Audio Engineering Society

Convention Express Paper 302

Presented at the AES 157th Convention
2024 October 8–10, New York, NY, USA

The sound of storytelling: An exploratory study of sound design and music in film drama

Ruby Crocker[1], Nelly Garcia[1], Joshua Reiss[1], and György Fazekas[1]

[1]*Queen Mary University of London, Centre for Digital Music*

Correspondence should be addressed to Ruby Crocker and Nelly Garcia (`r.o.n.crocker@qmul.ac.uk` or `n.v.a.garcia-sihuay@qmul.ac.uk`)

ABSTRACT

The integration of sound effects and music plays a central role in shaping the audience's emotional engagement and narrative comprehension in film. This study analyzed ten scenes from the drama genre, which emphasizes human emotions and narrative-driven storytelling. Participants were divided into three groups and were exposed to six combinations of audio and visual stimuli. They reported salient sounds and their interpretations, with a focus on context and emotional responses. One hypothesis suggests that effective sound design blurs the line between music and sound effects. When asked, 63% of participants felt music was more important for providing context, while 37% prioritized sound effects. The findings indicate that music alone emphasizes specific emotions, whereas sound effects alone create moderate variability between emotional response and sound identification.

1 Introduction

Sound in film can deepen the audience's understanding of the narrative, enhance immersion and enrich their interpretation of the emotions the filmmakers are trying to convey [1]. Sound effects make scenes feel more realistic, immersing the audience in the characters' experiences. Music provides contextual and emotional cues. It can convey inner feelings of characters and shape perceptions of the audience, such as foreshadowing threats or creating unease before visual cues appear.

Directors, composers and sound editors work together to create original scores that enhance the tone of a film, although non-original music is also used for cultural or artistic effects [2, 3].

Sound design, which includes sound effects but excludes music and dialogue, also plays a crucial role in shaping emotions [4]. The line between sound effects and music is often deliberately blurred for aesthetic reasons, yet evaluating sound design can be challenging due to the difficulty of accessing separate tracks. Modern filmmakers, like Christopher Nolan, seamlessly integrate speech, sound effects, and music across various

styles [5, 6].

The perception of emotions is subjective and based on personal experiences, thoughts and feelings [7]. Categorical approaches help to classify emotional perceptions and facilitate analysis [8, 9]. In emotion research, various models and adjective classifications, such as Russell's Circumplex Model and Hevner's Adjective Circle, have been utilized to standardize emotional descriptors and improve the consistency of emotional data [10, 11]. These classifications help recognize and generate emotion labels for tracks [12]. Familiarity with the context is essential in emotion perception tests [13].

This study focuses on drama films and explores how music enhances emotional arousal by enriching scene rhythm and feeling [14]. The drama genre in film critically explores the human experience, focusing on the everyday struggles and emotions of ordinary people [15].Self-reporting was used herein to capture the audience's emotional connection. Though psychological tests can often reveal deeper emotions, self-reporting is supported as a valid method[16, 17], and has been shown to be effective in large groups [18], [19].

Previous studies have focused on how music affects emotional responses, often due to limited access to sound design material. This study examines participants' responses to both music and sound design tracks in drama films, with the aim of understanding cinematic storytelling, how filmmakers evoke emotions, and the important role of sound design or music in evoking specific emotions.

2 Methods

To investigate the impact of visual stimuli on cognitive processing, 32 participants were divided into three control groups. Each group watched 10 different film scenes with varying combinations of audio elements (sound effects, score, or both) and visual components (with or without visuals). After each scene, participants completed a brief questionnaire assessing their perception of context and emotional response. In this context 'sound effects' included dialog, Foley sounds, ambient sounds and room tone. The choice of whether scenes included visuals or not aimed to assess their impact on cognitive processing, as visuals might obscure differences in perception. The combinations of audio and visual elements were randomized across groups for each scene to minimize bias.

2.1 Experiment design

The experiment took place in a studio designed to mimic a cinema environment, featuring a large projector, high-quality PA system, and dim lighting to enhance focus. Participants were given a brief introduction to the study's requirements, ensuring an immersive and undisturbed experience.

