



Audio Engineering Society

Convention Express Paper 10281

Presented at the AES 160th Convention
2026 May 28–30, Copenhagen, Denmark

This Express Paper was selected on the basis of a submitted synopsis that has been peer-reviewed by at least two qualified anonymous reviewers. The complete manuscript was not peer reviewed. This Express Paper has been reproduced from the author's advance manuscript without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. This paper is available in the AES E-Library (<http://www.aes.org/e-lib>), all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

EMORSION – Examining the Impact of Audio Parameters on Emotional Responses and Immersion in Film.

Nelly Garcia^{*1}, Ruby Crocker^{*1}, Bleiz M. Del Sette¹, Fabrizio Smeraldi¹, Charalampos Saitis¹, George Fazekas¹, and Joshua Reiss¹

^{*}*These authors contributed equally and share first authorship*

¹*Queen Mary University of London*

Correspondence should be addressed to Nelly Garcia (n.v.a.garcia-sihuay@qmul.ac.uk)

ABSTRACT

EMORSION is an exploratory proof-of-concept study examining how film audio design shapes audience emotion and immersion in cinema setting. Four film scenes were selected across the horror (2) and drama (2) genres, balanced between mainstream and independent productions. For each scene, multiple alternative audio mixes were created by systematically manipulating three core aspects of audio design, frequency (pitch), dynamics (loudness), and directionality (spatial placement). Three audience groups viewed the scenes, with each group exposed to one manipulated mix alongside a control mix for each scene. Audience responses were assessed through a triangulated multimodal framework combining self-reported emotion and immersion via a questionnaire, physiological measures including heart rate monitoring, and video-based motion tracking. The protocol successfully captured measurable, interpretable differences across audio conditions, indicating that even subtle changes in audio design can shape emotional perception and immersion. Unconventional mixes tended to produce greater variability in audience interpretation, while conventional immersive mixes were associated with stronger cross-audience agreement. These findings establish the feasibility of the EMORSION protocol and motivate larger-scale studies to characterise the role of specific audio parameters in shaping audience experience.

1 Introduction

Sound plays a central role in shaping emotional responses and immersion in film [1], enhancing the narrative and helping communicate the director's intended message [2]. While the role of music has been extensively studied, sound effects have received considerably

less empirical attention; Kock and Louven [3] showed that both music and sound design contribute significantly to perceived immersion and suspense, with their combination producing the strongest effect. Nevertheless, isolating the perceptual contribution of individual audio parameters remains methodologically challeng-

ing, particularly for sound effects, and existing work has rarely been conducted in ecologically valid contexts such as cinema environments, where the listening conditions differ substantially from typical laboratory setups [4]. To address this gap, we introduce EMORSION (Examining the Impact of Audio Parameters on Emotional Responses and Immersion in Film), an experimental protocol designed to investigate how audio augmentation in film mixes influences audience immersion, emotional interpretation, and affective response in a cinema setting, with participants acting as part of a live audience.

In this paper, 'immersion' refers to a viewer's state of deep mental involvement with an audiovisual experience, in which attention is strongly oriented toward the film and away from awareness of the surrounding physical environment [5]. Recent neuroscientific research highlights triangulation as a robust framework for studying perception, integrating physiological, behavioural, and self-report measures to capture multiple dimensions of human experience [6, 7]. A common quantification of immersion across these measures is *response similarity*: the closer participants' reactions are to those observed in real-world situations, the greater the inferred level of immersion [8]. In EMORSION, immersion is not treated as a directly measurable variable; instead we adopt a triangulated perspective, drawing on subjective, physiological, and behavioural indicators to characterise audience experience, acknowledging that immersive experience is inherently personal. This study serves as a proof of concept, demonstrating that triangulated measurement is feasible in a real cinema setting and that audio modifications produce measurable effects on audience experience.

2 Related Work

Sound design and music significantly shape audience perception, with growing research examining their role in immersive experiences [9, 10, 11]. When film scenes are ambiguous, viewers rely heavily on music to infer mood, narrative direction, and character traits [12]. While most studies focus on individual participants in controlled laboratory settings, recent work has begun to explore cinema-like environments and real theatres in order to capture collective audience experiences [13, 14].

