
Article:
Williams, Joseph, Francombe, Jon and Murphy, Damian Thomas orcid.org/0000-0002-6676-9459 (2024) Camera Sourced Heart Rate Synchronicity:A Measure of Immersion in Audiovisual Experiences. Applied Sciences. 7228. ISSN: 2076-3417

https://doi.org/10.3390/app14167228

---

applied sciences

MDPI

Article

Camera-Sourced Heart Rate Synchronicity: A Measure of Immersion in Audiovisual Experiences

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

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

Abstract: Audio presentation is often attributed as being capable of influencing a viewer's feeling of immersion during an audiovisual experience. However, there is limited empirical research supporting this claim. This study aimed to explore this effect by presenting a clip renowned for its immersive soundtrack to two groups of participants with either high-end or basic audio presentation. To measure immersion, a novel method is applied, which utilises a camera instead of an electroencephalogram (ECG) for acquiring a heart rate synchronisation feature. The results of the study showed no difference in the feature, or in the responses to an established immersion questionnaire, between the two groups of participants. However, the camera-sourced HR synchronicity feature correlated with the results of the immersion questionnaire. Moreover, the camera-sourced HR synchronicity feature was found to correlate with an equivalent feature sourced from synchronously recorded ECG data. Hence, this shows the viability of using a camera instead of an ECG sensor to quantify heart rate synchronisation but suggests that audio presentation alone is not capable of eliciting a measurable difference in the feeling of immersion in this context.

Keywords: audio presentation; multichannel audio; biosignals; psychophysiological methods; remote photoplethysmography

Citation: Williams, J.; Francombe, J.; Murphy, D. Camera-Sourced Heart Rate Synchronicity: A Measure of Immersion in Audiovisual Experiences. *Appl. Sci.* **2024**, *14*, 7228. https://doi.org/10.3390/app14167228

Academic Editor: Douglas O'Shaughnessy

Received: 30 April 2024
Revised: 5 August 2024
Accepted: 13 August 2024
Published: 16 August 2024

1. Introduction

This paper sets out to apply and evaluate a novel means of measuring immersion using a camera within a study that aims to better understand the true influence of audio presentation methods on audiovisual experiences. The term *immersion* is applied here as a means of labelling the extent to which a participant feels they are placing their full attention within a mediated world. The term *audio presentation* refers to a holistic concept that includes the soundtrack, all aspects of the system utilised for its reproduction, and the acoustics of the room where it is presented.

The opening scene of the film *Gravity (2013)* [1] is consistently theorised as evoking a feeling of immersion when experienced with a faithful reproduction of the Dolby Atmos soundtrack [2–6]. In the original experimental work presented in this paper, we conduct a between-subjects study in which this scene is experienced with either high-end audio presentation or basic audio presentation. In the high-end condition, the moving image is accompanied by a lossless *Dolby Atmos* [7] version of the soundtrack, reproduced using 12 *Genelec* loudspeakers. In the basic audio condition, the same moving image is accompanied by a *Dolby Digital 2.0* (320kbps) stereo version of the soundtrack, reproduced using the built-in television (TV) loudspeakers [8].

To measure immersion in this study, we apply a novel camera-sourced physiological feature alongside an established self-report method. The camera-sourced feature captures the extent to which a participant's heart rate (HR) rises and falls synchronously with the rest of the cohort in response to the evolving narrative. This feature, known as HR

inter-subject correlation (ISC-HR), has been shown to index immersion [9,10], attentional engagement [11,12], and flow [13] when acquired using an electroencephalogram (ECG). However, no research has been conducted that attempts to obtain this feature using a camera. To properly evaluate the camera-sourced feature, an ECG is also used in this study. We propose that the camera may be deployed without an ECG in future research. The camera-based method is considered advantageous as it avoids the discomfort and obtrusiveness of physical sensors. Moreover, this method is promising due to its potential for simultaneously capturing other experientially indicative features, such as facial movements and gaze direction.

The overarching hypothesis for this study is that the increase in immersion influenced by a high-end audio presentation, compared with a basic audio presentation, can be measured using a camera-sourced ISC-HR feature. Retrospectively, the primary contribution of this research is the documentation of the camera-based method. This method may be used to inform research across various domains concerned with the assessment of the human psychological condition at a distance, including telemedicine and affective computing. While a detailed discussion of the broader applications and ethical considerations of using cameras for physiological data extraction is beyond the scope of this manuscript, this lab-based psychophysiological investigation, conducted with explicit consent, demonstrates the potential of non-contact methods for evaluating audiovisual experiences.

2. Background

2.1. Audio Technologies

There are various technological factors that influence whether an audio presentation may be considered high-end or basic. Many of these, such as audio compression bitrate, loudspeaker placement, and frequency response, are numerically described in Section 3.2. However, the difference in audio formats is more complex to characterise.

Dolby Atmos is an audio format that premiered with the release of *Brave (2012)* [14]. Unlike traditional channel-based audio systems, *Dolby Atmos* utilises an object-based approach. In channel-based systems, audio assets are confined to a fixed number of output channels, with each being delivered to a specific loudspeaker or set of loudspeakers in a predefined configuration. In contrast, *Dolby Atmos* encodes each sound asset as an audio object, complete with metadata that describes its position within a three-dimensional digital environment. During playback, object-based systems decode the audio objects to the available loudspeakers based on their locations in the configuration.

In a domestic context, *Dolby* has created guidelines for loudspeaker locations associated with thirty-four different configurations, with a minimum of three and a maximum of twenty loudspeakers [15]. The 7.1.4 loudspeaker configuration used in this study is composed of seven loudspeakers placed around the listener, one subwoofer, and four loudspeakers placed above the listener.

2.2. Immersion and Audiovisual Media

The term *immersion* is often inconsistently defined and applied in academia and other fields [16]. A filter model, as first proposed in [17], provides a useful reference for exploring these varying interpretations [18,19]. Three domains are conceptualised in this model: physical, sensory, and affective. Audiovisual experiences start in the physical domain, with physical characteristics such as sound pressure level or the number of loudspeakers being measurable objectively. Next, these physical changes pass through a sensory filter into the perceptual domain, where they can be described as perceptual attributes that we sense, such as loudness or envelopment. Finally, these perceptual attributes pass through a cognitive filter into the affective domain, where they can be described as global measures that we feel, such as fear or enjoyment.

Immersion is sometimes considered to be a perceptual attribute or physical characteristic. However, in this paper, a more common interpretation amongst the audio engineering community is applied: immersion is a *global measure* [18–20]. To describe this global mea-

sure in this study, we exercise the following definition: "placing one's full attention within the film world" [16]. Some sense of being perceptually located in the film world, rather than the real world, and a sense of focused attention on the narrative are both referenced in this definition. This definition of immersion was derived through an analysis of open-text response data collected from a sample of the public in 2022 [16]; hence, the outcomes of a single-question immersion assessment should reflect this interpretation. Moreover, the definition reflects Murray's interpretation [21], which is cited and applied in the development of the Film IEQ [22]. Notably, both the single-question approach and Film IEQ are considered to be valid approaches, which have been widely used for assessing immersion in audiovisual experiences, including in combination [9,23–26].

Immersion is also used synonymously with many other terms, such as presence and narrative engagement. This semantic issue is an ongoing point of discussion within various research domains, including audio research [16,20,27]. This cluster of related attributes has been applied in various publications investigating the influence of audio presentation on non-participatory audiovisual experiences, such as film viewing.

Lessiter et al. [28] found that presence increased with the addition of bass but not with an increase in channel count. Zielinski et al. [29,30] demonstrated that downmixing and filtering can significantly decrease basic audio quality. Lipscomb et al. [31] found no significant difference between stereo and a 5.1 presentation mode for music, but there was a tentative increase in *immersiveness* for film clips. Oramus et al. [32] showed that participants did not perceive Dolby Atmos as significantly more immersive, dynamic, or having better localisation or audio quality. Agrawal et al. [24] found no significant difference in immersion between groups with different channel counts, but there was a significant difference in the sense of envelopment for some stimuli with each incremental increase in channel count. Williams et al.'s [19] results suggest that sounds emanating from the rear surround loudspeakers in a 7.1 soundtrack can be distracting. Langiulli et al. [33] found that an increase in channel count correlated with higher scores on questions related to realism, emotional involvement, and physical immersion, and some indication of neural correlates were associated with differences in audio presentation.

Overall, these empirical studies highlight the complex relationship between audio presentation and non-participatory audiovisual experience as well as the lack of consistent methods for testing. Nevertheless, there seems to be some agreement that audio presentation can in some cases positively affect participants' experiences based on studies where they are exposed to more than one mode of audio presentation [24,28–30,33]. However, the ability to directly compare different audio presentations and the use of very short stimulus material negatively impact the ecological validity of the research outcomes when assessing global measures (e.g., immersion and presence), as these audiovisual experiences do not reflect real-world scenarios [34]. Using longer commercially available extracts and not allowing participants to make a side-by-side comparison of different audio technologies for the same stimuli has led to inconclusive results [19,24].

Reviewing research investigating the influence of audio presentation on interactive audiovisual experiences, Hedge et al. similarly concluded that there is a lack of consensus in terms of both methodology and specific "ways in which audio impacts immersive experience" [35]. However, again there are signs that audio presentations can influence the feeling of immersion from a less ecologically valid experiment that uses novel media and within-subject designs [36].