The decision to conduct the study in a cinema-like environment with groups, rather than individual controlled settings, was based on several considerations. While a controlled environment would minimize outside influences on the participants, it would also be time-consuming and less natural for watching a film. The dark, cinema-like setting enhances the immersion and emotion [20], aligning with the filmmakers' intentions without requiring separation of the audience.

Five questions were designed to gauge participants' interpretations of the narrative based on visual and auditory stimuli. These five questions were asked to each participant at the end of watching each scene. These questions, outlined in table 1 and accessible via a Google Form [1], focused on emotion perception, identifying key sounds and understanding their impact on the meaning of the film. Some questions used tickbox options for controlled responses, while others were open-ended to allow creative freedom. This balance enabled us to draw meaningful conclusions while capturing diverse perceptions.

Participants selected emotions from a tick-box list, making the data categorical and easier to analyze. Restricting choices to the top three emotions per scene ensured that the results remained manageable and comparable, even with a smaller sample size. While this approach reduces flexibility and may influence responses, it simplifies the evaluation of subjective experiences like emotions and music. Categorical data also minimizes variability and errors, ensuring participants understood the questions and responded consistently. Although broad data from open-ended responses can offer depth, categorical data with a smaller sample size helps to identify patterns and improves generalizability, striking a balance between detailed insight and broader applicability.

To ensure consistency, each scene used the same set of 28 emotion adjectives for question 5, as shown in

---

[1]https://forms.gle/FBJPx2454LU4gDDt9

Table 1. For question 4, 16-21 sound options were provided. This approach allowed for uniform assessment across scenes and facilitated clearer comparisons. The emotion adjectives were selected based on established models such as Russell's Circumplex Model, Hevner's Adjective Circle, and revisited versions of the models [10, 11]. They were chosen for their relevance to the drama genre and film descriptors, balancing the need for variety with ease of use for participants. For an example, see the emotion adjectives for Scene 10 in figure 1.

The questions were inspired by studies of how composers manipulate emotions through film music and the role of music as an emotional catalyst in film [1, 21, 22, 23]. To explore participants' perceptions of emotion and narrative in film clips, different test groups were exposed to different combinations of sound effects and music, as shown in Table 3. IMG' refers to the visuals in the scene, 'SFX' refers to all environmental, atmospheric, action and realism-based sounds, such as footsteps, and 'MSC' refers to non-diegetic music in the scene. The six combinations tested were: 1) sound effects only, 2) music only, 3) music and image, 4) sound effects and image, 5) sound effects and music, and 6) sound effects, music and image.

2.2 Scene selection

The selection process considered availability, relevance, cultural diversity, award status, and a variety of styles in music, sound, and narrative. Films were chosen from the IMDb movie database from the last ten years [2] and Academy Awards lists [3] based on nominations or wins for Best Sound Effects and Best Original Score in the past decade. The presentation order and selected scenes are detailed in Table 2. Table 3 shows the combination of stimuli (IMG, SFX, MSC) each group was shown for each film scene. This approach ensured a diverse range of films, each nominated for at least two international awards like the Golden Reel, Critics' Choice, or BAFTA in these categories.

Films were chosen from a range of different countries and cultures to ensure a diverse representation. The majority of films selected were still from the USA with influence and languages from other countries, See Table 2. For example 'Minari' is based and from the US but is primarily focusing on a Korean-American

---

[2] https://www.imdb.com/
[3] https://www.oscars.org/oscars/ceremonies

family. Considering that participants may have biases toward films from their own countries and varying reactions depending on language, it's essential to note that different nationalities may not always experience the same emotional responses to music and cinema. Just as individuals have diverse responses and interpretations of narratives, the aim is to ensure inclusion and acknowledge this essential diversity [24]. However, cultural differences may weaken the results because of this diversity and so the majority of films selected were still from the US/Western world.

The presence of sound effects was crucial to evaluate and compare how music and sound design impact an audience. Recognizing the necessity of isolating music and sound effects (both diegetic and non-diegetic) to assess their individual impact, the analysis was conducted by segregating these elements accordingly.