Building on this body of work, immersion has been quantified through three complementary set of measures: subjective, physiological, and behavioural [8,

15]. Subjective measures rely primarily on self-report questionnaires, which remain a central method for assessing emotional expression, perception, and induction in music research [16]. Physiological measures capture autonomic responses linked to emotional engagement; electrocardiography (ECG), which measures the heart's electrical activity, is among the most widely used, and commercially available heart rate monitors such as the Polar H10 have demonstrated high accuracy and reliability in validation studies [17]. Furthermore, Rooney et al. [18] further linked decreases in heart rate to increased immersion, consistent with the idea that film immersion is an absorbed state characterised by calm focus. Behavioural measures have employed movement analysis [19], with prior film research associating stillness and interpersonal synchrony with immersion, whereas hand movements or coughing may signal audience response or disengagement [14]. Subtler cues such as nods or taps can also indicate musical immersion [7], and spatialised sound can evoke orienting responses, such as head or body movements toward a sound source, that indicate attentional orientation [20].

Beyond the choice of measures, the temporal design of immersive experiences also matters. Immersion does not scale linearly with content duration: prior work suggests that approximately seven minutes may be optimal for spatial immersive experiences, a phenomenon referred to as duration neglect [21]. This finding directly informed the stimulus design of EMORSION, which combines subjective, physiological, and behavioural measures with bounded clip durations to support meaningful comparison across audio conditions in a live cinema setting.

3 Methodology

We conducted three sessions at BLOC Studios, a cinema facility with a 36-speaker Dolby Atmos system and 4K projection.¹ In each session, participants viewed four film scenes, each presented twice — once as a control mix and once as an augmented mix (eight presentations total). Reflecting the triangulation approach, data were gathered across three modalities: physiologically, via a Polar H10 chest-strap sensor² for continuous heart rate monitoring; behaviourally, via two stationary cameras capturing movement proxies such

¹<https://www.qmul.ac.uk/bloc/>

²<https://www.polar.com/uk-en/sensors/h10-heart-rate-sensor>

Table 1: Selected film scenes and augmented mix assignment per session.

Film	Genre	Timeline	Duration	Target Emotion	Session 01	Session 02	Session 03
Ford vs Ferrari (FVF)	Adventure/Suspense	2h02–2h10	8 min	Tense, wonder	Dynamics	Directionality	Frequency
A Quiet Place (AQP)	Horror	5:00–10:00	5 min	Sad, tense	Frequency	Directionality	Dynamics
I Saw the TV Glow (ISTVG)	Horror	58:45–1h04	5 min	Intrigue, tense	Frequency	Dynamics	Directionality
Decision to Leave (DTL)	Suspense	1h35–1h46	10 min	Tense, intrigue	Directionality	Frequency	Dynamics

**Fig. 1:** Participant setup illustrating behavioural tracking (reflective wristbands), physiological monitoring (sensor strap), and self-report data collection via mobile device.

as stillness and fidgeting, supported by reflective wristbands for manual motion analysis; and subjectively, via a six-item self-report questionnaire³ completed on participants' mobile devices after each scene, measuring emotional response and perceived immersion. Each session opened with a 15-minute introduction covering study objectives, participant expectations, and informed consent for physiological and video data collection, and concluded with an open group discussion.

3.1 Participants

A total of 40 participants took part in the study (17 male, 22 female, and 1 non-binary). Session 1 included 13 participants (5 male, 8 female); Session 2 included 13 participants (4 male, 8 female, and one non-binary); Session 3 included 14 participants (9 male and 5 female). There was a diverse range of nationalities among participants, including English (21), European (9), Chinese (5), Mexican (2), and one participant each for Iran, India, Egypt, and Turkey.

3.2 Film Scene Selection and Audio Modifications

The four films scenes were selected following consultation with experts from Queen Mary School of Drama

and two professional sound engineers. Selection criteria required scenes with a strong balance of music and sound effects (including Foley and environmental sound), a stand-alone narrative, and a suitable emotional range for immersive viewing. Horror and drama were chosen to limit stylistic variability while retaining within-genre diversity. Both mainstream and independent productions were included, as independent films have demonstrated comparable or greater immersive and emotional impact than mainstream counterparts [22]. One independent and one mainstream scene were selected for each genre, and all scenes ran between 5 and 10 minutes. Prior familiarity with the selected films was low across participants, with only two or three recognising each of the films presented.