The lack of irrefutable evidence supporting the hypothesis that audio presentation can change naturalistic audiovisual experiences, in terms of a different feeling of immersion and other related global measures, may indicate that there is simply no significant effect. However, research has shown that a more dramatic change in presentation, such as specifically viewing a film in the cinema compared with at home, can create a stronger feeling of immersion [37]. Moreover, research has shown that a between-groups method can be applied to show the influence of visual presentation, specifically screen size, on psychophysiological reactions to audiovisual media [38].

2.3. Psychophysiological Methods and Audiovisual Media

Psychophysiological methods explore the interrelation between psychological phenomena and physiological processes. A variety of these methods, which often include both self-report questionnaires and physiological monitoring, have been applied for the assessment of audiovisual experiences. Common physiological features applied in this domain include heart rate variability (HRV), which is considered to be an indicator of stress and emotional response [39].

In this research, we concentrate on ISC-HR analyses. The principle underlying ISC analyses is to assess the degree of synchrony in physiological responses among participants. Hence, these measures can only be used when subjects' experiences are time-locked, such as non-participatory film and TV viewing. In some cases, ISC features are calculated by comparing the physiological response of one participant with the average response of all participants. However, more commonly, ISC features are calculated by taking the average pairwise correlation of one participant with all other participants [40]. This second approach is considered advantageous as it allows for commonality amongst clusters of participants to be represented in the ISC feature rather than the ISC feature only representing the similarity of a participant's response with the overall trend [41]. The ISC is commonly calculated for each participant i within a set of length N as the mean of all pairwise Pearson correlation coefficients (PCCs) [42]. Hence, in the equation below, variable j is used to index all other participants in the set, excluding participant i.

$$\mathrm{ISC}_i = \frac{1}{N-1} \sum_{j \neq i}^{N} \mathrm{PCC}(HR_i, HR_j) \quad (1)$$

Research by Madsen et al. showed that a viewer's level of attentional engagement, defined as the combination of attracting the attention of the eye and the cognitive processing of the target of this attention, correlates with ISC features associated with electroencephalography (ECG), gaze position, pupil size, and heart rate (ISC-HR) [11]. Notably, in this study, each participant viewed the media individually within this study, showing that these physiological modalities synchronise even when participants are not co-present.

It has been found that heart rate synchronises during media experiences in accordance with an individual's level of immersion. This statement is supported by experimental data showing how ISC-HR extracted from ECG data correlates with existing immersion measures, including the Film IEQ and a single-item approach [9]. Assessing both audiovisual and audio-only stimulus across a set of experiments, Perez et al. showed that ISC-HR is reduced when subjects are distracted from the narrative, and a higher ISC-HR predicts a better recall of the narrative [12].

Perez also revealed that slow changes in heart rate were most indicative of whether viewers were exposed to a distraction or not. Specifically, frequency analyses suggest that this difference is greatest when ISC is calculated using an ECG-sourced HR signal bandpass filtered between 0.08 Hz and 0.153 Hz [12]. This finding is also supported by Madsen et al. [11], although their data seem to suggest that even lower frequencies may also be effective indicators of distraction. Gibbs et al. [13] applied this consideration to the processing of their heart rate data when assessing the relationship between ISC-HR and flow during musical performance, low-pass filtering their interbeat interval signal with a cutoff of 0.05 Hz.

The frequency of synchronisation found to be indicative of the immersion-related measures documented in these studies may have been affected by the pacing and duration of the stimulus used [41]. Madsen et al. used 5 to 10 min educational videos [11], Perez used 1 to 11 min audio and audiovisual stories [12], and Gibbs used performances of length 5 to 13 min [13].

Commonly used statistical tests, such as the t-test [43] and Pearson correlation [42], assume that observations are independent. However, ISC-HR values violate this assumption because the magnitude of each value depends not only on the participant's HR but also

on the HR of all other participants in the dataset. Hence, forms of permutation testing are often applied, albeit not ubiquitously [9–12,41]. The statistical methods used in this study can be found in Section 3.5.

The use of HR data captured using a camera method for ISC-HR analyses has not been previously explored. However, there is a larger body of research that has shown the feasibility of using a camera to evaluate stress in a variety of application conditions, solely based on cardiological signals extracted from facial videos [44–46]. In the context of ISC-HR measurement, downsampling and low-pass filtering are common parts of the HR extraction process [9–13]. A similar *moving average* HR signal can be sourced from a camera [47]. For information regarding the method used to obtain HR data using a camera, see Section 3.3.

3. Methodology

3.1. Ethical Considerations

The video footage captured in this study could not be anonymised as participants' faces had to be visible in the recordings. Hence, with the sensitivity of these data in mind, a storage period of 120 days was defined. Subsequently, these output data were maintained with no way of associating the identity of the participant with their data. Moreover, throughout the research project, only the first author of this paper was permitted to view any video footage.

There was no financial reward for participating in this study. 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 that 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: Williams20230406).

3.2. Audiovisual Media

This study posits that audio presentation influences a viewer's feeling of immersion. Hence, the audiovisual stimuli and playback systems were chosen to accentuate this effect whilst still delivering an ecologically valid result. According to the filmmakers, *Gravity* is designed for *Dolby Atmos* [6]. Moreover, it has been described how the *Dolby Atmos* version of the soundtrack is an integral part of the film, which creates a strong feeling of immersion [2–5]. All of these sources mention the opening scene, which is considered exemplary in its use of spatial sound techniques. Hence, a reproduction of the Dolby Atmos version of this opening 16:35 min clip using an industry-grade loudspeaker set-up was chosen as the stimuli for those in the high-end group. This condition was designed to be the most accurate possible reproduction, and hence the delivery of the filmmaker's intent, that we could facilitate. For the basic presentation, we chose to deliver the audio through the built-in TV speakers. This condition was considered to be representative of a standard baseline presentation where there has been no additional upgrade to the audio playback setup.

3.3. Audiovisual Setup

The basic audio presentation condition was implemented using the built-in loudspeakers of the *Mitsubishi LDT422V* TV [48] (purchased in the UK) . This 42-inch monitor was selected to fit between the existing loudspeakers configured in the listening room. In the high-end presentation: for the centre channel, a *Genelec 8010A* [49] model was used; for the height channels, *Genelec 8030Cs* [50] were used; and for the subwoofer, a *Genelec 7040A* [51] was used. All of the remaining loudspeakers in the configuration were *Genelec 8040As* [52]. All loudspeakers were purchased in the UK. The location of the loudspeakers in the 7.1.4 system matched the *Dolby* guidelines [15]. As the distance from the listening position to the height channels (1545 mm) was slightly more than the distance to the other loudspeakers

in the array (1250 mm), a delay compensation of 0.86 ms was applied for phase alignment. Figure 1 shows a diagram of the audiovisual setup.

Each loudspeaker was calibrated at a loudness of 80 dB SPL (A-weighted) using *Room EQ Wizard (V5.30.9)* [53], an *Earthworks M30* (purchased in the UK)microphone [54], and the built-in calibration switches to achieve a flat frequency response. An 80 dB SPL was chosen as the calibration loudness, as this is a reasonable and realistic listening level. The subwoofer crossover was set to 85 Hz, meaning frequencies below this threshold were produced by a dedicated bass speaker. The frequency response of each channel, excluding the low-frequency effect channel, was measured about a flat target response of 80 dB SPL from 20 Hz and 20 kHz. For all of these channels, the MAE between the measured and target response was less than 1 dB SPL. The media was decoded from the special edition version of the *Gravity* Blu-ray, which contains both stereo sound encoded with *Dolby Digital* at 320 kbps and *Dolby Atmos* sound encoded with *Dolby TrueHD* (48 kHz, 24 Bit) as versions of the soundtrack [8]. To enable this decoding, the *Dolby Reference Player* was used [55].

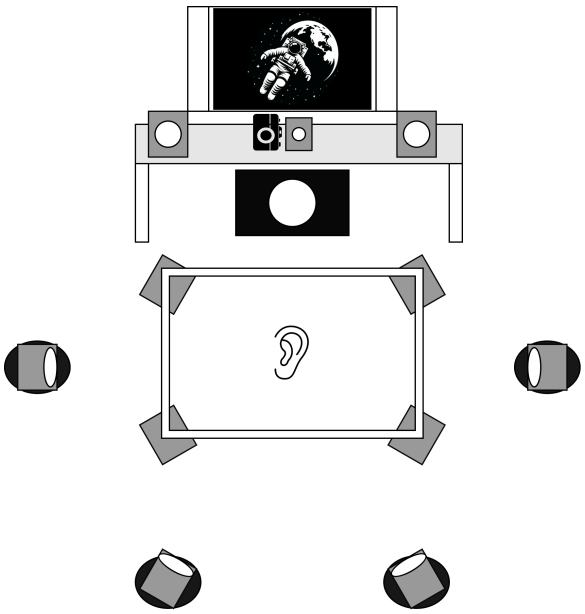

Figure 1. Diagram of the listening room where the study took place. The subwoofer was placed under the desk, seven loudspeakers were placed around the listener 1.3 m above the ground, and four loudspeakers were placed above the listener, mounted on a frame attached to the ceiling. The angles subtended between the loudspeakers and the listener aligned with the *Dolby Atmos 7.1.4* specifications [15]. The camera was mounted at desk height. The TV loudspeakers were on either side of the screen.

