







Article

Perception of Audio–Visual Synchronization in Olfactory-Enhanced 360-Degree Video

Aleph Campos da Silveira ^{1,2,3,*} , Roope Raisamo ^{2,*} , Fotios Spyridonis ³ , Alexandra Covaci ⁴ ,
Gheorghita Ghinea ³  and Celso Alberto Saibel Santos ¹ 

¹ Department of Informatics, Federal University of Espirito Santo, Vitoria 29075-910, ES, Brazil; saibel@inf.ufes.br

² TAUCHI Research Center, Tampere University, Tampere 33014, Finland

³ Department of Computer Science, College of Engineering, Design and Physical Sciences, Brunel University London, Uxbridge UB8 3PH, UK; fotios.spyridonis@brunel.ac.uk (F.S.); george.ghinea@brunel.ac.uk (G.G.)

⁴ College of Engineering and Digital Design, University of Kent, Kent CT2 7NX, UK; a.covaci@kent.ac.uk

* Correspondence: aleph.camposdasilveira@tuni.fi (A.C.d.S.); roope.raisamo@tuni.fi (R.R.)

Abstract

This study examines the impact of olfactory stimuli on user experience (UX) metrics in 360-degree videos under varying levels of audio–visual (AV) skew. Subjective responses and questionnaire results revealed that scents helped stabilize enjoyment and artifact tolerance scores, particularly under severe AV skews, compared to non-olfactory conditions. However, the stationary nature of the scent delivery device decreased the intensity of olfactory stimuli, limiting their potential impact. Objective analyses highlighted a masking effect in 360-degree videos, where participant visual exploration reduced sensitivity to AV skews. Despite these challenges, olfactory stimuli demonstrated resilience to AV skews, suggesting their potential to buffer negative effects and enhance immersive experiences. However, they did not significantly improve overall video quality ratings. The study underscores the need for advances in olfactory display technology, such as head-mounted scent emitters and dynamic sensory integration, to enhance multimedia experiences.

Keywords: mulsemmedia; olfaction; perceived quality; synchronization



Academic Editor: Andrea Prati

Received: 19 August 2025

Revised: 21 September 2025

Accepted: 22 September 2025

Published: 25 September 2025

Citation: da Silveira, A.C.; Raisamo, R.; Spyropoulos, F.; Covaci, A.; Ghinea, G.; Santos, C.A.S. Perception of Audio–Visual Synchronization in Olfactory-Enhanced 360-Degree Video. *Appl. Sci.* **2025**, *15*, 10414. <https://doi.org/10.3390/app151910414>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Multimedia content typically comprises video and audio elements, which are intrinsically interconnected [1–3]. Over time, technological advances have enabled the seamless integration of multimedia into computer systems, facilitated by high-speed communication channels, broadband networks (such as 5G), and advanced multimedia devices, including head mounted displays (HMDs) and CAVE systems, all of which contribute significantly. These innovations have facilitated the large-scale production and dissemination of multimedia content, generating opportunities for content creators and consumers.

In real-time multimedia playback, whether localized or within a distributed framework, it is imperative to preserve temporal coherence among disparate media sources to ensure seamless playback [2]. However, multimedia synchronization content involves significant complexities [4]. Furthermore, the perception of audio synchronization in videos augmented with multimedia content remains insufficiently explored. Elucidating user experiences in these contexts can yield critical insights for advancing multimedia synchronization methodologies and enhancing the holistic viewing experience.

Although audio and video synchronization has previously been examined [5–7], assessing synchronized media augmented with additional sensory content, such as olfactory and haptic stimuli, constitutes a growing field of inquiry, particularly with the proliferation of HMDs [8,9]. This presents a unique challenge for researchers, as the integration of additional sensory stimuli has the potential to alter user comprehensive sensory experiences and media perception [10,11]. Consequently, it is crucial to explore how the synchronization of these diverse sensory modalities may influence user perception and engagement with mulsemmedia content.

The ongoing research on mulsemmedia Quality of Experience (QoE) involves synchronization of mulsemmedia data with AV content, as well as the capture and rendering of mulsemmedia signals [12]. Any difference in the presentation times of data objects related to the media of different media is called an interstream “skew” [6]. When multiple media streams are perfectly synchronized, there is no skew, and the value is 0 ms. However, when it comes to olfactory devices, achieving 0 skew is difficult due to the persistent nature of olfaction [13,14].

Notwithstanding the extensive examination of audiovisual (AV) skews in conventional multimedia, the inquiry into 360-degree videos remains scant. To bridge this gap, we performed an empirical study on 360-degree videos subjected to varying degrees of skewness and augmented with olfactory stimuli. Contrary to traditional multimedia, where AV skewness typically exhibits a low tolerance threshold and is promptly discernible by viewers, we posited that the immersive nature of 360-degree videos, when enhanced with mulsemmedia content, could potentially offer a distinct advantage, allowing for more lenient AV synchronization compared to traditional two-dimensional multimedia. The comprehensive visual engagement in a 360-degree video may obscure certain auditory discrepancies, thereby allowing for a broader margin of synchronization without impairing the viewer’s overall experience. Consequently, we embarked on a study posited on two foundational research questions:

- RQ1. To what extent do AV skews impact the QoE in 360-degree mulsemmedia environments?
- RQ2. Does mulsemmedia exhibit a masking effect on QoE in 360-degree mulsemmedia contexts in the presence of AV skews?

As elaborated upon in subsequent sections, our experimental configuration incorporated 360-degree video recordings depicting three distinct scenarios with three levels of AV dynamism: the serene environment of a *Coffee Shop*, a mild *Fireworks* scenery, and dynamic high-intensity action of a *Kung Fu* scene. Uniform AV skews were methodically applied to each video, introducing controlled time delays between video and audio components.

Although our study focuses primarily on QoE, it is important to highlight its relationship with User Experience (UX), especially in the context of multisensory media. QoE traditionally refers to the user’s satisfaction with a service or application, often influenced by technical parameters such as delay, jitter, media quality, and intermedia synchronization. In contrast, UX encompasses a broader range of factors, including emotional, cognitive, and contextual elements that extend beyond the system’s performance. The boundaries between QoE and UX can be fluid in immersive environments, such as 360-degree mulsemmedia. For this work, QoE is employed as a measurable proxy of user-perceived quality, with the understanding that it captures only a subset of the broader UX.

The structure of this manuscript is as follows. Section 3 outlines the experimental setups and devices used. Section 4 presents our findings. These findings are further discussed in Section 5. Finally, Section 6 provides a comprehensive summary and conclusion of our research.

2. Physiological Signals

Although surveys and interviews are useful tools for collecting data on how users interact with a product or service, they exhibit inherent limitations in effectively capturing the real-time dynamics of human–machine interactions. This approach is susceptible to inaccuracies, largely due to its reliance on user memory, which can introduce errors and biases in recollection [15]. To mitigate these issues, there is growing interest in physiological monitoring as an alternative method for capturing user behavior during the interaction process.

Physiological signals are metrics of human physiological processes and can be classified into two main categories: physical and physiological [16–19]. Physical signals include outputs of muscular activity, such as body tension, pupil dilation, ocular movements, blinking, posture, respiration, facial expressions, and vocal production. However, physiological signals are directly associated with the Autonomic Nervous System (ANS) and encompass cardiac rhythms, neural activities, and muscular excitability, which are evaluated by electrocardiography. These empirical measures are increasingly used for evaluations as they are immune to the biases inherent in subjective measures, such as questionnaires, which can be prone to inaccuracies [20]. To substantiate the evaluation of how mulsemmedia content can augment the QoE and UX, we have employed an approach centered on these objective metrics, avoiding exclusive reliance on post-experiment surveys.

The multimedia industry will benefit from these innovations. The rapid evolution of multimedia in terms of technical sophistication and cultural significance has provoked debates about its impact on human behavior. In response, technological advances in sensor technology and machine learning are now being utilized to model user experiences during multimedia perception. Multiple physiological processes can be monitored by biofeedback instruments that measure them and transform the results into straightforward and immediately available comprehensible data.

For example, Galvanic Skin Response (GSR), although not without its flaws [21], has become a reliable indicator of emotional arousal and stress [22]. Also referred to as ElectroDermal Activity (EDA), GSR captures changes in the electrical properties of the skin that fluctuate in response to sweat gland activity. Because the sympathetic nervous system controls sweat gland activity, GSR is a direct measure of physiological arousal. During stressful or emotionally charged moments in a game, GSR readings increase as skin conductance increases. These variations are captured through sensors attached to the skin, objectively measuring the player's emotional state.

By evaluating GSR data in real time, game developers can acquire insights into periods of increased emotional engagement or stress [23]. These data can be utilized to modify game challenges, offer immediate feedback, or craft personalized experiences that align with the player's emotional condition. Such dynamic modifications can enhance the overall user experience, rendering the game more immersive and attuned to the player's needs and capabilities.

Alongside GSR, heart rate variability (HRV), derived from electrocardiogram (ECG) data, constitutes another pertinent physiological measure. HRV reflects the temporal fluctuations between heartbeats and is a crucial indicator of ANS functionality. Regarding mental state identification, decreased HRV is often associated with lower levels of stress and engagement, while elevated HRV generally signifies increased engrossment or stress [22]. Consequently, HRV offers additional information on the player's emotional state, which enhances the depth of UX assessments during gameplay.

