three minutes to complete each task. The flight booking task received an additional minute for data entry, flight selection, and traveller information, as determined in the pilot study.

To subjectively report whenever a participant felt frustrated, they were advised to click on the circular button on their wristband when they feel frustrated. This information will be used as the basis for further investigation. After the completion of each task, participants had to set relaxed for a couple of minutes while listening to the same relaxing music. During that time, they had to rate the task on a 5-point Likert scale that indicates how frustrating the task was.

3.3.2 Post-study questionnaire

After the task completion, the participants were asked to order the tasks, from least to most. Additionally, they filled out a familiarity questionnaire to determine prior exposure to the presented webpages. This step aimed to identify any potential bias in the data and ensure accurate analysis. Lastly, participants engaged in a semi-structured interview where they shared experiences of encountering frustration while using technology, providing valuable insights for further investigation.

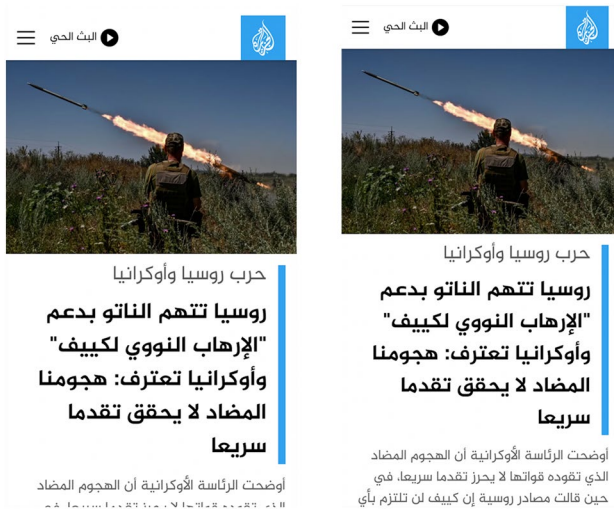


Fig. 1 Safari browser (left), GearPro Browser (right)

3.4 Instruments

3.4.1 Materials

An Empatica E4 wristband was used to capture the participant's GSR data. In addition, a frame grabber was utilized to capture the screen. An iPhone 12 was used to interact with websites that run on the iOS operating system. Gear Pro¹ was selected as it is one of the browsers that support the installation of user scripts in iOS, unlike Safari and Chrome. Gear Pro is a browser that looks similar to common browsers such as Safari. Figure 1 shows a screenshot of the GearPro browser in comparison to the Safari browser.

3.4.2 Webpages used

To identify the webpages to be explored, we visited Alexa's top visited webpages in Qatar, to try to control for the lack of familiarization [52]. Alexa is an Amazon webpage that shows website visits in the world. While choosing the webpages, each website had to meet the following criteria: (a) belongs to a local organization, (b) must have an Arabic version, (c) the task must not require any personal information, (d) had to be accessible checked via Mada's monitor for accessibility.²

¹ <https://gear4.app>

² <https://monitor.mada.org.qa>

3.4.3 Stimuli and task selection

Table 2 shows the tasks assigned to the VI participants, the website, the task description, and the form of induced frustration, followed by Table 3, which shows the tasks assigned to sighted participants. The Task ID is divided into three parts, the first is the number of the task (e.g., T1), the second is the participant's group (e.g., V for VI and S for sighted), the final is the type of the task (e.g., N for the non-frustrating task and F for the frustrating task).

Tasks order was counter-balanced to eliminate any order effects. Groups were generated and participants were assigned one of these groups. For example, a governmental webpage is always followed by a non-governmental webpage. In addition, a frustrating task is always followed by a non-frustrating task and vice versa.

3.5 Data analysis

Figure 2 summarizes the data analysis steps conducted. The data analysis starts by dividing the GSR file into the different tasks completed by the participant. This is followed by artefact removal. The GSR signal is then decomposed into phasic and tonic signals. The data is then fed into sparsEDA [53] for peak detection. Finally, a statistical analysis of the peak count is conducted.

3.5.1 Data pre-processing

We used a data-cleaning approach and pre-processed the GSR data before analyzing it. This is a common practice since the GSR data is often noisy and sensitive to motion and temperature changes. We used the same pre-processing approach as [54] with some alterations such as dividing the GSR file based on the tasks. The pre-processing included the following steps: (1) Task Division, (2) elimination of artifacts, and (3) decomposition.

3.5.1.1 Task division The procedure followed when recording the GSR data using the Empatica E4 wristband is that an event is recorded at the start and end of each task to denote it as the start and end of the task. So, the first step when pre-processing the data is to divide each task by start, end, and valence markers.

3.5.1.2 Elimination of artifacts Peaks caused by motion artifacts can be mistaken for a peak and thus the quality of data analysis may be at risk. As a result, removing artifacts is a necessary step in data pre-processing. We utilized a machine learning algorithm that automatically identifies areas of artifacts proposed by [53] and also employed by [55, 56]. A segment of the 5-s window was classified as either noisy (includes artifacts) or clean. This filter elimi-

Table 2 Experiment tasks for VI participants

Website	Task description	Simulated frustration
Educational	Search for an expedition	None
Governmental	Navigate to another page	None
Flight Booking	Book a trip from London to Doha	None
News	Find a program name	None
Educational	Finding a school	A long list with no search
Governmental	Navigate to another page	Not accessible drop-down
Flight Booking	Book a trip from Jordan to Doha	Session timeout
News	Find some news	Many pop-ups

nates motion artifacts while maintaining the GSR peaks. This process was carried out for each GSR data in an online tool called EDA Explorer [53]. To eliminate motion arti-

facts, the GSR signal in that time slot is replaced by linear regression of the previous slot with a non-artifact signal and the following slot with a non-artifact signal [57]. An example of motion artifact detection is seen in Fig. 3.

Table 3 Experiment tasks for sighted participants

Website	Task description	Simulated frustration
Educational	Search for an expedition	None
Health	Seeking information	None
News booking	Book a trip from London to Doha	None
News	Find a program name	None
Educational	Search for an expedition	Page reloads
Health	Seeking information	Slow performance
News booking	Book a trip from Jordan to Doha	Session timeout
News	Find some news	Many pop-ups

3.5.1.3 Decomposition GSR signals are composed of different types of signals, phasic and tonic signals. From the phasic signals, the ER-SCR and NS-SCR can be extracted. The tonic signals can be decomposed into skin conductance levels (SCL). In this study, the ER-SCR is utilized, which is a peak that occurs as a result of an event. The event in this study is the appearance of accessibility and usability issues that are programmed into the tasks. The sparsEDA algorithm [57], was used in the study to decompose GSR signals.