For each film scene four distinct mixes were created: an original control mix (7.1.2 Dolby Atmos) and three augmented mixes. Augmented mixes varied across three conditions: frequency, directionality, and dynamics. Each augmented mix was modified solely along its respective axis, and this resulted in a total of 16 unique audio mixes. The specific audio parameters modified for each mix are summarised and described below:

- **Dynamics:** Manipulation of level and dynamic range via compressors, limiters, and expanders, controlling contrast between soft and loud events.
- **Frequency:** Modifications to spectral and pitch-related characteristics, brightness, timbral weight, and tonal centre, using equalization, saturation, distortion, and key transposition.
- **Directionality:** Alteration of spatial audio distribution via stereo and 5.1 Atmos panning, affecting sound source localisation and spatialisation.

All mixes were produced in Reaper and DaVinci Resolve using factory plug-ins. Selected scenes and timelines are presented in Table 1. Scene order and augmentation selection was counterbalanced.

³<https://shorturl.at/EvXGO>

Table 2: Key self-report results by film and mix condition. Statistically significant immersion p-values in bold.

Film	Mix	Dominant Emotion	Immersion (p)	Most Salient
Ford vs Ferrari (FVF)				
Original	S3	Tense (45.9%)	0.01	SFX (57.1%)
Frequency	S3	Calm (28.9%)	0.01	SFX (50.0%)
A Quiet Place (AQP)				
Original	S2	Tense (69.2%)	0.002	SFX (53.8%)
Directionality	S2	Tense (69.2%)	0.002	SFX (64.3%)
Decision to Leave (DTL)				
Original	S3	Tense (35.7%)	0.02	SFX (50.0%)
Directionality	S3	Tense (42.9%)	0.02	SFX (71.4%)
I Saw the TV Glow (ISTVG)				
Original	S2/S3	Disgust (30.8–42.9%)	0.03 / 0.0006	SFX / Visuals
Dynamics	S3	Disgust, Distress (35.7%)	0.0006	SFX (50.0%)
Frequency	S2	Disgust (38.5%)	0.03	SFX (53.8%)

4 Results

Following the triangulation framework (see Section 2), self-report, behavioral, and physiological data were analysed. Additional materials and secondary analysis report are available at <https://emorsion.netlify.app>.

4.1 Self-Report Measures

For each scene, participants completed a five-item questionnaire assessing emotional response and immersion, comparing original and augmented mixes. ANOVA was applied to emotional intensity ratings and salient element identification; chi-square tests assessed emotion selection and perceived emotional change over time; and a paired t-test evaluated immersion differences. Intensity change p-values were not statistically significant. Presentation order, most frequent responses, p-values, and percentages are reported in Table 2, with statistically significant values in bold.

Ford vs Ferrari: Session 1 audiences reported higher excitement for the dynamics mix and greater tension for the original. Dynamics were most frequently identified as the modified parameter. Sound effects were most salient in the directionality mix, while music was most salient for session 2 audiences (50%). The frequency mix was perceived as perceptually distinct by 50% of participants.

A Quiet Place: Session 1 audiences found audio effects most prominent in the original mix and music

more prominent in augmented mixes. Session 2 audiences reported consistent tension across both mixes, with a statistically significant immersion change in the directionality mix. In session 3, nearly all participants (13/14) reported emotional changes with the dynamics mix, with sound effects remaining most prominent.

Decision to Leave: Session 1 associated the augmented mix with distress and foregrounded visuals, while the original emphasised tension, sound effects, and music, with pitch identified as the primary modification. Session 2 found the original mix evoked calmer, darker emotions, whereas the dynamics mix increased curiosity; 46.2% identified pitch as the main modification. In session 3, sound effects were dominant (71.4%) and the directionality mix showed a statistically significant immersion change ($p = 0.02$).

I Saw the TV Glow: Emotional patterns were broadly similar across mixes. Session 1 foregrounded sound effects; session 2 reported disgust most frequently in the frequency mix (38.5%); session 3 found the dynamics mix eliciting disgust and distress (35.7%), with the original predominantly eliciting disgust, and visuals and audio effects most noticeable.

Notably, the self-reported questionnaire revealed a significant effect on perceived immersion. Across all sessions, most participants reported increased immersion for at least one augmented mix compared to the original. Specifically, frequency-based mixes (ISTVG and FVF) and directionality mixes (DTL and AQP) were associated with higher immersion ratings, as reflected in the corresponding immersion-change p-values.