In this study, to improve and better understand our user evaluations, we used GSR and HRV to augment the evaluation of QoE and UX, thereby providing a more comprehensive perspective on users' perception of multimedia and mulsemmedia.

3. Methodology

Physiological data collection was implemented as an improved evaluation methodology for mulsemmedia environments. As mentioned above, biofeedback encompasses the acquisition and preparation of physiological data through specialized instruments. These instruments measure physiological activities, transform the data into interpretable information, and comprehensively present them.

3.1. Apparatus

The following instrumentation was used for collecting data in our study:

1. Oculus/Meta Quest 2 128 GB Standalone Wireless All-In-One VR Headset System.
2. Polar H10 and Verity Sense Heart Rate Sensors.
3. NGW-1pc Grove GSR Sensor 3.3 V/5 V.
4. ExHalia SBi4.

3.1.1. Oculus/Meta Quest 2

The HMD Meta Quest 2 [9,24] is a head-mounted display that is economically accessible and user-friendly. Initially marketed as Oculus Quest 2 (<https://developer.oculus.com/blog/introducing-oculus-quest-2-the-next-generation-of-all-in-one-vr/>, Accessed 24/09/2025), it was developed by Meta Platforms (formerly Facebook, Inc.) and became the HMD with the highest sales in 2022 (<https://www.statista.com/statistics/1222146/xr-headset-shipment-share-worldwide-by-brand/>, Accessed 24/09/2025)).

3.1.2. Polar H10

The Polar H10 chest strap collects and processes HRV measurements by detecting electrical signals from the heart. Following the recommendation of [25], Polar H10 electrodes were moistened with water at room temperature before being positioned on the participant and subsequently fitted around the participant's chest, just below the pectoral muscles, with the HR sensor located at the xiphoid process of the sternum. Velcro was affixed to the reverse side of the chest strap to allow for an optimal fit around the participant. Before HRV recordings, all participants' biosignals were monitored in an idle state. Following the placement of the HRV, the participants remained seated for 2 min.

3.1.3. Grove GSR

The Grove GSR module enables the detection of such pronounced emotional states by applying two electrodes to two fingers on one hand. Following the guidelines of [26], the GSR electrodes were affixed to the index and middle fingers of the participants. The GSR sensor measures micro voltages (MV) across the digits using the affixed electrodes. Additionally, the sensor computes skin resistance (SR) in ohms based on the MV input. The sensor manufacturer provides the SR calculation formula based on MV. The sensor was interfaced with an Arduino Uno, and the collected data were transferred to a computer at a sampling rate of 192 Hz.

3.1.4. ExHalia SBi4

The ExHalia SBi4 Personal Scent Emission Device, interfaced via USB connection, delivered the olfactory stimuli. Exhalia offers an array of additional scent emission devices designed to enhance scented atmospheres, point-of-sale environments, and scented objects [27]. For this experiment, only a cartridge of lavender scent was used to achieve a pleasant aroma. The operation of the device was managed using Unity (version 2024.1.0) in conjunction with PlaySEM [28].

Participants engaged in 360-degree content using a Meta Quest 2 headset. A Unity application, integrated with the Oculus application and interfaced with Meta 2 via Oculus Link, organized the video sequence and adjusted its skew. This configuration was operated on a Dell Precision Tower 3620 desktop, equipped with a quad-core Intel Core i7-7700 HQ processor, 16 GB of RAM, 500 GB SSD, and an AMD RX560 graphics card, in conjunction with a Dell Latitude 3490 laptop.

3.2. Participants

We used purposive sampling to recruit participants, organizing them into two distinct groups as outlined below. Purposive sampling is a type of nonprobability method in which researchers intentionally select individuals based on specific characteristics or criteria, relying on their judgment to identify those most relevant to the study.

1. Group 1 (G₁-NonOlfactory) comprised 15 participants who experienced the 360-degree videos with AV skews but without the olfactory stimuli;
2. Group 2 (G₂-Olfactory) comprised 30 participants who experienced the 360-degree videos with AV skews and olfactory stimuli.

In the study, a mixed factorial design was employed with a factor between subjects (olfactory stimuli: present vs. absent) and a factor within subjects (AV skew levels: low, medium, high), where 45 participants, consistent with similar experiments as shown by [29], were randomly assigned to the olfactory or non-olfactory group and experienced all skew levels. Late analyses combined audiovisual skew values of equal magnitude but opposite directions into three categories: LS—low skew (± 1 s); MS—medium skew (± 3 s); and HS—high skew (± 5 s).

This decision was driven by theoretical and statistical considerations. From a theoretical point of view, our primary interest was in the magnitude of temporal asynchrony rather than its direction. Empirically, preliminary models treating each signed skew separately resulted in sparse data per condition, unstable parameter estimates, and no statistically significant effects, despite similar response patterns for lead and lag of the same magnitude.

By collapsing across directions, we increased the number of observations per condition, improved the stability of the mixed-effect estimates, and enhanced statistical power. This aggregation revealed robust and internally consistent effects in both heart rate and skin conductance measures, which were not apparent in the more fragmented and direction-specific analysis.

All subjects self-reported normal vision and were pre-screened for contraindications (e.g., epilepsy, psychoactive drug treatment) pertinent to Virtual Reality (VR) usage. The sample size and sampling methodology are consistent with the established principles of experimental design [30]. Ethical approval for this study was granted by the Ethics Committee of Brunel University London (Review Number: 40020-LR-Oct/2022-41826-3) on 24 October 2022.

It is crucial to underscore the count of participants, as articulated by [31]. The intricate nature of biodevices and the management of the resultant data complicate the execution of experiments involving real users in practical environments. Frequent calibration of biodevices is imperative, as signal noise and interference can invalidate samples. Capturing physiological signals on a scale poses considerable difficulties, as highlighted by [31], with reviewed studies involving fewer than 30 participants and most experimental setups collecting simultaneous data from only about six subjects. Consequently, data collection often spans several weeks. Experiments are typically conducted in rigorously controlled settings to mitigate these complexities. Furthermore, while numerous studies address ethical considerations, the ongoing handling and protection of user data pose persistent

issues. Researchers must thoroughly define the protocols for data storage and security, as well as the measures used to ensure the ethical use of the data.

3.3. Materials

To address the two research questions of our study, three 360-degree videos of varying AV dynamism were selected, ranging from low to high, as proposed by [32]. The *LD* (Figure 1a) shows the process of making coffee, accompanied by dubbed background sounds. Considering the straightforward audio context, we expected that viewers would be less likely to notice AV skews in this video. At the opposite end of the spectrum, the *HD* video (Figure 1c) features a sequence of fight scenes in First-Person View (FPV) requiring meticulous AV synchronization to align sound effects with visual movements on the screen. We anticipated that any AV misalignment in this clip would be easily noticeable to participants. In the middle is the *MD* video (Figure 1b), which contains the spectacle of a fireworks show. Given that the scene is set against a night sky, punctuated by the display of fireworks and their accompanying sounds, this video's dynamism is situated between the two mentioned scenarios. All videos were encoded at AAC (LC) (mp4a/0x6134706D), 48 KHz, stereo, fltp, 128 Kbps (default) and H.264 (high), and vp9.



Figure 1. Screenshots of the three video stimuli: (a) *Coffee Shop*, a low-dynamism scene with minimal audiovisual activity; (b) *Fireworks*, a medium-dynamism display with moderate visuals and sound effects; (c) *Kung Fu*, a high-dynamism martial arts sequence with rapid movements and intense audio.

To investigate the research questions of our study, we manipulated the audio tracks of the videos to introduce delays and hastening effects ranging from 1 to 5 s. Each video was edited to a duration of 60 s. The videos were selected according to the study of [32], which employs the videos in a sequential order that can be classified from static (coffee preparation) to mild and high dynamism (*Fireworks* and *Kung Fu*, respectively).

3.4. Experimental Protocol

The experiments were conducted within a specialized facility. Upon entry and before the start of the experimental protocol, participants were evaluated for neurological or psychological disorders that could interfere with their responses to video stimuli. Participants were seated in a swivel chair, allowing them to experience 360-degree video content with rotational mobility. Comprehensive information on the purpose and procedures of the study was presented to participants to ensure their understanding and obtain informed consent. The participants then adjusted the VR headset, the chest strap, and the GSR collector to achieve optimal comfort. Subsequently, the participants were exposed to a series of videos, with the presentation sequence being randomized as illustrated in Table 1 to control for order effects and achieve a balanced experimental design for all groups.

Table 1. Experimental design showing the six video clips viewed by each participant. Each clip was adapted from three original videos and varied by motion dynamism (low dynamism (LD), medium dynamism (MD), and high dynamism (HD)) and by temporal audio offset and audio skews, ranging from -5 s to $+5$ s.

| Participant | Video 1 | Video 2 | Video 3 | Video 4 | Video 5 | Video 6 |
|-------------|-----------|-----------|-----------|-----------|-----------|-----------|
| 1 | LD -5 s | MD -3 s | HD -1 s | LD $+1$ s | MD $+3$ s | HD $+5$ s |
| 2 | HD -5 s | LD -3 s | MD -1 s | HD $+1$ s | LD $+3$ s | MD $+5$ s |
| 3 | MD -5 s | HD -3 s | LD -1 s | MD $+1$ s | HD $+3$ s | LD $+5$ s |

3.5. Research Instruments

Following each exposure session to a 360-degree video, participants were invited to answer questions about their experience. Consequently, this questionnaire addressed RQ1 and RQ2, evaluating the participant's QoE and providing critical insights into their subjective perception and satisfaction with the video.