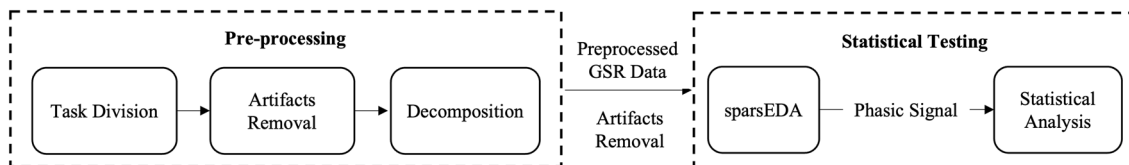


Fig. 2 Overall method for GSR signal analysis

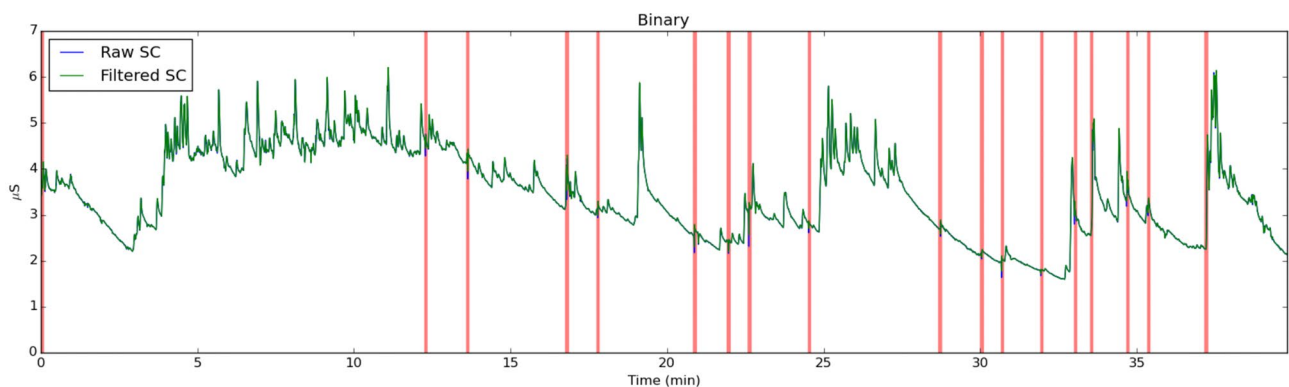


Fig. 3 Example of motion artefact detection

Table 4 Mean and standard deviation of peaks per task for sighted and VI users

	Sighted		VI	
	Mean	Sd	Mean	Sd
Flight booking (F)	3.063	3.889	3.846	3.648
Flight booking (NF)	0.938	0.998	3.154	3.976
News (F)	1.25	1.571	1.692	2.097
News (NF)	1.188	1.377	1.462	2.145
Educational (F) (Page refresh)	1	1.366		
Educational (NF) (Page refresh)	2	2.733		
Health (F)	2.188	2.857		
Health (NF)	1.313	2.272		
Governmental (F)			1.769	2.421
Governmental (NF)			3.154	3.848
Educational (F) (Inaccessible menu)			0.923	1.935
Educational (NF) (Inaccessible menu)			2	3.162

3.5.2 Statistical testing

To understand the distribution of the data, first, the data is cleaned, and the mean, and standard deviation are calculated. The measures of spread were also used to understand the normality of the data and its distribution. According to [58], the top three statistical characteristics to identify stress are the maximum peak amplitude, the number of peaks (SCR), and the average of the GSR signal. In this study, a similar methodology is adopted, employing the number of peaks for statistical analysis. The GSR peaks can be calculated using sparsEDA [57]. After the phasic signal is decomposed, the peaks are identified and segmented to correspond with each task. The statistical tests have been conducted using R Studio.

The Wilcoxon Signed-Rank statistical test is used to understand the effect of the mode of interaction i.e., frustrating, and non-frustrating, on the VI and sighted participants separately. The dependent variable is the number of peaks, and the independent variables are the type of interaction i.e., frustrating, and non-frustrating. This answers the first hypothesis. To answer the second hypothesis, the Mann–Whitney test is used to compare the level of arousal between the sighted and VI participants. The dependent variable is the number of peaks, and the independent variables are the frustrating task peaks for each participant group. To answer the third hypothesis, the Wilcoxon Signed-Rank statistical test compares the different frustrating tasks and identifies which task is most frustrating. The dependent variable is the number of peaks, and the independent variables are the frustrating task peaks for the sighted and VI participants separately. This statistical test was previously used by [21].

Table 5 Statistical testing for sighted users' tasks

Task	Usability issue	Wilcoxon signed-ranks test	Significance result
Educational task	Page refresh	$Z = -2.07$, $p = .98$	Not significant
Health task	Slowed Internet	$Z = .84$, $p = .22$	Not Significant
News task	Popup Ads	$Z = -1.68$, $p = .96$	Not significant
Flight booking task	Session timeout	$Z = 2.72$, $p = .009$	Significant

Table 6 Perceived frustration score for sighted users

Task	Average perceived frustration score
Educational task	Frustrating task: 4.38 Non-Frustrating task: 0.44
Health task	Frustrating task: 3.12 Non-Frustrating task: 2.18
News task	Frustrating task: 4.56 Non-Frustrating task: 0.31
Flight booking task	Frustrating task: 4.69 Non-Frustrating task: 1.5

4 Results

4.1 Task information

Descriptive statistics, including the mean and standard deviation of the peaks per task, were calculated. Among VI users, the flight booking task in both frustrating and non-frustrating interactions, along with the non-frustrating interaction of the governmental task, exhibited the highest mean of peaks as seen Table 4. This suggests that these tasks elicited greater arousal levels compared to other tasks for VI users. For sighted users, the flight booking task during frustrating interaction resulted in the highest mean of peaks. This highlights the need for further investigation into the effects of session timeout on the participants and their physiological responses.

4.2 Answering study hypotheses

The Wilcoxon Signed-Ranks Test examined the relationship between recorded peaks and interaction type (frustrating vs. non-frustrating) to address **H1**. Table 5 presents task details, Wilcoxon Signed-Ranks Test results, and significance. Table 6 shows participants' perceived frustration scores, measured on a 5-point Likert scale completed after the task.

Table 7 Statistical testing for VI users' tasks

Task	Usability issue	Wilcoxon signed-ranks test	Significance result
Educational task	Unsearchable List	$Z = -2.29$, $p = .99$	Not significant
Governmental task	Non-responsive Menu	$Z = -0.94$, $p = .85$	Not significant
News task	Popup ads	$Z = -1.51$, $p = .94$	Not significant
Flight booking task	Session timeout	$Z = 0.63$, $p = .28$	Not significant

Considering a significance level of 0.05, the results indicated that the flight booking task had significantly higher peaks during the frustrating interaction ($Z = 2.72$, $p = 0.009$). However, the educational, health, and news tasks did not significantly differ in peak count between frustrating and non-frustrating interactions ($Z = -2.07$, $p = 0.98$), ($Z = 0.84$, $p = 0.22$), ($Z = -1.68$, $p = 0.96$). This supports H1, suggesting that only the flight booking task led to significantly higher arousal levels in the frustrating interaction for sighted users. The participants' perceived frustration scores aligned with the interaction type. The frustrating tasks received higher scores, while non-frustrating tasks received lower scores.

There were no significant differences between types of interaction for VI participants across any of the tasks. Using a significance level of 0.05, the educational, governmental, news, and flight booking tasks did not significantly differ in peak count between frustrating and non-frustrating interactions ($Z = -2.29$, $p = 0.99$), ($Z = -0.94$, $p = 0.85$), ($Z = -1.51$, $p = 0.94$), ($Z = 0.63$, $p = 0.28$). The participants' feedback, such as perceived frustration scores, contradicted the interaction type. For instance, a frustrating task like the educational task received a low score, while a non-frustrating task like the flight booking task received a high score. These findings reject the null hypothesis (H1) and demonstrate that none of the tasks significantly increased arousal during the frustrating interaction for VI users as seen in Table 7.

