Article:

Williams, Joseph, Francombe, Jon and Murphy, Damian Thomas orcid.org/0000-0002-6676-9459 (2023) Evaluating the Influence of Room Illumination on Camera-Based Physiological Measurements for the Assessment of Screen-Based Media. Applied Sciences. 8482. ISSN 2076-3417

https://doi.org/10.3390/app13148482

applied sciences

MDPI

Article

Evaluating the Influence of Room Illumination on Camera-Based Physiological Measurements for the Assessment of Screen-Based Media

Joseph Williams [1,*,†], Jon Francombe [2] and Damian Murphy [1,†]

[1] AudioLab, School of Physics Engineering and Technology, University of York, York YO10 5DD, UK; damian.murphy@york.ac.uk
[2] Bang & Olufsen a/s, 7600 Struer, Denmark; jofr@bang-olufsen.dk
[*] Correspondence: jw1858@york.ac.uk
[†] Current address: School of Physics Engineering and Technology, University of York, Heslington, Genesis 6, York YO10 5DQ, UK.

Abstract: Camera-based solutions can be a convenient means of collecting physiological measurements indicative of psychological responses to stimuli. However, the low illumination playback conditions commonly associated with viewing screen-based media oppose the bright conditions recommended for accurately recording physiological data with a camera. A study was designed to determine the feasibility of obtaining physiological data, for psychological insight, in illumination conditions representative of real world viewing experiences. In this study, a novel method was applied for testing a first-of-its-kind system for measuring both heart rate and facial actions from video footage recorded with a single discretely placed camera. Results suggest that conditions representative of a bright domestic setting should be maintained when using this technology, despite this being considered a sub-optimal playback condition. Further analyses highlight that even within this bright condition, both the camera-measured facial action and heart rate data contained characteristic errors. In future research, the influence of these performance issues on psychological insights may be mitigated by reducing the temporal resolution of the heart rate measurements and ignoring fast and low-intensity facial movements.

Keywords: biosensors; psychophysiological methods; remote photoplethysmography (rPPG); facial action coding (FAC); multi-modal physiological data; heart rate (HR)

check for updates

Citation: Williams, J.; Francombe, J.; Murphy, D. Evaluating the Influence of Room Illumination on Camera-Based Physiological Measurements for the Assessment of Screen-Based Media. *Appl. Sci.* **2023**, *13*, 8482. https://doi.org/10.3390/app13148482

Academic Editors: Douglas O'Shaughnessy, Paula Viana, Pedro Carvalho and Teresa Chambel

Received: 20 April 2023
Revised: 11 July 2023
Accepted: 19 July 2023
Published: 22 July 2023

1. Introduction

Research into camera-based physiological measurement systems has flourished in recent years. Open-source software is now available which allows for the collection of data associated with face, head, and eye movements, as well as heart rate, using only a camera [1,2]. These software packages can be applied to a wide range of video signals, not just those captured by specialist cameras, and have been used in a variety of domains including deep fake recognition [3,4]. However, the primary objective of camera-based physiological measurement systems is to be a viable alternative to contact sensors or manual video annotation techniques.

Physiological measurements can provide insights into psychological states. Research investigating these mind–body interactions is commonly referred to as psychophysiological research. The methods employed in this field exhibit significant variability, which can be categorized based on input modalities and analytic techniques. In terms of input modalities, researchers often strive to gather as much insightful data as possible without interfering with the subjects' naturalistic experience. Regarding analytic techniques, researchers often aim to leverage all of the collected data to test their hypotheses. The focus of this study is specifically on the application of psychophysiological methods for assessing screen-based media viewing experiences. This area of research is highly active for two primary reasons: screen-based media viewing is a significant sociological phenomenon, as evidenced by a 2022 survey conducted by Ofcom, which reported that individuals in the United Kingdom spend an average of 5 h and 16 min per day watching television and video content across all devices [5]; and screen technologies offer a convenient means of delivering experimental stimuli.

Studies have demonstrated that psychophysiological methods can provide insights that traditional methods cannot when assessing audiovisual quality [6]. Wilson [7] observed a significant increase in heart rate (HR) and galvanic skin response associated with viewing screen-based media rendered at 5 frames per second (FPS), compared to 25 FPS. This was despite only 16% of the participants noting a degradation in quality. Moreover, Wilson also states that audio degradation was detectable through similar analyses of these physiological modalities. These results exemplify why many believe psychophysiological methods hold promise as a better approach for subjective audiovisual quality assessment, compared to customary methods which rely solely on self-assessment, and therefore may enable the development of more effective objective audiovisual quality assessment methods [8–13].

Narrative engagement can also be effectively assessed using psychophysiological methods. Research has demonstrated that when individuals watch screen-based media, changes in physiological features such as heart rate and gaze position correlate between those who feel engaged [14–16]. Psychophysiological methods have also been used extensively for assessing experiences in the domain of extended reality research [17]. Hinkle et al. [18] applied contact sensors to measure facial muscle activity, body movement, eye movements, skin temperature, and respiratory, cardiovascular, and neural activity. Various feature extraction techniques and machine learning algorithms were then applied to these input modalities, with the aim of developing a system capable of predicting emotion. The results suggested that emotions could reliably be predicted using this approach, however, the choice of features had a significant effect on the accuracy of prediction.

Camera-based physiological measurement systems have also been used in psychophysiological research. In Leong et al.'s 2023 review of the use of visual data in affective computing, 30 quality-screened studies associated with facial expression and/or body gesture-based emotion recognition were found [19]. The application of camera-based heart rate measurement systems has also been explored by researchers investigating emotional responses to playing games [20] and viewing pictures [21].

This camera-based approach to psychophysiological research may be considered advantageous when compared to methods which use contact sensors, as a camera can be a more discrete way of capturing insightful data. This suggests that camera-based methods have the potential to provide more ecologically valid results when compared to contact sensors. Another positive aspect of using a camera-based system is the ease of making a video recording, compared to a contact sensor recording. One negative aspect of using camera-based systems is that they are considered to have poor robustness to many interpersonal and environmental factors—such as low room illumination (see Section 2).

Within the broader field of psychophysiological research, multiple sensors are often used to provide a rich dataset for gaining psychological insight [6,7,15,17,18,22–26]. However, few studies use multimodal data extracted from a camera despite the technological feasibility. For example, out of the 11 studies explored by Leong et al., only 3 applied a multimodal camera-based physiological measurements solution—recording both body gestures and facial expressions [19].

In some cases, researchers may wish to apply camera-based psychophysiological methods for the assessment of cinematic viewing experiences. These experiences are defined here as those in which the viewer is placed in near-darkness and the screen inhabits much of their field of view. The motivation behind the reproduction of such conditions in experimental work may be a wish to recreate a real-world viewing experience, as many choose to view film or television shows with the lights turned down. Researchers may also wish to create a cinematic viewing experience for exploring physiological signs of emotional responses, as darker ambient lighting has been shown to induce a greater sense of arousal [27]. Notably, these cinematic conditions are also recommended when assessing image quality [28]. Moreover, a cinematic viewing distance is recommended for accompanying images with sound from a multichannel audio system [29].

The aim of this research is to explore the feasibility of using a multimodal camera-based physiological measurement solution for the assessment of cinematic viewing experiences. There is currently no existing research in this application area. Due to the conflict in recommendations regarding illumination, this factor is investigated in depth, considering how it impacts the physiological measurements obtained. This publication introduces a new approach for assessing the performance of facial action coding systems, which is significantly less labour-intensive than the existing method (see Section 3.1). Moreover, it describes the first documented solution for obtaining both heart rate and facial action measurements using a single camera, leveraging existing software systems.

2. Background

To be able to evaluate the feasibility of using a multimodal camera-based physiological measurement solution for the assessment of cinematic viewing experiences, such a solution must first be implemented. To achieve this, investigations were made into open-source software which could be leveraged. In the first sub-section, *PyVHR* [2,3], a software package designed for obtaining heart rate (HR) from video, is described and factors which influence its performance are discussed. In the second sub-section, *OpenFace 2* [1,30], a software package applied for the automatic facial action coding (FAC) of video, is described in a similar manner. These technical factors help inform the camera configuration described in Section 3.3.

2.1. *PyVHR*

PyVHR is a framework for assessing the performance of remote photoplethysmography (rPPG) algorithms. rPPG is a technique developed to estimate human cardiac activity from a video signal. The version of *PyVHR* evaluated in this paper is version 1.0.2 which can be accessed via the *PyVHR* releases page on the GitHub repository [3]. It is stated in the publication associated with *PyVHR*, that this framework is unique in its capabilities for providing a standardized and reproducible way of implementing full rPPG pipelines, with a specific focus on obtaining continuous heart rate measurements from a video signal, and benchmarking them against multiple datasets [3].

Conceptually, *PyVHR* takes the critical sub-process used in an established rPPG algorithm, which convert an array of red, green and blue (RGB) digital pixels to an estimate of blood volume pressure (BVP) for a region of skin, and places it in a standardised pipeline. Hence, the RGB to BVP process can be compared without the influence of variations in any other sub-processes. Examples of such variations include differences in how the colour of the skin is sampled and how the heart rate is calculated from the BVP signal. Out of the ten RGB to BVP sub-processes implemented, the Hann and Jeannes chrominance approach (CHROM) [31] is shown to be most effective on the two datasets [32] tested in the associated publication for *PyVHR* [3].

CHROM is a digital signal-processing algorithm which was developed with the aim of improving motion robustness, so that remote heart rate measurements may be captured effectively on subjects using gym equipment [31]. Despite the algorithm showing poor performance within this difficult application area, it has become one of the most popular approaches for rPPG. Notably, Haugg et al. showed a drop in performance from an average error of 1.91 beats per minute (BPM) with the subject resting to 14.81 BPM with the subject using the gym [33]. The algorithm adds sophistication to previously proposed methods based on blind source separation, by considering the difference between light reflected off the surface of the skin and light which has travelled through the skin and therefore contains information associated with the subject's cardiac cycle. It was developed using 117 one-minute uncompressed facial videos, as well as an undefined number of videos of subjects using a stepping machine and static bicycle in a gym. Both datasets were recorded with a specialised camera in parallel with a pulse-oximeter finger clip [31].

There are multiple documented ways of using the *PyVHR* software framework, including a graphical user interface (GUI). Within the demo workbook, functions from the software pipeline are called individually. This segmentation of the overall process, from video to heart rate, avoids extended run times. The CHROM method is used in the demo workbook implementation [3] and hence hypothetically provides a reproducible system performance benchmark with minimal usability hurdles. Details of the implementation from the demo workbook can be found in Table 1.

Table 1. Description of each sub-process used to estimate heart rate from a video signal within the *PyVHR* pipeline. Note that the demo workbook implementation was used to obtain heart rate data in the study described later in this manuscript, with the 6-patch approach rather than the 100-patch approach for skin extraction.

Name of Subprocess	Description in *PyVHR* Documentation [3]	Implementation in *PyVHR* Demo Workbook [2]
Skin extraction	"The goal of this first step is to perform a face skin segmentation in order to extract PPG-related areas; the latter are subsequently collected in either a single patch (holistic approach) or a bunch of "sparse" patches covering the whole face (patch-wise approach)."	A set of 100 equidistant patches are selected across the face. Alternatively, code is provided for the selection of 6 patches: 3 on the forehead, 1 on the nose, and 1 on each cheek.
RGB signal processing	"The patches, either one or more, are coherently tracked and are used to compute the average colour intensities along overlapping windows, thus providing multiple time-varying RGB signals for each temporal window"	The RGB signals are obtained by taking the mean value of each patch across cascading 8-s windows, with a stride of 1-s.
Pre-filtering	"Optionally, the raw RGB traces are pre-processed via canonical filtering, normalization or de-trending; the outcome signals provide the inputs to any subsequent rPPG method."	Pixel values below 0 or above 230 are removed from patches. Then a sixth-order Butterworth bandpass filter with a passband between 0.75 Hz and 4 Hz is applied.
BVP extraction	"The rPPG method(s) at hand is applied to the time-windowed signals, thus producing a collection of heart rate pulse signals (BVP estimates), one for each patch."	The CHROM method is applied [31].
Post-filtering	"The pulse signals are optionally passed through a narrow-band filter in order to remove unwanted out-of-band frequency components."	A sixth-order Butterworth bandpass filter with a passband between 0.75 Hz and 4 Hz is applied.
BPM estimation	"A BPM estimate is eventually obtained through simple statistics relying on the apical points of the BVP power spectral densities."	The power spectral density (PSD) of each window of each patch of the filtered BVP estimate is calculated. Then, the median of the PSD peak frequency for all patches for that window is converted into a BPM value.

There are two versions of *PyVHR*, one for CPU-only run times and the other for computers with graphical processing units (GPUs). The GPU version of the package is reported to run in real-time for 30 FPS high-definition video. The CPU version is also reported to run in real-time when using the holistic skin segmenting sub-process. However, when utilising the *patches* skin extraction functionality of the system (see Table 1) the CPU version takes around 0.12 s to process a single frame. These results are based on the following hardware configuration: *"Intel Xeon Silver 4214R 2.40 GHz"* (CPU), *"NVIDIA Tesla V100S PCIe 32 GB"* (GPU) [3].