QoE Questionnaire

This questionnaire contains four (G_1 -NonOlfactory) to seven (G_2 -Olfactory) questions, depending on the group. These questions, shown in Table 2, serve as a tool to collect qualitative feedback, allowing us to gain insights into the QoE of users and understand their preferences and perceptions related to 360-degree video content.

Table 2. QoE questions for 360-degree videos across the two groups. G_1 -NonOlfactory participants viewed AV content only; G_2 -Olfactory viewed the same content with added olfactory effects. Items address enjoyment, artifact perception, tolerance; for G_2 -Olfactory, they address olfactory annoyance, distraction, and intensity. Ratings were collected immediately after each stimulus. Ratings were made on a five-point Likert scale.

| Group | QoE ID | Question |
|--|--------|---|
| G_1 -NonOlfactory and G_2 -Olfactory | QoE-Q1 | I enjoyed watching the 360-degree video. |
| | QoE-Q2 | I noticed artifacts in the 360-degree video. |
| | QoE-Q3 | I do not mind artifacts in the 360-degree video. |
| | QoE-Q4 | Rate the overall quality of the 360-degree video. |
| G_2 -Olfactory | QoE-Q5 | The olfactory effects are annoying. |
| | QoE-Q6 | The olfactory effects are distracting. |
| | QoE-Q7 | Rate the intensity of the olfactory effects. |

The statements cover different aspects of the user's viewing experience. For Q1–Q3 and Q5–Q6, participants were asked to provide their responses on a five-point Likert scale, ranging from (1) *Strongly disagree* to (5) *Strongly Agree*. Q4 and Q7 were similarly coded, where each participant gave an overall rating on a five-point scale ranging from (1) *Very Bad* to (5) *Very Good*.

3.6. Experimental Design

The protocol of this research follows four parts of the experiment session, which are listed below, based on the experiment of [33,34]:

1. Initial GSR and ECG measurements were performed while participants were instructed to relax and clear their minds. This phase aimed to capture baseline biosignal readings during the brain's resting state for a subsequent comparison with data collected in later experimental stages.
2. The participants were then exposed to a series of videos, each approximately 1 min in duration. Table 1 details the specific sequence of videos. This video sequence

was structured to guarantee that an equivalent number of participants evaluated each video.

3. At the end of each video, participants were asked to remove the HMD and complete the QoE questionnaire (Table 2).
4. Subsequently, the participants were instructed to wear the HMD and prepare for the following video presentation.

3.7. Data Pre-Processing and Inferential Strategy

Physiological signals (ECG/HR and GSR) were pre-processed with a deliberately lightweight and transparent procedure. Importantly, no formal filtering (e.g., band-pass or low-pass) or automatic artifact rejection was applied to the signals. Instead, raw traces were visually inspected for gross motion artifacts or missing segments, and no trials were excluded as a result of this inspection. Therefore, the analyses are based on the complete data set without segment-level rejection. This decision simplifies reproducibility and avoids introducing filtering choices that might affect temporal dynamics; however, it also means that small-amplitude artifacts remain in the data. We acknowledge this as a limitation, since unfiltered noise can slightly inflate within-condition variability and dampen sensitivity to subtle effects. However, the main patterns we report proved robust under this inclusive approach, and no results hinge on selective trial exclusion.

For each participant, we computed a resting baseline (mean over the initial 2 min rest period) and subtracted this baseline from all subsequent trial recordings. The trial data were then clipped to the stimulus window and aggregated at the trial level by computing the mean of the baseline-corrected samples. To allow comparability between participants, these aggregated trial values were normalized within each subject using min–max scaling before group-level modeling.

We note a terminology clarification: although the Apparatus Subsection referred to heart rate variability (HRV), the actual dependent measure reported here is mean heart rate (HR) per trial. HR was selected because it provides a stable summary measure under our trial structure and because HRV requires longer artifact-free intervals than those available in these datasets. To avoid ambiguity, all results and tables were harmonized to consistently refer to HR (rather than HRV).

As detailed in the following subsections, statistical inference used linear mixed-effect models (LMMs) implemented with `statistics models`. Each dependent variable (`norm_HR`, `norm_GSR`) was modeled with fixed effects of group (olfactory vs. non-olfactory), skew (categorical: LS, MS, HS), the three videos, and their interactions, with a random intercept for the participant. The fit of the model was assessed using the AIC/BIC and Wald z tests for fixed effects. Where residual diagnostics suggested borderline normality, bootstrap 95% confidence intervals for key contrasts and Hedges' g were additionally reported. Pairwise post hoc contrasts were calculated as paired t tests between participants (within-subject contrasts), with Bonferroni correction as the primary multiple comparison adjustment and Benjamini–Hochberg as the secondary sensitivity check. Random-slope variants (e.g., random slopes for skew) were attempted where feasible; model convergence and potential overfitting were explicitly monitored and reported.

4. Results

1. High skew (HS) corresponds to AV skews of -5 s and $+5$ s.
2. Medium skew (MS) corresponds to AV skews of -3 s and $+3$ s.
3. Low skew (LS) corresponds to AV skews of $S - 1$ s and $S + 1$ s.

The respective pairs of skew values were applied to the three videos used in this study. The average values reported hereafter are the results of the mean of the evaluations of the

three videos individually. This was performed to allow us to visualize the influence of the skew intensity instead of observing only the skew itself in isolation.

4.1. *G₁-NonOlfactory—Skews Non-Olfactory*

In analyzing *G₁-NonOlfactory*, we investigated how AV skews interfered with the participants' perception, QoE, and UX experiences.

4.1.1. By Skews

Table 3 delineates the QoE results associated with various AV synchronization discrepancies (skews).

Table 3. *G₁-NonOlfactory* results (without olfactory stimulation): QoE statistics. Values are reported as mean (median, standard deviation). LS—low skew; MS—medium skew; HS—high skew.

| Metric | LS Mean (Med, SD) | MS Mean (Med, SD) | HS Mean (Med, SD) |
|---|-------------------|-------------------|-------------------|
| Quality of Experience Questions (Range: 1 to 5) | | | |
| QoE-Q1 | 4.10 (4.00, 0.79) | 4.20 (4.00, 0.74) | 3.63 (4.00, 1.18) |
| QoE-Q2 | 3.33 (4.00, 1.31) | 3.50 (4.00, 1.33) | 3.67 (4.00, 1.11) |
| QoE-Q3 | 2.87 (3.00, 1.39) | 2.73 (3.00, 1.41) | 3.00 (3.00, 1.46) |
| QoE-Q4 | 3.17 (3.00, 1.10) | 3.20 (3.00, 0.90) | 3.13 (3.00, 1.14) |

Exploring QoE-Q1 regarding enjoyment under varying levels of AV synchronization, we observe differences between conditions. In LS, the average enjoyment score was 4.10 (SD = 0.79), suggesting a generally positive experience with moderate consistency between participants. In MS, enjoyment peaked with the highest score of 4.20 (SD = 0.74), indicating that this condition provided the most enjoyable experience with minimal variability. However, in the HS condition, enjoyment decreased significantly to 3.63 (SD = 1.18), reflecting greater diversity in audience reactions, which was likely caused by the pronounced effects of AV skews.

Analyzing QoE-Q2, which measures the noticeability of artifacts, LS scored 3.33 (SD = 1.31), showing moderately low noticeability with significant variability, hinting at varied perceptions among participants. In the MS condition, the noticeability of the artifact decreased slightly to 3.50 (SD = 1.33), maintaining high variability and suggesting a mixed viewer experience. Surprisingly, in HS, the score improved to 3.67 (SD = 1.11), reflecting slightly less noticeable artifacts, but still significant variability.

For QoE-Q3, which explores tolerance for artifacts, the results reveal an intriguing trend. LS scored the lowest at 2.87 (SD = 1.39), indicating limited tolerance to artifacts and substantial variability in feedback. MS showed a marginal improvement with a score of 2.73 (SD = 1.41); however, this condition remained the most tolerated overall. In contrast, HS scored the highest at 3.00 (SD = 1.46), indicating a relatively better tolerance to artifacts, although the variability remained notable. This suggests that as the prominence of the artifact increases with higher skew levels, the participants' expectations may adjust, resulting in slightly improved tolerance.

Examining QoE-Q4, which reflects perceived quality, LS received a moderate rating of 3.17 (SD = 1.10), indicating a balanced perception of quality with low variability. MS performed slightly better, with a score of 3.20 (SD = 0.90), indicating improved quality perception and greater consistency. However, HS performed slightly worse on the test with a score of 3.13 (SD = 1.14).

4.1.2. Physiological Response

The graphs below represent the aggregated and averaged physiological data across the three video formats: LD, MD and HD. The horizontal axis X-axis) denotes time in seconds, spanning from 0 s to 60 s, which corresponds to the duration of each video. The vertical axis Y-axis) displays the average normalized physiological responses (i.e., each data point plotted reflects the mean value derived from the three video formats), scaled between 0 and 1.

The physiological data collected (see Figures 2 and 3) indicate that low and medium levels of AV skew are associated with a reduction in HR, as evidenced by the negative slope of the data during the measurement period. In contrast, HS is correlated with an increase in HR. Concerning the collected GSR data, all groups exhibited a noticeable increase in their values, with a similar rate of change observed across each group.