This could be attributed to the accessibility challenges experienced by VI users during the non-frustrating interactions. For example, in the flight booking task, users encountered a modal dialog: a window that opens within a website. Within this task, users needed to fill in trip information through a modal dialog, such as filling out a form. The lack of updates from the screen reader for the modal dialog might have frustrated users, as they were not made aware of its presence. Consequently, participants proceeded to complete the data entry without interacting with the session timeout, which is programmed to appear after completing the data entry page and clicking the submit button. It is worth noting that both frustrating and non-frustrating interactions were treated equally. When asked about prior experience with the flight booking webpage, only two participants reported

Table 8 Perceived frustration score for VI users

Task	Perceived frustration score
Educational task	Frustrating task: 0.23 Non-frustrating task: 1.38
Health task	Frustrating task: 2.77 Non-frustrating task: 1.58
News task	Frustrating task: 4.38 Non-frustrating task: 1.85
Flight booking task	Frustrating task: 3.62 Non-frustrating task: 3.62

Table 9 Comparison between participants in the flight booking task

	Type of interaction	Mann–Whitney test	Significance
Flight booking task	Frustrating	$W = 34$, $p = .65$	Not significant
	Non-frustrating	$W = 9$, $p = .03$	Significant

Table 10 Comparison between participants in the news task

	Type of Interaction	Mann–Whitney Test	Significance
News task	Frustrating	$W = 12$, $p = .83$	Not significant
	Non-frustrating	$W = 6$, $p = .44$	Not significant

having previous experience, indicating that the majority were still exploring the page for the first time. This observation may explain why the perceived frustration scores were similar for both types of interactions, as shown in Table 8.

The Mann–Whitney test compared arousal levels between sighted and VI participants during the flight booking task, addressing H2. No significant difference was found in the frustrating interaction ($W = 34$, $p = 0.65$). However, in the non-frustrating task, VI individuals had significantly higher arousal levels compared to sighted individuals ($W = 9$, $p = 0.03$). Thus, H2 is confirmed, indicating that VI

Table 11 Results of the Wilcoxon test comparing arousal between each task for Sighted Group

Group 1	Group 2	Z	P value
Flight booking	News task	1.69	0.10
Flight booking	Health task	1.72	0.262
Flight booking	Educational task	2.04	0.045*
News task	Health task	−.06	1.00
News task	Educational task	1.07	0.308
Health task	Educational task	1.01	0.343

* = $p < 0.05$ **Table 12** Results of Wilcoxon test comparing arousal between each task for VI Group

Group 1	Group 2	Z	P value
Flight booking	News task	1.778	.041*
Flight booking	Governmental task	.889	.199
Flight booking	Educational task	1.956	.028*
News Task	Governmental task	−.267	.622
News Task	Educational task	1.244	.118
Governmental Task	Educational task	1.014	.176

* $p < .05$

individuals exhibit significantly higher arousal levels compared to sighted individuals.

Similarly, the Mann–Whitney Test was used to compare the level of arousal between the sighted and VI participants while completing the news task to answer **H2**. There is no significant difference in the frustrating and non-frustrating type of interaction for the news task with a Mann–Whitney result of ($W = 12, p = 0.83$) and ($W = 6, p = 0.44$). This answers **H2** by rejecting the null hypothesis and demonstrating that VI users did not exhibit significantly higher arousal levels compared to sighted users when completing the news task (Tables 9, 10).

A Wilcoxon Signed-Ranks Test was used to analyze the difference between the frustrating tasks to answer **H3**. The following tables show the results for the sighted and VI participants respectively. The pairwise comparison of the websites is demonstrated in Table 11 for the sighted participants. The results revealed that the arousal level i.e., the number of peaks, is significantly higher in the flight booking task than in the educational task with a Wilcoxon Signed-Ranks of ($p = 0.045$). Thus, it cannot be deduced that one of the tasks is considerably more arousing than the other tasks. This answers **H3** by rejecting the null hypothesis and demonstrating that slow webpage performance will not yield the highest levels of arousal when compared to other usability problems for the sighted.

The pairwise comparison of the websites is displayed in Table 12 for the VI participants. The results revealed that the flight booking task is significantly more arousing than the news task and the education task with a Wilcoxon Signed-Ranks of ($p = 0.041$) and ($p = 0.028$). This answers **H3** by accepting the null hypothesis and demonstrating that the flight booking task yields the highest levels of arousal when compared to other usability problems for the sighted.

5 Discussion

Creating usable and accessible webpages is integral to a positive user experience. Over the past decades, cross-disciplinary research, spanning HCI, interactive design, accessibility, and human factors, has enriched our understanding of user interactions with technology. Advancements in digital devices enable objective assessment of user experiences, using tools like GSR, Eye Tracking, and ECG, uncovering effects on human well-being. These tools reveal the broader impact of accessibility issues, extending beyond interaction to harm the human body. Global mandates now require website and product accessibility, holding organizations accountable for compliance. Despite frameworks for assessment, many entities overlook accessibility, leading to user frustration during online tasks.

Frustration and stress have been assessed by wearables during smartphone interaction that yielded a diverse range of applications. Notably, Sano et al. introduce an inventive stress recognition scheme that seamlessly combines wearables and smartphones to conduct a comprehensive stress analysis [59]. Similarly, Reimer et al. contribute a stress recognition pilot system that merges physiological signals such as HRV with contextual information through wearables [60]. This innovative approach enables a nuanced assessment of stress levels during smartphone interactions. Expanding on this theme, L. Zhu et al. explore the utilization of wrist-based EDA signals from wearables to predict stress levels, achieving high accuracy rates through the implementation of machine learning classifiers [61]. Additionally, Ng et al. (2022) undertake the development of a machine-learned model, which forecasts physiological and perceived stress for the following day, integrating sensor-based and ecological momentary assessment (EMA)-based features. Furthermore, T. Kim et al. introduce MindScope, a mobile app incorporating personalized stress prediction algorithms rooted in smartphone data, facilitating enhanced stress comprehension and management [62]. This comprehensive exploration is complemented by an investigation into the impact of algorithmic explainability on stress reduction and user self-reflection. The following sections describe the ongoing accessibility challenges and highlight the impact

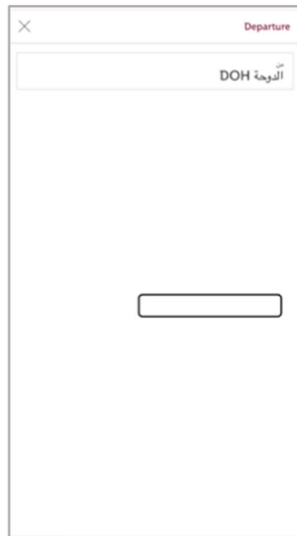


Fig. 4 Accessibility issue in the modal dialog

of usability issues on user frustration, underscoring a vital area for improving user experiences through objective measurement.

5.1 Accessibility problems for VI users

Despite encountering various issues, the findings revealed that the most frustrating experiences for VI users were related to accessibility. Even in non-frustrating and feasible tasks, VI users faced significant challenges due to numerous accessibility issues encountered during the tasks. These issues often hindered them from successfully completing the tasks. For instance, in the flight booking task, accessibility issues such as inaccessible menu items, combo boxes, and insufficient feedback hindered completion, resulting in a 0% success rate for some non-frustrating tasks.