The room in which the system was set up is an acoustically treated studio control room. The background noise level was measured at 30.6 LAEq (15 min) using an NTI Audio XL2 analyser [56]. Participants were seated in the system's acoustically optimal area, commonly known as the *sweet spot*, with the 42-inch television screen being placed 1.61 m away. This distance was chosen based on guidance from the International Telecommunication Union (ITU) for pairing moving images with multichannel audio [57].

Measurements of the stimuli loudness were performed in both conditions, initially objectively using a *Tenma 72-860A* [58] loudness meter (purchased in the UK) and then subjectively using two expert listeners. Using the objective approach, loudness was initially set to peak at 85 dB SPL (A-weighted). Subsequently, an expert listener was tasked with switching between the two conditions and adjusting them so they were both at a comfortable listening level and also had no perceivable difference in loudness. After the loudness had been matched perceptually, a binaural recording was made using a *Neumann KU 100* [59] dummy head microphone (purchased in the UK) to allow for an objective

comparison of the frequency content. The frequency analyses of these two recordings can be seen in Figure 2.

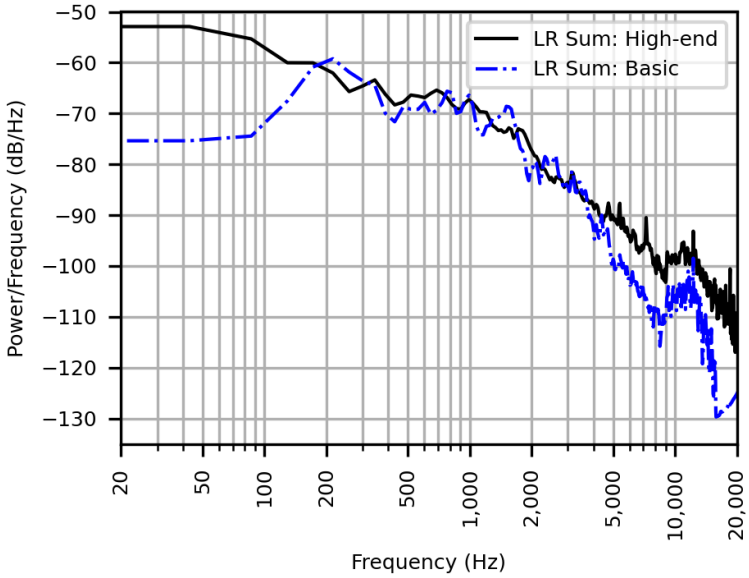

Figure 2. Power spectral density of the full 16:35 non-normalised binaural audio recordings of the audiovisual stimuli in each condition. This analysis shows the difference in the frequency spectrum influenced by the presentation mode, notably at frequencies below 200 Hz and above 15 KHz.

3.4. Recording Heart Rate

A *Canon 700D* camera [60] (purchased in the UK) was mounted in portrait orientation underneath the TV screen. This is not a state-of-the-art or specifically chosen model; it was released in 2013, and it may be considered an entry-level digital lens single-reflex camera based on its technical specifications and retail price. Previous research indicates that higher levels of video compression can negatively impact the performance of heart rate extraction from video data [61]. Hence, the camera was set to the mode with the lowest compression ratio for the camera. This led to the selection of a 1280 by 720 pixel resolution with progressive scanning (also known as 720p) and a frame rate of 50 FPS. As lower resolution negatively influences the accuracy of remote heart rate measurements [62], the camera's optical zoom was used to ensure that the participant's face took up as much of the frame as possible. Aperture and shutter speed were then set to create a bright and blur-free image. The aperture parameter was set to /5.6F. The shutter speed parameter was set at the maximum value possible for shooting at 50 FPS, 1/60s. Finally, the sensitivity (ISO) setting was raised to 1600 based on visual inspection to further brighten the video footage.

LosslessCut (V3.54.0) [63] was used to cut the videos of the participants to 30 s before and after the film clip. Subsequently, start and end timestamps were calculated by obtaining the time the video file was created using *Exiftool (V12.60)* [64] and adding the cut times from *LosslessCut* [63]. The electrocardiogram (ECG) data were captured by a *Polar H10* [65] chest strap at a sample rate of 130 Hz and transmitted to the *Polar Sensor Logger App* [66], along with a time stamp for each sample. This recording was then truncated using the aforementioned video start and end timestamps. Hence, the ECG data and video data were synchronised using the recording device clocks. The synchrony of the clocks was assessed before each session to ensure they aligned to the nearest second.

PyVHR (V2.0) [47] was used to estimate the heart rate during each video. The code used to process the video footage can be found in the code repository published alongside this manuscript [67]. The software operates as follows:

- Skin patches are extracted from the facial video footage.

- Pre-filtering is applied to each pixel signal within the patches.
- Each patch is processed to obtain a blood volume pulse (BVP) signal.
- Post-filtering is applied to each of the BVP signals.
- A rectangular windowing function is applied to each BVP signal.
- For each window of the BVP signals, the peak of the power spectral density is obtained. Each of these values corresponds to the estimated HR for the corresponding patch and window.
- The median of the HR values, across all patches, is obtained for each window.

To allow the videos to be processed using *PyVHR* [47], they had to be rotated so that the face was the right way up, as the camera had been mounted on its side. This rotation was implemented through a change in metadata, using *ExifTool* [64]; hence, there was no loss of information within the video file. Each video was processed with the default filter parameters, and the median heart rate was obtained. To improve the accuracy of the camera-sourced HR data, this was then used to inform a second pass, where the videos were reprocessed with a low cutoff frequency set corresponding to 25 BPM below this value.

To configure *PyVHR* [47], a decision had to be made regarding an appropriate stride and window length for the HR data. It is suggested that the insightful frequency range within ISC features is dependent on the stimuli [41]. As our stimulus has a slow-moving narrative spread across almost 17 min, a window length of 60 s and a stride of 1 s is deemed viable. This balances the necessary window length for accurate camera method heart rate measurements with the consideration of appropriate frequency content for assessing ISC. Notably, Wiliams et al. suggest that an 8 s window is not long enough to capture heart rate data of appropriate accuracy for psychophysiological research [68]. McDuff et al. showed a low mean absolute error (MAE) of 2.68 BPM when implementing a window length of 60 s, with similar compression to the camera used in this study [61].

With the *Polar H10* [65] ECG data, the instantaneous heart rate is obtained from the time between each heart beat (RR intervals). This series of heart rate measurements is then processed with the same window length and stride as the camera-measured data. Notably, each 1 s output sample is calculated using the combined past and subsequent 30 s of input data. This produces an equivalent ground-truth ECG HR signal and subsequent ECG ISC-HR feature for comparison with the camera-sourced data. The code and visualisations of this process can be found in the project *GitHub* repository [67].

Bright and consistent room illumination is crucial for accurate video-based heart rate measurement [68]; therefore, lighting conditions were standardised across all trials. Based on previous research and the lighting control system available, lighting similar to a well-lit domestic setting was recreated [68]. Brightest and darkest measurements were recorded at the listening position corresponding to the screen when turned off and when displaying a white screen. Overhead lighting, incident upon the viewer's face, was measured at 357 and 373 lux. Frontal lighting was measured at 121 and 163 lux, incident on the viewer's face. Notably, research by Rooney et al. also suggests that darker room illumination influences a greater feeling of media engagement and appreciation [69]. Hence, this further motivates the necessity for maintaining these constant lighting conditions across all trials.

3.5. Permutation Testing Implementation

To test the significance of any observed difference in the average ISC-HR value between groups, a permutation test was implemented [70]. The test works by initially calculating the observed difference in the median and then randomising the group labels and comparing the difference in the median for this new permutated version of the dataset. This process of randomisation and comparison is repeated 100,000 times to generate a consistently reproducible likelihood value, which serves as an alternative to the p-value obtained from a Mann–Whitney U test. The median is used rather than assessing the difference in terms of mean, as it is a way of generating an average that is more robust to outliers.

To calculate the significance of any observed relationship between ISC-HR values and any continuous feature measured using the survey, a similar permutation correlation test was implemented [70]. The test works by initially calculating the observed Spearman's rank correlation coefficient (SCC) [71] between ISC-HR and the chosen feature and then randomising the order of the ISC-HR values before comparing the difference SCC for this new permutated version of the dataset. This process of randomisation and comparison is repeated 100,000 times to generate a consistently reproducible likelihood value, which serves as an alternative to the p-value obtained from a single Spearman correlation test. SCC is used instead of PCC as it is a more robust way of assessing correlation because it is less sensitive to outliers.

3.6. Protocol

One participant at a time was invited into the listening room to watch the film clip. Participants were told to scan a QR code on their mobile device, which directed them to a *Qualtrics* form, which delivered all of the instructions and survey questions.