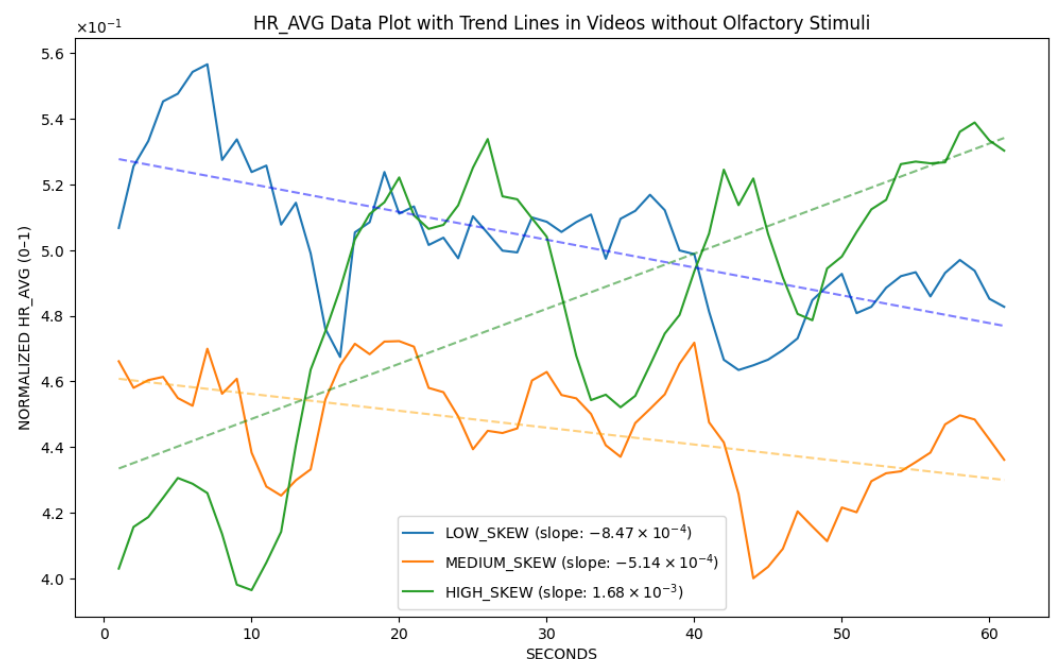


Figure 2. HR data collected from G_1 -NonOlfactory (participants without olfactory stimuli). Different colors represent the different AV skew levels. LS is shown with blue lines, MS with orange lines, and HS with green lines. The straight lines indicate the slope calculated using simple linear regression, indicating a slight increase for HS versus near-flat or marginal declines for LS and MS.

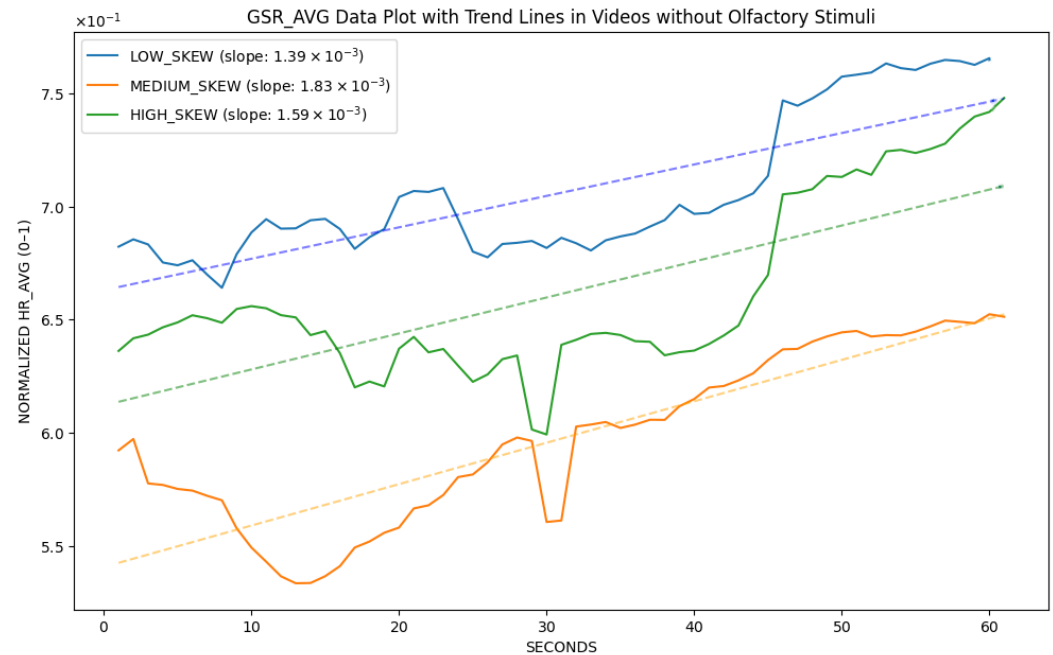


Figure 3. GSR data collected from G_1 -NonOlfactory (participants without olfactory stimuli). Different colors represent various AV skew levels. LS is shown with blue lines, MS with orange lines, and HS with green lines. The straight lines indicate the slope calculated using simple linear regression. Compared to G_2 -Olfactory (with olfactory stimuli), these slopes are steeper, which implies that videos alone without olfactory stimuli evoke stronger physiological responses, or, in this case, elevated level of stress induction.

4.2. G_2 -Olfactory—Skews and Olfactory

4.2.1. By Skew

In table 4, we examine how the results of the QoE evaluation vary by different levels of skewness with the presence of olfactory stimuli, indicating degrees of AV synchronization mismatches.

Table 4. G_2 -Olfactory results (participants with olfactory stimuli): QoE statistics. Values are reported as mean (median, standard deviation). LS—low skew; MS—medium skew; HS—high skew.

| Metric | LS Mean (Med, SD) | MS Mean (Med, SD) | HS Mean (Med, SD) |
|---|-------------------|-------------------|-------------------|
| Quality of Experience Questions (Range: 1 to 5) | | | |
| QoE-Q1 | 3.50 (4.00, 0.88) | 3.50 (4.00, 0.88) | 3.73 (4.00, 0.90) |
| QoE-Q2 | 3.43 (4.00, 1.16) | 3.65 (4.00, 1.04) | 3.60 (4.00, 1.12) |
| QoE-Q3 | 3.43 (3.40, 0.97) | 3.65 (3.60, 1.09) | 3.60 (3.60, 1.03) |
| QoE-Q4 | 2.67 (2.70, 0.98) | 2.58 (2.60, 1.05) | 2.69 (2.70, 1.05) |
| QoE-Q5 | 2.27 (2.00, 0.92) | 2.47 (2.20, 1.06) | 2.33 (2.00, 1.02) |
| QoE-Q6 | 2.44 (2.00, 0.95) | 2.75 (3.00, 0.96) | 2.46 (2.00, 0.94) |
| QoE-Q7 | 2.81 (3.00, 1.07) | 2.81 (3.00, 1.08) | 2.79 (3.00, 1.09) |

Beginning with enjoyment (QoE-Q1), the scores remained broadly stable across the various magnitudes of skew. The mean enjoyment was 3.50 (Med = 4.0, SD = 0.88) for both the low skeleton (LS) and medium skeleton (MS), and it modestly higher at 3.73 (Med = 4.0, SD = 0.90) for the high skeleton (HS). These values indicate a generally positive and consistent experience across conditions, with only small condition-wise changes in central tendency.

For artifact noticeability (QoE-Q2) and artifact tolerance (QoE-Q3), the olfactory group reported moderate noticeability and fairly high tolerance across skew levels. The notice-

ability means were 3.43 (LS, SD = 1.16), 3.65 (MS, SD = 1.04) and 3.60 (HS, SD = 1.12). The tolerance means followed a similar pattern: 3.43 (LS, SD = 0.97), 3.65 (MS, SD = 1.09) and 3.60 (HS, SD = 1.03). In general, participants perceived artifacts but remained moderately tolerant of them, with only small differences between the skew buckets.

Overall quality (QoE-Q4) remained relatively low to moderate and stable across skews: LS = 2.67 (Med = 2.7, SD = 0.98); MS = 2.58 (Med = 2.6, SD = 1.05); and HS = 2.69 (Med = 2.7, SD = 1.05). Thus, olfactory exposure did not produce a large upward shift in global video quality ratings, suggesting that content/production factors dominated these judgments.

Specific olfactory items (annoyance, distraction, intensity) generally show low-to-moderate responses. The scent annoyance (QoE-Q5) score averages were 2.27 (LS, SD = 0.92), 2.47 (MS, SD = 1.06), and 2.33 (HS, SD = 1.02), indicating only mild annoyance at most. The scent distraction (QoE-Q6) was also low to moderate with means of 2.44 (LS, SD = 0.95), 2.75 (MS, SD = 0.96), and 2.46 (HS, SD = 0.94). The perceived intensity of the scent (QoE-Q7) was moderate and consistent: 2.81 (LS, SD = 1.07), 2.81 (MS, SD = 1.08), and 2.79 (HS, SD = 1.09).

Taken together, these descriptive results show that exposure to olfactory scents produced only modest changes in self-reports: artifact noticeability and tolerance cluster around the mid-high range (3.4–3.7), while olfactory-specific complaints (annoyance, distraction) are low (2.3–2.8). Condition differences are small in relation to within-condition variability, so inferential tests are required before claiming reliable effects. (See Section 4.3 for mixed model analyses and post hoc contrasts.)