In the governmental task, VI participants encountered an inaccessible menu, which impeded completion of the non-frustrating task yielding less than half of the VI participants completing the task. Difficult navigation through extensive menus with multiple levels of sub-menus was identified as a high-severity accessibility issue in previous research [5, 13, 64]. VI users face challenges in locating and navigating hamburger menus and long lists with subheadings. Accidental activation of submenus often leads to an undesired webpage navigation [13]. These issues place a cognitive load on VI users, requiring careful navigation and hindered by insufficient heading levels in content. The results align with existing literature [5, 13, 64, 65].

The combo box posed accessibility issues in multiple tasks. In the flight booking form, the combo box for "to" and "from" destinations, along with the date picker and drop-down list, presented challenges. Clicking on the

combo box triggered a modal dialog where users could enter the destination. However, this modal dialog proved inaccessible to screen readers. Despite the dialog being open, the screen reader continued to focus on the underlying webpage content, as depicted in Fig. 4. Consequently, users were unable to type the destinations and faced significant barriers in completing the task. This accessibility issue is classified as severe according to [13] and hinders the user from completing the task. Similarly, in the educational task, an inaccessible combo box restricted users from listening to drop-down options, limiting their ability to interact with the task effectively. V01 has found this to be frustrating; he described:

“What if I don’t know what to type? You’ve already given me the requirement for study purposes, but in a natural setting, it will be difficult to complete this task, especially since I don’t know the options.” — [V01]

Additionally, combo boxes lack feedback and screen readers fail to report changes within them. Offering a list of options in combo boxes may enhance accessibility for VI individuals and reduce the potential for errors. This issue has been highlighted in previous studies [13, 66–68], where inadequate feedback was frequently mentioned as a problem. In addition, problems faced by sighted and VI users can have a more severe impact on VI users, such as inadequate feedback [5].

In the educational task, when users type the first letter of a word, the combo box auto-fills with the first matching word. However, this change is not rendered by the screen reader. For example, when typing "elementary school," users only need to type "e" and the remaining term will be added to the combo box. While this feature is helpful for sighted users, it proved frustrating for VI users as they were unaware of the term being added. Participants would attempt to input the second letter, only to be surprised by a complete word already occupying the combo box. The lack of sufficient feedback resulted in frustration for VI participants [13]. Technical enhancement recommended by the WCAG 2.1 has long proposed multiple solutions to address this issue.

Date pickers present challenges due to their complex controls and the multitude of values to select [13, 69, 70]. In this study, only one participant managed to add the correct date during their second attempt, indicating the high level of inaccessibility of the date picker. To enhance accessibility, date pickers should prioritize the month view as the default, with the option to switch to the year view using a pinch gesture and provide either a spinner or an editable text box for entering the day [69]. Flight booking webpages can adopt a similar user interface design to improve accessibility for VI individuals. However, the combination of an inaccessible modal dialog, destination text boxes, and the date picker made this task unattainable in both frustrating

and non-frustrating formats, resulting in a frustrating user experience.

Despite the presence of inaccessible UI components like an unsearchable drop-down list, participants did not perceive them as frustrating. The unsearchable drop-down list refers to a lengthy list that VI users must navigate sequentially to find their desired option, such as a list of countries. For instance, in the educational task, participants V07 and V12 expressed their opinions and did not express dissatisfaction with the long drop-down list. For example, V07 responded:

“I’m used to the long drop-down; I don’t find it frustrating.” — [V07]

Furthermore, V12 added:

“To reduce the possibility of error, I prefer having a list of options to choose from rather than typing and searching.” — [V12]

Although rated as a moderate issue by VI individuals [13], this task reveals that participants did not perceive the accessibility problem as frustrating, contrary to previous findings in the literature [13, 71]. For a more accessible approach, it is recommended to show all options types of drop-down menus [72]. This can also be attributed to participants becoming accustomed to such accessibility issues and no longer finding them troublesome [73]. Additionally, participant V12 noted that having a pre-defined list is preferable to manually type a country. Implementing a similar approach in the flight booking task, particularly for selecting destination and departure countries, could enhance the ease of form completion for more VI users.

In the flight booking task, none of the VI participants experienced a session timeout due to the accessibility issues previously mentioned. The session timeout was programmed to occur upon completing the data entry page, which includes destination selection and date inputs. No significant difference in frustration levels was found between sighted participants facing usability issues (session timeout) and VI participants facing accessibility problems (Mann–Whitney test: $W = 34, p = 0.65$). However, in the non-frustrating task, VI participants exhibited significantly higher arousal levels than sighted participants, highlighting the impact of accessibility issues on arousal despite the absence of usability problems. This emphasizes that the primary concern lies in accessibility rather than usability. This is consistent with the literature where blind users were more severely impacted by accessibility problems than sighted users [5, 74].

5.2 Usability problems for the sighted users

During the simulations of various usability problems, such as pop-up ads, page refresh, slow network, and session timeout, only the session timeout proved to be frustrating. The

session timeout differed in design compared to the other tasks, appearing after participants completed the data entry page, initially giving the impression of a non-frustrating task without obvious usability issues. In contrast, the systematic occurrence of pop-up ads and page refresh every 3 s contributed to the session timeout being reported as more frustrating.

Extensive research has investigated pop-up ads as profit-yielding strategies, emphasizing the significance of the first 10 s of a page visit in users' decision to stay or leave [75]. Users' heightened skepticism during this initial period, resulting from previous encounters with poorly designed webpages, amplifies the likelihood of departure. Conducting a preliminary study to observe participants' behavior can help identify the optimal timing for personalized pop-up ads, enhancing their natural flow and enabling a more thorough examination of the effect of popup ads on user frustration.

In the educational task, participants anticipated and accepted the systematic page refresh, resulting in a lack of frustration. Similarly, in the news task, participants were initially surprised by the first ad occurrence but systematically closed subsequent ads without waiting for loading. These observations align with the concept of priming, where prior information unconsciously influences behavior [76, 77]. Participants' low expectations of technology performance in these tasks reduced their susceptibility to frustration compared to those with high expectations [78]. Thus, it can be inferred that low expectations played a role in minimizing frustrations. It is important to note that this usability problem lacks ecological validity, meaning its applicability beyond the controlled lab environment is limited [79].

Moreover, conducting the experiment in a controlled environment led participants to anticipate encountering technological problems. Despite the careful selection of tasks that resemble participants' daily activities, five participants specifically mentioned that after experiencing the first usability problem, they anticipated encountering further issues in subsequent tasks. S07 noted:

“I understood what’s happening in the study; you’re altering the webpages. [S07]

In a previous study comparing various usability issues (e.g., session timeout, system failure, popup ads, and mouse malfunction), session timeout elicited the highest level of arousal when no specific time limit was set for task completion [21]. This can be attributed to the relatively longer duration required to complete the session timeout task compared to other tasks, leading to a cumulative effect on arousal levels [21]. Intrigued by these findings, we sought to investigate whether setting a time limit for task completion would yield similar results. Our findings confirm that even when the time to complete tasks is controlled, session timeout maintains a significant difference between frustrating and non-frustrating

interactions, as evidenced by the Wilcoxon Signed-Rank test result ($Z = 2.72$, $p = 0.009$).