- An information sheet, including a description of the aims of the study and how their data would be handled.
- Confirmation of consent to participate in the study.
- A series of demographic questions, including investigators' assessment of skin tone using the Monk scale [72].
- A diagrammatic and written description of how to self-fit the moistened *Polar H10* ECG sensor, in alignment with the best practices described in the user manual [65]. The investigator left the room whilst the participant fitted the sensor and, upon returning, monitored the ECG signal for quality assurance purposes.
- Information regarding the film clip they were about to watch.
- Two questions regarding their familiarity with and initial impression of the film.
- A prompt for the investigator to start the film clip.
- A series of questions used to assess their level of immersion, composed of the 24 Film IEQ items [22] plus one single-item immersion measure ("While watching the film I felt completely immersed") [25]. Each item is scored on an integer precision continuous 0–100 scale.
- An opportunity to write a comment regarding both their audiovisual experience and their overall impressions of the study.
- A series of questions regarding how the camera, lighting, and ECG sensor, influenced their audiovisual experience.
- A chance to write any final comments.

At no point in the study were participants made aware of the two conditions being assessed. In the information sheet, they were told that the aim was purely to test a method for assessing audiovisual experience. The earlier questions regarding the anticipation were included as possible covariates for analyses. The later questions helped to assess how comfortable and naturalistic viewers felt their experience was. Moreover, alongside the open response comments, they help to provide a richer dataset for understanding each participant's experience. The demographic questions were selected to help profile our sample as well as to provide information that may be indicative of the performance of the camera-based heart rate measurement. The protocol was designed to allow participants' heart rates to stabilise at a baseline level before viewing the film clip. This consideration was especially important as many participants walked or cycled to the building in which the study took place. Moreover, to create privacy and therefore promote honesty, the investigator stood around 3 m to the side of the participants whilst they filled out the survey.

3.7. Participants

Emails and invitational flyers were used to recruit participants. Exclusion criteria were established to disqualify participants who were under the age of 18 years or who had known cardiovascular conditions. These were included due to concerns regarding

the ability to provide consent and capture unrepresentative heart rate data, respectively. Additionally, individuals with known auditory or uncorrected visual impairments were excluded from recruitment due to concerns that these factors could affect the audiovisual experience and potentially produce outlying idiosyncratic responses to the stimuli. Finally, peers who had knowledge of this research project were also considered ineligible. In total, 37 eligible participants completed the study.

One participant was excluded from subsequent analyses for medical reasons, disclosed at the end of the session. In four instances, the investigator was unable to record the ECG data with the ECG sensor due to equipment failure. Therefore, the following analysis is based on a sample of 32 participants.

The majority of participants were male, with 12 in the high-end and 11 in the basic group, alongside 4 females in each group and 1 non-binary individual in the high-end group only. Participants reported recent consumption of nicotine (3 per group), caffeine (9 in high-end, 7 in basic), or engagement in exercise (2 in high-end, 1 in basic), while 7 participants in each group reported none of the above. Media qualifications were more common in the basic group (8 versus 6 in the high-end). Glasses were worn by 4 participants in each group, facial hair was equally present (7 per group), and makeup was worn by fewer participants (1 in high-end, 2 in basic). The age of participants ranged from 18 to 45 years in the high-end group, with a mean of 25.4 years, and 20 to 47 years in the basic group, with a mean of 29.6 years. Skin tone, assessed using the Monk scale, varied from 1 to 5 with a mean of 2.8 in the high-end group, and from 1 to 6 with a mean of 3.6 in the basic group.

4. Results

4.1. Analysis of Participant Comments

Out of the 25 participants who provided comments, 16 offered insights that were not addressed by the questions in the survey. Among these 16 participants, the aural experience was specifically mentioned by 4 participants in the basic audio group and 5 in the high-end audio group. Notably, there were no questions specifically addressing the soundtrack included in the survey (see Section 3.6).

Out of the four participants who left sound-related comments in the basic audio presentation group, one highlighted how the music at the start made them feel immersed. Another made a more general comment labelling the soundtrack as compelling. Finally, two participants explicitly observed that the sound originated solely from the TV loudspeakers and not from any other loudspeakers in the room.

Out of the five participants who left sound-related comments in the high-end audio presentation group, three highlighted positive aspects of the sound or sound system (e.g., "The surround sound made it very impressive"). Of the two remaining participants, one highlighted the intensity of the soundtrack and how they felt that the heart rate sounds affected their cardiovascular activity. The final participant within this subset commented that they felt a question should have been included about how the audio setup affected immersion. To reiterate, all participants were unaware that participants received different aural conditions within this study.

Three comments offered critiques of specific survey questions. Two of these comments referenced specific items within the Film IEQ, Item 12: "To what extent did you find the concepts and themes of the film clip challenging?"; and Item 20: "At any point did you find yourself become so involved that you wanted to speak to the film clip directly?".

The analysis of the open-response question that accompanied the familiarity and initial impression questions showed that 4 participants had not heard of the film and that 12 participants had heard of the film but had not seen it. Of the 16 participants who had reported having seen the film, none described having seen it more recently than a "few years ago".

Of the remaining comments, which did not relate to any questions included in the survey, one described how they felt "the viewing angle was a little uncomfortable". Another

described how they found reflections on the screen "a little distracting". One participant made a negative comment about the media, reflecting the sentiment that they did not enjoy the film clip. Finally, another participant mentioned that they were tired during the experiment, as they had not slept well.

4.2. Analyses of Survey

In this section, the term *immersion score* is used to describe the sum of the 24-item Film IEQ [22] and single-item immersion assessment methods [25]. The range of possible immersion scores is −300 to 2200. Upon inspection of the distribution of the immersion scores, there appeared to be an extreme outlier who recorded a significantly lower immersion score than all other participants in the study. As this self-report score represented a highly idiosyncratic opinion, it was considered not conducive to evaluating the between-group effects of audio presentation and was therefore removed from the statistical tests described in this section. However, their response can be seen in Figure 3. Notably, this was the same participant who made a negative comment about the media.

Prior to viewing the film clip, participants were asked to assess their familiarity with and initial impression of the content—how excited they were to watch the clip based on their taste in films, if they had seen it before, and how they were feeling at the time. The Shapiro–Wilk test showed that the immersion scores were normally distributed, but both the familiarity and initial impression scores were not. Hence, two Spearman correlation tests were applied to determine the relationship between responses to the familiarity and initial impression questions, using the *immersion score*. Note that the immersion score is calculated by adding the score for the Film IEQ to the single immersion item. Neither familiarity nor initial impression showed a statistically significant correlation with the overall immersion score; familiarity (SCC = −0.118, p = 0.527) and initial impression (SCC = 0.044, p = 0.815). Hence, the subsequent analyses do not apply responses to these questions as covariates.

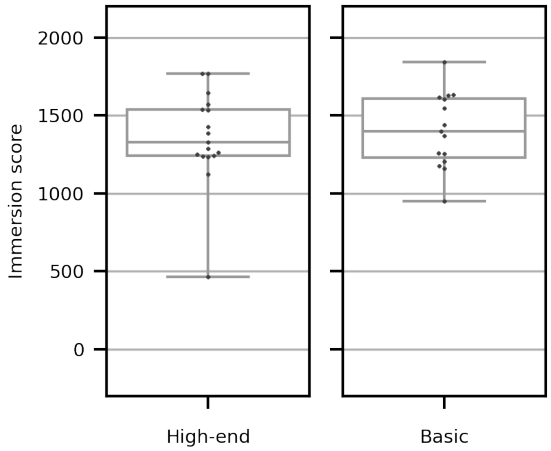

Figure 3. Box–Swarm plot showing the distribution of immersion scores for each group. The limits of the plot represent the highest and lowest possible scores associated with the immersion questionnaire, −300 and 2200.

For both groups, immersion scores were normally distributed, as assessed by the Shapiro–Wilk test. The mean immersion score was higher in the high-end audio group (1411.625) compared with the basic audio group (1404.800). There was a homogeneity of variances, as assessed by Levene's test for equality of variances (p = 0.617). There was no statistically significant difference in the mean immersion score between high-end audio and basic audio groups, t(30) = 0.086 and p = 0.868. Hence, this analysis suggests there was no difference in the extent to which participants felt immersed between the two conditions.

Figure 4 shows the participant's responses to the question, regarding how the camera, lighting, and ECG sensor, influenced their audiovisual experience. These results show that

these factors had a variable impact on the participant's viewing experience. To assess if there was a significant difference in the average response value for the questions associated with the negative impact of the ECG sensor, camera, and lighting, pairwise comparisons were made. As the Shapiro–Wilk test assessed that the responses were not all normally distributed, the Mann–Whitney U test was applied. These tests revealed no significant difference in median response between the questions associated with the negative impact of the ECG sensor and camera ($U(31) = 384.5$, $p = 0.178$). However, when comparing these distributions with the responses to the question associated with the negative impact of the room illumination, in both cases, a significant difference was found (camera: $U(31) = 169.5$, $p < 0.001$; ECG: $U(31) = 204.0$, $p < 0.001$).