4.2 Physiological Data

The Polar H10 sensor recorded heart rate (HR, bpm) and RR intervals (ms) at 1 Hz for each participant during each scene. One participant was excluded for declining to wear the sensor. Each remaining participant viewed 8 scenes (4 control, 4 augmented). We first excluded incomplete recordings due to mid-session dropouts and removed physiologically implausible values (HR: 46–200 bpm; RR: 300–1300 ms) [23]. Afterwards, we performed interpolation via Piecewise Cubic Hermite Interpolating Polynomial [24] on all time series and further excluded windows where interpolation exceeded 30% of samples.

Time- and frequency-domain metrics were then calculated for each remaining series (see Table 3) [25, 23]. Data were aggregated into 12 subsets by pairing each augmented condition with its corresponding control, and pairwise t-tests were conducted per metric with Benjamini-Hochberg correction for multiple comparisons.

Frequency-domain measures yielded no significant results after correction, except a marginal effect for total power in the DTL directionality mix ($p = 0.035$). Time-domain metrics were more informative: SDNN showed significant effects for dynamic mixes of DTL ($p = 0.014$) and AQP ($p < 0.001$); HR standard deviation similarly for DTL ($p = 0.035$) and AQP ($p < 0.001$), with an additional effect for the DTL frequency mix ($p = 0.035$). HR interquartile range showed significant effects for dynamic mixes of DTL ($p = 0.04$), AQP ($p = 0.01$), and ISTVG ($p = 0.01$).

4.3 Movement Tracking

Audience movement was analysed using video-based motion tracking (Figure 2). Skeletal movement was extracted using OpenPose via OpenPifPaf⁴, which estimates 2D body keypoints with associated confidence scores. Only keypoints exceeding a fixed confidence threshold (above %10) were included and assigned to participants using manually defined bounding boxes corresponding to seating positions. To reduce noise and computational load, analysis was performed on temporally subsampled frames at approximately 1 Hz (480 frames for an 8-minute scene). Out of focus, blurred or excessively low light footage was excluded

⁴<https://openpifpaf.github.io/intro.html>

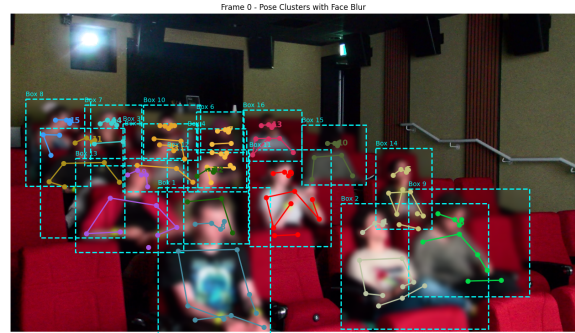


Fig. 2: Pose skeletal keypoints for Movement detection of participants with bounding boxes

from the analysis. Total movement was quantified as the sum of frame-to-frame skeletal keypoint displacements across the scene, normalised by bounding-box size, weighted by keypoint confidence, and aggregated to produce a per-participant movement magnitude. Additionally, we computed four metrics to capture other salient aspects of the viewing experience. **Mean and SD Movement** summarise total movement across participants - in other words, how active the audience was during the scene: a high mean indicates that participants moved more overall, while a low mean indicates that they remained largely still, and the standard deviation captures between-participant variability. **Mean and SD Synchrony** track the pairwise alignment of participants' movements, measured via cosine similarity of skeletal vectors on a scale from 0 to 1, where 1 indicates strong synchrony and 0 indicates none; the standard deviation reflects variability in alignment across participants. The most salient results are shown in Table 4, a complete summary of the analysis can be found at <https://emorsion.netlify.app>.