Our results show that viewer enjoyment and usability remain moderately stable across varying AV skews, with slight increases in scores for certain questions in higher skew groups. The perception of artifacts and memory structure showed a slight decrease in sensitivity to synchronization deviations, particularly in the olfactory stimuli scenario.

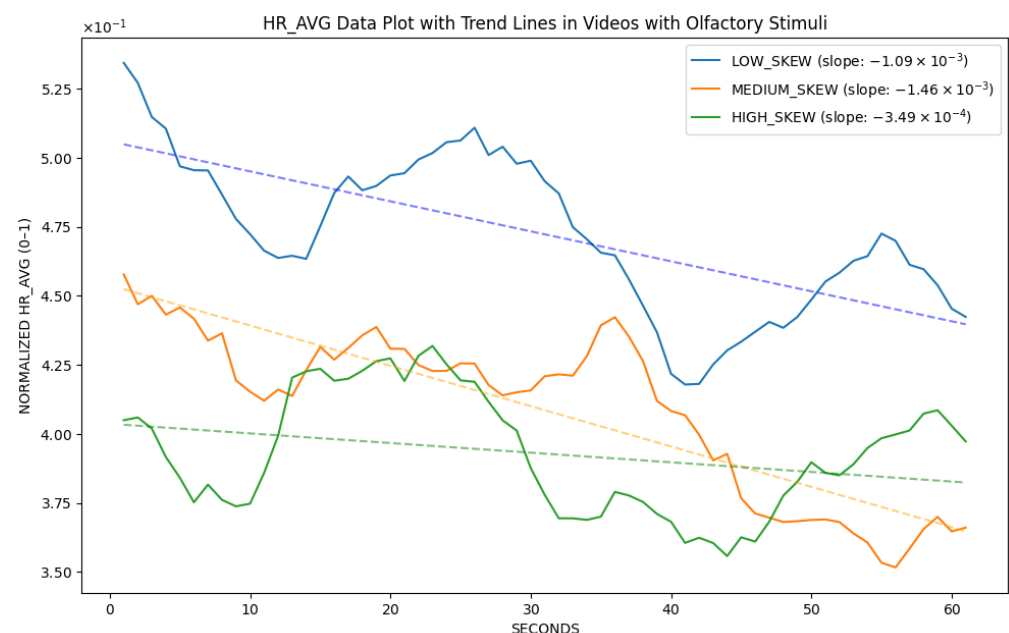


Figure 4. HR data collected from G₂-Olfactory (participants with olfactory stimuli). Different colors represent various AV skew levels. LS is shown with blue lines, MS with orange lines, and HS with green lines. The straight lines indicate the slope calculated using simple linear regression. Within this normalized scale, the downward trends are modest in all skew levels. Medium skew exhibits the most pronounced decrease in HR over time, high skew is comparatively stable, and low skew shows an intermediate trend. This suggests that the AV skew level may modulate the trajectory of heart-rate responses during olfactory-augmented video viewing, with MS producing the strongest downward drift.

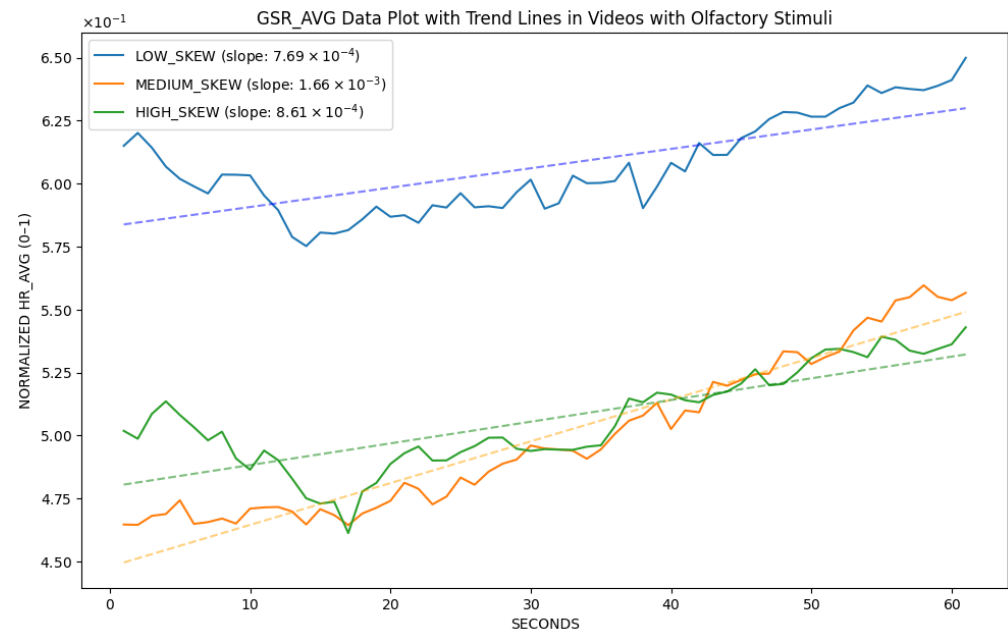


Figure 5. GSR data collected from G_2 -Olfactory (participants with olfactory stimuli). Different colors represent various AV skew levels. The straight lines indicate the slope calculated using simple linear regression. LS is shown with blue lines, MS with orange lines, and HS with green lines. The results imply that medium AV skew produced the strongest escalation in emotional arousal, as measured by skin conductance, while high and low skew had weaker but still noticeable effects.

4.2.2. Physiological Responses

Contrary to the findings of the previous group, all groups exhibited a reduction in heart rate in this scenario (see Figure 4), which involved olfactory stimuli. However, a high degree of AV skewness resulted in a decreased slope compared to low and medium degrees of AV skewness. The GSR data mirrored those of the preceding group (see Figure 5); all groups showed a rise in values, with a comparable rate of change between the two groups.

4.3. Mixed-Effect Analysis of HR and GSR on Olfactory Conditions

We analyzed normalized heart rate (*norm_hr*) and skin conductance (*norm_gsr*) metrics using linear mixed-effect models (LMM) with random intercepts for participants. Fixed effects included group (G_1 -Non-Olfactory vs. G_2 -Olfactory), skew (collapsed to LS = ± 1 s, MS = ± 3 s, HS = ± 5 s), video (1–3), and their interactions. The models were fitted to the participant-level condition means ($n = 126$ observations from 42 participants).

For *norm_hr*, the LMM returned no significant main effects of group, skew, or video and no significant higher-order interactions (all Wald $p > 0.18$). However, paired post hoc contrasts between subjects on the collapsed skew factor revealed that LS produced larger physiological responses than MS and HS. Specifically, LS vs. HS: $t(41) = 5.80$, $p < 0.001$, Hedges' $g \approx 1.09$, mean difference = 0.190 (95% bootstrap CI [0.126, 0.255], $n = 42$); LS vs. MS: $t(41) = 5.41$, $p < 0.001$, Hedges' $g \approx 0.91$, mean difference = 0.159 (95% bootstrap CI [0.104, 0.216], $n = 42$).

For *norm_gsr*, the LMM revealed a significant group \times video interaction, $\beta = -0.429$, 95% CI $[-0.839, -0.020]$, $p = 0.040$, indicating a reduced GSR in the Olfactory group specifically during the Fireworks videos. The post hoc contrasts on skew mirrored the HR pattern: LS vs. HS: $t(41) = 3.43$, $p = 0.004$ (corrected for Bonferroni), Hedges' $g \approx 0.77$, mean difference = 0.223 (95% bootstrap CI [0.092, 0.350]); LS vs. MS: $t(41) = 2.77$, $p = 0.025$, $g \approx 0.62$, mean difference = 0.177 (95% bootstrap CI [0.044, 0.299]).

Table 5 lists the paired post hoc contrasts (LS vs. MS, LS vs. HS) with bootstrap CI and Hedges' g . Tables 6 and 7 report the LMM fixed-effect estimates, standard errors, Wald z statistics, p values, and 95% CIs. All pairwise tests were Bonferroni-corrected.

Table 5. Post hoc pairwise contrasts (collapsed skew: LS, MS, HS). Mean differences are LS—other condition. Bootstrap 95% CIs, Bonferroni-corrected p , and Hedges' g are shown.

| DV | Contrast | Mean Diff | 95% Bootstrap CI | t (df) | p_{bonf} | Hedges' g |
|----------|----------|-----------|------------------|------------|-------------------|-------------|
| norm_HR | LS—HS | 0.1900 | [0.1262, 0.2550] | 5.803 (41) | <0.001 | 1.09 |
| | LS—MS | 0.1588 | [0.1038, 0.2161] | 5.414 (41) | <0.001 | 0.91 |
| norm_GSR | LS—HS | 0.2232 | [0.0920, 0.3495] | 3.428 (41) | 0.004 | 0.77 |
| | LS—MS | 0.1766 | [0.0437, 0.2986] | 2.767 (41) | 0.025 | 0.62 |

Table 6. Fixed effects from the mixed linear model for normalized heart rate (*norm_hr*). Model: *norm_hr* ~ Group * Skew * Video + (1 | participant). Coef = estimate; SE = standard error; z = Wald z ; CI = 95% confidence interval.

| Term | Coef | SE | z | p | 95% CI |
|-------------------------------------|--------|-------|--------|--------|-----------------|
| Intercept | 0.493 | 0.081 | 6.080 | <0.001 | [0.334, 0.651] |
| Group (Olfactory) | −0.130 | 0.096 | −1.357 | 0.175 | [−0.318, 0.058] |
| Skew (LS vs. ref) | 0.128 | 0.115 | 1.120 | 0.263 | [−0.096, 0.353] |
| Skew (MS vs. ref) | −0.045 | 0.115 | −0.389 | 0.697 | [−0.269, 0.180] |
| Video (Firefighter vs. Coffee Shop) | −0.141 | 0.115 | −1.228 | 0.219 | [−0.365, 0.084] |
| Video (Kung Fu vs. Coffee Shop) | −0.152 | 0.115 | −1.326 | 0.185 | [−0.377, 0.073] |
| Group × Video (Olf × 2) | 0.145 | 0.136 | 1.072 | 0.284 | [−0.120, 0.411] |

Model fit: AIC = −79.049, BIC = −22.324, LogLik = 59.525.