Exploring the robustness of rPPG systems is an active area of research. Nowara et al.'s [34] results suggest that CHROM is relatively robust to different skin tones, compared to the other four algorithms tested [35–38]. However, they found that for the darkest skin tone category in their dataset performance was poor, with an average error of 13.58 BPM compared to all other skin tones for which the average error was between 2.14 BPM and 4.09 BPM [34]. There are also many other interpersonal factors which still have an unclear influence on performance, such as age and wearing make-up. In a study assessing responses to screen-based media with a large sample of participants, many of these influencing factors are likely to be relevant.

Illumination has also been investigated. Yang et al. suggested that a drop from a medium to dark condition causes a drop in average error from 1.04 BPM to 3.60 BPM for their best-performing algorithm—a CHROM-based solution [39]. Moreover, Yin et al. explored how unstable room illumination may influence CHROM performance, finding an average error of 1.04 BPM under stable lighting compared to an average error of 2.45 BPM under varying lighting [40]. Tohma et al. suggested that a bright and stable light greater than 500 lux should be cast on the subject's face, when studying the influence of illumination on the measurement of heart rate variability from a video signal [41]. Moreover, they suggest that subjects should remain static while measurements are being taken.

McDuff et al. showed that video compression can also influence CHROM performance, decreasing from an average error of 1.78 BPM for uncompressed to an average error of 9.31 BPM for the most compressed videos evaluated [42]. Notably, all publications discussed in this section use specialised scientific cameras [31,41] or specific webcam models [32,39,40] to avoid video compression. However, hypothetically this may not be necessary in some use cases as McDuff et al. also showed that light video compression only influences a small decrease in average error for heart rate measurements when calculating heart rate from a one-minute window of the BVP signal [42].

Other technical features of the video signal may also hypothetically influence the performance of rPPG systems. Video frame rate is one of these factors which has been investigated, researchers have suggested a minimum frame rate of 30 FPS [41,43]. Less research has been conducted to explore the influence of video resolution on rPPG system performance. However, Blackford et al. suggest that both frame rate and resolution have a minimal impact on the performance of their rPPG system at obtaining heart rate estimations [44].

In *PyVHR*, the inclusion of each sub-process is supported by its positive influence on performance as shown in the publications in which they are proposed [3]. The patches approach for skin region is of interest is of particular relevance to this project as it has has been shown to negate some of the negative influence of unstable illumination on system performance, allowing for the selection of optimal regions of interest which facilitate the extraction of a BVP signal with a higher signal to noise ratio (SNR) [32,40,45]. A low SNR can lead to the true heart rate frequency not being the peak value of the power spectral density function for the measurement window, hence leading to an incorrect HR estimate (see Table 1).

2.2. *OpenFace 2*

OpenFace 2 is the second iteration of the *OpenFace* toolkit, developed for the measurement of multi-modal physiological data from a video signal. The version of *OpenFace 2* evaluated in this paper is version 2.2.0. This can be accessed via the releases page on the *OpenFace 2* GitHub repository [1]. The framework allows for facial landmark detection: the prediction of where parts of the face are located (e.g, nose, mouth, eyebrows); and facial action coding (FAC): the assessment of the expression of a subject using Ekman's standardised notation [46]. Moreover, it is also capable of gaze and head rotation estimation. Example use cases for *OpenFace 2* utilise the multiple physiological modalities measured to allow robot teachers to adapt and react to students' levels of engagement [25], or make continuous predictions of emotion [26].

There are two ways of using *OpenFace 2*: via the command line interface (CLI) and the graphical user interface (GUI). The first affords the user a series of high-level functions, capable of performing each task individually or together (e.g., head pose estimation, gaze estimation, automatic FAC) with a single line of code. The second approach encapsulates a series of graphical applications which allow for the configuration and use of *OpenFace*, either offline or in real-time through interfacing with a local camera (e.g., a webcam), without the need to write any code at all. Both approaches allow for the same functionality and control in terms of available parameters. *OpenFace 2* comes pre-trained and does not require any supplementary software, or parameter tuning, to extract physiological data from a video signal.

OpenFace 2 packages together solutions from existing publications, for each respective task, and then applies some optimisation before configuring them into a pipeline with an easy-to-use interface. In terms of FAC, the face is initially recognised [47], landmarks are detected and tracked [48], and then these landmark locations are turned into a set of FAC intensities [49]. In terms of speed of execution, the most computationally expensive task of facial landmark detection and tracking is capable of running at "*30–40 Hz frame rates on a quad-core 3.5 G Hz Intel i7-2700K processor, and 20 Hz frame rates on a Surface Pro 3 laptop with a 1.7 GHz dual-core Intel core i7 4650U processor, without any GPU support when processing 640 × 480 pixel videos*" [30]. Hence, it may be interpreted that when processing higher resolution or frame-rate videos, or using a computer exhibiting lower computational performance, real-time processing may be implausible.

OpenFace 2 is described as being more robust to occlusion and low illumination conditions than the original *OpenFace* toolkit [30]. This improvement was made by including examples of these challenging conditions in the training dataset used to develop the machine learning models applied in the facial landmark detection and face tracking subsystems. However, the performance of the FAC system under low illumination cannot be validated. This is due to the fact that there are, to the best of our knowledge, no readily available FAC datasets which contain any examples of dark or changing illumination. One noted limitation of the system is that faces less than 100 pixels across are not accurately assessed by the facial landmark detection system. For more information regarding the training data and technical implementation used in *OpenFace 2*, as well as performance benchmarks for all physiological modalities, refer to the associated publication [30].

Researching factors which influence the performance of automatic FAC solutions is complex. This is due to the difficulty associated with obtaining ground truth data. Commonly, this process requires a team of skilled researchers to manually assess and transcribe the facial movements from a video recording. This process takes a trained facial action coder around 50–60 min for every minute recorded. Hence, there are a limited number of examples of videos, with associated FAC, captured under challenging conditions available to researchers. The FAC recognition system in *OpenFace 2* was trained on DISFA [50], SEMAINE [51], BP4D [52], UNBC-McMaster [53], Bosphorus [54] and FERA 2011 [55].

3. Methodology

The aim of this research is to explore the feasibility of using a multimodal camera-based physiological measurement solution for the assessment of cinematic viewing experiences. As dark room illumination is a feature of a cinematic viewing experience but bright room illumination is recommended for camera-based physiological measurement systems, this factor is investigated in depth. Hence, we are investigating the optimal solution in which the data obtained is viable for psychological insight and the room illumination does not detract from the viewing experience. Based on existing research, we hypothesise that increasing room illumination will positively influence the performance of the camera-based physiological measurement solution.

To evaluate this hypothesis, a comparison is made between camera-measured signals, extracted from a single video file per participant, and expected signals. This is done for heart rate (HR) and facial action coding (FAC) in three different illumination conditions. To obtain an expected signal for comparison with heart rate measurements from *PyVHR*, a *Polar H10* chest strap heart rate monitor sensor, purchased from the Polar UK website [56], was used. To obtain an expected signal for the facial actions, to be compared to each participant's camera-measured response, a novel instructional film was created as a means to reduce the common labour-intensive approach of manually annotating each video.

3.1. The Instructional Film

The institutional film prompts the replication of a series of predetermined facial movements as exemplified by an on-screen model. Hence, the expected signal was obtained by analysing the movements of this model. Specifically, this was done by manually coding the video based on Ekman's FAC system through cross-referencing information found in the FAC Manual [46]. Therefore, the instructional film method compares the participant's movements, as measured by the camera-based FAC solution, to a transcription of the model's movements, made by the investigator, to gain insight into the performance of the camera-based FAC solution.

This instructional film approach is considered desirable, as the common alternative of manually coding each video is highly labour-intensive. As over 4 h of footage was captured in this study, this would have taken approximately 220 h. However, inherently there is a trade-off associated with this approach. The discrepancy between the expected response signal and the camera-measured signal is influenced by two factors: system performance, reflecting the accuracy of the system in measuring the individual's physiological state; and task performance, involving the individual's ability to synchronously reproduce the prompted action in accordance with on-screen instructions.

To produce a dataset Cosker et al. used a similar protocol in which a group of participants were asked to recreate facial actions to the best of their ability [57]. Moreover, Gosselin et al. showed that over 70% of their participants tested were able to activate target action units—although they often co-activated other non-target action units when trying to do so [58]. Therefore, it was considered feasible that participants should be able to recreate facial action adequately enough to provide insight into differences in system performance for *OpenFace* 2, between the illumination conditions as long as only the target expressions are analysed. This is, however, based on the assumption that participants perform the facial actions task consistently between trials.

There are also some benefits of using an instructional film for assessing the performance of the camera-based heart rate measurement system. When assessing responses to screen-based media, facial expressions are often elicited. Hence, the task of reproducing facial movements can be considered representative of these spontaneous movements. Therefore, the influencing factor of motion is accounted for using this method when assessing the system performance. Moreover, as the instructional film is delivered via a screen, the changes in illumination on the skin are also representative of those which may be elicited from stimuli used for exploring psychophysiological responses to screen-based media.

To create the instructional film, video content provided courtesy of IMotion was used to prompt the 18 facial actions *OpenFace 2* is capable of recognising [59]. These short clips, each showing a different facial action, were repeated three times in a row before moving on to the next one. The repetitions of each action were sequenced rhythmically and consistently, with the expression onsets starting every three seconds throughout the film so that the timing of the movement was predictable. To allow for the synchronisation of the video footage with the instructional film, a visual marker was placed in the instructional film in the form of a white flash at the start, which created a clear change in illumination visible on the viewer's skin. This was followed by a black screen which lasted 15 s, and then the instructions began. Figure 1 shows the overall brightness of the instructional film, showing the bright flash at the start followed by a dark screen which precedes the facial action stimuli. The instructional film was 177 s long and rendered at 30 FPS at 720p.

It should be noted that further tasks were also included in the instructional film, which continued after the aforementioned 177 s, notably prompting head and eye movements. The associated data is not analysed in this paper.

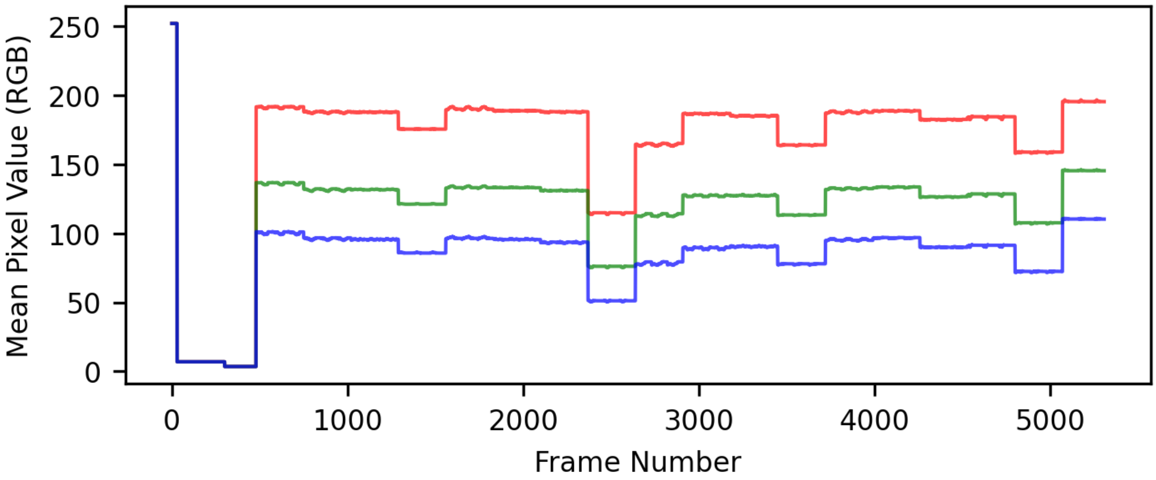

Figure 1. Plot of mean pixel values for the red, green, and blue channels in the instructional film. Note the bright screen at the start of the instructional film, followed by the short dark section, and then the series of facial action stimuli that the participants are tasked with reproducing.

3.2. Experimental Protocol

The experimental protocol can be seen in Figure 2. Trial counterbalancing was used to avoid participants' improving task performance influencing the outcomes of the study. Specifically, this meant randomising the order of the illumination conditions for each participant. This was done to help address concerns regarding the assumption underlying the use of the instructional film: that participants perform the task consistently between trials. By randomising the order of the illumination conditions, the influence of any hypothetical improvements in task performance with each subsequent trial is reduced. Note, that the term *trial* is used here to refer to each attempt at the instructional film.