Across all scenes, mean audience movement ranged from 122 to 535 (global mean = 330). Horror scenes (ISTVG, AQP) consistently elicited lower-than-average mean movement, indicating group physical stillness across participants, whereas FVF and DTL produced higher movement levels overall. Temporal variability in movement (SD Movement) differed across films, reflecting whether activity was sustained or concentrated in brief moments. Overall movement levels differed across films and sessions, though this pattern was not consistent across all stimuli. Synchrony values were generally higher during second viewings, suggesting increased collective alignment over time. Directional-

RR Time Domain Metrics	
SDNN	Standard deviation of RR intervals
RMSSD	Root mean square of successive differences,
RR Mean	Mean of RR intervals
pNN20 and pNN50	Proportion of successive pairs that differ by more than 20/50 ms divided by total number of RR
HR Time Domain Metrics	
Mean and Median HR	Overall mean and median values for HR
HR STD	Standard deviation of HR
IQR HR	Interquartile Range (i.e., range of the middle 50% of measurements) of HR
Mean Difference	Mean difference of HR
RR Frequency Domain Metrics	
VLF, LF, and HF Power	Power in the very low (0-0.04 Hz), low (0.04-0.15 Hz), and high (0.15-0.4 Hz) frequencies
Total Power	Sum of the power in all frequency bands
LF/HF	Ratio between low frequencies and high frequencies

Table 3: Description of the extracted metrics from RR and HR measurements

ity mixes increased mean movement in comparison to the original mixes.

Despite relatively high overall movement, *FVF* showed stable synchrony and low variation between participants (0.81-0.83), suggesting that audiences tended to move in similar ways. By contrast, *ISTVG* exhibited lower movement (0.000-0.110) with higher synchrony (0.86), consistent with a more still and collectively absorbed viewing style, reinforcing participants claiming to be more immersed. Between-participant variability (SD Movement) was also high ($SD = 349.03$), suggesting that some audience members moved extensively while others remained comparatively still.

5 Discussion

In this study we combined self-report, physiological, and behavioural measures, using a triangulated method to assess immersion in a live cinema setting.

Self-report provided the strongest evidence: participants reported higher immersion for at least one augmented mix per scene (see Table 2). Directionality manipulations had the greatest impact for DTL and AQP. For AQP, increased spatialisation aligned with reduced movement; for DTL, a slight heart rate increase suggested heightened engagement, with sound effects most frequently reported as salient. This supports prior work showing surround sound can function as a focusing mechanism, drawing attention inward through enhanced spatial cues [26].

Physiological measures were most sensitive to dynamics mixes across DTL, AQP, and ISTVG. AQP showed reduced heart rate variability alongside self-reports of

stress and tension, suggesting sustained arousal. DTL and ISTVG exhibited heart rate increases, indicating heightened reactivity rather than a uniform response, consistent with evidence that dynamically intense audio can elicit aversive responses [27]. For ISTVG, increased movement co-occurred with higher immersion ratings and stronger reports of *nervousness* and *disgust*, suggesting movement reflected emotional discomfort rather than disengagement [28, 29, 30, 31].

DTL, AQP, and ISTVG are all tension- or suspense-driven scenes, and all three showed measurable changes across mix conditions. This suggests suspense-driven films are particularly well suited to multimodal immersion studies, though dominant scene emotions (e.g. disgust in ISTVG) should be carefully considered in both design and analysis.

The FVF frequency mix increased self-reported immersion without corresponding physiological or movement changes, suggesting limited emotional impact. Perceptual biases such as peak and recency effects may have shaped ratings, as gradual frequency changes are less likely to register as salient moments [21].

Overall, subtle audio augmentations, particularly dynamics and directionality, can meaningfully influence subjective engagement, though different parameters elicit responses in different ways depending on the film.

6 Limitations and Future Work

Despite these initial encouraging results, limitations should be acknowledged in the current data collection pipeline and stimulus design. Movement tracking proved the most challenging measure: one session

Table 4: Summary of head movement and synchrony by film and mix.

Mix	Mean Move.	SD Move.	Rel. Activity	Mean Sync.	SD Sync.
Ford vs Ferrari (FVF)					
Original	1.000	349.03	Very High	0.83	0.02
Dyn	0.814	353.65	High	0.81	0.02
A Quiet Place (AQP)					
Original	0.396	245.23	Below Avg.	0.81	0.04
Dir	0.404	240.33	Below Avg.	0.77	0.05
Decision to Leave (DTL)					
Original	0.928	487.92	High	0.81	0.03
Dyn	0.905	512.79	High	0.81	0.04
I Saw the TV Glow (ISTVG)					
Original	0.019	111.31	Very Low	0.82	0.07
Freq	0.110	165.30	Low	0.82	0.03

yielded almost no usable data, restricting cross-session comparisons, and frame rate reduction to 1 fps limited temporal resolution. Stimuli selection and mix precision were constrained by limited access to commercial audio stems, while sensor availability and camera coverage restricted sample size and tracking accuracy at peripheral seating positions.