Table 7. Fixed effects from the mixed linear model for normalized skin conductance (*norm_gsr*). Model: *norm_gsr* ~ Group * Skew * Video + (1 | participant). Bold indicates $p < 0.05$.

| Term | Coef | SE | z | p | 95% CI |
|---------------------------------|---------------|--------------|---------------|--------------|-------------------------|
| Intercept | 0.522 | 0.125 | 4.179 | <0.001 | [0.277, 0.766] |
| Group (Olfactory) | −0.032 | 0.148 | −0.220 | 0.826 | [−0.322, 0.257] |
| Skew (LS vs. ref) | 0.237 | 0.177 | 1.344 | 0.179 | [−0.109, 0.583] |
| Skew (MS vs. ref) | 0.018 | 0.177 | 0.104 | 0.917 | [−0.328, 0.364] |
| Video (Kung Fu vs. Coffee Shop) | 0.269 | 0.177 | 1.524 | 0.128 | [−0.077, 0.615] |
| Video (Kung Fu vs. Coffee Shop) | −0.064 | 0.177 | −0.363 | 0.717 | [−0.410, 0.282] |
| Group × Video (Olf × 2) | −0.429 | 0.209 | −2.055 | 0.040 | [−0.839, −0.020] |

Model fit: AIC = 45.330, BIC = 102.056, LogLik = −2.665.

We inspect the standard diagnostics for both LMMs in Table 8. Residual normality (Shapiro–Wilk) was borderline for *norm_hr* ($p = 0.050$) and indicated a mild departure from normality for *norm_gsr* ($p = 0.032$). The Breusch–Pagan tests did not indicate strong heteroskedasticity (HR $p = 0.072$, GSR $p = 0.136$). Given these findings, we present bootstrap 95% confidence intervals for the paired contrasts and report effect sizes (Hedges' g). Attempts to fit random slope models (e.g., random slope for Skew) were made; however, these models did not converge robustly given the current data and factor structure, so the reported models use random intercepts only.

Table 8. Model fit and diagnostic statistics for mixed linear models (random intercepts for participants).

| Model | Observations | Groups (n) | AIC | BIC |
|---------------------------------|---------------------------|------------|---------|---------|
| norm_HR ~ Group*Skew*Video | 126 | 42 | −79.049 | −22.324 |
| norm_GSR ~ Group*Skew*Video | 126 | 42 | 45.330 | 102.056 |
| Residual Shapiro <i>p</i> (HR) | 0.050 (borderline normal) | | | |
| Residual Shapiro <i>p</i> (GSR) | 0.032 (mild departure) | | | |
| Breusch–Pagan <i>p</i> (HR) | 0.072 | | | |
| Breusch–Pagan <i>p</i> (GSR) | 0.136 | | | |

In summary, physiological data show a pattern of increased responses for low-skew (LS) compared to MS and HS in both HR and GSR. At the same time, olfactory manipulation produced a content-specific modulation of GSR (Olfactory \times Fireworks video). Interpretive caveats and methodological limitations (e.g., stationary scent delivery, 360-degree masking effects, absence of spatial audio) are discussed in Section 5.

5. Discussion

5.1. Subjective User Evaluation Results

Figure 6 presents the group means and standard deviations for the main QoE items. Enjoyment (QoE-Q1) remained relatively stable between skew levels in both groups. G_1 -NonOlfactory showed a peak under MS conditions (mean = 4.20, SD = 0.74) and a decrease under HS (mean = 3.63, SD = 1.18). In contrast, G_2 -Olfactory reported enjoyment scores that were more consistent with skew levels (mean range 3.50 to 3.73), suggesting that olfactory signals reduced the variability in self-reported enjoyment.

For artifact noticeability (QoE-Q2), the group means were comparable. G_1 -NonOlfactory ranged from 3.34 to 3.77, while G_2 -Olfactory ranged from 3.43 to 3.70. However, in artifact tolerance (QoE-Q3), G_2 -Olfactory consistently reported higher tolerance (means 3.53 to 3.70) compared to G_1 -NonOlfactory (means 2.73 to 2.87). For overall quality (QoE-Q4), the scores were relatively stable across the skew levels but systematically higher for G_1 -NonOlfactory (≈ 3.20) than for G_2 -Olfactory (≈ 2.70).

To complement these descriptive results, linear mixed-effect models (random intercepts for participants) were fitted to each of the four primary QoE measures.

For enjoyment (QoE-Q1), the models revealed a group \times skew interaction (Table 9). In the G_1 -NonOlfactory, enjoyment was higher in LS and MS than in HS; however, this advantage was significantly attenuated in the olfactory group. This can indicate that while enjoyment degraded with increasing skew in the non-olfactory group, the presence of scent in the olfactory group stabilized ratings across skew levels. The olfactory group reported significantly higher artifact tolerance (QoE-Q3) ($\beta = 0.75$, $p = 0.001$) while lower overall quality (QoE-Q4) ($\beta = -0.76$, $p < 0.001$).

Taken together, these findings imply that olfactory stimuli can act as a buffer: certain scents mitigated some adverse impacts of skews on enjoyment, leading to more consistent ratings across various skew levels. However, this did not result in an improvement in overall inherent quality scores.

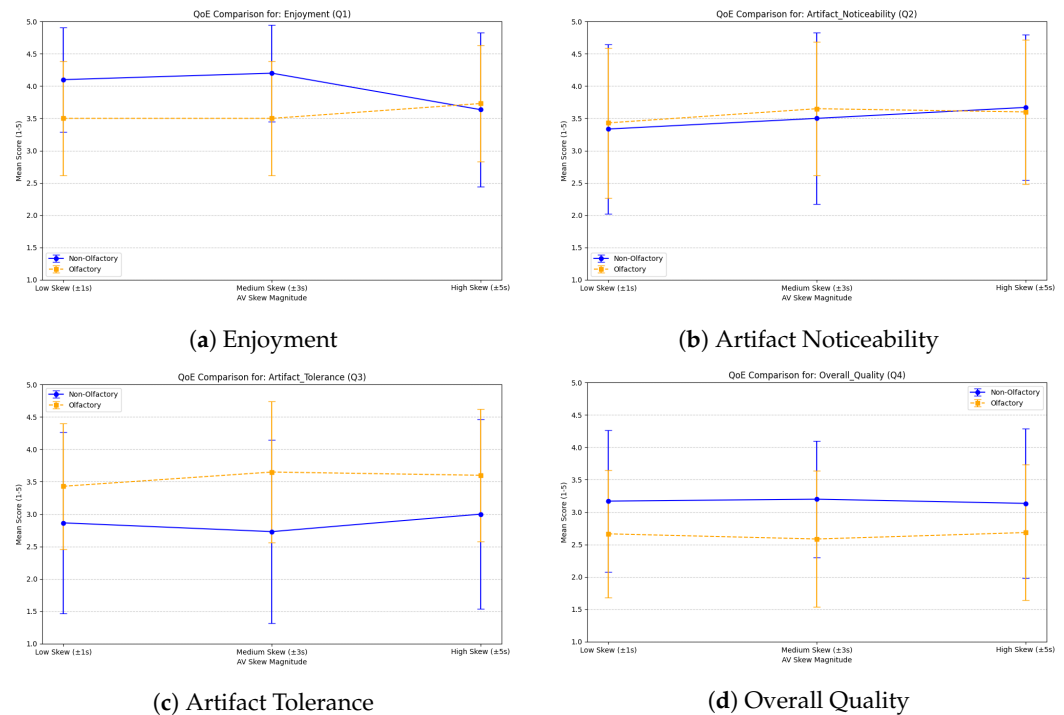


Figure 6. Subjective comparison of QoE metrics for the two groups (G_1 -NonOlfactory/blue line and G_2 -Olfactory/orange line). Each panel shows mean ratings (error bars = SD) for the Low (LS), Medium (MS) and high (HS) AV skew buckets. (a) QoE-Q1 (enjoyment); (b) QoE-Q2 (artifact noticeability); (c) QoE-Q3 (artifact tolerance); and (d) QoE-Q4 (overall quality).