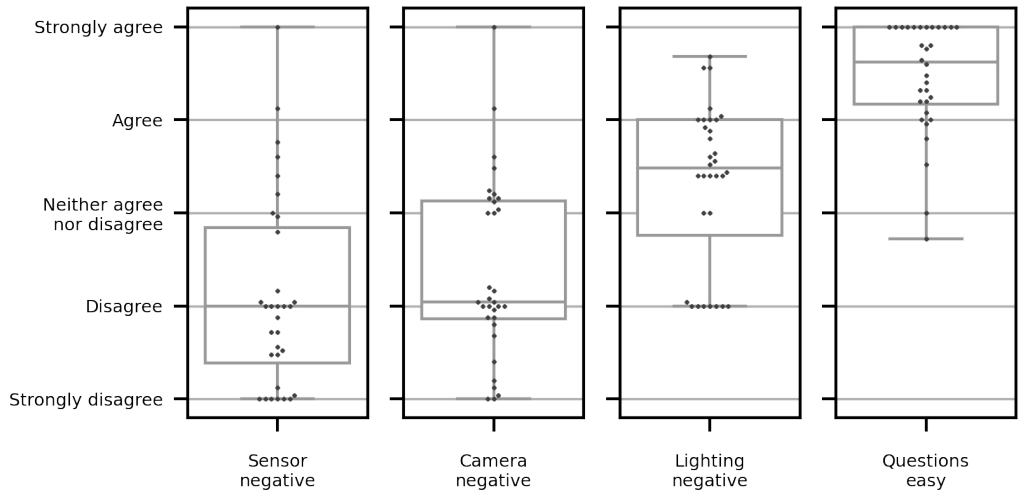

Figure 4. Distribution of the response to questions asked after the audiovisual experience. The questions are shortened; in the survey, they appear as: "I found the presence of the heart rate monitor negatively impacted my viewing experience", "I found the presence of the camera negatively impacted my viewing experience", "I found the brightness of the room negatively impacted my viewing experience", and "I found all the questions easy to understand".

4.3. Physiological Analyses

In this section, the camera-sourced heart rate is initially analysed. Subsequently, this analysis is reproduced using the ECG sensor heart rate data. Finally, a post hoc analysis is performed on both the camera-sourced data and ECG-sourced data.

The ISC-HR was calculated using Equation (1), shown in Section 2.3. Figure 5 shows the ISC-HR value distributions for both groups. The median ISC-HR value was higher in the group receiving the high-end audio presentation. However, using the permutation test described in Section 3.5 , this difference was assessed as being not statistically significant (difference in median = 0.020, $p = 0.491$).

Visual inspection of the camera-sourced ISC-HR data shown in Figure 5 suggested that there was one outlier within the high-end group. This participant had a much lower ISC-HR value than the others within the same group and had the highest average heart rate out of all participants in the study (103 BPM). Moreover, their HR appeared to be slowly decelerating throughout the audiovisual experience. With the participant removed, the permutation test evaluated that the difference in median was still not statistically significant (difference in median = 0.027, $p = 0.360$). All heart rate data, including that pertaining to the statistical outlier, is visualised in the code repository [67].

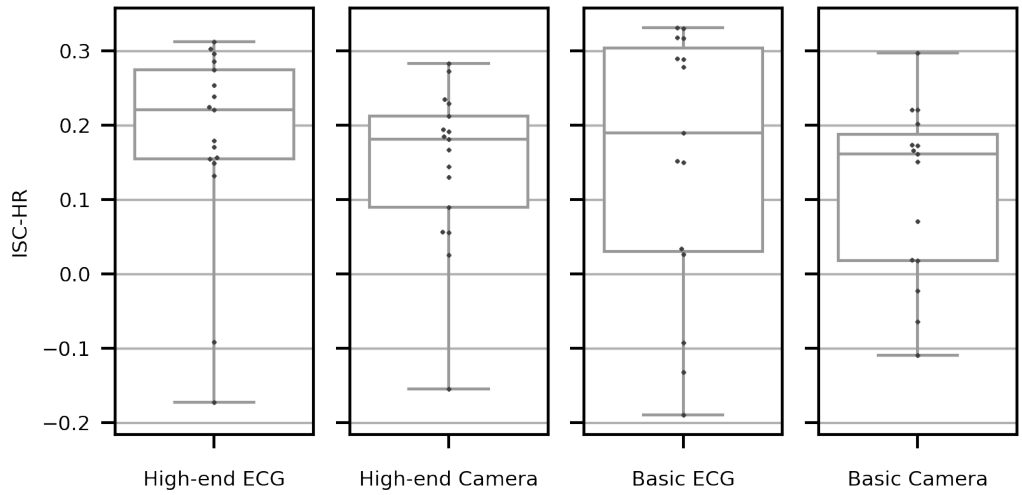

Figure 5. ISC-HR values obtained using both the camera method and ECG sensor method, split between the two experimental groups.

The ISC-HR feature was recalculated using the ECG-sourced HR data and compared between groups. The median ISC-HR value was higher in the group receiving the high-end audio presentation. However, using the permutation test described in Section 3.5, this difference in the median was assessed as being not statistically significant (difference in median = 0.032, $p = 1.000$). Visual inspection of the ECG-sourced ISC-HR values in Figure 5 appears to show the same outlier found using the camera-sourced data, as well as another statistical outlier who also had a much lower ISC-HR value than the other participants in the high-end audio presentation group. With the common outlier removed, the ISC-HR feature was recalculated, and the permutation test again assessed the between-group differences in the median as insignificant (difference in median = 0.034, $p = 0.780$). With both of these outliers removed, the difference in median ISC-HR between groups is not significant when using either the camera-measured heart rate data (difference in median = 0.0312, $p = 0.314$) or ECG sensor heart rate data (difference in median = 0.044, $p = 0.773$).

To further assess the performance of the camera approach, MAE and PCC values were calculated through a comparison of the camera and ECG-sourced HR data for each participant. The mean MAE and PCC values were 1.522 BPM and 0.782. The range of MAE and PCC values were 0.650 BPM to 4.420 BPM and 0.330 to 0.960, respectively. The participant with the lowest PCC value, between ECG and camera-sourced HR, was the ISC-HR outlier in the high-end group, who only appeared when analysing the ECG-sourced feature. The PCC between the ISC-HR values calculated using the camera-sourced heart rate data and the ECG-sourced heart rate data was calculated as 0.868 and 0.907 for the participants in the high-end and basic conditions, respectively.

Seeking an alternative means of validation regarding the relationship between immersion and heart rate synchronicity, ISC-HR values were compared with the immersion scores from the questionnaire. Figure 6 visualises this comparison, highlighting two clear outliers who recorded the highest and lowest self-report immersion scores using the questionnaire. One of these outliers was the participant who left a negative comment regarding the film media. For the other, there is no clear indication as to why their rate did not match the trend associated with other participants who assessed themselves as feeling highly immersed. Using the permutation correlation test, described in Section 3.5, the positive relationship between immersion score and ISC-HR was assessed as being insignificant, with the outliers being kept in the dataset (camera ISC-HR: SCC = 0.290, $p = 0.108$, ECG ISC-HR: SCC = 0.339, $p = 0.057$). With the two outliers removed and ISC-HR recalculated, the correlation was assessed as significant (camera ISC-HR: SCC = 0.472, $p = 0.009$, ECG ISC-HR: SCC = 0.441, $p = 0.0148$).

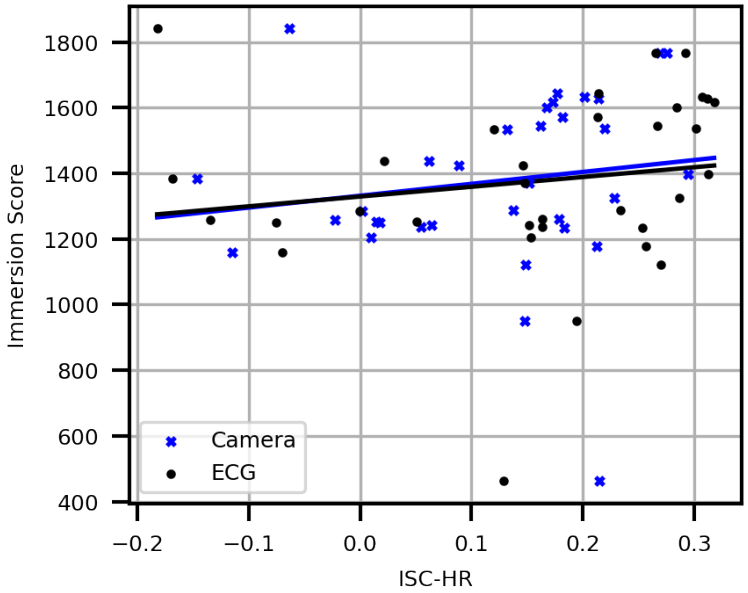

Figure 6. The positive relationship between immersion score and ISC-HR is shown. The same two statistical outliers are visible in the top-left and bottom-right quadrants. Linear regression lines have been added using the least squares error method, and their overlapping highlights the similarity of the two distributions.

4.4. Post Hoc Analysis

As described in Section 3.7, there was a difference in mean age between the two experimental groups. Hypothetically, this difference may have had a confounding effect on the negligible influence of audio presentation found in this study, as age is known to affect media preference [73], which is related to immersion. A Spearman correlation test revealed that there was no correlation between immersion score and age (SCC = 0.013, p = 0.942). A permutation correlation test of age and the camera-source ISC-HR feature and ECG-source ISC-HR feature also showed no significant correlation (camera ISC-HR: SCC = 0.104, p = 0.566, ECG ISC-HR: SCC = 0.236, p = 0.186). Moreover, the difference in median age between the two groups was assessed as being not statistically significant using a Mann–Whitney U test (U: 87.5, p-value: 0.0824). A Mann–Whitney U test was used because the age distributions of participants in the two groups were not assessed as being normally distributed using the Shapiro–Wilk test.