Future work should address these constraints along two complementary directions. First, integrating qualitative movement data could strengthen the empirical links between movement analysis and experiential states, while a multi-camera or infrared sensor array would improve tracking reliability in the low-light conditions typical of cinema. Second, in terms of stimuli, securing stem access through student or independent productions would enable more precise manipulation, and parameters beyond volume and spatialisation - such as timbre via synthesis or machine-learning transfer - could be explored; collaboration with professional mixers would also help ensure that augmented mixes meet industry standards.

7 Conclusions

The findings of our proof of concept confirm the validity of the triangulation method, from which we observed patterns and results consistent with those reported in previous controlled studies. Furthermore, we demonstrate that subtle changes in the audio domain can be effectively explored with this approach to assess their influence in an ecologically valid environment. The multimodal analysis showed that self-reported immersion was most sensitive to audio manipulations, while physiological and behavioural measures provided complementary evidence whose variability itself offers

insight into the heterogeneity of audience response. Together, these results indicate that the EMORSION protocol is sufficiently sensitive to detect the effects of subtle audio manipulations on audience experience, motivating larger-scale studies to characterise how specific audio parameters shape narrative engagement.

References

- [1] Chion, M. and Gorbman, C., "Audio-Vision: Sound on Screen," 1994.
- [2] Garner, T., *Sonic Virtuality: Sound as Emergent Perception*, 2015, ISBN 9780199392834, doi:10.1093/acprof:oso/9780199392834.001.0001.
- [3] Kock, M. and Louven, C., "The Power of Sound Design in a Moving Picture: an Empirical Study with emoTouch for iPad," *Empirical Musicology Review*, 13(3-4), 2019, ISSN 1559-5749, doi:10.18061/emr.v13i3-4.6572.
- [4] Greene, L. and Kulezic-Wilson, D., editors, *The Palgrave Handbook of Sound Design and Music in Screen Media*, Palgrave Macmillan UK, London, 2016, ISBN 978-1-137-51679-4 978-1-137-51680-0, doi:10.1057/978-1-137-51680-0.
- [5] Agrawal, S., Simon, A., Bech, S., Bærentsen, K., and Forchhammer, S., "Defining immersion: Literature review and implications for research on immersive audiovisual experiences," *Journal of AES*, 68(6), pp. 404–417, 2019.
- [6] Hodges, D., "Can Neuroscience Help Us Do a Better Job of Teaching Music?" *General Music Today*, 23, pp. 3–12, 2010, doi:10.1177/1048371309349569.
- [7] Ronan, D., Reiss, J., and Gunes, H., "An empirical approach to the relationship between emotion and music production quality," 2018.
- [8] Zhang, C., "The why, what, and how of immersive experience," *IEEE Access*, 8, pp. 90878–90888, 2020.
- [9] Anestis, A. and Goussios, C., "How Cinema Sounds Affect the Perception of a Motion Picture," *Universal Journal of Psychology*, 3, pp. 147–152, 2015, doi:10.13189/ujp.2015.030503.