A familiarisation task was also included in the protocol to reduce variation in task performance with each subsequent trial. Specifically, subjects were shown the video clips associated with each facial action included in the instructional film, via the Qualtrics form, and prompted to replicate each one until they felt they could reproduce it to the best of their ability. Participants were allowed as much time as they needed, which was typically around 15 min (exact data was not recorded). To ensure participants did not rush through this familiarisation task, they were asked to assess the difficulty of replicating each individual facial action. This data could then also be used to help evaluate the efficacy of the instructional film approach. To aid the practice of producing the facial actions, participants were given a mirror for the duration of the familiarisation task. After practising each movement, the participants were also given the chance to familiarise themselves with the speed and style of the instructional video through the viewing of a short preview. After this preview, they were asked to assess the pacing of the activity.

As part of the protocol participants were also asked to sit quietly and wait for three minutes before each trial. This length of time was chosen based on the following criteria:

- To allow participants to approach a baseline heart rate, hence providing a similar cardiovascular signal between trials allowing for a fair comparison of system performance.
- To reduce the intensity of the experiment session, allowing participants to relax and not become overwhelmed to the point they cannot maintain a consistent level of task performance.
- To not reduce the intensity of the experiment session to the point where participants are bored, which may also affect their task performance.
- To provide enough time for the investigator to prepare for the next trial.

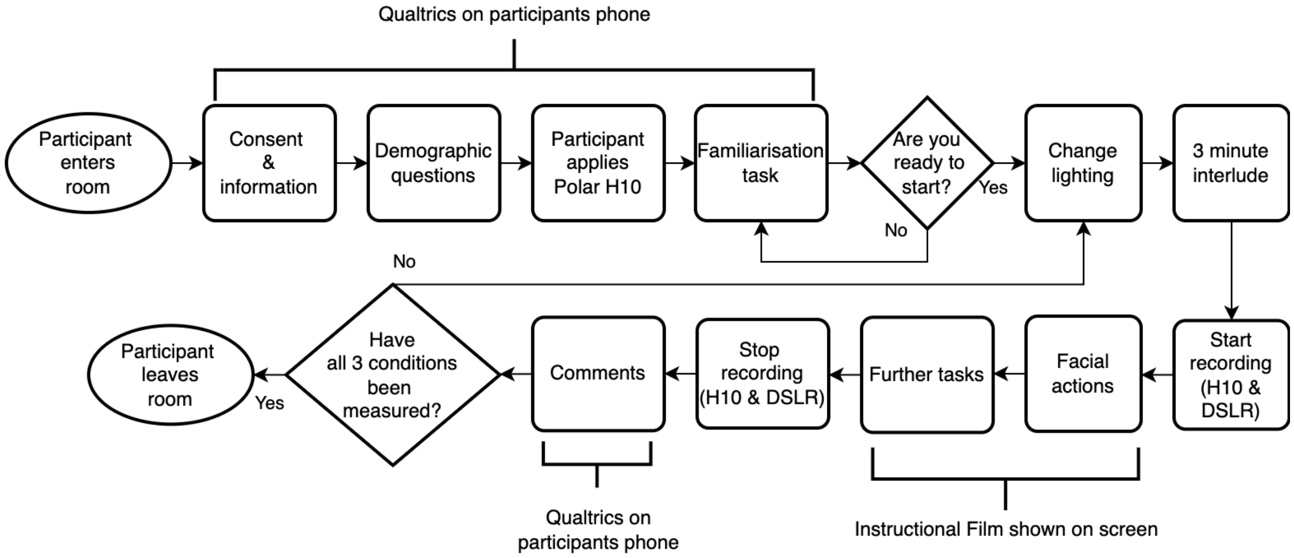
Qualtrics on participants phone
Participant enters room
Consent & information
Demographic questions
Participant applies Polar H10
Familiarisation task
Are you ready to start?
Yes
No
Change lighting
3 minute interlude
Start recording (H10 & DSLR)
Facial actions
Further tasks
Stop recording (H10 & DSLR)
Comments
Have all 3 conditions been measured?
No
Yes
Participant leaves room
Qualtrics on participants phone
Instructional Film shown on screen

Figure 2. Experimental protocol flow chart, showing each step in the study for each participant.

The video data captured in this study could not be anonymised, as participants' faces had to be clearly visible in the recordings. Hence, based on legal requirements, the decision was made to only keep this footage for the necessary period for processing—evaluated as 60 days by the researchers involved [60]. After this point in time, the data extracted from the footage could subsequently be maintained with no way of associating the identity of the participant with their data. This was achieved by destroying all communications between participants and investigators, such as replies to recruitment emails. Moreover, the randomly generated participant codes from Qualtrics were used to label the data pertaining to each subject. To ensure that the data were secure, password protection was used on the folders in which the data were stored.

Participants were all employees of Bang & Olufsen in Struer, Denmark. There was no financial reward for being a part of the study. To be eligible, participants had to be over the age of 18 and have no known heart problems. Informed consent was obtained from all subjects involved in the study and participants were told they could leave the study at any point, without reason, and their data would be destroyed. The protocol, handling of data, and all other ethical factors were approved by the University of York, School of Physics, Technology & Engineering ethics committee (reference: Williams291121).

In total 22 participants were recruited. However, two participants were later discounted due to missing data. In one trial the investigator failed to press record on the camera; in another, the *Polar H10* did not record during one trial due to a loss of Bluetooth connection. Hence, the subsequent analysis is made on the remaining 20 participants, for which all data were obtained.

The demographic questions asked in this study help to describe who was in the sample in terms of possible influencing factors applicable to camera-based physiological measurement. Participants were asked to report: their gender; whether they had facial hair; whether they were wearing glasses; and if they had performed vigorous exercise or consumed caffeine, alcohol, or nicotine, in the past 3 h. The results of these questions can be seen in Table 2. Participants were also asked to note their age and assess their skin tone using the Monk scale [61]. The maximum, minimum, and mean ages of participants were 18, 41 and 63, respectively. The maximum, minimum, and mean skin tone was 1, 3, and 7, respectively.

Table 2. Description of the sample ascertained from responses to the questions presented to each participant at the start of the study.

Participant Question	Option	Frequency	%
Gender	Female	3	15
	Male	17	85
Factor in last 3 h	Nicotine	2	10
	Alcohol	1	5
	Caffeine	13	65
	Exercise	1	5
	None	5	25
Glasses	Yes	11	55
	No	9	45
Facial hair	Yes	7	35
	No	13	65

3.3. Experimental Setup

To recreate a cinematic viewing experience, a suitable viewing distance had to be chosen. The International Telecommunication Union standard for the subjective assessment of the quality of television images (ITU-R BT.500-14) [28] recommends a viewing distance of 3.2 times the height of the screen and this condition was replicated for this study. This recommendation relates to the optimum distance for assessing image quality, in which the screen inhabits a wide field of view. For the 46-inch (corner to corner) screen installed in the room where the experiment took place, this meant that the viewing position was 183 cm away from the screen. Notably, the ITU recommendation for the subjective assessment of multi-channel stereophonic sound systems with accompanying images (ITU-R BS.775-4) [29], similarly recommends 3 times the height of the screen for such studies.

To decide on a suitable set of room illuminations for testing, recommendations from ITU-R BT.500-14 were used, alongside considerations for the practicality of reproduction in the room where the study took place. Three room-illumination levels were chosen to correspond to the 'maximum', 'minimum', and 'off' settings deliverable using the lighting interface within the room. The darkest setting was representative of the illumination level considered optimal for the assessment of image quality, the medium setting was representative of a dimly lit domestic environment, and the bright setting was representative of a well-lit domestic environment [28].

The illumination conditions were measured by pointing a lux meter directly upwards towards the ceiling and then forward towards the television screen from the viewing position. For these two directions, readings were taken for the maximum and minimum light intensity, corresponding to a white and black image being shown on the screen, in all three illumination conditions. The screen brightness did not change between conditions or trials. All of these readings can be seen in Table 3. All lighting conditions were created by controlling two white light LED panels the centres of which are 1.2 m above, 1 m to the left and right, and 83 cm in front of the viewer's head. There only other light source in the room was the television screen. The experimental setup can be seen in Figure 3.

Researchers commonly use specific webcams or specialised scientific cameras when trying to record heart rate from video (see Section 2.1). However, in this study participants were recorded using a *Panasonic Lumix DMC-FZ1000* digital single-lens reflex (DSLR) camera (purchased in Denmark), placed discretely underneath the screen on a fixed tripod. This camera was chosen as it has an optical zoom, allowing for video to be captured at a distance without decreasing the resolution. Moreover, DSLR cameras can generally open to a wider aperture and have larger, more sensitive, sensors than webcams. This means they are more suited to capturing video footage in darker settings. This is also considered a reproducible solution, as the *Panasonic Lumix DMC-FZ1000* is an entry-level DSLR camera which has been commercially available since 2016. Many other DSLR cameras are capable of recording with the same, or similar, technical parameters. The 13.2 × 8.8 mm sensor found in the *Panasonic Lumix DMC-FZ1000* is typical of entry-level DSLR cameras. One trade-off with using a DSLR camera is that video compression is unavoidable. As noted in Section 2.1, compression is known to reduce the performance of rPPG solutions. However, compression is also very useful for ensuring the manageability of data collected in terms of both storage capacity and playback compatibility.

Table 3. Description of the illumination conditions, made up of the illumination incident on the participant due to the screen and the room lighting. The minimum illumination measurement was made with the television displaying a black screen and the maximum with a white screen.

Illumination	Direction	Maximum (lux)	Minimum (lux)
Bright	Upward	370	349
	Forward	270	192
Medium	Upward	60	46
	Forward	93	23
Dark	Upward	5	0
	Forward	20	0

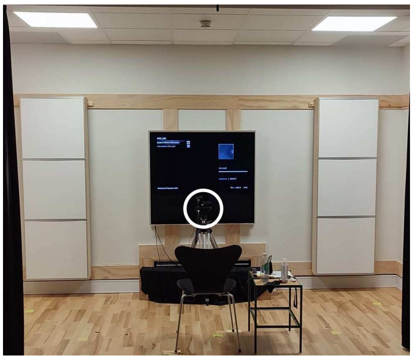

Figure 3. Photograph taken in the room where the experiment took place, showing the participant's chair, lighting, and position of the camera on a tripod under the television (highlighted by the white circle).

The parameters chosen for shooting were: high resolution (1080p), lowest video compression ratio available (28 Mbps H.264), zoom to frame face (50 mm), highest aperture possible at the zoom position (f/3.4), and slowest feasible shutter speed to create the longest exposure possible whilst shooting at 50 FPS (1/60 s). The decision was made to use a progressive encoding format, rather than an interlaced format (e.g., 1080i), due to the fact *PyVHR* and *OpenFace 2* both use frame-wise processing. The shutter speed was deemed appropriate as no noticeable motion blur was found when recording facial actions in a pre-study test, as assessed using frame-by-frame inspection. The only difference in camera settings between the three room illuminations was the ISO setting, which relates to the sensitivity of the sensor to incoming light. The bright, medium, and dark sensitivity settings were set to ISO400, ISO1600, and ISO6400, respectively. Maintaining the ISO400 sensitivity would render the participant's face invisible in all but the bright illumination.

The data compression applied by the camera provides a roughly 100:1 reduction in data, as uncompressed video at the same specification would have a bit rate of 2.98 Gbit/s. This value is obtained by multiplying the number of channels (3) by the bit depth of each channel (8) and then multiplying this by the resolution (1080 × 1920) and the frame rate (50). McDuff et al. found that applying a similar compression ratio artificially led to an increase in average error from 1.78 BPM to 2.68 BPM. However, it should be noted that their system did not apply skin patch processing and the correlation between rPPG and contact sensor readings is not reported in this publication [42].

4. Analysis

The null hypothesis for this study is that room illumination has no influence on the mean absolute error (MAE) and Pearson correlation (PCC) performance metrics. There is also another null hypothesis, associated with the assumption of consistent task performance discussed in Section 3.1, which states that there is a difference in MAE and PCC metrics for the FAC between trials. To obtain the PCC and MAE metrics, the camera-measured signals, outputted by *OpenFace 2* and *PyVHR*, are compared to the expected signals, associated with perfectly replicating the model's movements and data recorded by the contact sensor (*Polar H10*), respectively. Subsequently, these metrics are assessed for each physiological modality using statistical testing to evaluate if they have statistically significant different mean values. As well as assessing if there is a difference in the performance metrics between the three illumination conditions, the analysis in this section also aims to explore the nature of these differences.

4.1. Preprocessing

To make the analysis in this study possible, the camera-measured signals and expected signals first had to be obtained, synchronised, and compiled into a dataset. An outline of this process can be seen in Figure 4.