5. Discussion

The hypothesis for this study is that the increase in immersion influenced by a high-end audio presentation, compared with a basic audio presentation, can be measured using a camera-sourced ISC-HR feature. The results suggest that no such influence was elicited, as assessed using a camera-sourced ISC-HR feature, an equivalent ECG-sourced feature, and an immersion self-report questionnaire. However, the intercorrelation of these three measures suggests that camera-sourced ISC-HR is a viable means for assessing immersion. Specifically, the statistically significant correlation between the camera-sourced and the equivalent ECG-sourced ISC-HR features shows that the camera approach was accurate in this study. The statistically significant correlation between both ISC-HR features and the validated immersion self-report score, once two statistical outliers are removed, shows that the measures index each other.

These findings supplement previous research demonstrating correlations between ISC-HR features and self-report scores for immersion [9,10], attentional engagement [11,12], and flow [13]. However, this study introduces a novel contribution by using a camera to acquire ISC-HR data, a method not previously explored. To validate this approach further, additional research should assess the generalisability of using camera-sourced

ISC-HR features for measuring immersion across diverse populations. In particular, it is crucial to explore the effectiveness of this method for participants with darker skin tones, as this demographic was underrepresented in our sample and darker skin tones can affect the accuracy of camera-sourced HR measurements [74]. When considering this research, alternative and emerging methods for obtaining heart rates from video may be used to enhance robustness and reliability. Some of these methods may be implementable using PyVHR, the software used in this study, as it is still under active development [47].

The camera-based approach is desirable as it is fully non-contact. It is also desirable as it has the potential to unlock alternative features that provide insight into patrons' feelings of immersion. Both the ISC of facial actions [75] and ISC of gaze location [76] are promising emerging measures for assessing audiovisual experiences, which may be captured using the camera configuration used in this study. However, although the self-report data revealed that the camera and the ECG sensor affect the viewing experience to a similar extent, the bright lighting required for the camera's operation has a markedly more adverse impact than the ECG sensor. Hence, we suggest that future research should focus on how the camera may be deployed more discretely as well as the viability of obtaining psychologically indicative features in low-light conditions. Addressing these considerations will allow for a more naturalistic assessment of audiovisual experiences, where participants feel more comfortable.

Regarding the validity of the measurement instruments, there are some concerns that the questionnaire may not allow for the reliable self-assessment of immersion. Only eleven respondents strongly agreed with the statement "I found all questions easy to understand" and Items 12 and 20 from the Film IEQ were highlighted as sub-optimal. The presence of outliers and the imperfect correlation between the ISC-HR features and the immersion self-report scores also raise some concerns and suggest that ISC-HR is not a replacement for an immersion questionnaire. Upon reflection, immersion may not be the closest associated attribute to ISC features, as they are more objective measures of shared experience [75]. However, construct validity issues [77] may not be the primary effect; instead, the outliers' responses may have been caused by difficult-to-manage confounding factors. For example, they may have felt a demand characteristic to report themselves as deeply immersed, or their cardiovascular activity may have been idiosyncratic due to stress [39].

Within the analysis, there was some indication of a difference within the open-response comments. Five participants in the high-end group attributed the audio presentation using words such as immersive, intense, or impressive, whereas only two participants made similar comments within the basic audio presentation group. However, this single minor between-group differences cannot be considered conclusive. One interpretation for the lack of change in the dependent variables is that the difference between a high-end and basic audio presentation alone cannot change a spectator's feeling of immersion. This finding aligns with the results of some similar empirical studies [19,24] but not with the theory-based literature, which describes audio presentation as being capable of profound experiential differences [2,78].

An alternative interpretation of the lack of between-group differences is that the methodology applied is not capable of demonstrating the hypothesised effect. Without changing to a less ecologically valid protocol that allows for the direct comparison of audio presentations, such as those used in many previous studies [24,28–30,33], the scope for viable methodological adaptations for future research in this domain can be summarised as increasing the sample size, changing the stimuli, and improving research instruments (e.g., physiological features, questionnaire). Notably, fully utilising the dynamic range of the high-end system by increasing the loudness in this condition may be a viable option.

Focusing on the research instrument, possible alternative psychophysiological HR features may be explored using the dataset and code repository [67] associated with this publication. To elaborate, an interesting approach may involve the development of a novel inter-subject correlation (ISC) feature that considers similarities in the magnitude of HR changes. This could involve replacing the Pearson correlation coefficient (PCC)

function in the ISC-heart rate (HR) equation with an alternative function. This function may be something simple such as root-mean-square deviation or more complex such as a novel multivariate function that takes into account individual external factors such as age and gender. Exploring the ISC of heart rate variability (HRV) measures presents another promising direction [79]. Furthermore, assessing ISC features in the context of salient narrative events, as discussed in [9], could yield insightful findings. It would also be valuable to investigate the feasibility of deriving these features from both the ECG and camera data. However, such work is beyond the scope of this paper as it warrants careful attention to avoid the discovery of spurious correlations, a risk heightened by extensive data exploration without specific hypotheses.

Focusing on stimuli, one could argue that the lack of differences between groups found in this study may be partly attributed to the poor translation of *Gravity (2013)* [1] to anything other than a high-end cinema. It is possible that the high-end condition in this study did not deliver the same experience evaluated as immersive by the theoreticians [2–6]. Notably, the dimensions of the cinema screen may provide better spatial congruency between the moving image and the soundtrack. Ensuring this audiovisual matching using Dolby's domestic setup guides [15] is problematic as the size of the screen is not defined alongside the positions of the loudspeakers. In the future, it may be advantageous to consider running a similar study to the one documented in this manuscript, again using *Gravity (2013)*, in a cinema. Notably, such a study could be an opportunity to evaluate the effectiveness of a camera-based approach for assessing many participants at once.

If the study is to be repeated in a similar setting to the one in this study, rather than in a cinema, a film such as *Joker (2019)* [80] may be considered. Stylistically, this film utilises the spatial sound design concepts embedded into formats such as *Dolby Atmos* in a different way to *Gravity (2013)*. *Joker (2019)* [80] uses richly detailed spatial sound design to provide a representative backdrop to the action [81] rather than trying to emulate a congruent virtual audiovisual environment as in *Gravity (2013)* [1]. Prospectively, this media may translate better from the commercial cinema to the home cinema. Moreover, the sound design of *Joker (2019)* [80] may be considered more representative of common modern mixing techniques [82]. Hence, analysing it may provide more generalisable insight into the true influence of audio presentation methods on audiovisual experiences.

6. Conclusions

A camera-sourced ISC-HR feature, quantifying the similarity of each participant's heart rate with all other participants in terms of *minute-by-minute* changes, was found to have a statistically significant correlation with responses to a post-experience immersion questionnaire once two significant outliers were removed. Moreover, the accuracy of this camera-sourced ISC-HR feature was confirmed using an equivalent ground-truth feature obtained from an ECG signal. Based on these results, we conclude that camera-sourced heart rate synchronicity is a viable means of indexing immersion for our chosen stimuli. This is a novel method and a distinct research contribution. Considering the existing research that utilises similar ISC-HR features acquired using an ECG sensor, we expect that the measure may also generalise to other audiovisual stimuli. However, there are concerns regarding the robustness of the camera-based approach to dark-skinned individuals, different cameras, and preferable darker viewing conditions. To confirm these assertions, further research is required.

The results of the study suggest that audio presentation alone may not be capable of changing a spectator's feeling of immersion during a film-viewing experience. This is considered to be a research contribution. We now consider that the experiential value of new audio technologies is overstated in this context. In the future, alternative stimuli and a larger sample size should be applied to empirically confirm or deny the hypothetical positive experiential outcomes associated with high-end audio presentation. When discussing the results of future research efforts, we suggest consider how the specific media translates between playback scenarios.

Our research should also be supplemented with more exploratory analysis, considering the difference between the two groups and the relationship between alternative physiological features and the self-report scores, using the dataset from this study, which is available upon request. Notably, work should be conducted to explore the ECG data, which are only considered here as a ground truth for comparison with the camera-measured feature. To process these data, the GitHub repository associated with this manuscript should be utilised [67].

Author Contributions: Conceptualisation, J.W., D.M., and J.F.; methodology, J.W., D.M., and J.F.; software, J.W.; validation, J.W.; formal analysis, J.W.; investigation, J.W.; resources, J.W.; data curation, J.W.; writing—original draft preparation, J.W.; writing—review and editing, J.W., D.M., and J.F.; visualisation, J.W.; supervision, D.M. and J.F.; project administration, D.M. and J.F.; 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; and in part by a University of York-funded PhD studentship, with additional support from Bang & Olufsen.

Institutional Review Board Statement: This study was conducted in accordance with the Declaration of Helsinki and approved by the University of York School of Physics, Technology & Engineering ethics committee (protocol code: Williams20230406, date of approval: 25 April 2023).