In the health task, which involved a usability problem of a slow internet connection, participants did not express frustration. Further analysis of participant backgrounds and the overall internet status during the experiment revealed that 50 percent ($n = 8$) of the participants were campus students who conducted the experiment on their university campus. The university had been experiencing internet connectivity issues before and during the experiment. Consequently, many participants were primed by the prevalent internet issues, perceiving the problem in the task as a common occurrence rather than something specific to the experiment. For example, S17 commented:

“I’ve been experiencing similar issues on campus, once I was attending a conference, and it kept disconnecting” [S17]

Although users did not express frustration in this scenario, extensive literature indicates that prolonged waiting times can indeed lead to frustration. Guidelines for enhancing user experience emphasize the significance of minimizing loading times [80]. Additionally, it is important to consider that regular users tend to be more tolerant of loading times compared to new users [81]. Prolonged waiting times can adversely affect user performance and attitude [80] including longer download delays, where delays of 30 and 60 s have been found to increase frustration [82]. Therefore, the inclusion of diverse participants may yield different outcomes. Overall, poor usability diminishes user productivity and contributes to user attrition [83].

5.3 Other factors that influence frustration

Despite the study being conducted in a controlled setting, other factors, such as the individual's personality, demographics can contribute to the user's frustration. Personality traits play a huge role in participants' frustration tolerance [84]. According to the findings, extrovert participants outperformed those with high conscientiousness or high neuroticism when they were frustrated [85]. In this study, personality traits were not considered, which can play a role in the participants' reaction to accessibility and usability problems.

The demographics of the users can also play a role in the users' physiological feedback. Age, gender, and ethnicity affect the physiological signals of the users [86]. For example, the age group for recruiting the participants is from 18 to 65. The physiological response of younger users is different from that of older users [86]. Older individuals report higher levels of arousal than young adults [87]. Thus, to limit external factors, a specific age group is needed. To provide more precise results, the personality and demographics should be considered and preferably controlled.

6 Limitation and future work

6.1 Limitations

The study's main limitation is the frequency and design of some usability problems (pop-up ads and page refresh). The frequency of page refresh and pop-up ads is systematically set to occur at 3-s intervals. Although previously tested to be significant [21], where participants experienced pop-ups at one-second intervals, the high frequency of the pop-up ads is not very natural. This gives participants the impression of an experiment.

6.2 Future work

Although some applications and software, such as Mada accessibility monitor,³ exist to identify accessibility and usability issues, a comprehensive application for developers is required to improve the user experience. Taking the results of the physiological signals into account can bring a better, more objective understanding of the accessibility and usability issues that occurred and hence produce better tools for monitoring accessibility and usability.

In this study, the GSR was the only physiological signal used. Using additional physiological signals to detect valence such as ECG can result in more valuable findings. Furthermore, when a study has a usability issue, it was assumed that this is what caused the frustration while there might be other usability or accessibility issues. Tracking tools such as mouse tracking or eye-tracking can help determine the area sighted participants were looking at or interacting with when a peak occurred. When analysing the data, the accumulated peaks per task were used as the metric for analysis because peaks are related to the frustration and arousal [9, 46, 47]. Other GSR metrics can be found and used from the literature such as the average GSR [48], accumulated GSR [88], and the peak amplitude [47] are some of the most used analysis strategies.

In addition, this study focused only on participants with visual impairments. However, this study can be tailored to any other population group where frustration can negatively affect their well-being. Such groups may include older people [9], adults with autism [89], and people with mental health problems related to chronic depression, and other relevant problems in which frustration can have a considerable effect on their well-being [90].

³ <https://monitor.mada.org.qa>

6.3 Implications for use

This paper has implications for policy, practice, and web evaluation regarding accessibility. The study provides evidence-based results showing that VI users experience higher levels of frustration with inaccessible content compared to usability problems. This emphasizes the importance of addressing accessibility issues and linking this research to policy makers and product designers. In the realm of accessibility, this research holds significant implications for advancing accessibility, shaping policy decisions, and fostering inclusive design in digital experiences. Additionally, the paper contributes to evaluating accessibility by introducing objective measures, challenging the subjective nature of current evaluation practices. Additionally, the proposed system of logging user frustration and its causes can inform future user experience improvements.

7 Conclusion

Frustration occurs when an individual is unable to achieve the desired goal. This emotion is common in user interaction, particularly when developers overlook the user when designing webpages. Frustration can lead to a variety of psychological issues, including low self-esteem and, in extreme cases, aggressive behavior so it is an emotion that developers should avoid. In this study, the impact of accessibility and usability issues on people's frustration was investigated. This study provides evidence-based findings that confirm the impact of accessibility issues on VI users' frustration and hence their overall user experience. It showed that accessibility issues can be more frustrating than the common usability problem for VI users. This in turns re-emphasized the importance of adhering to accessibility guidelines and employing different methods of access evaluation beyond the standard guideline compliance techniques currently used in practice. In this study, the sighted individuals were only physiologically frustrated by the session timeout. More research is needed to streamline the use of objective assessment in user experience evaluation research. This includes supporting the study design, the tasks selection, and the choice of the objective assessment method.

Acknowledgements “This Project outcomes report for the general public is displayed verbatim as submitted by the Graduate Student Supervisor(s) for this award. Any opinions, findings and conclusions or recommendations expressed in this report are those of the Submitter(s) and do not necessarily reflect the views of the Qatar National Research Fund; QNRF has not approved or endorsed its content”.

Authors' contribution YA and DA contributed to conceptualizing and designing of the study. YA conducted the study with the participants. YA performed the analysis. DA reviewed the results. YA and DA wrote

the main manuscript text. YA prepared all figures and tables. Both authors reviewed the manuscript.

Funding Open Access funding provided by the Qatar National Library. This publication was made possible by the GSRA grant # GSRA7-2-0524-20096 from the Qatar National Research Fund (a member of Qatar Foundation). The findings achieved herein are solely the responsibility of the author[s].

Data availability The datasets generated during and/or analyzed during the current study are not publicly available due to information that could compromise the privacy of research participants but are available from the corresponding author at a reasonable request. However, the results of data analyzed during this study are included in this paper.

Code availability Not applicable.

Declarations

Conflict of interest All authors declare that they have no conflict of interest.

Ethical approval All procedures performed in studies involving human participants followed the ethical standards of the institutional and/or national research committee. We received approval from the Qatar Biomedical Research Institute's Institute of Research Board for all users who participated in this thesis (IRB Application QBRI-IRB 2021-07-100).