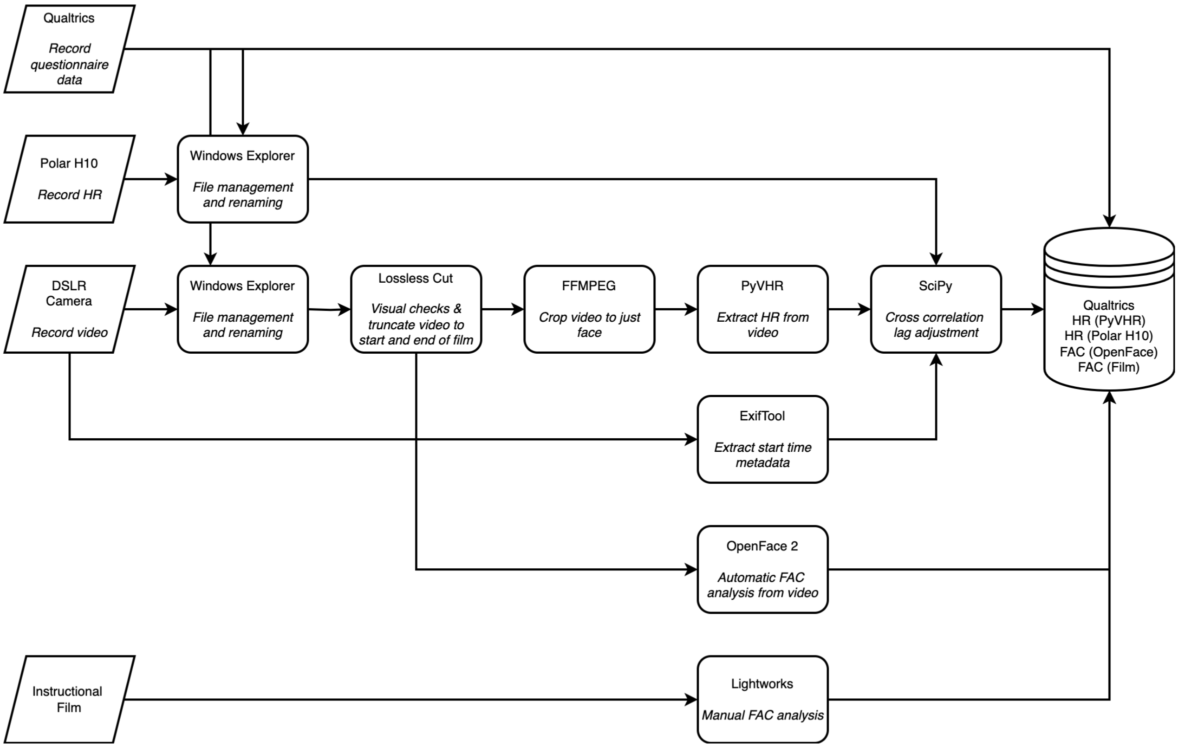

Figure 4. Data preprocessing flow chart showing how all the data were collected for the study and processed ready for further analysis (see Section 4).

Initially, the instructional film was manually coded to obtain the expected signal for FAC. For each movement, the first, last, and peak frame number reference was transcribed. Next, the intensity of the model's movement at this peak was also assessed, through cross-examination with examples within Ekman's FAC Manual [46]. Using this data, an array of expected facial movement intensities could be produced which matched the sample rate of the recorded videos (50 Hz). This expected response signal was univariate, only describing the activation of the target expression being demonstrated by the on-screen model in the instructional film.

The video footage was cut so that the recordings began at the start of the instructional film. This was done with *LosslessCut* [62], a software solution for segmenting video files without re-encoding, using the visual marker discussed in Section 3.1. This also allowed for visual inspection, confirming that all the footage was in focus and all participants had made an attempt at the instructions prompted in the film. The truncated video was then processed with *OpenFace 2 (2.2.0)* using the CLI *FeatureExtraction* command, to output the camera-measured signal [1]. This was then sliced to create a univariate signal containing only the target expressions.

The *Polar H10* is able to process electrocardiogram (ECG) data in real-time on the device and deliver a heart rate value each second via Bluetooth, as well as a timestamp, hence no pre-processing was needed to obtain this expected heart rate signal. To obtain the camera-measured signal, *PyVHR* was used to process the video footage from the start of the instructional film until the end of the facial actions task. The specific rPPG process used is described in the demo file available via the associated GitHub repository [2]. Notably, the six-region approach was preferred to the one-hundred-region approach for facial patch sampling as it was considered to reduce the influence of facial hair and glasses by not sampling these regions.

When processing the video data using *PyVHR (1.0.2)*, it became clear that processing any more than one minute of video with the patches extraction subprocess enabled, led to a 'memory full' error. To address this issue, each video was cropped using the *FFmpeg* library at lossless quality [63]. The crop window removed much of the background, whilst still ensuring that none of the participants' faces moved out of frame at any point in any of the videos. Specifically, the video was cropped to a 700 pixel by 1080 pixel frame, centred in the middle of the original 1920 pixel by 1080 pixel frame. Reducing the resolution in this way caused the file sizes to increase, due to the lossless re-encoding, but led to faster execution times in *PyVHR* and successful processing without error.

After the video data had been processed using *PyVHR*, to obtain the camera-measured signal, the *Polar H10* data was synchronised with this data by aligning the associated timestamps. To obtain these millisecond precision timestamps from the file's metadata, *ExifTool* was used [64]. However, the alignment of the two signals appeared to be incorrect using this method. Hence, the synchronisation was reassessed by analysing the maximum value of the cross-correlation function, with the expected (*Polar H10*) and camera-measured (*PyVHR*) signals as inputs. This assessment was made for all videos in bright condition. Implementation notes for this process are accessible via the *SciPy* library documentation [65]. The outcome of this process suggested that the internal clock of the *Polar H10* sensor was 54 s behind the timestamps obtained from the camera. Subsequently, this delay was accounted for.

During the preliminary analysis of the data, it became apparent that for some video frames no measurement had been made by the camera-based systems. This occurred when the face was barely illuminated at the start of the facial action instructions in the darkest setting. Another, less common reason for no measurement being made was during a facial occlusion event. For example, one participant adjusted their glasses at the end of the instructional film and two participants stood up at the end of their third and final trial. As no facial occlusion events occurred during the facial actions task this was deemed to have had no influence on the outcomes of this study.

4.2. Qualtrics Analysis

The difficulty assessment made during the familiarisation task can be seen in Figure 5. These data show that some facial actions were considered to be harder to reproduce than others. They also show that the perceived difficulty of the facial actions varied between participants. Participants also commented on the difficulty of performing the first two facial actions separately: AU01 and AU02. However, it is clear that on average the facial actions were not assessed as being difficult. A question relating to the pacing of the facial actions was also presented to the participants, for which all but one assessed the speed and time between actions as optimal (not too fast or too slow).

A built-in feature in *Qualtrics* was used for counter-balancing, randomising the order of illumination conditions for each participant. However, the counterbalancing was not applied evenly. Notably, more participants had their first trial in the bright illumination compared to the other two conditions. In Table 4, the outcomes of the automatic counterbalancing process are shown in full.

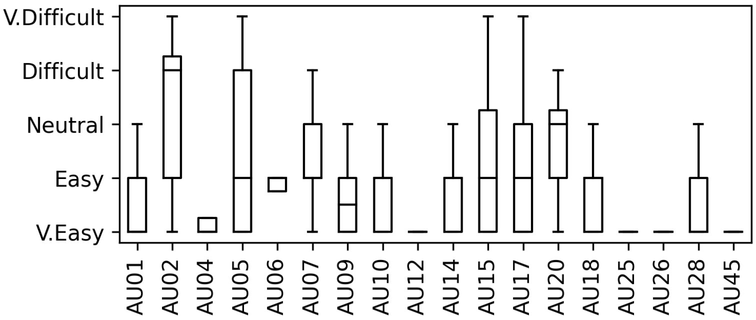

Figure 5. Box plot showing the distribution of difficulty scores for each facial action as assessed by the participants in the study. The features of the box plot represent the minimum, maximum, and quartiles of the response distributions.

Table 4. Table showing the order in which the illumination conditions were presented. Note the imperfect outcomes of the counterbalancing process automatically managed by *Qualtrics*. Ideal counterbalancing would lead to all values in the frequency column being the same.

Illumination	Trial Order	Frequency	%
Bright	First	9	45
	Second	5	25
	Third	6	30
Medium	First	4	20
	Second	6	30
	Third	10	50
Dark	First	7	35
	Second	9	45
	Third	4	20

4.3. Facial Action Analysis

These analyses compare the 177 s of FAC data extracted from each video by *OpenFace 2* of participants responding to the instructional film, with the 177 s of data manually coded from the movements of the model in the instructional film. The sample rate for both of these univariate signals is 50 Hz (see Section 4.1). The relationship between the camera-measured target expression and the expected target expression, from the manual coding, can be quantified by calculating the MAE and PCC for each video. Both of these signals are visualised in Figure 6. In Figures 7 and 8 the distribution of these performance metrics is shown, for each illumination condition.

The distribution of each metric in each illumination condition was tested for normality using the Shapiro–Wilk test [66]. All six distributions were found to be normally distributed, so parametric statistics were used in the subsequent analysis. For both the MAE and PCC distributions the assumption of sphericity was met, as assessed using Mauchly's test of sphericity [67]. Hence, a repeated measures ANOVA [68] was applied revealing that there were significant differences in both MAE and PCC between the illumination conditions (MAE: F = 8.251, p = 0.001) (PCC: F = 9.427, p < 0.001).

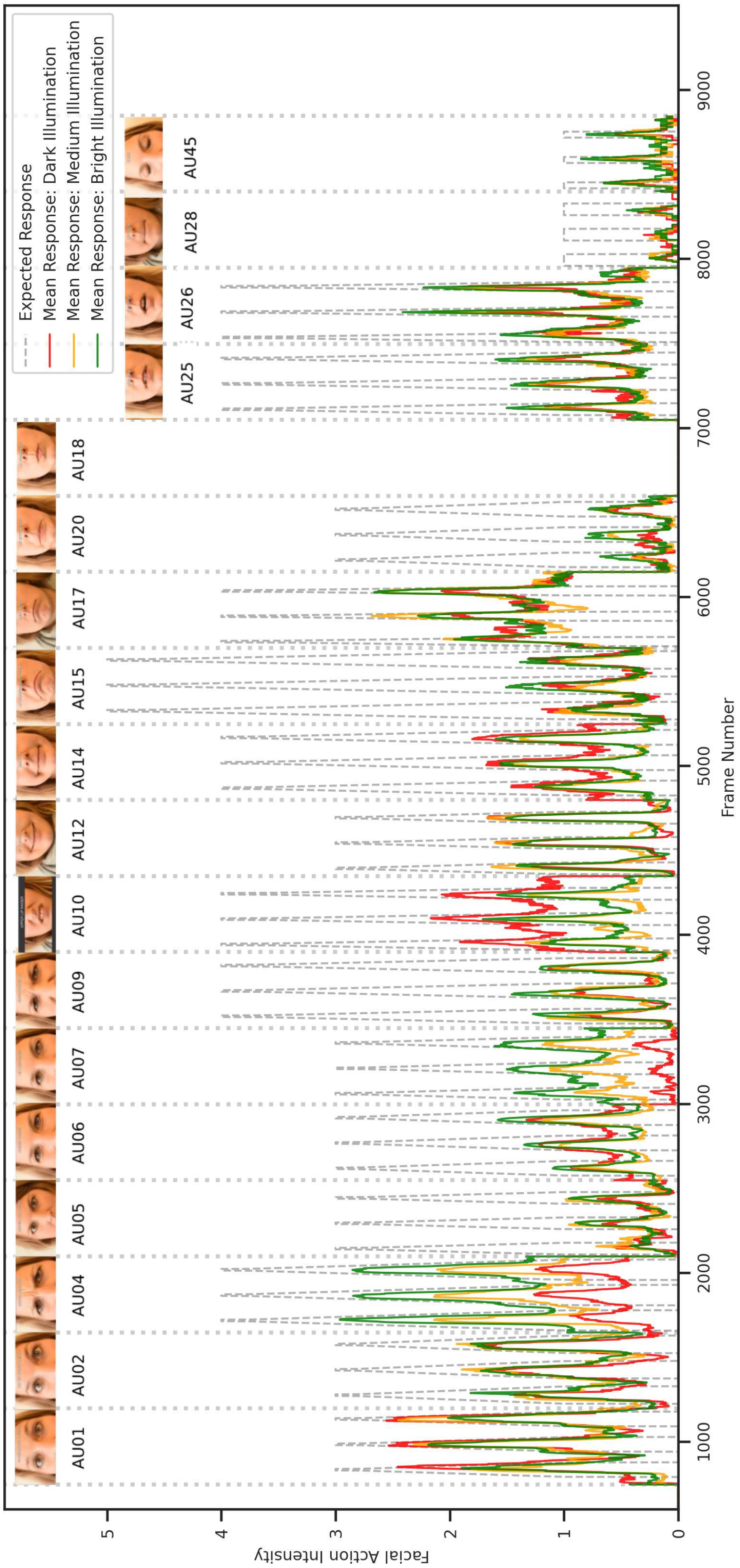

Figure 6. Figure showing the mean of the camera-measured target facial action for each illumination condition for each frame, alongside the expected response associated with perfectly replicating the instructional film. Note that AU18 was accidentally included in the film despite not being recognised by *OpenFace 2*, hence data captured during this time was ignored.