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

Data Availability Statement: A dataset containing skin segments extracted from the video footage, ECG data, and immersion questionnaire results is available upon request; please email Damian Murphy (damian.murphy@york.ac.uk) and Joseph Williams (jw185@york.ac.uk). A code repository associated with this publication is publicly available on GitHub [67].

Acknowledgments: The authors would like to thank Hugo Hammond for his support in discussing the analytic methods applied in this paper. Acknowledgement is also given to Jay Harrison, David Geary, Constantin Popp, and Jasmine Williams for their ongoing support in discussing the themes of this project.

Conflicts of Interest: Author Jon Francombe was employed by the company Bang & Olufsen a/s. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

MAE	Mean absolute error
ISC	Inter-subject correlation
HR	Heart rate
PCC	Pearson's correlation coefficient
SCC	Spearman's correlation coefficient
ECG	Electrocardiogram
SPL	Sound pressure level
LAEq	Equivalent continuous sound pressure level

References

1. Cuarón, A. Gravity. Motion Picture by Warner Brothers Pictures, 2013.
2. Fernando, I.Z.R.; Simon, v.d.B.; Loépez, M.; Pauletto, S. Immersive Continuity: Long Takes, 3-D Sound, and the Impression of Reality in the Cinema of Alfonso Cuaroén. Ph.D. Thesis, University of York, York, UK, 2021.
3. Liang, D. Sound, space, gravity: A kaleidoscopic hearing (part I). *New Soundtrack* **2016**, *6*, 1–15. https://doi.org/10.3366/sound.2016.0079.
4. Liang, D. Sound, space, gravity: A kaleidoscopic hearing (part II). *New Soundtrack* **2016**, *6*, 191–202. https://doi.org/10.3366/sound.2016.0091.
5. Mera, M. Towards 3-D sound: Spatial presence and the space vacuum. In *The Palgrave Handbook of Sound Design and Music in Screen Media*; Palgrave Macmillan: London, UK, 2016; pp. 91–111. https://doi.org/10.1057/978-1-137-51680-0_7.

6. Cuarón, A. Gravity: Director Alfonso Cuarón on Dolby Atmos. Available online: https://www.youtube.com/watch?v=c7dKt80o0qE (accessed on 14 August 2024 YouTube video, 2014.
7. Dolby Atmos—Official Site. Available online: https://www.dolby.com/technologies/dolby-atmos/ (accessed on 14 August 2024).
8. Gravity Blu-Ray (Special Edition). Available online: https://www.blu-ray.com/movies/Gravity-Blu-ray/138954/ (accessed on 14 August 2024).
9. Hammond, H. Developing Continuous Measures of Audience Immersion. Ph.D. Thesis, University of Bristol, Bristol, UK, 2023.
10. 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.
11. Madsen, J.; Parra, L.C. Cognitive processing of a common stimulus synchronizes brains, hearts, and eyes. *PNAS Nexus* **2022**, *1*, pgac020. https://doi.org/10.1093/pnasnexus/pgac020.
12. 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*, 109289. https://doi.org/10.1016/j.celrep.2021.109289.
13. Gibbs, H.J.; Czepiel, A.; Egermann, H. Physiological synchrony and shared flow state in Javanese gamelan: positively associated while improvising, but not for traditional performance. *Front. Psychol.* **2023**, *14*, 1214505.
14. Mark Andrews, B.C. Brave. Motion Picture by Walt Disney Pictures and Pixar Animation Studios, 2012.
15. Dolby Laboratories. Speaker Setup Guides. 2024. Available online: https://www.dolby.com/about/support/guide/speaker-setup-guides/ (accessed on 14 August 2024).
16. Williams, J.; Shepstone, S.; Murphy, D. Understanding Immersion in the Context of Films with Spatial Audio. In Proceedings of the Audio Engineering Society Conference: 2022 AES International Conference on Audio for Virtual and Augmented Reality, Redmond, VA, USA, 15–17 August 2022.
17. Bech, S.; Zacharov, N. *Perceptual Audio Evaluation—Theory, Method and Application*; John Wiley & Sons: Hoboken, NJ, USA, 2006. https://doi.org/10.1002/9780470869253.
18. Agrawal, S.R.; Bech, S. Immersion in audiovisual experiences. In *Sonic Interactions in Virtual Environments*; Springer International Publishing: Cham, Switzerland, 2022; pp. 319–351. https://doi.org/10.1007/978-3-031-04021-4_11.
19. Williams, J.; Francombe, J.; Murphy, D. Exploring the influence of multichannel soundtracks on film immersion. In Proceedings of the Audio Engineering Society Conference: AES 2023 International Conference on Spatial and Immersive Audio, Huddersfield, UK, 23–25 August 2023; Audio Engineering Society: New York, NY, USA, 2023.
20. Agrawal, S.; Simon, A.; Bech, S.; Bæntsen, K.; Forchhammer, S. Defining immersion: Literature review and implications for research on audiovisual experiences. *J. Audio Eng. Soc.* **2020**, *68*, 404–417. https://doi.org/10.17743/jaes.2020.0039.
21. Murray, J.H. In *Hamlet on the Holodeck: The Future of Narrative in Cyberspace*; The MIT Press: Cambridge, MA, USA, 1998; p. 99.
22. Rigby, J.M.; Brumby, D.P.; Gould, S.J.J.; Cox, A.L. Development of a questionnaire to measure immersion in video media: The Film IEQ. In Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video, Salford, UK, 5–7 June 2019; Association for Computing Machinery: New York, NY, USA, 2019; TVX '19, pp. 35–46. https://doi.org/10.1145/3317697.3323361.
23. Agrawal, S.; Bech, S.; Bærentsen, K.; De Moor, K.; Forchhammer, S. Method for subjective assessment of immersion in audiovisual experiences. *J. Audio Eng. Soc.* **2021**, *69*, 656–671. https://doi.org/10.17743/jaes.2021.0013.
24. Agrawal, S.; Bech, S.; De Moor, K.; Forchhammer, S. Influence of Changes in Audio Spatialization on Immersion in Audiovisual Experiences. *J. Audio Eng. Soc.* **2022**, *70*, 810–823. https://doi.org/10.17743/jaes.2022.0034.
25. Reysen, S.; Plante, C.N.; Roberts, S.E.; Gerbasi, K.C. Initial validation and reliability of the single-item measure of immersion. *Creat. Ind. J.* **2019**, *12*, 272–283. https://doi.org/10.1080/17510694.2019.1621586.
26. Choi, Y.; Kim, J.; Hong, J.H. Immersion Measurement in Watching Videos Using Eye-tracking Data. *IEEE Trans. Affect. Comput.* **2022**, *13*, 1759–1770. https://doi.org/10.1109/TAFFC.2022.3209311.
27. Zhang, C. The Why, What, and How of Immersive Experience. *IEEE Access* **2020**, *8*, 90878–90888. https://doi.org/10.1109/ACCESS.2020.2993646.
28. Lessiter, J.; Freeman, J.; Davidoff, J. Really hear? The effects of audio quality on presence. In Proceedings of the Fourth Annual International Workshop on Presence, Philadelphia, PA, USA, 21–23 May 2001; pp. 288–324.
29. Zieliński, S.; Rumsey, F.; Bech, S. Effects of down-mix algorithms on quality of surround sound. *AES J. Audio Eng. Soc.* **2003**, *51*, 780–798.
30. Zieliński, S.; Rumsey, F.; Bech, S. Effects of Bandwidth Limitation on Audio Quality in Consumer Multichannel Audiovisual Delivery Systems. *J. Audio Eng. Soc.* **2003**, *51*, 475–501. https://doi.org/10.AESE-Library10278.
31. Lipscomb, S.D.; Kerins, M. An empirical investigation into the effect of presentation mode in the cinematic and music listening experience. In Proceedings of the 8th International Conference on Music Perception & Cognition, Evanston, IL, USA, 3–7 August 2004.
32. Oramus, T.; Neubauer, P. Comparison Study of Listeners' Perception of 5.1 and Dolby Atmos. In Proceedings of the Audio Engineering Society Convention 147, New York, NY, USA, 16–19 October 2019; Audio Engineering Society: New York, NY, USA, 2019.