2.3 Participants

The experiment involved 32 participants, all selected for their expertise in sound and music studies. This specialist focus provided deeper insights into how sound and music influence context and emotion in film, and established a baseline for future studies with broader audiences (see section 5). Most participants (90%) were aged between 25 and 34, with uneven group sizes due to availability: Group 1 had 8 participants, Group 2 had 10 and Group 3 had 14. Although a larger sample size might have increased reliability, the time constraints made that impractical. Of the participants, 20 identified as male and 12 as female, representing 15 nationalities, with Chinese, Mexican and British being the most common. Most were audio researchers or postgraduate students.

3 Results

Statistical tests analysed the relationship between sound effects and music composition. While 63% of respondents felt that music provided more context than sound effects, 37% felt the opposite. Many noted that both elements were equally important, depending on the style and genre of the film.

Of the 10 films shown, most participants were unfamiliar with them. Scene 9 from Parasite was the most familiar, with 17 participants indicating some recognition out of 32. Conversely, Scene 3 from *Loveless* and Scene 6 from *Minari* was the

Q1:	If you had to decide which is more important for contextual purposes in film, would you choose sound effects or musical composition?
Q2:	What part of the sound of this scene is most prominent to you? (give it in seconds, if you can ?)
Q3:	Describe the prominent sound from the question above:
Q4:	Choose 5 most prominent sound effects/features that you could identify in the scene:*
Q5:	At the end of the scene, which emotions do you feel are most prominent?*

Table 1: Questions asked to participants. *Options were provided for participants to choose from.

	Film Title	**Director**	**Country**	**Year**	**Scene Length**	**Scene Timestamp**
Scene 1	*EO*	Jerzy Skolimowski	Poland-Italy	**2022**	4m 02s	30m 42s – 34m 45s
Scene 2	*Close*	Lukas Dhont	Belgium-France-Netherlands	**2022**	3m 16s	01h 08m 50s – 4h 08m 16s
Scene 3	*Loveless*	Andrey Zvyagintsev	Russia-France-Germany	**2017**	1m 44s	01m 16s – 1m 18s
Scene 4	*The Power of The Dog*	Jane Campion	USA	**2021**	1m 59s	01m 45s – 03m 42s
Scene 5	*Moonlight*	Barry Jenkins	USA	**2016**	3m 24s	48m 11s – 51m 34s
Scene 6	*Minari*	Lee Isaac Chung	USA-Korea	**2020**	1m 14s	18m 37s – 19m 52s
Scene 7	*Marriage Story*	Noah Baumbach	USA	**2019**	1m 26s	1h 23m 10s – 1h 24m 38s
Scene 8	*Killers of the Flower Moon*	Martin Scorsese	USA	**2023**	2m 26s	20m 51s – 30m 09s
Scene 9	*Parasite*	Bong Joon-ho	South Korea	**2019**	1m 49s	1h 16m 35s – 1h 18m 26s
Scene 10	*Nomadland*	Chloé Zhao	USA	**2020**	1m 30s	44m 46s – 46m 17s

Table 2: Film Scenes Selected

least familiar, with only 1 participant reporting some familiarity out of 32 participants. Out of 320 familiarity ratings, 53 participants indicated they were familiar with the scenes, 19 rated their familiarity as 'somewhat,' and 248 reported no familiarity. This suggests that most participants had not seen these scenes before.

Ideally, an experiment would use clips that are entirely unfamiliar to all participants. However, the selection was limited by the use of award-winning films, which are essential for their proven effectiveness in conveying emotions through sound and music. Since more participants lead to stronger conclusions and diverse perceptions of the scenes under different stimuli combinations (IMG, MSC, SFX), entries were not excluded based on prior familiarity.

An ANOVA test on the multiple choice questions regarding sound identification and associated emotions

showed no significant differences between groups 1, 2 and 3. A Kruskal-Wallis test also gave H-values above 0.05, indicating no statistical significance. Nevertheless, there were variations in the choice of sound and emotion for each stimulus, suggesting nuanced responses. The top 3 sounds and emotions are listed in table 4. Each group experienced a unique mix of sound, music and visuals, so scene-specific analysis is required for accurate interpretation.