Table 9. Mixed linear model—fixed effects for enjoyment (QoE-Q1). Reference levels: Group = non-olfactory, Skew = HS.

| Term | Coef | SE | z | p | 95% CI |
|-------------------------|--------|-------|-------|--------|------------------|
| Intercept (G_1 , HS) | 3.629 | 0.172 | 21.12 | <0.001 | [3.292, 3.966] |
| Group (Olfactory) | 0.054 | 0.163 | 0.33 | 0.739 | [−0.266, 0.374] |
| Skew (LS vs. HS) | 0.467 | 0.181 | 2.58 | 0.010 | [0.113, 0.821] |
| Skew (MS vs. HS) | 0.567 | 0.181 | 3.14 | 0.002 | [0.213, 0.921] |
| Group \times LS | −0.650 | 0.221 | −2.94 | 0.003 | [−1.084, −0.216] |
| Group \times MS | −0.717 | 0.221 | −3.24 | 0.001 | [−1.150, −0.283] |

5.2. Objective User Evaluation Results

The comparison of QoE scores across skew levels presented in Figure 6a–d for both non-olfactory and olfactory conditions reveals some patterns in participants' responses to different aspects of the 360-degree video experience.

Regarding QoE-Q1, for non-olfactory conditions, the enjoyment of the participants in G_1 remains relatively stable across skew levels, with a slight decline in the desynchronization HS levels. However, the median scores are still around 4.0. In olfactory conditions, enjoyment scores also remain fairly stable, though they show a slightly lower mean (around 3.50) compared to non-olfactory conditions at most skew levels. This suggests that the addition of olfactory elements did not significantly enhance enjoyment.

Participants in both groups reported moderate artifact awareness, with scores slightly increasing from low to high skews, but maintaining overall consistency between all groups.

The tolerance for artifacts remained consistent in non-olfactory and olfactory conditions. In the olfactory condition, the participants demonstrated a slightly higher tolerance for artifacts at all skew levels, with scores remaining around 3.50. This could suggest that the scent might have acted as a compensating factor, making artifacts less distracting or

annoying. However, this trend would benefit from further exploration considering the results of the previous questions.

An important finding from our reliability analysis in Section 4.3 is the pronounced susceptibility of the Fireworks video (MD) to the olfactory effect. We attribute this to the inherent characteristics of the video, which featured fireworks that exploded directly in front of the viewer. This visual alignment naturally directed the user's attention toward the olfactory device positioned in front of them, enhancing the sensory impact.

Finally, for overall quality, the non-olfactory scores for general quality were moderate, averaging around 3.00, and did not vary significantly between skew levels. This suggests that synchronization problems had only a limited impact on perceived overall quality in both conditions.

5.3. Limitations and Interpretation

This study investigated whether olfactory signals modulate subjective tolerance and physiological responses to audio–video desynchronization in 360-degree videos. The combined evidence indicates a nuanced effect: olfactory stimulation did not uniformly increase QoE ratings or perceived video quality but altered how participants responded to temporal AV mismatch. Convergent physiological results showed larger responses for low-magnitude skews (LS) relative to medium and high skews (MS/HS) in both HR and GSR, and a content-specific olfactory \times video interaction for GSR (see Table 5 and Section 4.3). Subjectively, enjoyment was more stable across skew levels in the olfactory group, while the non-olfactory group exhibited larger skew-dependent swings (LS/MS > HS).

Several methodological and contextual factors limit the generalization. First, the olfactory device was stationary and produced relatively weak percepts for many participants; this likely attenuated potential olfactory effects and reduces ecological validity. Future work should test directional or head-mounted scent delivery and stronger, well-controlled concentrations. Second, the 360-degree format (and the absence of spatialized audio) allows for a visual exploration that can mask audio desynchronization and shift attention away from temporal cues; integrating 3D audio and counterbalancing exploration strategies would improve sensitivity to AV skew. Third, some participants reported uncertainty about the experimental task (demand/attention effects), which may depress self-reported detection rates; physiological measures are an important complementary source of evidence.

Despite modest effect sizes on some subjective measures, the study offers three practical contributions: (i) a reproducible biosignal pre-processing and aggregation pipeline, (ii) principled mixed-effect inference that respects repeated measures, and (iii) bootstrap CIs and Hedges' g for key contrasts. All of these increase robustness and replicability for future QoE work. Together, the results suggest that olfactory stimuli tend to stabilize subjective responses to AV skew (reducing condition-to-condition variability) while selectively interacting with content to modulate arousal. This is scientifically valuable because it shifts the narrative from 'olfaction always improves experience' to a more precise claim: olfaction reshapes how temporal mismatches are processed, which can be exploited when designing resilient multisensory experiences.

We recommend follow-up studies that (i) employ stronger/directional scent delivery and spatial audio, (ii) examine content-specific interactions with larger samples (to allow us to obtain random slope models and carry out more complex skew \times content tests), and (iii) combine signed and magnitude-based skew predictors as sensitivity analyses. These steps will clarify when olfactory augmentation enhances tolerance and when it increases sensitivity to technical artifacts.

6. Conclusions

This study examined whether olfactory signals modulate subjective tolerance and physiological responses to audio–video desynchronization in 360-degree videos. Combining questionnaire data with HR and GSR measures, we find a consistent and interpretable pattern: low-magnitude skews (LS; ± 1 s) elicit physiological responses greater than medium and high skews (MS/HS), and olfactory stimulation alters how participants respond to skew rather than uniformly increasing the reported quality. In particular, olfactory exposure was associated with a reduced condition-to-condition variability in enjoyment while producing a content-specific olfactory \times video effect on GSR (see Section 4.3 and Table 5).

Methodologically, the paper contributes a reproducible biosignal pre-processing pipeline and a principled inferential workflow (linear mixed-effect models with participant-level random intercepts, bootstrap confidence intervals, and Hedges' g for key contrasts). These analytic choices improve robustness by combining subjective and physiological QoE measures in immersive media research.

The study has important practical limitations that likely attenuated effect sizes: the scent delivery was stationary (reducing perceived intensity), the audiovisual content lacked 3D spatial audio, and the 360-degree format allowed participants to visually explore their surroundings away from audio events. These factors constrain generalizability and suggest that stronger, directional scent delivery and spatial audio are promising directions for clarifying when olfaction meaningfully enhances the tolerance of AV mismatches.

A key limitation of the present work is that due to sample size constraints, we summarized audio–video (AV) asynchronies into magnitude bins (low, medium, high) rather than modeling lead and lag directions as separate factors. Although this approach increases statistical power, it necessarily combines theoretically non-equivalent conditions and may obscure direction-specific effects. We acknowledge that this choice limits the verifiability and interpretability of lead–lag asymmetries. Future studies should explore this direction as an explicit factor or report exploratory lead–lag contrasts. However, our primary objective in this aggregation was to demonstrate that large AV skews—regardless of direction—were attenuated in their impact on enjoyment by olfactory stimulation. The observed buffering effect, alongside content-specific physiological modulation, suggests that scent can reshape perceptual tolerance to temporal mismatch, even if the precise directional dynamics remains to be fully characterized.

However, overall results should be read as a nuanced contribution: olfactory cues do not simply raise QoE across the board but can reshape perceptual tolerance to temporal mismatches and interact selectively with content to modulate arousal. Future work should (a) adopt head-mounted or directional scent systems and spatial audio, (b) increase sample sizes to enable random-slope models and finer-grained signed-skew analyses, and (c) evaluate multisensory combinations (olfaction + haptics + spatial audio) to determine how sensory redundancy can be engineered to produce more resilient, immersive experiences.

Author Contributions: Conceptualization, A.C.d.S.; formal analysis, A.C.d.S.; funding acquisition, R.R. and C.A.S.S.; investigation, A.C.d.S.; methodology, A.C.d.S.; project administration, C.A.S.S.; software, A.C.d.S.; supervision, R.R., G.G. and C.A.S.S.; validation, A.C.d.S. and R.R.; visualization, A.C.d.S.; writing—original draft preparation, A.C.d.S., G.G. and C.A.S.S.; writing—review and editing, R.R., F.S., A.C., G.G. and C.A.S.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Coordination for the Improvement of Higher Education Personnel (CAPES, Brazil) under Finance Codes 88887.570688/2020-00 and 88881.689984/2022-01; the National Council for Scientific and Technological Development (CNPQ, Brazil) under Finance Code 307718/2020-4; the Fundação de Amparo à Pesquisa e Inovação do Espírito Santo (FAPES, Brazil) under Finance Code 2021-GL60J; and the Research Council of Finland, grant number 355575.

Institutional Review Board Statement: The study was approved by the Ethics Committee of Brunel University London (protocol code: 40020-LR-Oct/2022-41826-3; date of approval: October 2022).

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

Acknowledgments: The authors would like to thank all study participants for their time and valuable feedback. The authors also thank the staff of the Federal University of Espirito Santo, Tampere University, and Brunel University London for their assistance with the installation of the equipment used and data collection. During the preparation of this manuscript, Overleaf Writefull was used for grammar checking and refinement, and ChatGPT (OpenAI) was used to help with table formatting.