- [10] Saroka, V., "The Role of Sound in the Immersive Experience," *Avant*, 14, 2024, doi:10.26913/ava3202406.
- [11] Crocker, R., Garcia, N., Reiss, J., and Fazekas, G., "The sound of storytelling: An exploratory study of sound design and music in film drama," in *157th AES Convention*, AES, 2024.
- [12] Ansani, A., Marini, M., D'Errico, F., and Poggi, I., "How soundtracks shape what we see: Analyzing the influence of music on visual scenes through self-assessment, eye tracking, and pupillometry," *Frontiers in Psychology*, 11, p. 556697, 2020.
- [13] Zulato, E., "From cinema to the lab: Psychological experiments as liminal affective technologies," *Theory Psychology*, 2025, doi:10.1177/09593543251391140.
- [14] Theodorou, L., Healey, P. G., and Smeraldi, F., "Engaging with contemporary dance: What can body movements tell us about audience responses?" *Frontiers in Psychology*, 10, p. 71, 2019.
- [15] Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., and Walton, A., "Measuring and defining the experience of immersion in games," *International Journal of Human-Computer Studies*, 66(9), pp. 641–661, 2008.
- [16] Juslin, P. and Laukka, P., "Expression, Perception, and Induction of Musical Emotions: A Review and a Questionnaire Study of Everyday Listening," *Journal of New Music Research*, 33, pp. 217–238, 2004.
- [17] Gilgen-Ammann, R., Schweizer, T., and Wyss, T., "RR interval signal quality of a heart rate monitor and an ECG Holter at rest and during exercise," *European Journal of Applied Physiology*, 119(7), pp. 1525–1532, 2019.
- [18] Rooney, B., Hennessy, E., and 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," *Society for Cognitive Studies of the Moving Image*, 2014.
- [19] Gonzalez, V. and Żelechowska, A., "Correspondences Between Music and Involuntary Human Micromotion During Standstill," *Frontiers in Psychology*, 9, 2018, doi:10.3389/fpsyg.2018.01382.
- [20] Leman, M. and Maes, P.-J., "The Role of Embodiment in the Perception of Music," *Empirical Musicology Review*, 9, pp. 236–246, 2015, doi:10.18061/emr.v9i3-4.4498.
- [21] Zhang, C., Hoel, A. S., Perkis, A., and Zadtootaghaj, S., "How long is long enough to induce immersion?" in *10th QoMEX*, pp. 1–6, IEEE, 2018.
- [22] Aditya, D., "Why Independent Films Matter?" *ResearchGate. Artikkeli. Julkaistu*, 26, p. 2024, 2024.
- [23] Bruin, J., Stuldreher, I. V., Perone, P., Hogenelst, K., Naber, M., Kamphuis, W., and Brouwer, A.-M., "Detection of Arousal and Valence from Facial Expressions and Physiological Responses Evoked by Different Types of Stressors," *Frontiers in Neuroergonomics*, 5, 2024, ISSN 2673-6195, doi:10.3389/fnrgo.2024.1338243.
- [24] Benchekroun, M., Chevallier, B., Zalc, V., Istrate, D., Lenne, D., and Vera, N., "The Impact of Missing Data on Heart Rate Variability Features: A Comparative Study of Interpolation Methods for Ambulatory Health Monitoring," *IRBM*, 44(4), p. 100776, 2023, ISSN 1959-0318, doi:10.1016/j.irbm.2023.100776.
- [25] Bahameish, M., Stockman, T., and Requena Carrión, J., "Strategies for Reliable Stress Recognition: A Machine Learning Approach Using Heart Rate Variability Features," *Sensors*, 24(10), p. 3210, 2024, ISSN 1424-8220, doi:10.3390/s24103210.
- [26] Mendonça, C. and Korshunova, V., "Surround Sound Spreads Visual Attention and Increases Cognitive Effort in Immersive Media Reproductions," in *Proceedings of the 15th Audio Mostly*, pp. 16–21, ACM, Graz Austria, 2020, ISBN 978-1-4503-7563-4, doi:10.1145/3411109.3411118.
- [27] Dimitriev, D., Indeykina, O., and Dimitriev, A., "The Effect of Auditory Stimulation on the Nonlinear Dynamics of Heart Rate: The Impact of Emotional Valence and Arousal," *Noise and Health*, 25(118), p. 165, 2023, ISSN 1463-1741, doi:10.4103/nah.nah_15_22.
- [28] Ottaviani, C., Mancini, F., Petrocchi, N., Medea, B., and Couyoumdjian, A., "Autonomic Correlates of Physical and Moral Disgust," *International Journal of Psychophysiology*, 89(1), pp. 57–62, 2013, ISSN 0167-8760, doi:10.1016/j.ijpsycho.2013.05.003.
- [29] Kreibitz, S. D., "Autonomic Nervous System Activity in Emotion: A Review," *Biological Psychology*, 84(3), pp. 394–421, 2010, ISSN 0301-0511, doi:10.1016/j.biopsycho.2010.03.010.
- [30] Kurakata, K., Mizunami, T., and Matsushita, K., "Sensory unpleasantness of high-frequency sounds," *Acoustical Science and Technology*, 34(1), pp. 26–33, 2013.
- [31] Ringer, H., Rösch, S. A., Roeber, U., Deller, J., Escera, C., and Grimm, S., "That sounds awful! Does sound unpleasantness modulate the mismatch negativity and its habituation?" *Psychophysiology*, 61(2), p. e14450, 2024.