Informed consent Informed consent was obtained from all the participants included in the study.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Statista: 101 Mobile Marketing Statistics And Trends For 2020 (2022) Accessed Aug 28 2021 (Online). <https://quoracreative.com/article/mobile-marketing-statistics>
2. Mitchell.: Definitions of Accessibility (1995). <https://trid.trb.org/view/716768>
3. Pribeanu, C., Marinescu, R.-D., Iordache, D., Gheorghe-Moisii, M.: Exploring the usability of municipal web sites: a comparison based on expert evaluation results from four case studies. *Inform. Econ. J.* **14**, 1 (2010)
4. Sahni, S., Dubey, S.K.: Web usability: issues, challenges and solutions. *Int. J. Adv. Eng. Res. Sci.* **1**(2), 6 (2014)
5. Carvalho, M.C.N., Dias, F.S., Reis, A.G.S., Freire, A.P.: Accessibility and usability problems encountered on websites and applications in mobile devices by blind and normal-vision users. In: *Proceedings of the 33rd Annual ACM Symposium on Applied*

- Computing, in SAC '18, pp. 2022–2029. Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/3167132.3167349>.
6. Seo, K.-K., Lee, S., Chung, B.D., Park, C.: Users' emotional valence, arousal, and engagement based on perceived usability and aesthetics for web sites. *Int. J. Human-Computer Interact.* **31**(1), 72–87 (2015). <https://doi.org/10.1080/10447318.2014.959103>
 7. Berkowitz, L.: *Aggression: A Social Psychological Analysis*. McGraw-Hill (1962).
 8. Lazar, J., Jones, A., Shneiderman, B.: Workplace user frustration with computers: an exploratory investigation of the causes and severity. *Behav. Inf. Technol.* **25**(3), 239–251 (2006). <https://doi.org/10.1080/01449290500196963>
 9. Lunn, D., Harper, S.: Using galvanic skin response measures to identify areas of frustration for older web 2.0 users. In *Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A)*, in W4A '10, pp. 1–10. Association for Computing Machinery, New York, NY, USA (2010). <https://doi.org/10.1145/1805986.1806032>.
 10. Spina, C.: WCAG 2.1 and the current state of web accessibility in libraries. *Weave J. Libr. User Exp.* **2**(2), 1 (2019). <https://doi.org/10.3998/weave.12535642.0002.202>
 11. Silverman, B., Elizabeth, R., Penny, R.C., Bleach, K.: *Technology and Accommodations: Employment Experiences of U.S. Adults Who Are Blind, Have Low Vision, or Are Deafblind* (2022).
 12. Pascual, A., Ribera, M., Granollers, T., Coiduras, J.L.: Impact of accessibility barriers on the mood of blind, low-vision and sighted users. *Procedia Comput. Sci.* **27**, 431–440 (2014). <https://doi.org/10.1016/j.procs.2014.02.047>
 13. Alajarmeh, N.: The Extent of Mobile Accessibility Coverage in WCAG 2.1: Sufficiency of Success Criteria and Appropriateness of Relevant Conformance Levels Pertaining to Accessibility Problems Encountered by Users Who are Visually Impaired. *Univers. Access Inf. Soc.* (2021). <https://doi.org/10.1007/s10209-020-00785-w>.
 14. Whitney, G., Kolar, I.: Am I missing something? *Univers. Access Inf. Soc.* **19**(2), 461–469 (2020). <https://doi.org/10.1007/s10209-019-00648-z>
 15. Wentz, B., Lazar, J.: Exploring the Impact of Inaccessible Redesign and Updates. In: Langdon, P., Lazar, J., Heylighen, A., Dong, H. (eds.) *Designing Around People*, pp. 3–12. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-29498-8_1.
 16. Cohen, A.H., Fresneda, J.E., Anderson, R.E.: How inaccessible retail websites affect blind and low vision consumers: their perceptions and responses. *J. Serv. Theory Pract.* **33**(3), 329–351 (2023). <https://doi.org/10.1108/JSTP-08-2021-0167>
 17. Sullivan, T., Matson, R.: Barriers to use: usability and content accessibility on the Web's most popular sites, pp. 139–144 (2000). <https://doi.org/10.1145/355460.355549>.
 18. Theofanos, M.F., Redish, J.: Bridging the gap: between accessibility and usability. *Interactions* **10**(6), 36–51 (2003). <https://doi.org/10.1145/947226.947227>
 19. Russell, J.: A circumplex model of affect. *J. Pers. Soc. Psychol.* **39**, 1161–1178 (1980). <https://doi.org/10.1037/h0077714>
 20. Wascher, C.A.F.: Heart rate as a measure of emotional arousal in evolutionary biology. *Philos. Trans. R. Soc. B Biol. Sci.* **376**(1831), 20200479 (2021). <https://doi.org/10.1098/rstb.2020.0479>
 21. Matthews, O., Davies, A., Vigo, M., Harper, S.: Unobtrusive arousal detection on the web using pupillary response. *Int. J. Hum.-Comput. Stud.* **136**, 1 (2019). <https://doi.org/10.1016/j.ijhcs.2019.09.003>
 22. Eijlers, E., Boksem, M.A.S., Smidts, A.: Measuring neural arousal for advertisements and its relationship with advertising success. *Front. Neurosci.* **14**, 1 (2020). <https://doi.org/10.3389/fnins.2020.00736>
 23. Vitasari, P., Wahab, M.N.A., Herawan, T., Sinnadurai, S.K., Othman, A., Awang, M.G.: Assessing of physiological arousal and cognitive anxiety toward academic performance: the application of catastrophe model. *Procedia - Soc. Behav. Sci.* **30**, 615–619 (2011). <https://doi.org/10.1016/j.sbspro.2011.10.119>
 24. Cusveller, J., Gerritsen, C., de Man, J.: Evoking and Measuring Arousal in Game Settings, pp. 165–174. https://doi.org/10.1007/978-3-319-05972-3_17
 25. Rodic, B., Stanojevic, G., Sapic, R., Labus, A., Bogdanović, Z.: Wearable solution for assessing physiological arousal towards students' interest and engagement in the classroom (2016).
 26. Hlas, A., Neyers, K., Molitor, S.: Measuring student attention in the second language classroom. *Lang. Teach. Res.* **23**, 136216881771376 (2017). <https://doi.org/10.1177/1362168817713766>
 27. Brennan, D.: Signs of Frustration. WebMD, 2020. <https://www.webmd.com/mental-health/signs-frustration>. Accessed 23 March 2022
 28. Johnson, C.M., Larson, H.A., Conn, S.R., Estes, L.A., Ghibellini, A.B.: The Impact of Relaxation Techniques on Third Grade Students' Self-perceived Levels of Test Anxiety, p. 9 (2009).
 29. Nagai, Y., Jones, C.I., Sen, A.: Galvanic skin response (GSR)/electrodermal/skin conductance biofeedback on epilepsy: a systematic review and meta-analysis. *Front. Neurol.* **10**, 2019 (2022). <https://doi.org/10.3389/fneur.2019.00377>
 30. Suhaimi, N.S., Mountstephens, J., Teo, J.: EEG-based emotion recognition: a state-of-the-art review of current trends and opportunities. *Comput. Intell. Neurosci.* **2020**, e8875426 (2020). <https://doi.org/10.1155/2020/8875426>
 31. Nikolova, D., Petkova, P., Manolova, A., Georgieva, P.: ECG-based emotion recognition: overview of methods and applications. In: ANNA '18; *Advances in Neural Networks and Applications 2018*, pp. 1–5 (2018).
 32. Maulsby, R.L., Edelberg, R.: The interrelationship between the galvanic skin response, basal resistance, and temperature. *J. Comput. Physiol. Psychol.* **53**(5), 475–479 (1960). <https://doi.org/10.1037/h0041916>
 33. Westeyn, T., Presti, P., Starner, T.: ActionGSR: a combination galvanic skin response-accelerometer for physiological measurements in active environments. In: *2006 10th IEEE International Symposium on Wearable Computers*, pp. 129–130 (2006). <https://doi.org/10.1109/ISWC.2006.286360>.
 34. Darrow, C.W.: Problems in the use of the galvanic skin response (GSR) as an index of cerebral function: implications of the latent period. *Psychophysiology* **3**(4), 389–396 (1967). <https://doi.org/10.1111/j.1469-8986.1967.tb02723.x>
 35. Dawson, M., Schell, A., Fillion, D.: *The Electrodermal System* (2007).
 36. Critchley, H.D.: Review: electrodermal responses: what happens in the brain. *Neuroscientist* **8**(2), 132–142 (2002). <https://doi.org/10.1177/107385840200800209>
 37. Alinia, P., et al.: Associations between physiological signals captured using wearable sensors and self-reported outcomes among adults in alcohol use disorder recovery: development and usability study. *JMIR Form. Res.* **5**(7), e27891 (2021). <https://doi.org/10.2196/27891>
 38. Hollenstein, T., McNeely, A., Eastabrook, J., Mackey, A., Flynn, J.: Sympathetic and parasympathetic responses to social stress across adolescence. *Dev. Psychobiol.* **54**(2), 207–214 (2012). <https://doi.org/10.1002/dev.20582>
 39. Spear, L.P.: Heightened stress reactivity and emotional reactivity during pubertal maturation: Implications for psychopathology. *Dev. Psychopathol.* **21**(1), 87–97 (2009). <https://doi.org/10.1017/S0954579409000066>