Table 4 shows that 'alarmed' is chosen more often in the absence of visuals, possibly because participants rely more on sound, which increases alertness. It also shows that sound effects are more often identified as salient when visuals are included, despite the audio being unchanged, suggesting that visuals enhance the recognition of sound effects.

Question 4 of the evaluation determined when participants identified the most salient sound. ANOVA and

	Group 1	Group 2	Group 3
EO	IMG+ SFX+MSC	SFX+MSC *	SFX+MSC *
Close	IMG+MSC	IMG+SFX+MSC	I M
Loveless	IMG+MSC	IMG+SFX+MSC	SFX *
TPOTD	IMG+SFX+MSC	SFX+MSC *	IMG+SFX+MSC
Moonlight	MSC *	SFX *	IMG+ SFX+MSC
Minari	IMG+MSC	IMG+SFX+MSC	SFX+MSC *
Marriage Story	IMG+SFX+MSC	SFX+ MSC *	MSC *
KOFTM	IMG+MSC	IMG+SFX+MSC	SFX+MSC *
Parasite	IMG+MSC	IMG+SFX+MSC	SFX *
Nomadland	SFX+MSC *	IMG+MSC	MSC *

Table 3: Test groups and stimuli, * = No Image, IMG = Image, MSC= Music, SFX = Sound Effects

Kruskal-Wallis tests revealed significant values in only 1 out of 10 scenes, showing no major statistical differences. In *Minari* ($p = 0.03$, $h = 0.01$), participants with both sound effects and music (SFX+MSC) identified the prominent sound within 50-100 seconds, while those with music only (IMG+MSC) did so at the beginning of the scene.

4 Discussion

Emotions were particularly influenced by the presence of visuals, with music playing a greater role in conveying context than sound effects. Participants experienced a wider range of emotions when listening to audio alone. Interestingly, certain scenes elicited specific responses; for example, participants laughed with visuals, but were silent without them. Table 2 shows, as expected, greater ambiguity in the choice of emotion when only sound effects were presented. Figures 1 and 2 show less emotional diversity in Group 1 compared to Groups 2 and 3, indicating a stronger emotional resonance with music. When sound effects were added, more participants chose "sinister". Standard deviation, kurtosis and skewness analyses supporting these results are shown in table 5.

Based on Table 5 for Parasite, Group 2 (SFX+MSC+IMG) had the least variability in sound identification, reflecting a balanced listening experience. Group 1 (MSC+IMG) had the highest skewness and kurtosis, suggesting a strong focus on specific sounds, while Group 3 (SFX only) had moderate variability and skewness. These results suggest that combining sound effects, music and visuals leads to more consistent sound identification, while music alone highlights specific emotions and sound effects alone introduces moderate variability. This finding may help filmmakers to better balance audiovisual elements to shape audience perception.

Although randomizing the order helps reduce bias, the sequence of scenes might still affect participants' responses. For example, an early dramatic or intense scene could influence how subsequent scenes are perceived.

Furthermore, in drama scenes where music dominated over sound effects, participants associated emotions such as calm, happiness and peace, as seen in Close (scene 2), Minari (scene 8) and Nomadland (scene 10). In contrast, scenes with prominent sound effects, such as Parasite (scene 7), led to more intense emotional responses such as tension, in line with Michel Chion's concept of rhythm in a scene [14]. Group 3, who experienced more scenes without visuals, showed greater agreement with prominent sounds and consistent identification of emotional moments.

5 Conclusions and Future Work

Sound design and film music play a nuanced, contextual role in evoking emotion. Directors control sound effects or emphasise music to shape the emotional tone of a scene. Our results suggest that music typically guides the intended emotion, while sound effects broaden the emotional range.