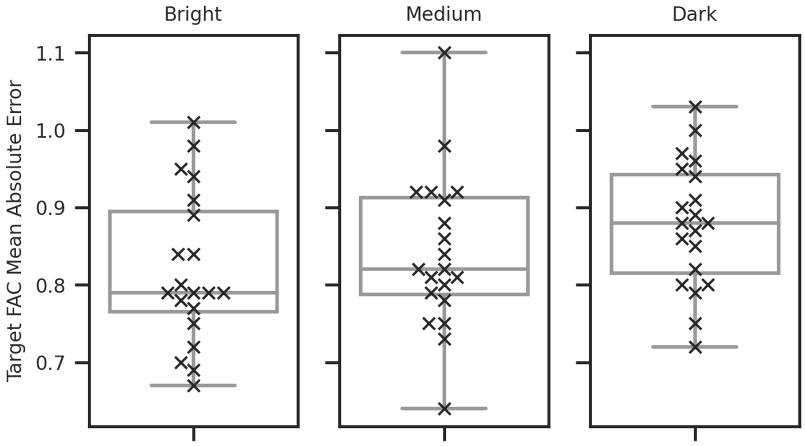

Figure 7. Box plot showing the distribution of MAE values quantifying the difference between the expected target facial action signal and the signal camera-measured signal for each illumination condition. Each cross, in the overlaid swarm plot, shows an individual value within the distribution. The means and standard deviation of the distributions are as follows: Bright (mean: 0.82, std: 0.10), Medium (mean: 0.84, std: 0.10), Dark (mean: 0.88, std: 0.08).

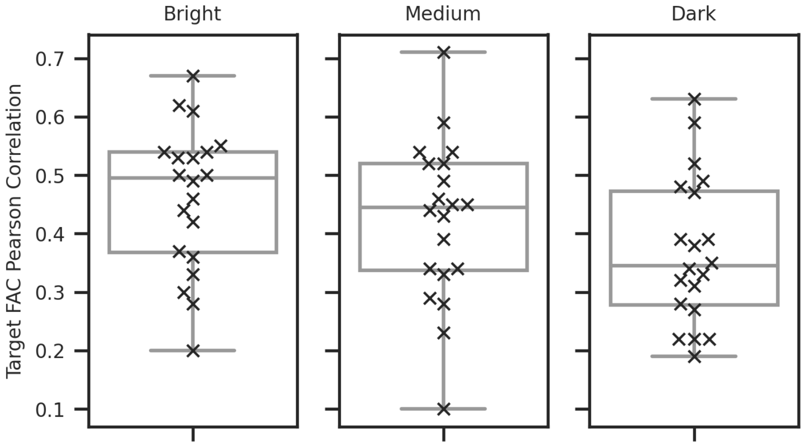

Figure 8. Box plot showing the distribution of PCC values quantifying the relationship between the expected target facial action signal and the camera-measured signal for each illumination condition. Each cross, in the overlaid swarm plot, shows an individual value within the distribution. The means and standard deviation of the distributions are as follows: Bright (mean: 0.46, std: 0.12), Medium (mean: 0.42, std: 0.14), Dark (mean: 0.37, std: 0.12).

As post hoc analyses, pairwise comparisons were made to assess the significance of the difference in mean PCC and MAE between each illumination condition. This was done using the dependent *t*-Test. As multiple comparisons are made Bonferroni adjustment [69,70] is applied, leading to an adjusted significance threshold of $p < 0.008$. These results can be seen in Table 5. They show a significant difference between the Bright and Dark conditions for both MAE and PCC.

To test the assumption that participants perform the facial actions task consistently between trials, the differences between the MAE and PCC metrics for each participant's first, second, and third trials were assessed. All six distributions were found to be normally distributed using the Shapiro–Wilk test [66], so parametric statistics were used in the subsequent analyses. The two sets of three distributions of performance metrics, corresponding to the MAE and PCC in each illumination condition, were found to meet the assumption of sphericity using Mauchly's test [67]. A repeated measures ANOVA [68] was applied revealing that neither the MAE nor PCC approached statistically significant differences in mean value between the illumination conditions (MAE: F = 0.553, $p = 0.580$) (PCC: F = 0.614, $p = 0.547$). The data analysed in this paragraph can be seen in Figures 9 and 10.

Table 5. Results of dependent t-tests assessing differences in mean for the performance metric distributions between illumination conditions, for automatic FAC. Statistically significant differences in mean are shown in bold.

FAC Distributions for Comparison			*t*-Test (A vs. B)	
Metric	**Condition A**	**Condition B**	***t***	***p***
PCC	Dark	Medium	−2.299	0.033
	Dark	Bright	−4.109	**<0.001**
	Medium	Bright	−2.156	0.044
MAE	Dark	Medium	2.428	0.025
	Dark	Bright	3.977	**0.001**
	Medium	Bright	1.566	0.134

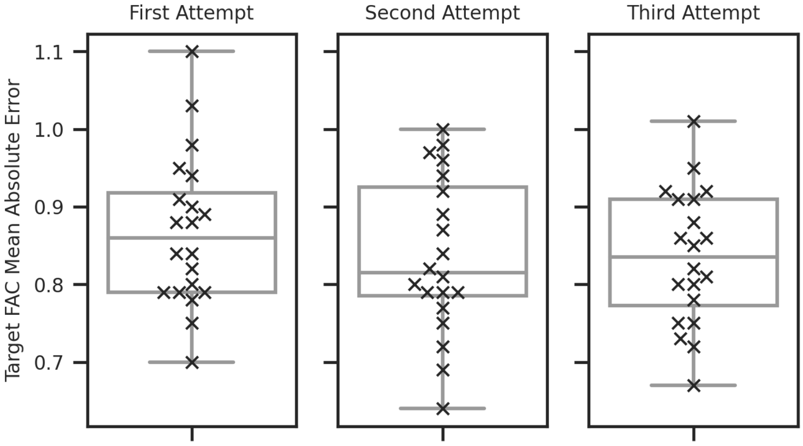

Figure 9. Box plot showing the distribution of MAE values quantifying the difference between the expected target facial action signal and the camera-measured signal for each attempt at the facial actions task instructed by the film. Each cross, in the overlaid swarm plot, shows an individual value within the distribution. The means and standard deviation of the distributions are as follows: First (mean: 0.39, std: 0.13), Second (mean: 0.43, std: 0.14), Third (mean: 0.43, std: 0.12).

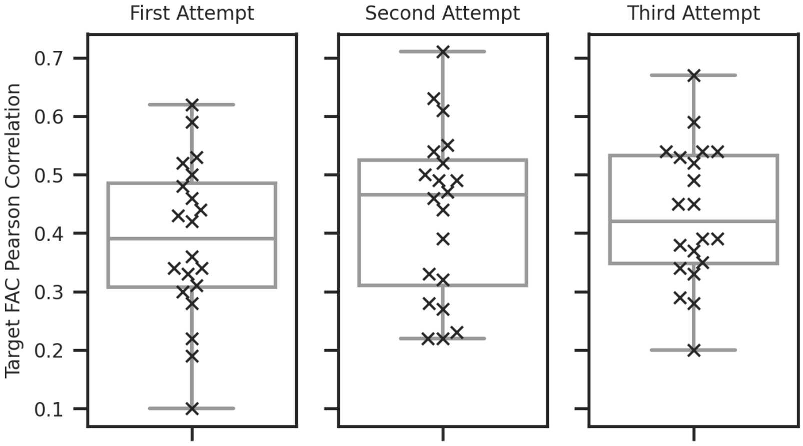

Figure 10. Box plot showing the distribution of PCC values quantifying the relationship between the expected target facial action signal and camera-measured signal for each attempt at the facial actions task instructed by the film. Each cross, in the overlaid swarm plot, shows an individual value within the distribution. The means and standard deviation of the distributions are as follows: First (mean: 0.87, std: 0.10), Second (mean: 0.84, std: 0.10), Third (mean: 0.84, std: 0.09).

Through visual inspection of the camera-measured signals, it appeared that there are three characteristic errors. First, on average, the camera-measured facial action was of lower intensity than that expressed by the model. Second, *OpenFace 2* often measures facial actions with an intensity between zero and one, whereas the expected signal is never assessed in this range of values. This is due to Ekman's system defining a magnitude of one being the minimum observable intensity [46]. In Figure 11, both of these characteristic errors can be seen.

The third characteristic error is that the camera-measured signal appears to make small erratic changes, even when the facial expression does not appear to change in the associated video recording. In Figure 12, the power spectral density of the target signal measured by *OpenFace 2* is shown for all data collected under each of the three illumination settings. It has been suggested that the fastest micro-expressions last no longer 65 ms [71]. In all three cases, there appear to be frequency components representative of movements lasting less than this duration (15.38 Hz). The power spectral density was calculated using the default implementation in the associated *matplotlib* function [72], with an increased window length of 1024 applied to optimise frequency resolution for visualisation, applied to the concatenated target expressions signals for each condition. Power spectral density was chosen for this visualisation as it allows for a clear comparison of the power of each frequency integer band.

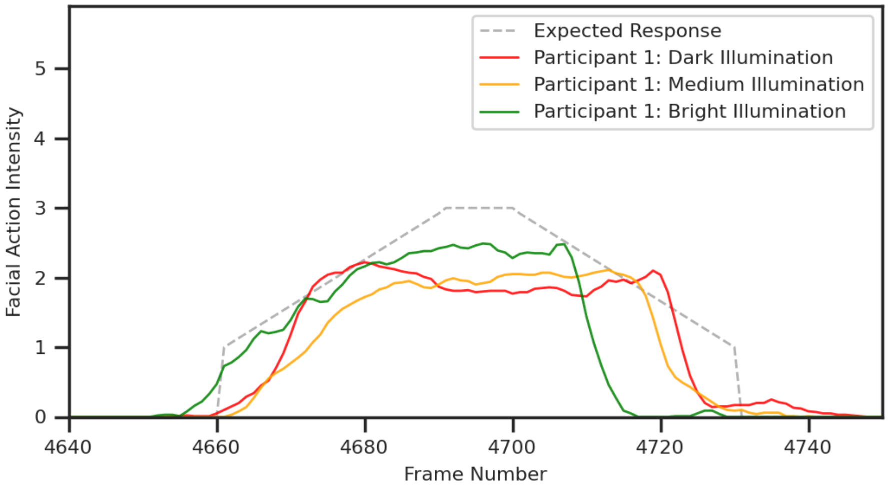

Figure 11. Figure showing the camera-measured response for participant one to the AU12 facial action for each illumination condition, alongside the expected response associated with perfectly replicating the instructional film. Note that the camera-measured data from all three conditions are lower in magnitude than the expected response, contain values between zero and one, and contain small erratic changes in magnitude.

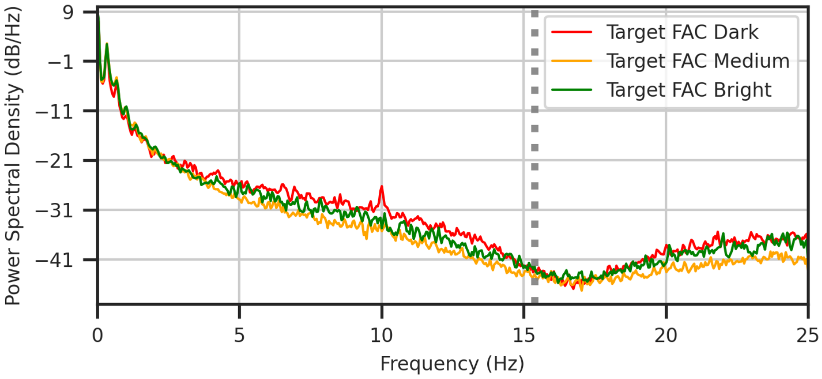

Figure 12. Power spectral density of recorded facial action code data from *OpenFace 2*, for the target expression. Research into the speed of microexpressions suggests that there should be no frequency components to the right of the grey line (15.38 Hz) [71].

4.4. Heart Rate Analysis

These analyses compare the 177 s of HR data extracted from the video footage by *PyVHR* of participants responding to the instructional film, with the 177 s of HR data measured by the *Polar H10* contact sensor. The sample rate of both these univariate signals is 1 Hz. The relationship between the camera-measured signal and the expected heart rate, from the *Polar H10* contact sensor, can be quantified by calculating the average MAE and PCC for each video. In Figures 13 and 14 the distribution of these performance scores is shown, for each illumination condition.