40. Arroyo, I., Cooper, D., Burleson, W., Woolf, B., Muldner, K., Christopherson, R.: Emotion Sensors Go To School. Presented at the Frontiers in Artificial Intelligence and Applications (2009). <https://doi.org/10.3233/978-1-60750-028-5-17>.
41. Di Lascio, E., Gashi, S., Santini, S.: Unobtrusive assessment of students' emotional engagement during lectures using electrodermal activity sensors. In Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. **2**(3), 103:1–103:21 (2018). <https://doi.org/10.1145/3264913>.
42. Senthil, S., Lin, W.: Measuring Students' Engagement Using Wireless Heart Rate Sensors, pp. 699–704 (2017). <https://doi.org/10.1109/SmartTechCon.2017.8358461>.
43. Wang, C., César, P.: Physiological measurement on students' engagement in a distributed learning environment. *PhyCS* (2015). <https://doi.org/10.5220/0005229101490156>
44. Thammasan, N., Stuldreher, I.V., Schreuders, E., Giletta, M., Brouwer, A.-M.: A Usability Study of Physiological Measurement in School Using Wearable Sensors. *Sensors* **20**(18), 18 (2020). <https://doi.org/10.3390/s20185380>
45. Ceaparu, I., Lazar, J., Bessiere, K., Robinson, J., Shneiderman, B.: Determining causes and severity of end-user frustration (2002). *Int. J. Hum.-Comput. Interact.* **17**, 1 (2003). https://doi.org/10.1207/s15327590ijhc1703_3
46. Bruun, A., Ahm, S.: Mind the Gap! Comparing Retrospective and Concurrent Ratings of Emotion in User Experience Evaluation. In Abascal, J., Barbosa, S., Fetter, M., Gross, T., Palanque, P., Winckler, M. (eds). *Human-Computer Interaction – INTERACT 2015* pp. 237–254. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-22701-6_17
47. Nourbakhsh, N., Chen, F., Wang, Y., Calvo, R.: Detecting Users' Cognitive Load by Galvanic Skin Response with Affective Interference. *ACM Trans. Interact. Intell. Syst.* **7**, 1–20 (2017). <https://doi.org/10.1145/2960413>
48. Horvers, A., Tombeng, N., Bosse, T., Lazonder, A.W., Molenaar, I.: Detecting emotions through electrodermal activity in learning contexts: a systematic review. *Sensors* **21**(23), 23 (2021). <https://doi.org/10.3390/s21237869>
49. Li, T., Lajoie, S.: Predicting aviation training performance with multimodal affective inferences. *Int. J. Train. Dev.* **25**(3), 301–315 (2021). <https://doi.org/10.1111/ijtd.12232>
50. Demmin, D.L., Silverstein, S.M.: Visual impairment and mental health: unmet needs and treatment options. *Clin. Ophthalmol. Auckl. NZ* **14**, 4229–4251 (2020). <https://doi.org/10.2147/OPHTH.S258783>
51. Google Data: How mobile latency impacts publisher revenue. Think with Google (2016). <https://www.thinkwithgoogle.com/intl/en-154/marketing-strategies/app-and-mobile/need-mobile-speed-how-mobile-latency-impacts-publisher-revenue/>. Accessed 22 Jan. 2022.
52. Alexa: Alexa - Top Sites in Qatar - Alexa (2021). <https://www.alexa.com/topsites/countries/QA>. Accessed 22 Jan 2022.
53. Taylor, S., Jaques, N., Chen, W., Fedor, S., Sano, A., Picard, R.: Automatic Identification of Artifacts in Electrodermal Activity Data. **2015**, 1937 (2015). <https://doi.org/10.1109/EMBC.2015.7318762>.
54. Hernandez, J., Riobo, I., Rozga, A., Abowd, G., Picard, R.: Using Electrodermal Activity to Recognize Ease of Engagement in Children during Social Interactions. Presented at the UbiComp 2014—Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Sep. 2014. <https://doi.org/10.1145/2632048.2636065>.
55. Kong, Y., Posada-Quintero, H.F., Chon, K.H.: Sensitive physiological indices of pain based on differential characteristics of electrodermal activity. *IEEE Trans. Biomed. Eng.* **68**(10), 3122–3130 (2021). <https://doi.org/10.1109/TBME.2021.3065218>
56. Posada-Quintero, H.F., et al.: Using electrodermal activity to validate multilevel pain stimulation in healthy volunteers evoked by thermal grills. *Am. J. Physiol.-Regul. Integr. Comp. Physiol.* **319**(3), R366–R375 (2020). <https://doi.org/10.1152/ajpregu.00102.2020>
57. Hernando Gallego, F., Luengo, D., ArtésRodríguez, A.: Feature extraction of galvanic skin responses by non-negative sparse deconvolution. *IEEE J. Biomed. Health Inform.* **1**, 1 (2017). <https://doi.org/10.1109/JBHI.2017.2780252>
58. Lee, H., Kleinsmith, A.: Public Speaking Anxiety in a Real Classroom: Towards Developing a Reflection System, pp. 1–6 (2019). <https://doi.org/10.1145/3290607.3312875>.
59. Sano, A., Picard, R.W.: Stress recognition using wearable sensors and mobile phones. In 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, pp. 671–676 (2013). <https://doi.org/10.1109/ACII.2013.117>.
60. Reimer, U., Maier, E., Laurenzi, E., Ulmer, T.: Mobile Stress Recognition and Relaxation Support with SmartCoping: User-Adaptive Interpretation of Physiological Stress Parameters (2017). <https://doi.org/10.24251/HICSS.2017.435>.
61. Zhu, L., et al.: Stress detection through wrist-based electrodermal activity monitoring and machine learning. *IEEE J. Biomed. Health Inform.* **27**(5), 2155–2165 (2023). <https://doi.org/10.1109/JBHI.2023.3239305>
62. Kim, T., et al.: Prediction for Retrospection: Integrating Algorithmic Stress Prediction into Personal Informatics Systems for College Students' Mental Health. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, in CHI '22. Association for Computing Machinery, New York, NY, USA (2022). <https://doi.org/10.1145/3491102.3517701>.
63. Ng, A., et al.: Predicting the Next-Day Perceived and Physiological Stress of Pregnant Women by Using Machine Learning and Explainability: Algorithm Development and Validation. *JMIR MHealth UHealth* **10**(8), 1. <https://doi.org/10.2196/33850>.
64. Raufi, B., Ferati, M., Zenuni, X., Ajdari, J., Ismaili, F.: Methods and techniques of adaptive web accessibility for the blind and visually impaired. *Procedia - Soc. Behav. Sci.* **195**, 1999–2007 (2015). <https://doi.org/10.1016/j.sbspro.2015.06.214>
65. Whitenoton, K.: Minimize Cognitive Load to Maximize Usability. Nielsen Norman Group, 2013. <https://www.nngroup.com/articles/minimize-cognitive-load/>. Accessed 23, Mar. 2022
66. Jarry, A., Chapdelaine, C., Kurniawan, S., Wittich, W.: Blind Adults' Perspectives on Technical Problems and Solutions When Using Technology (2017) Accessed: July 11, 2023 (Online). <https://nfb.org/images/nfb/publications/jbir/jbir17/jbir070102.html>
67. Petty, R.E.: Technology Access in the Workplace and Higher Education for Persons with Visual Impairments An Examination of Barriers and Discussion of Solutions (2005).
68. Savva, A.: Understanding Accessibility Problems of Blind Users on the Web (2017).
69. Mehta, Y., Joshi, A., Joshi, M., Jadhav, C.: Accessibility of Date Picker for Touchscreens. In: Proceedings of the 8th Indian Conference on Human Computer Interaction, in IHCI '16, pp. 64–69. Association for Computing Machinery, New York, NY, USA, (2016). <https://doi.org/10.1145/3014362.3014368>.
70. Bhagat, S., Joshi, P.: Evaluation of accessibility and accessibility audit methods for e-Governance portals. In: Proceedings of the 12th International Conference on Theory and Practice of Electronic Governance, pp. 220–226. ACM, Melbourne, VIC (2019). <https://doi.org/10.1145/3326365.3326394>.
71. Kurt, S.: The accessibility of university web sites: the case of Turkish universities. *Univ. Access Inf. Soc.* **10**(1), 101–110 (2011). <https://doi.org/10.1007/s10209-010-0190-z>