33. Langiulli, N.; Calbi, M.; Sbravatti, V.; Umiltà, M.A.; Gallese, V. The effect of Surround sound on embodiment and sense of presence in cinematic experience: A behavioral and HD-EEG study. *Front. Neurosci.* **2023**, *17*, 1222472. https://doi.org/10.3389/fnins.2023.1222472.
34. Kerins, M. Understanding the impact of surround sound in multimedia. In *The Psychology of Music in Multimedia*; Oxford University Press: Oxford, UK, 2013; pp. 365–388. https://doi.org/10.1093/acprof:oso/9780199608157.003.0016.
35. Hedges, J.; Sazdov, R.; Johnston, A. Measuring the influence of audio on immersive experience in extended reality and digital games: A systematic review. In Proceedings of the 2023 Immersive and 3D Audio: From Architecture to Automotive (I3DA), Bologna, Italy, 5–7 September 2023; pp. 1–13. https://doi.org/10.1109/I3DA57090.2023.10289267.
36. Potter, T.; Cvetković, Z.; De Sena, E. On the relative importance of visual and spatial audio rendering on vr immersion. *Front. Signal Process.* **2022**, *2*, 904866.
37. Fröber, K.; Thomaschke, R. In the dark cube: Movie theater context enhances the valuation and aesthetic experience of watching films. *Psychol. Aesthet. Creat. Arts* **2021**, *15*, 528–544. https://doi.org/10.1037/aca0000295.
38. Reeves, B.; Lang, A.; Kim, E.Y.; Tatar, D. The effects of screen size and message content on attention and arousal. *Media Psychol.* **1999**, *1*, 49–67.
39. Giannakakis, G.; Grigoriadis, D.; Giannakaki, K.; Simantiraki, O.; Roniotis, A.; Tsiknakis, M. Review on Psychological Stress Detection Using Biosignals. *IEEE Trans. Affect. Comput.* **2019**, *13*, 440–460.
40. Lahnakoski, J.; Change, L. Intersubject Correlation—Naturalistic Data Analysis. Available online: https://naturalistic-data.org/content/Intersubject_Correlation.html (accessed on14 August 2024 2020).
41. Nastase, S.A.; Gazzola, V.; Hasson, U.; Keysers, C. Measuring shared responses across subjects using intersubject correlation. *Soc. Cogn. Affect. Neurosci.* **2019**, *14*, 667–685. https://doi.org/10.1093/scan/nsz037.
42. Laerd Statistics. Pearson Correlation Coefficient Statistical Guide. Available online: https://statistics.laerd.com/statistical-guides/pearson-correlation-coefficient-statistical-guide.php (accessed on 14 August 2024).
43. Laerd Statistics. Independent t-Test Using SPSS Statistics. Available online: https://statistics.laerd.com/spss-tutorials/independent-t-test-using-spss-statistics.php (accessed on 14 August 2024).
44. Morales-Fajardo, H.M.; Rodríguez-Arce, J.; Gutiérrez-Cedeño, A.; Viñas, J.C.; Reyes-Lagos, J.J.; Abarca-Castro, E.A.; Ledesma-Ramírez, C.I.; Vilchis-González, A.H. Towards a non-contact method for identifying stress using remote photoplethysmography in academic environments. *Sensors* **2022**, *22*, 3780. https://doi.org/10.3390/s22103780.
45. Sabour, R.M.; Benezeth, Y.; De Oliveira, P.; Chappe, J.; Yang, F. Ubfc-phys: A multimodal database for psychophysiological studies of social stress. *IEEE Trans. Affect. Comput.* **2021**. https://doi.org/10.1109/TAFFC.2021.3056960.
46. Zhou, K.; Schinle, M.; Weimar, S.; Gerdes, M.; Stock, S.; Stork, W. End-to-End Deep Learning for Stress Recognition Using Remote Photoplethysmography. In Proceedings of the 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Las Vegas, NV, USA, 6–8 December 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 1435–1442. https://doi.org/10.1109/BIBM55620.2022.9995577.
47. 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. https://doi.org/10.7717/peerj-cs.929.
48. Mitsubishi Electric. Mitsubishi LDT422V LCD Display. Product Manual. 2009. Available online: https://www.manuals.co.uk/mitsubishi/ldt422v/manual (accessed on 14 August 2024).
49. Genelec. Genelec 8010A Studio Monitor. Product Information. Available online: https://www.genelec.com/8010a (accessed on 14 August 2024).
50. Genelec. Genelec 8030C Studio Monitor. Product Information. Available online: https://www.genelec.com/8030c (accessed on 14 August 2024).
51. Genelec. Genelec 7040A Studio Subwoofer. Product Information. Available online: https://www.genelec.com/7040a (accessed on 14 August 2024).
52. Genelec. Genelec 8040A Studio Monitor. Product Information. Available online: https://www.genelec.com/previous-models/8040a (accessed on 14 August 2024).
53. Mulcahy, J. REW Room Acoustics and Audio Device Measurement and Analysis Software. Available online: https://www.roomeqwizard.com/ (accessed on 6 November 2023).
54. Earthworks Audio. M30 Measurement Microphone. Product Brochure. 2023. Available online: https://earthworksaudio.com/wp-content/uploads/2018/07/M30-Data-Sheet-2018.pdf (accessed on 14 August 2024).
55. Dolby Reference Player. Available online: https://professional.dolby.com/product/ (accessed on 6 November 2023).
56. NTi Audio XL2 Audio and Acoustic Analyzer. Available online: https://www.nti-audio.com/en/products/xl2-sound-level-meter (accessed on 30 April 2024).
57. International Telecommunication Union. Recommendation ITU-R BS.775-4 Multi-Channel Stereophonic sound System with and without Accompanying Picture. 2012. Available online: https://www.itu.int/rec/R-REC-BS.775-4-202212-I/en (accessed on 15 August 2024).
58. Tenma 72-860A Loudness Meter. Available online: https://www.tenma.com (accessed on 30 April 2024).
59. Neumann KU 100 Dummy Head Microphone. Available online: https://www.neumann.com/homestudio/en/ku-100 (accessed on 30 April 2024).

60. Canon Inc. Canon EOS 700D. Digital SLR Camera. 2013. Available online: https://www.canon.co.uk/support/consumer/products/cameras/eos/eos-700d.html (accessed on 14 August 2024).
61. 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; IEEE: Piscataway, NJ, USA, 2017; pp. 63–70. https://doi.org/10.1109/FG.2017.17.
62. 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; SPIE: Bellingham, WA, USA, 2015; Volumr 9417, pp. 639–652. https://doi.org//10.1117/12.2083940.
63. Softonic. LosslessCut GitHub. Available online: https://github.com/mifi/lossless-cut (accessed on 26 June 2023).
64. Harvey, P. Exiftool. Available online: https://exiftool.org/ (accessed on 14 August 2024).
65. Polar Electro. *Polar H10 Heart Rate Sensor User Manual*. 2023.Available online: https://support.polar.com/e_manuals/h10-heart-rate-sensor/polar-h10-user-manual-english/manual.pdf (accessed on 14 August 2024).
66. Jukka Happonen. Polar Sensor Logger. Google Play Store. 2023. Available online: https://play.google.com/store/apps/details?id=com.j_ware.polarsensorlogger (accessed on 14 August 2024).
67. AudioLab. Github: Camera-Sourced-Heart-Rate-Synchronicity, Nov. 2023. Available online: https://github.com/AudioLabYork/Camera-Sourced-Heart-Rate-Synchronicity (accessed on 14 August 2024).
68. 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.
69. Rooney, B.; Hennessy, E.; Bálint, K. Viewer versus film: Exploring interaction effects of immersion and cognitive stance on the heart rate and self-reported engagement of viewers of short films. In Proceedings of the Poster presentation at the Society for Cognitive Studies of the Moving Image, Franklin & Marshall College, Lancaster, PA, USA, 11–14 June, 2014.
70. Pesarin, F.; Salmaso, L. The permutation testing approach: A review. *Statistica* **2010**, *70*, 481–509. https://doi.org/10.6092/issn.1973-2201/3599.
71. Laerd Statistics. Spearman's Rank-Order Correlation Using SPSS Statistics. Available online: https://statistics.laerd.com/spss-tutorials/spearmans-rank-order-correlation-using-spss-statistics.php (accessed on 14 August 2024).
72. Monk, E. Monk Scale. Available online: https://skintone.google/the-scale (accessed on 26 June 2023).
73. Statista. Favorite Movie Genres in the U.S. by Age. 2018. Available online: https://www.statista.com/statistics/949810/favorite-movie-genres-in-the-us-by-age/ (accessed on 14 August 2024).
74. 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. https://doi.org/10.1109/CVPRW50498.2020.00150.
75. Cheong, J.H.; Molani, Z.; Sadhukha, S.; Chang, L.J. Synchronized affect in shared experiences strengthens social connection. *Commun. Biol.* **2023**, *6*, 1099.
76. Bühler, B.; Bozkir, E.; Deininger, H.; Gerjets, P.; Trautwein, U.; Kasneci, E. On Task and in Sync: Examining the Relationship between Gaze Synchrony and Self-Reported Attention During Video Lecture Learning. *arXiv* **2024**, arXiv:2404.00333.
77. Tomarken, A.J. A psychometric perspective on psychophysiological measures. *Psychol. Assess.* **1995**, *7*, 387.
78. Kerins, M. *Beyond Dolby (Stereo): Cinema in the Digital Sound Age*; Indiana University Press: Bloomington, IN, USA, 2010; pp. 92–95. https://doi.org/10.2979/bdlb.2010.1.
79. Coutinho, J.; Pereira, A.; Oliveira-Silva, P.; Meier, D.; Lourenço, V.; Tschacher, W. When our hearts beat together: Cardiac synchrony as an entry point to understand dyadic co-regulation in couples. *Psychophysiology* **2021**, *58*, e13739.
80. Phillips, T. Joker. Motion Picture by Warner Bros. Pictures, 2019.
81. Kerins, M. Hearing reality in Joker. *New Rev. Film. Telev. Stud.* **2021**, *19*, 89–100.
82. Bolen, K. Mixing 3D audio for film. In *3D Audio*; Routledge: London, UK, 2021; pp. 274–282.