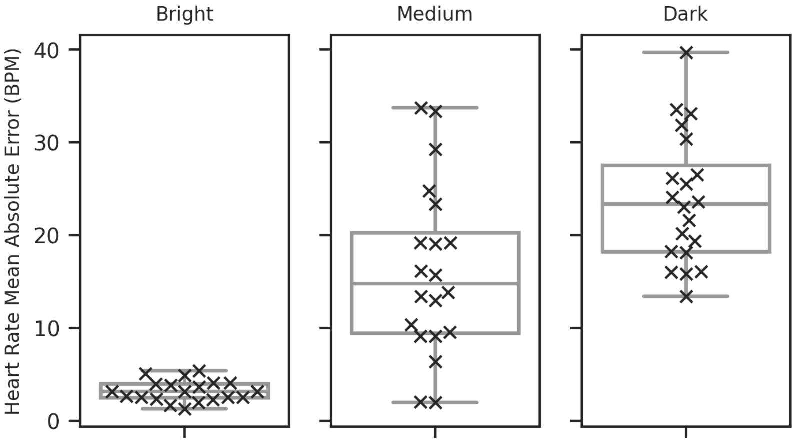

Figure 13. Box plot showing the distribution of MAE values quantifying the difference between the expected target heart rate signal from the contact sensor and the camera-measured signal for each illumination condition. Each cross, in the overlaid swarm plot, shows an individual value within the distribution. The means and standard deviation of the distributions are as follows: Bright (mean: 3.20, std: 1.11), Medium (mean: 16.12, std: 9.01), Dark (mean: 23.81, std: 6.89).

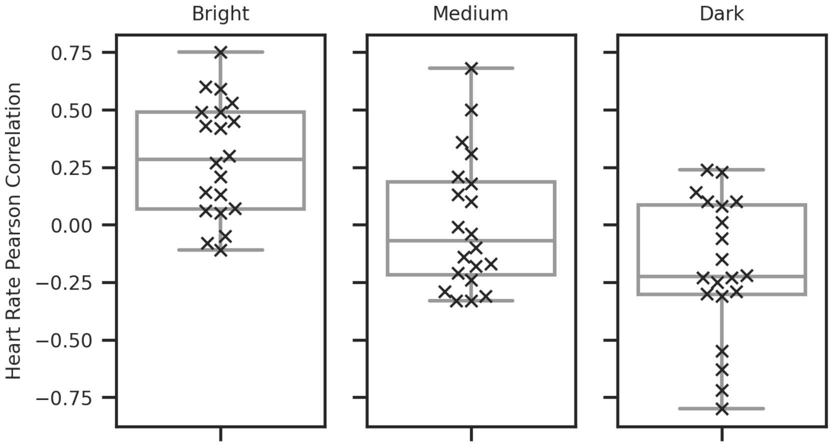

Figure 14. Box plot showing the distribution of PCC values quantifying the relationship between the expected target heart rate signal from the contact sensor and the camera-measured signal for each illumination condition. Each cross, in the overlaid swarm plot, shows an individual value within the distribution. The means and standard deviation of the distributions are as follows: Bright (mean: 0.28, std: 0.24), Medium (mean: 0.01, std: 0.29), Dark (mean: 0.19, std: 0.30).

The distribution of each metric in each illumination condition was tested for normality using the Shapiro–Wilk test [66]. All six distributions were shown to be normally distributed, so parametric statistics were used in subsequent analysis. The PCC distributions were found to meet the assumption of sphericity, using Mauchly's test of sphericity [67]. However, the MAE distributions did not meet the assumption of sphericity, so the Greenhouse–Geisser correction was applied [73]. A repeated measures ANOVA [68] was applied revealing that there were significant differences in both mean MAE and PCC values between the illumination conditions (MAE: F = 43.415, $p < 0.01$) (PCC: F = 14.716, $p < 0.01$).

As post hoc analyses, pairwise comparisons were made to assess the significance of the difference in mean MAE and PCC between each illumination condition. This was done using the dependent *t*-test. As multiple comparisons are made Bonferroni adjustment [69,70] is applied, leading to an adjusted significance threshold of $p < 0.008$. These results can be seen in Table 6. They show a statistically significant difference for all pairwise comparisons, apart from between Dark and Medium conditions for PCC.

Table 6. Results of dependent *t*-tests assessing differences in mean for the performance metric distributions between illumination conditions for heart rate estimation. Statistically significant differences in mean are shown in bold.

HR Distributions for Comparison			*t*-Test (A vs. B)	
Metric	**Condition A**	**Condition B**	***t***	***p***
PCC	Dark	Medium	−1.901	0.036
	Dark	Bright	−5.788	**<0.001**
	Medium	Bright	−3.327	**0.002**
MAE	Dark	Medium	2.724	**0.007**
	Dark	Bright	12.044	**<0.001**
	Medium	Bright	6.403	**0.001**

In Figure 15, the residual error, the difference between the heart rate measurements from the contact sensor and the camera-based solution, is shown. Visual inspection of this graph suggests the system commonly overestimates the heart rate in medium and dark conditions. The prevalence of frames recorded with a heart rate of zero during the dark introduction part of the film can also be seen in this figure. Through further inspection of the two signals, it appeared that the camera-based measurement would also often increase and decrease by large increments over the course of only a few seconds. This characteristic *spiking* can be seen in Figure 16. In Figure 17, the relation between the heart rate values recorded with the *Polar H10* is compared to the data obtained using *PyVHR*, for each corresponding sample, using a two-dimensional histogram. This visualisation highlights both the tendency for the camera-based system to overestimate as well as the tendency to make measurements of zero BPM in the dark illumination condition. Moreover, it highlights the commonality of erroneous readings between 120 BPM and 130 BPM in bright and medium illumination conditions.

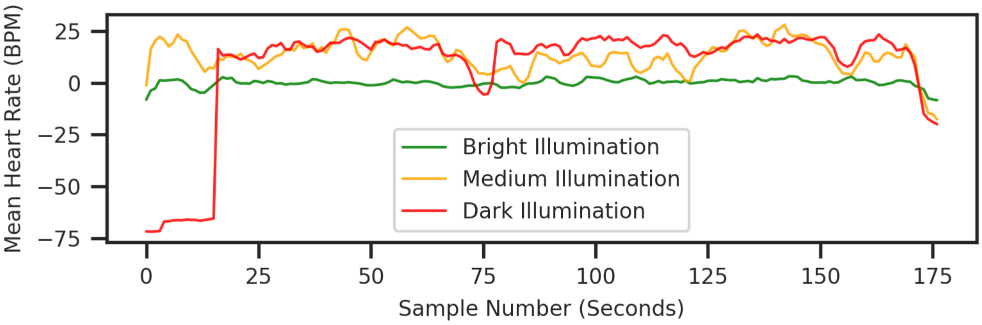

Figure 15. Mean difference between the camera-measured HR and the HR measured by the contact sensor for each sample in each illumination condition.

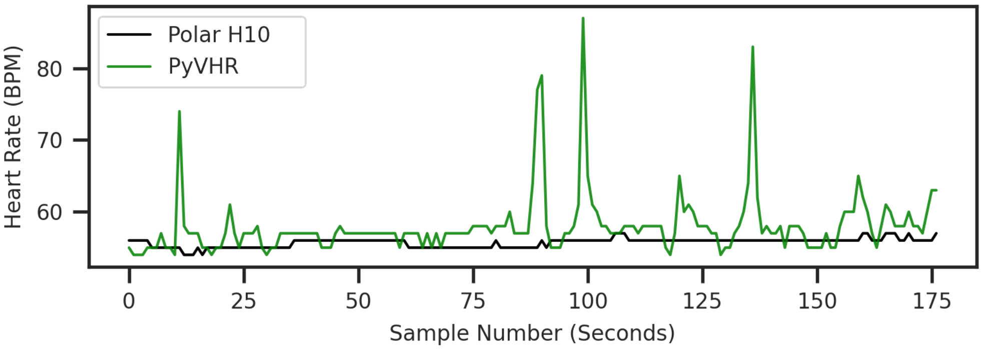

Figure 16. Example comparing the camera-measured signal and the physical heart rate sensor for a video captured in the bright illumination condition. Note the spiking behaviour in the camera-measured data.

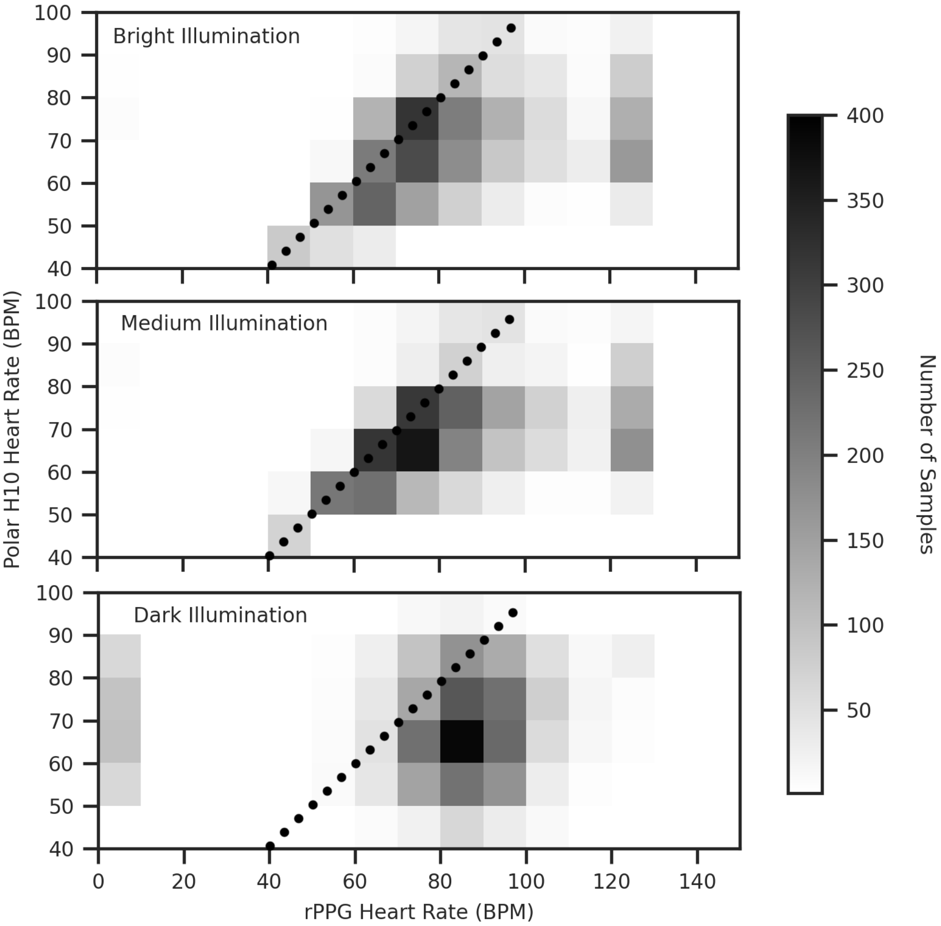

Figure 17. Two-dimensional histogram showing the relationship between the camera-measured heart rate values and the expected values as measured by the contact sensor, for each sample under each illumination condition. The dotted line describes where the two measures are equal. A bin size of ten helps characterise the relationship between the two variables clearly.

5. Discussion

In the methodology section, concerns were highlighted as to whether the relationship between the camera-measured signal (generated by *OpenFace 2*), and the expected signal (associated with perfectly replicating the model's movements) could be used to effectively assess system performance. Specifically, the concern was that the individual's improving performance in the facial action replication task, with each subsequent trial, may create a confounding effect between the illumination conditions which may lead to incorrect conclusions to be drawn from the study. Therefore, any finding of a between illuminations effect was considered to be based on an assumption that participants perform the facial actions task consistently between trials. To test this assumption a repeated measures ANOVA was used, showing there is no such difference in MAE or PCC between trials. Hence, we consider that under the conditions of this study, this assumption has been met and that the findings regarding the influence of illumination on the performance of the camera-based FAC solution can be considered conclusive.

To reduce any influence of inconsistent task performance, counterbalancing of the order of conditions should be applied and further considerations should be made to mitigate the influence of participants struggling to create certain facial actions (e.g., AU02). For example, the instructional film may be individualised to only prompt facial movements which participants can effectively recreate. Such optimisations could lead to an instructional film approach being a viable alternative to manually coded facial action datasets, for training and benchmarking camera-based FAC systems. In the case of this study, the results obtained should not be considered a comparable benchmark when compared to other datasets due to the influence of task performance on the MAE and PCC metrics. As the pacing of the instructional film was considered optimal by all but one of the participants, similar timings should be used in future research when applying this methodology.

Another limitation of this research relates to the diversity of participants, given that 13 (65%) of the sample were white males. Hence, this demographic is over-represented. However, some interindividual factors which may hypothetically influence the performance of remote camera-based physiological measurement are well represented within the datasets. For example, many of the participants were wearing glasses (55%) or had facial hair (35%). These factors likely influenced the performance of the camera-based physiological measurement solution. However, they are not explored in this publication due to scope limitations. In future work exploring the use of camera-based physiological solutions, researchers should strive to obtain a more varied sample of participants.