Conflicts of Interest: The authors declare no conflicts of interest. The sponsors of this study had no role in its design; in the collection, analysis or interpretation of the data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Korbar, B.; Tran, D.; Torresani, L. Cooperative Learning of Audio and Video Models from Self-Supervised Synchronization. *arXiv* **2018**, arXiv:1807.00230.
2. Ehley; Furht; Ilyas. Evaluation of multimedia synchronization techniques. In Proceedings of the IEEE International Conference on Multimedia Computing and Systems, Boston, MA, USA, 14–19 May 1994; pp. 514–519. <https://doi.org/10.1109/MMCS.1994.292498>.
3. Heyse, J.; Carlier, S.; Verhelst, E.; Vander Linden, C.; De Backere, F.; De Turck, F. From Patient to Musician: A Multi-Sensory Virtual Reality Rehabilitation Tool for Spatial Neglect. *Appl. Sci.* **2022**, *12*, 1242. <https://doi.org/10.3390/app12031242>.
4. Silveira, A.; Santos, C. Ongoing Challenges of Evaluating Mulsemedia QoE. In Proceedings of the 2nd Workshop on Multisensory Experiences—SensoryX’22, Porto Alegre, Brazil, 22 June 2022. <https://doi.org/10.5753/sensoryx.2022.20005>.
5. Ghinea, G.; Ademoye, O.A. Perceived synchronization of olfactory multimedia. *IEEE Trans. Syst. Man Cybern.-Part A Syst. Hum.* **2010**, *40*, 657–663.
6. Murray, N.; Qiao, Y.; Lee, B.; Karunakar, A.K.; Muntean, G.M. Subjective Evaluation of Olfactory and Visual Media Synchronization. In Proceedings of the MMSys ’13: 4th ACM Multimedia Systems Conference, New York, NY, USA, 28 February–1 March 2013; pp. 162–171. <https://doi.org/10.1145/2483977.2483999>.
7. da Silveira, A.C.; Spyridonis, F.; Raisamo, R.; Covaci, A.; Ghinea, G.; Santos, C.A.S. On Perceived AV Synchronization in 360° Multimedia. *IEEE MultiMed.* **2024**, *31*, 7–17. <https://doi.org/10.1109/MMUL.2024.3415116>.
8. Saleme, E.B.; Covaci, A.; Mesfin, G.; Santos, C.A.; Ghinea, G. Mulsemedia DIY: A survey of devices and a tutorial for building your own mulsemedia environment. *ACM Comput. Surv.* **2019**, *52*, 1–29.
9. Javerliat, C.; Elst, P.P.; Saive, A.L.; Lavoué, G.; Baert, P. Nebula: An Affordable Open-Source and Autonomous Olfactory Display for VR Headsets. In Proceedings of the VRST ’22: 28th ACM Symposium on Virtual Reality Software and Technology, New York, NY, USA, 29 November–1 December 2022. <https://doi.org/10.1145/3562939.3565617>.
10. Yuan, Z.; Chen, S.; Ghinea, G.; Muntean, G.M. User quality of experience of mulsemedia applications. *ACM Trans. Multimed. Comput. Commun. Appl.* **2014**, *11*, 1–19.
11. Covaci, A.; Trestian, R.; Saleme, E.B.; Comsa, I.S.; Assres, G.; Santos, C.A.; Ghinea, G. 360 Mulsemedia: A way to improve subjective QoE in 360 videos. In Proceedings of the 27th ACM International Conference on Multimedia, Nice, France, 21–25 October 2019; pp. 2378–2386.
12. Akhtar, Z.; Siddique, K.; Rattani, A.; Lutfi, S.L.; Falk, T.H. Why is Multimedia Quality of Experience Assessment a Challenging Problem? *IEEE Access* **2019**, *7*, 117897–117915. <https://doi.org/10.1109/ACCESS.2019.2936470>.
13. Amsellem, S.; Höchenberger, R.; Ohla, K. Visual–Olfactory Interactions: Bimodal Facilitation and Impact on the Subjective Experience. *Chem. Senses* **2018**, *43*, 329–339. <https://doi.org/10.1093/chemse/bjy018>.
14. Ghinea, G.; Ademoye, O.A. Olfaction-enhanced multimedia: Perspectives and challenges. *Multimed. Tools Appl.* **2011**, *55*, 601–626.
15. Iriarte, A.A.; Erle, G.L.; Etxabe, M.M. Evaluating user experience with physiological monitoring: A systematic literature review. *Dyna New Technol.* **2021**, *8*, 21.
16. 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. <https://doi.org/10.1109/TAFFC.2019.2927337>.
17. McKee, M.G. Biofeedback: An overview in the context of heart-brain medicine. *Cleveland. Clin. J. Med.* **2008**, *75*, S31–S31. https://doi.org/10.3949/ccjm.75.suppl_2.s31.
18. Mateos-García, N.; Gil-González, A.B.; Luis-Reboredo, A.; Pérez-Lancho, B. Driver stress detection from physiological signals by virtual reality simulator. *Electronics* **2023**, *12*, 2179.

19. Petrescu, L.; Petrescu, C.; Mitruț, O.; Moise, G.; Moldoveanu, A.; Moldoveanu, F.; Leordeanu, M. Integrating biosignals measurement in virtual reality environments for anxiety detection. *Sensors* **2020**, *20*, 7088.
20. Ohata, M.; Tanaka, T. Estimation of Stress Level Based on Biosignals in Response to Emotional Stimuli. In Proceedings of the 2022 IEEE 4th Global Conference on Life Sciences and Technologies (LifeTech), Osaka, Japan, 7–9 March 2022; pp. 580–581. <https://doi.org/10.1109/LifeTech53646.2022.9754934>.
21. Babaei, E.; Tag, B.; Dingler, T.; Velloso, E. A Critique of Electrodermal Activity Practices at CHI. In Proceedings of the Proceedings of the CHI '21: 2021 CHI Conference on Human Factors in Computing Systems, Online, 8–13 May 2021. <https://doi.org/10.1145/3411764.3445370>.
22. Dillen, N.; Ilievski, M.; Law, E.; Nacke, L.E.; Czarnecki, K.; Schneider, O. Keep Calm and Ride Along: Passenger Comfort and Anxiety as Physiological Responses to Autonomous Driving Styles. In Proceedings of the Proceedings of the CHI '20: 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 25–30 April 2020; pp. 1–13. <https://doi.org/10.1145/3313831.3376247>.
23. Kan John, P.; Zhu, X.; Gedeon, T.; Zhu, W. Evaluating Human Impressions of an Initiative-taking Robot. In Proceedings of the CHI EA '22: 2022 CHI Conference on Human Factors in Computing Systems, New Orleans, LA, USA, 29 April–5 May 2022; <https://doi.org/10.1145/3491101.3519633>.
24. Abdalkarim, D.; Di Luca, M.; Aves, P.; Maaroufi, M.; Yeo, S.H.; Miall, R.C.; Holland, P.; Galea, J.M. A methodological framework to assess the accuracy of virtual reality hand-tracking systems: A case study with the Meta Quest 2. *Behav. Res. Methods* **2024**, *56*, 1052–1063. <https://doi.org/10.3758/s13428-022-02051-8>.
25. Speer, K.E.; Semple, S.; Naumovski, N.; McKune, A.J. Measuring Heart Rate Variability Using Commercially Available Devices in Healthy Children: A Validity and Reliability Study. *Eur. J. Investig. Health Psychol. Educ.* **2020**, *10*, 390–404. <https://doi.org/10.3390/ejihpe10010029>.
26. Seo, J.; Laine, T.H.; Sohn, K.A. An Exploration of Machine Learning Methods for Robust Boredom Classification Using EEG and GSR Data. *Sensors* **2019**, *19*, 4561. <https://doi.org/10.3390/s19204561>.
27. Murray, N.; Ademoye, K.; Ghinea, G.; Muntean, G.M. A Tutorial for Olfaction-Based Multisensorial Media Application Design and Evaluation. *ACM Comput. Surv.* **2017**, *50*, 1–30. <https://doi.org/10.1145/3108243>.
28. Saleme, E.a.B.; Santos, C.A.S. PlaySEM: A Platform for Rendering MulSeMedia Compatible with MPEG-V. In Proceedings of the WebMedia '15: 21st Brazilian Symposium on Multimedia and the Web, Manaus, Brazil, 27–30 October 2015; pp. 145–148. <https://doi.org/10.1145/2820426.2820450>.
29. Niso, G.; Romero, E.; Moreau, J.T.; Araujo, A.; Krol, L.R. Wireless EEG: A survey of systems and studies. *NeuroImage* **2023**, *269*, 119774. <https://doi.org/https://doi.org/10.1016/j.neuroimage.2022.119774>.
30. Greene, J.M.; D'Oliveira, M. Learning to Use Statistical Tests in Psychology. In *Learning to Use Statistical Tests in Psychology*; The Open University: Milton Keynes, UK, 1982.
31. Calvo-Morata, A.; Freire, M.; Martínez-Ortiz, I.; Fernández-Manjón, B. Scoping review of bioelectrical signals uses in videogames for evaluation purposes. *IEEE Access* **2022**, *10*, 107703–107715.
32. Comşa, I.S.; Saleme, E.B.; Covaci, A.; Assres, G.M.; Trestian, R.; Santos, C.A.S.; Ghinea, G. Do I Smell Coffee? The Tale of a 360° Mulsemmedia Experience. *IEEE MultiMed.* **2020**, *27*, 27–36. <https://doi.org/10.1109/MMUL.2019.2954405>.
33. Kosiński, J.; Szklanny, K.; Wiczorkowska, A.; Wichrowski, M. An Analysis of Game-Related Emotions Using EMOTIV EPOC. In Proceedings of the 2018 Federated Conference on Computer Science and Information Systems (FedCSIS), Poznań, Poland, 9–12 September 2018; pp. 913–917.
34. Murray, N.; Qiao, Y.; Lee, B.; Muntean, G.M.; Karunakar, A.K. Age and gender influence on perceived olfactory & visual media synchronization. In Proceedings of the 2013 IEEE International Conference on Multimedia and Expo (ICME), San Jose, CA, USA, 15–19 July 2013; pp. 1–6. <https://doi.org/10.1109/ICME.2013.6607467>.

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.