72. Nielsen, J., Li, A.: Mega Menus Work Well for Site Navigation. Nielsen Norman Group (2017). <https://www.nngroup.com/articles/mega-menus-work-well/>. Accessed 11 July 2023
73. McDermott, M.L.: Voting cues in low-information elections: candidate gender as a social information variable in contemporary united states elections. *Am. J. Polit. Sci.* **41**(1), 270–283 (1997). <https://doi.org/10.2307/2111716>
74. Petrie, H., Kheir, O.: The Relationship Between Accessibility and Usability of Websites, pp. 397–406 (2007). <https://doi.org/10.1145/1240624.1240688>.
75. Liu, C., White, R.W., Dumais, S.: Understanding web browsing behaviors through Weibull analysis of dwell time. In: Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 379–386. ACM, Geneva (2010). <https://doi.org/10.1145/1835449.1835513>.
76. Ferreri, N., Mayhorn, C.B.: That’s not what we expected: examining technology expectations and malfunctions on frustration. *Ergon. Des.* (2021). <https://doi.org/10.1177/10648046211007709>
77. Ferrario, M.A., et al.: Computing and mental health: intentionality and reflection at the click of a button. In: Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare, pp. 1–10. ACM, Barcelona (2017). <https://doi.org/10.1145/3154862.3154877>.
78. Ferreri, N., Mayhorn, C.B.: Examining frustration and performance when priming user expectations and providing a technology malfunction. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **64**(1), 1846–1850 (2020). <https://doi.org/10.1177/1071181320641444>
79. Brunswik, E., Kamiya, J.: Ecological cue-validity of ‘proximity’ and of other gestalt factors. *Am. J. Psychol.* **66**(1), 20–32 (1953). <https://doi.org/10.2307/1417965>
80. Galletta, D.F., Henry, R.M., McCoy, S., Polak, P.: When the wait isn’t so bad: the interacting effects of website delay, familiarity, and breadth. *Inf. Syst. Res.* (2006). <https://doi.org/10.1287/isre.1050.0073>
81. Tsiaousis, A.S., Giaglis, G.M.: Evaluating the effects of the environmental context-of-use on mobile website usability. In: 2008 7th International Conference on Mobile Business, pp. 314–322 (2008). <https://doi.org/10.1109/ICMB.2008.25>.
82. Selvidge, P.R., Chaparro, B.S., Bender, G.T.: The world wide wait: effects of delays on user performance. *Int. J. Ind. Ergon.* **29**(1), 15–20 (2002). [https://doi.org/10.1016/S0169-8141\(01\)00045-2](https://doi.org/10.1016/S0169-8141(01)00045-2)
83. Shitkova, M., Holler, J., Heide, T., Clever, N.: Towards Usability Guidelines for Mobile Websites and Applications
84. Ramirez-Castillo, D., Garcia-Roda, C., Guell, F., Fernandez-Montalvo, J., Bernacer, J., Morón, I.: Frustration tolerance and personality traits in patients with substance use disorders. *Front. Psychiatry* **10**, 1 (2019). <https://doi.org/10.3389/fpsy.2019.00421>
85. El Nashar, Z., El Bolock, A., Salah, J., Herbert, C., Abdennadher, S.: Investigating the effect of personality traits on performance under frustration. In: Liapis, A., Yannakakis, G.N., Gentile, M., Ninaus, M. (eds.) Games and Learning Alliance. Lecture Notes in Computer Science. Springer, Cham, pp. 595–604 (2019). https://doi.org/10.1007/978-3-030-34350-7_57
86. Jeronimus, B., Laceulle, O.: Frustration, pp. 1–8 (2017). https://doi.org/10.1007/978-3-319-28099-8_815-1
87. Fernández-Aguilar, L., Ricarte, J., Ros, L., Latorre, J.M.: Emotional Differences in Young and Older Adults: Films as Mood Induction Procedure. *Front. Psychol.* **9**, 1 (2018). <https://doi.org/10.3389/fpsyg.2018.01110>
88. Hossain, D., et al.: Measurement of cognitive load for writing tasks using galvanic skin response. In: Proceedings of the 6th International Conference on Networking, Systems and Security, in NSysS ’19, pp. 122–126. Association for Computing Machinery, New York, NY (2019). <https://doi.org/10.1145/3362966.3362976>.
89. Northrup, J.B., et al.: Observed emotional reactivity in response to frustration tasks in psychiatrically hospitalized youth with autism spectrum disorder. *Autism Int. J. Res. Pract.* **24**(4), 968–982 (2020). <https://doi.org/10.1177/1362361320908108>
90. Cowan, B.R., Jack, M.A.: Exploring the wiki user experience: the effects of training spaces on novice user usability and anxiety towards wiki editing. *Interact. Comput.* **23**(2), 117–128 (2011). <https://doi.org/10.1016/j.intcom.2010.11.002>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.