Analyses of the data collected suggest that the camera-based automatic FAC solution implemented performs better under brighter room illumination, as hypothesized. Although, the difference between the performance metrics for the medium and bright conditions is not considered statistically significant. Based on these results, we recommend that when using a similar camera-based FAC system, the room illumination should be at least as bright as the medium condition and should be kept consistent between participants.

In all three conditions, characteristic performance issues were found in the camera-measured FAC data: erratically changing values containing frequency content associated with facial movements faster than the fastest possible microexpression [71]; values between zero and one, which corresponds to an impossible score within Ekman's coding method [46]; values of lower intensity than the actions demonstrated by the model in the instructional film. However, it is unclear if the values between zero and one are due to some undocumented adaptation of the coding method, or a performance issue. Moreover, it is unclear if this underestimation is due to participants' task performance (high-intensity actions may be harder to reproduce) or system performance (*OpenFace 2* underestimates the magnitude of high-intensity facial actions). To avoid these issues influencing the outcomes of psychophysiological research, the recommendation is made to avoid drawing inferences based on microexpressions or low-intensity expressions, between zero and one, when using a similar camera-based FAC system. It should be noted that there is no conclusive indication that these characteristic errors were more prevalent with each incrementally darker room illumination, based on the analyses made.

The analyses of the data collected suggest that the camera-based heart rate measurement solution implemented performs better under brighter room illumination. Notably, the difference in MAE between the setting was large and highly significant and the performance of the system was very poor in all but the bright condition. Hence, when using a similar camera-based HR measurement system, we strongly suggest the recreation of illumination conditions similar to the bright setting. Or if possible, brighter and with a smaller component of light emanating from the screen.

The residual error between the camera-measured signal and the expected signal from the contact sensor can be partly characterised by distinct increases and decreases in heart rate occurring over a short time period. These upwards *spikes* appear to be part of a wider trend of higher heart rate predictions being made by the camera-based system, compared to the contact sensor. These *spikes* also contributed to the poor correlation between the camera-based heart rate measurements and the *Polar H10* measurement. Even in the bright condition, the correlation was found to be weak (mean PCC = 0.28). Hence, the suggestion is made that in psychophysiological research, features relating to short-term changes in heart rate (e.g, once every second) should not be obtained using a similar camera-based HR measurement system. Instead, the recommendation is made to obtain sparse measurements over a longer duration, reducing the temporal resolution of the signal. For example, before, during, and after the viewing of screen-based stimuli, or once every minute. To achieve this, adaptations to the stride and window length settings within the *PyVHR* configuration should be considered. This reduction in error, associated with an extended window length in depth, is explored in depth by Shin et al. [74].

Occlusion events were identified as a source of erroneous measurements for both the camera-based HR measurement and FAC systems. When a face is not recognized, which can happen due to low illumination or obstructions, no measurement can be obtained. In our study, data, it is evident that the initial illumination during the instructional film contributed to such occurrences in the dark condition (refer to Figure 15). While no occlusion events were observed during the facial actions task of the instructional film, the decision to exclude data analysis from the head movement and eye movement tasks, as mentioned in Section 3.1, was made due to the prevalence of facial occlusion events. Specifically, rotating the head to look away from the screen, in any direction, was found to cause a loss of tracking.

A DSLR camera was chosen due to its ease of procurement and convenience, as well as advantages when shooting at a distance and in dark conditions compared to common rPPG video capture devices such as specialised scientific cameras and specific model webcams (see Section 3.3). There was limited control over some camera settings, for example, the ISO setting was chosen based on visual inspection and may hypothetically have influenced the data obtained in an unexpected way. To better understand how camera settings, interpersonal factors, and environmental conditions influence the quality of physiological data encoded in a video signal, future research should explore the development of a quality assessment metric. This video quality metric should be developed by considering existing research into image and video quality assessment [10,75], as well as the research described in this manuscript pertaining to factors which influence the performance of remote physiological measurement solutions (Section 2). Provisionally, such a model would allow researchers to predict if a video signal is viable for physiological analysis, using standard tools such as *OpenFace 2* [30] and *PyVHR* [3].

Creating such a video quality metric may also provide insight into how image enhancement may be applied to improve the extraction of physiological data. Such techniques have shown promising results in some publications [76–78]. Moreover, similar intelligent systems, such as driverless cars, have benefited from image enhancement to enable their camera-based systems to perform in challenging conditions [79]. In the future, we expect that new camera-based physiological measurement software solutions will become available that utilise video enhancement, as well as other technological advancements. Provisionally, these will help improve the feasibility of using camera-based physiological measurement solutions in challenging conditions—such as for assessing cinematic viewing experiences.

6. Conclusions

In this study, the feasibility of using a multimodal camera-based physiological measurement solution for the assessment of cinematic viewing experiences was considered. A system was implemented that extracts HR and performs FAC using only open-source software and a basic DSLR camera. It was then tested under three illumination conditions, revealing that a reduction in room illumination has a significant negative effect on system performance. Hence, the recommendation is made to apply illumination similar to, or brighter and less variable than, the bright room condition in this study (192 to 270 lux incident on the viewer's face) when using similar camera-based physiological measurement solutions in the future. This recommendation is made even though it represents a sub-optimal reproduction of a cinematic viewing experience.

When analysing the data obtained using the camera-based physiological measurement solution, characteristic performance issues were found under all three illuminations tested. This led to the conclusion that researchers should consider reducing the temporal resolution of the HR measurements obtained from a camera when trying to make inferences about a person's psychological state. For FAC, characteristic performance issues were found which suggests that the system implemented should not be used to assess micro-expressions, or low-intensity facial actions when trying to make inferences about a person's psychological state. If all of these recommendations are followed, we consider the application of the camera-based multimodal physiological solution implemented in this study to be feasible for assessing screen-based media viewing experiences.

To gain this insight, a novel instructional film was used. This approach proved to be a time-effective way of gathering insightful data for evaluating the camera-based FAC system. Notably, the validation of illumination as a performance influencing factor, as described in the *OpenFace 2* publication [30], is considered to be a novel research contribution. The method saves time as only the instructional film requires manual FAC, rather than the videos of every participant in the dataset. Future research should be conducted to develop this method as a means for testing camera-based physiological measurement solutions.

No similar solution, which is capable of FAC and HR measurements from a single camera, has been designed and tested in research to date. In future, researchers should aim to test and improve the robustness of this solution. There is also potential to improve this solution for capturing physiological data related to screen-based media viewing experiences by investigating additional capabilities, such as head and eye tracking, as well as incorporating audio recording. These advancements can further enhance the comprehensiveness and accuracy of physiological measurements, thereby enriching our understanding of the psychological processes involved in media consumption.

Author Contributions: Conceptualization, J.W.; methodology, J.W., D.M. and J.F.; software, J.W.; validation, J.W.; formal analysis, J.W.; investigation, J.W.; resources, J.W., D.M. and J.F.; data curation, J.W.; writing—original draft preparation, J.W.; writing—review and editing, J.W., D.M. and J.F.; visualization, J.W.; supervision, D.M. and J.F.; project administration, D.M.; funding acquisition, D.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported in part by the UK Arts and Humanities Research Council (AHRC) XR Stories Creative Industries Cluster project, grant no. AH/S002839/1, in part by a University of York funded PhD studentship, with additional support from Bang & Olufsen, Denmark

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki. The protocol, handling of data, and all other ethical factors were considered and approved by the University of York, School of Physics, Technology & Engineering ethics committee (reference: Williams291121).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The performance metric data presented in this study are available on request from the corresponding author. Participants' physiological data are not available, as agreed with the sample when informed consent was given.

Acknowledgments: Thanks are given to IMotions for allowing the usage of their facial action code and examples as stimuli. Thanks are also given to Pablo Martínez-Nuevo and Sven Shepstone at Bang & Olufsen for their supervision and discussion regarding the topic of this paper.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

FAC	Facial Action Coding
rPPG	Remote Photoplethysmography
MAE	Mean Absolute Error
PCC	Pearson Correlation Coefficient
DSLR	Digital Single Reflex Camera
BPM	Beats Per Minute
FPS	Frames Per Second
HR	Heart Rate

References

1. Baltrusaitis, T. OpenFace 2.2.0 GitHub. Available online: https://github.com/TadasBaltrusaitis/OpenFace (accessed on 26 June 2023).
2. Phuselab. PyVHR GitHub. 2022. Available online: https://github.com/phuselab/pyVHR/blob/master/notebooks/ (accessed on 26 June 2023).
3. Boccignone, G.; Conte, D.; Cuculo, V.; D'Amelio, A.; Grossi, G.; Lanzarotti, R.; Mortara, E. pyVHR: A Python framework for remote photoplethysmography. *PeerJ Comput. Sci.* **2022**, *8*, e929. [CrossRef]
4. Yang, X.; Li, Y.; Lyu, S. Exposing deep fakes using inconsistent head poses. In Proceedings of the 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2019), Brighton, UK, 12–17 May 2019; pp. 8261–8265. [CrossRef]
5. Ofcom. Media Nations: UK 2022—Ofcom. 2022. Available online: https://www.ofcom.org.uk/ (accessed on 7 July 2023).
6. Baig, M.Z.; Kavakli, M. A survey on psycho-physiological analysis & Measurement Methods in Multimodal Systems. *Multimodal Technol. Interact.* **2019**, *3*, 37. [CrossRef]
7. Wilson, G.M. Psychophysiological indicators of the impact of media quality on users. In Proceedings of the CHI '01 Extended Abstracts on Human Factors in Computing Systems, Seattle, WA, USA, 31 March–5 April 2001; Association for Computing Machinery: New York, NY, USA, 2001; pp. 95–96. [CrossRef]
8. Bosse, S.; Brunnström, K.; Arndt, S.; Martini, M.G.; Ramzan, N.; Engelke, U. A common framework for the evaluation of psychophysiological visual quality assessment. *Qual. User Exp.* **2019**, *4*, 3 [CrossRef]
9. Min, X.; Gu, K.; Zhai, G.; Yang, X.; Zhang, W.; Le Callet, P.; Chen, C.W. Screen content quality assessment: Overview, benchmark, and beyond. *ACM Comput. Surv.* **2021**, *54*,187. [CrossRef]
10. Zhai, G.; Min, X. Perceptual image quality assessment: A survey. *Sci. China Inf. Sci.* **2020**, *63*, 211301. [CrossRef]
11. Min, X.; Zhai, G.; Zhou, J.; Farias, M.C.Q.; Bovik, A.C. Study of subjective and objective quality assessment of audio-visual signals. *IEEE Trans. Image Process.* **2020**, *29*, 6054–6068. [CrossRef]
12. Min, X.; Zhai, G.; Zhou, J.; Zhang, X.P.; Yang, X.; Guan, X. A multimodal saliency model for videos with high audio-visual correspondence. *IEEE Trans. Image Process.* **2020**, *29*, 3805–3819. [CrossRef]
13. Min, X.; Zhai, G.; Hu, C.; Gu, K. Fixation prediction through multimodal analysis. In Proceedings of the 2015 Visual Communications and Image Processing (VCIP), Singapore, 13–16 December 2015; pp. 1–4. [CrossRef]
14. Hammond, H.; Armstrong, M.; Thomas, G.A.; Gilchrist, I.D. Audience immersion: Validating attentional and physiological measures against self-report. *Cogn. Res. Princ. Implic.* **2023**, *8*, 22. [CrossRef]
15. Madsen, J.; Parra, L.C. Cognitive processing of a common stimulus synchronizes brains, hearts, and eyes. *PNAS Nexus* **2022**, *1*, pgac020. [CrossRef] [PubMed]
16. Pérez, P.; Madsen, J.; Banellis, L.; Türker, B.; Raimondo, F.; Perlbarg, V.; Valente, M.; Niérat, M.C.; Puybasset, L.; Naccache, L.; et al. Conscious processing of narrative stimuli synchronizes heart rate between individuals. *Cell Rep.* **2021**, *36*, 109692. [CrossRef]
17. Grassini, S.; Laumann, K. Questionnaire measures and physiological correlates of presence: A systematic review. *Front. Psychol.* **2020**, *11*, 349. [CrossRef]
18. Hinkle, L.B.; Roudposhti, K.K.; Metsis, V. Physiological measurement for emotion recognition in virtual reality. In Proceedings of the 2019 2nd International Conference on Data Intelligence and Security (ICDIS), South Padre Island, TX, USA, 28–30 June 2019; pp. 136–143. [CrossRef]
19. Leong, S.C.; Tang, Y.M.; Lai, C.H.; Lee, C. Facial expression and body gesture emotion recognition: A systematic review on the use of visual data in affective computing. *Comput. Sci. Rev.* **2023**, *48*, 100545. [CrossRef]
20. Dingli, A.; Giordimaina, A. Webcam-based detection of emotional states. *Vis. Comput.* **2017**, *33*, 459–469. [CrossRef]

21. Madan, C.R.; Harrison, T.; Mathewson, K.E. Noncontact measurement of emotional and physiological changes in heart rate from a webcam. *Psychophysiology* **2018**, *55*, e13005. [CrossRef]
22. Samadiani, N.; Huang, G.; Cai, B.; Luo, W.; Chi, C.H.; Xiang, Y.; He, J. A review on automatic facial expression recognition systems assisted by multimodal sensor data. *Sensors* **2019**, *19*, 1863. [CrossRef] [PubMed]
23. Dzedzickis, A.; Kaklauskas, A.; Bucinskas, V. Human emotion recognition: Review of sensors and methods. *Sensors* **2020**, *20*, 592. [CrossRef] [PubMed]
24. Egger, M.; Ley, M.; Hanke, S. Emotion recognition from physiological signal analysis: A review. *Electron. Notes Theor. Comput. Sci.* **2019**, *343*, 35–55. [CrossRef]
25. Ekman, P.; Hartmanis, E. Facial Activity Recognition as Predictor for Learner Engagement of Robot-Lead Language Cafes. Available online: https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1351886 (accessed on 18 July 2023).
26. Wu, S.; Du, Z.; Li, W.; Huang, D.; Wang, Y. Continuous emotion recognition in videos by fusing facial expression, head pose and eye gaze. In Proceedings of the 2019 International Conference on Multimodal Interaction, Suzhou, China, 14–18 October 2019; pp. 40–48. [CrossRef]
27. Porcu, S.; Floris, A.; Atzori, L. Towards the evaluation of the effects of ambient illumination and noise on quality of experience. In Proceedings of the 2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX), Berlin, Germany, 5–7 June 2019; pp. 1–6. [CrossRef]
28. ITU. ITU-R BT.500-14: Methodologies for the Subjective Assessment of the Quality of Television Images. Available online: https://www.itu.int/rec/R-REC-BT.500 (accessed on 26 June 2023).
29. ITU. ITU-R BS 775-2, Multi-Channel Stereophonic Sound System with and without Accompanying Picture. Available online: https://www.itu.int/rec/R-REC-BS.775/ (accessed on 26 June 2023).
30. Baltrusaitis, T.; Zadeh, A.; Lim, Y.C.; Morency, L.P. Openface 2.0: Facial behavior analysis toolkit. In Proceedings of the 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), Xi'an, China, 15–19 May 2018; pp. 59–66. [CrossRef]
31. De Haan, G.; Jeanne, V. Robust pulse rate from chrominance-based rPPG. *IEEE Trans. Biomed. Eng.* **2013**, *60*, 2878–2886. [CrossRef]
32. Bobbia, S.; Macwan, R.; Benezeth, Y.; Mansouri, A.; Dubois, J. Unsupervised skin tissue segmentation for remote photoplethysmography. *Pattern Recognit. Lett.* **2019**, *124*, 82–90. [CrossRef]
33. Haugg, F.; Elgendi, M.; Menon, C. GRGB rPPG: An efficient low-complexity remote photoplethysmography-based algorithm for heart rate estimation. *Bioengineering* **2023**, *10*, 243. [CrossRef]
34. Nowara, E.M.; McDuff, D.; Veeraraghavan, A. A meta-analysis of the impact of skin tone and gender on non-contact photoplethysmography measurements. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 284–285. [CrossRef]
35. Poh, M.Z.; McDuff, D.J.; Picard, R.W. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Opt. Express* **2010**, *18*, 10762–10774. [CrossRef]
36. Wang, W.; Den Brinker, A.C.; Stuijk, S.; De Haan, G. Algorithmic principles of remote PPG. *IEEE Trans. Biomed. Eng.* **2016**, *64*, 1479–1491. [CrossRef] [PubMed]
37. Chen, W.; McDuff, D. Deepphys: Video-based physiological measurement using convolutional attention networks. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 349–365.
38. Verkruysse, W.; Svaasand, L.O.; Nelson, J.S. Remote plethysmographic imaging using ambient light. *Opt. Express* **2008**, *16*, 21434–21445. [CrossRef] [PubMed]
39. Yang, Z.; Wang, H.; Lu, F. Assessment of deep learning-based heart rate estimation using remote photoplethysmography under different illuminations. *IEEE Trans. Hum.-Mach. Syst.* **2022**, *52*, 1236–1246 . [CrossRef]
40. Yin, R.N.; Jia, R.S.; Cui, Z.; Yu, J.T.; Du, Y.B.; Gao, L.; Sun, H.M. Heart rate estimation based on face video under unstable illumination. *Appl. Intell.* **2021**, *51*, 5388–5404. [CrossRef]
41. Tohma, A.; Nishikawa, M.; Hashimoto, T.; Yamazaki, Y.; Sun, G. Evaluation of remote photoplethysmography measurement conditions toward telemedicine applications. *Sensors* **2021**, *21*, 8357. [CrossRef]
42. McDuff, D.J.; Blackford, E.B.; Estepp, J.R. The impact of video compression on remote cardiac pulse measurement using imaging photoplethysmography. In Proceedings of the 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), Washington, DC, USA, 30 May–3 June 2017; pp. 63–70. [CrossRef]
43. Cerina, L.; Iozzia, L.; Mainardi, L. Influence of acquisition frame-rate and video compression techniques on pulse-rate variability estimation from vPPG signal. *Biomed. Eng./Biomed. Tech.* **2019**, *64*, 53–65. [CrossRef]
44. Blackford, E.B.; Estepp, J.R. Effects of frame rate and image resolution on pulse rate measured using multiple camera imaging photoplethysmography. In Proceedings of the Medical Imaging 2015: Biomedical Applications in Molecular, Structural, and Functional Imaging, Orlando, FL, USA, 24–26 February 2015; SPIE: Bellingham, WA, USA, 2015; Volume 9417, pp. 639–652. [CrossRef]
45. Wang, G. Influence of ROI selection for remote photoplethysmography with singular spectrum analysis. In Proceedings of the 2021 IEEE International Conference on Artificial Intelligence and Industrial Design (AIID), Guangzhou, China, 28–30 May 2021; pp. 416–420. [CrossRef]
46. Ekman, P.; Friesen, W.V. *Facial Action Coding System Volumes 1–2*; Consulting Psychologists Press: Washington, DC, USA, 1978.

47. Zhang, K.; Zhang, Z.; Li, Z.; Qiao, Y. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Process. Lett.* **2016**, *23*, 1499–1503. [CrossRef]
48. Zadeh, A.; Chong Lim, Y.; Baltrusaitis, T.; Morency, L.P. Convolutional experts constrained local model for 3d facial landmark detection. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Venice, Italy, 22–29 October 2017; pp. 2519–2528.
49. Baltrušaitis, T.; Mahmoud, M.; Robinson, P. Cross-dataset learning and person-specific normalisation for automatic action unit detection. In Proceedings of the 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Ljubljana, Slovenia, 4–8 May 2015; Volume 6, pp. 1–6. [CrossRef]
50. Mavadati, S.M.; Mahoor, M.H.; Bartlett, K.; Trinh, P.; Cohn, J.F. Disfa: A spontaneous facial action intensity database. *IEEE Trans. Affect. Comput.* **2013**, *4*, 151–160. [CrossRef]
51. McKeown, G.; Valstar, M.F.; Cowie, R.; Pantic, M. The SEMAINE corpus of emotionally coloured character interactions. In Proceedings of the 2010 IEEE International Conference on Multimedia and Expo, Singapore, 19–23 July 2010; pp. 1079–1084. [CrossRef]
52. Zhang, X.; Yin, L.; Cohn, J.F.; Canavan, S.; Reale, M.; Horowitz, A.; Liu, P.; Girard, J.M. Bp4d-spontaneous: A high-resolution spontaneous 3d dynamic facial expression database. *Image Vis. Comput.* **2014**, *32*, 692–706. [CrossRef]
53. Lucey, P.; Cohn, J.F.; Prkachin, K.M.; Solomon, P.E.; Matthews, I. Painful data: The UNBC-McMaster shoulder pain expression archive database. In Proceedings of the 2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG), Santa Barbara, CA, USA, 21–25 March 2011; pp. 57–64. [CrossRef]
54. Savran, A.; Alyüz, N.; Dibeklioğlu, H.; Çeliktutan, O.; Gökberk, B.; Sankur, B.; Akarun, L. Bosphorus database for 3D face analysis. In Proceedings of the Biometrics and Identity Management: First European Workshop (BIOID 2008), Roskilde, Denmark, 7–9 May 2008; Springer: Berlin/Heidelberg, Germany, 2008; pp. 47–56. [CrossRef]
55. Valstar, M.F.; Jiang, B.; Mehu, M.; Pantic, M.; Scherer, K. The first facial expression recognition and analysis challenge. In Proceedings of the 2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG), Santa Barbara, CA, USA, 21–25 March 2011; pp. 921–926. [CrossRef]
56. Polar H10. Available online: https://www.polar.com/uk-en/sensors/h10-heart-rate-sensor (accessed on 26 June 2023).
57. Cosker, D.; Krumhuber, E.; Hilton, A. A FACS valid 3D dynamic action unit database with applications to 3D dynamic morphable facial modeling. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 2296–2303. [CrossRef]
58. Gosselin, P.; Perron, M.; Beaupré, M. The voluntary control of facial action units in adults. *Emotion* **2010**, *10*, 266. [CrossRef] [PubMed]
59. iMotions; Martin, M.; Farnsworth, B. Homepage. Available online: https://imotions.com/ (accessed on 26 June 2023).
60. GDPR: Recital 39. Available online: https://gdpr-info.eu/recitals/no-39/ (accessed on 26 June 2023).
61. Monk, E. Monk Scale. Available online: https://skintone.google/the-scale (accessed on 26 June 2023).
62. Softonic. Losslesscut GitHub. Available online: https://github.com/mifi/lossless-cut (accessed on 26 June 2023).
63. Bellard, F. FFmpeg. Available online: https://ffmpeg.org/ (accessed on 26 June 2023).
64. Harvey, P. Exiftool. Available online: https://exiftool.org/ (accessed on 26 June 2023).
65. scipy.signal.correlation Lags Function SciPy v1.10.1. Available online: https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.correlation_lags.html (accessed on 18 July 2023).
66. Shapiro, S.S.; Wilk, M.B. An analysis of variance test for normality (complete samples). *Biometrika* **1965**, *52*, 591–611. [CrossRef]
67. Mauchly, J.W. Significance test for sphericity of a normal n-variate distribution. *Ann. Math. Stat.* **1940**, *11*, 204–209. [CrossRef]
68. Girden, E.R. *ANOVA: Repeated Measures*; Number 84; Sage: Newcastle upon Tyne, UK, 1992.
69. Bonferroni, C. Teoria statistica delle classi e calcolo delle probabilita. *Pubbl. R Ist. Super. Sci. Econ. Commericiali Firenze* **1936**, *8*, 3–62.
70. Armstrong, R.A. When to use the Bonferroni correction. *Ophthalmic Physiol. Opt.* **2014**, *34*, 502–508. [CrossRef]
71. Yan, W.J.; Wu, Q.; Liang, J.; Chen, Y.H.; Fu, X. How fast are the leaked facial expressions: The duration of micro-expressions. *J. Nonverbal Behav.* **2013**, *37*, 217–230. [CrossRef]
72. Matplotlib 3.7.1: Power Spectral Density. Available online: https://matplotlib.org/ (accessed on 18 July 2023).
73. Greenhouse, S.W.; Geisser, S. On methods in the analysis of profile data. *Psychometrika* **1959**, *24*, 95–112. [CrossRef]
74. Shin, Y.J.; Han, W.J.; Suh, K.H.; Lee, E.C. Effect of time window size for converting frequency domain in real-time remote photoplethysmography extraction. In Proceedings of the International Conference on Intelligent Human Computer Interaction, Kent, OH, USA, 20–22 Decembe 2021; Springer: Berlin/Heidelberg, Germany, 2021; pp. 145–149. [CrossRef]
75. Chen, Y.; Wu, K.; Zhang, Q. From QoS to QoE: A tutorial on video quality assessment. *IEEE Commun. Surv. Tutor.* **2014**, *17*, 1126–1165. [CrossRef]
76. Nowara, E.M.; McDuff, D.; Veeraraghavan, A. The benefit of distraction: Denoising camera-based physiological measurements using inverse attention. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 4955–4964.
77. Kim, J.H.; Jeong, J.W. Gaze in the dark: Gaze estimation in a low-light environment with generative adversarial networks. *Sensors* **2020**, *20*, 4935. [CrossRef]

78. Pebiana, S.; Widyanto, M.R.; Basaruddin, T.; Liliana, D.Y. Enhancing facial component analysis. In Proceedings of the 2nd International Conference on Software Engineering and Information Management (ICSIM '19), Bali, Indonesia, 10–13 January 2019; Association for Computing Machinery: New York, NY, USA, 2019; pp. 175–179. [CrossRef]
79. Min, X.; Zhai, G.; Gu, K.; Yang, X.; Guan, X. Objective quality evaluation of dehazed images. *IEEE Trans. Intell. Transp. Syst.* **2019**, *20*, 2879–2892. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.