As discussed in section 2.1, all groups experienced scenes with audio. Introducing scenes without audio could isolate the influence of visuals. Subtle audio adjustments (equalisation, reverb, dynamics) and new

Scene	Group 1		Group 2		Group 3	
	SFX	**Emotions**	**SFX**	**Emotions**	**SFX**	**Emotions**
EO	Synths 16%	Disturbed 17%	Synths 15%	Uncertainty 21%	Gunshot 19%	Suspense 21%
	Water 14%	Suspense 13%	Gunshot 15%	Tension 19%	Synths 15%	Alert 9%
	Gunshot 14%	Alert 8%	Strings 13%	Suspense 15%	Water 10%	Tension 15%
Close	Flute 19%	Warmth 17%	Footsteps 19%	Carefree 27%	Flute 26%	Hopeful 16%
	Piano 19%	Carefree 13%	Rustling 15%	Warmth 17%	Strings 15%	Calm 16%
	Strings 16%	Nostalgia 13%	Clothes 13%	Calm 13%	Piano 13%	Warmth 16%
Loveless	Synth 17%	Sorrow 21%	Crunches 17%	Uncertainty 20%	Footsteps 20%	Alert 19%
	Strings 17%	Uncertainty 21%	Footsteps 17%	Tension 17%	Breathing 13%	Uncertainty 11%
	Tremolo 11%	Mysterious 13%	Rustling 15%	Suspense 13%	Rustling 13%	Vulnerable 8%
TPOTD	Banjo 22%	Solitude 13%	Floor creek 16%	Intrigue 13%	Guitar 16%	Intrigue 14%
	Cows 16%	Calm 13%	Cows 16%	Mysterious 13%	Cows 14%	Calm 14%
	Footsteps 14%	Uncertainty 8%	Footsteps 16%	Alert 10%	Banjo 14%	Uncertainty 8%
Moonlight	Strings 26%	Tension 13%	Conversations 18%	Uncertainty 17%	Train 20%	Vulnerable 17%
	Woodwind 16%	Mysterious 13%	Smoke 14%	Vulnerable 13%	Conversations 16%	Joy 11%
	Drone 13%	Imaginative 8%	Lighter 14%	Curiosity 13%	Lighter 10%	Mysterious 8%
Minari	Vocals 23%	Carefree 21%	Fire crackle 20%	Warmth 23%	Vocals 17%	Warmth 18%
	String 19%	Nostalgia 13%	Shovel 18%	Carefree 17%	Guitar 13%	Carefree 17%
	Guitar 19%	Intimate 13%	Vocals 10%	Joy 13%	Footsteps 11%	Calm 10%
The H8F	Gunshot 29%	Suspense 26%	Gunshot 30%	Suspense 30%	Gunshot 33%	Alert 22%
	Strings 18%	Tension 13%	Clattering 11%	Alert 11%	Piano 14%	Suspense 22%
	Piano 15%	Alert 13%	Piano 11%	Dramatic 11%	Train 13%	Dramatic 11%
Drive My Car	Strings 28%	Joy 19%	Keys 21%	Warmth 21%	Piano 18%	Calm 13%
	Piano 17%	Intimate 14%	Car doors 17%	Joy 14%	Strings 18%	Warmth 11%
	Vocals 17%	Hopeful 11%	Piano 17%	Calm 10%	Synth 11%	Hopeful 11%
Parasite	Strings 21%	Tension 21%	Gunshot 20%	Alert 25%	Gunshot 16%	Suspense 24%
	Timpani 13%	Suspense 18%	Strings 15%	Suspense 15%	Strings 16%	Alert 13%
	Piano 10%	Uncertainty 10%	Water 12%	Tension 10%	Water 14%	Tension 13%
Banshees	Fiddle 22%	Nostalgia 13%	Sea 20%	Hopeful 17%	Fiddle 20%	Warmth 21%
	Strings 20%	Warmth 11%	Woodwind 17%	Warmth 14%	Strings 14%	Peace 10%
	Woodwind 14%	Peace 8%	Fiddle 14%	Nostalgia 11%	Piano 11%	Hopeful 8%

Table 4: The top three sounds and emotions associated with each scene across the three groups.
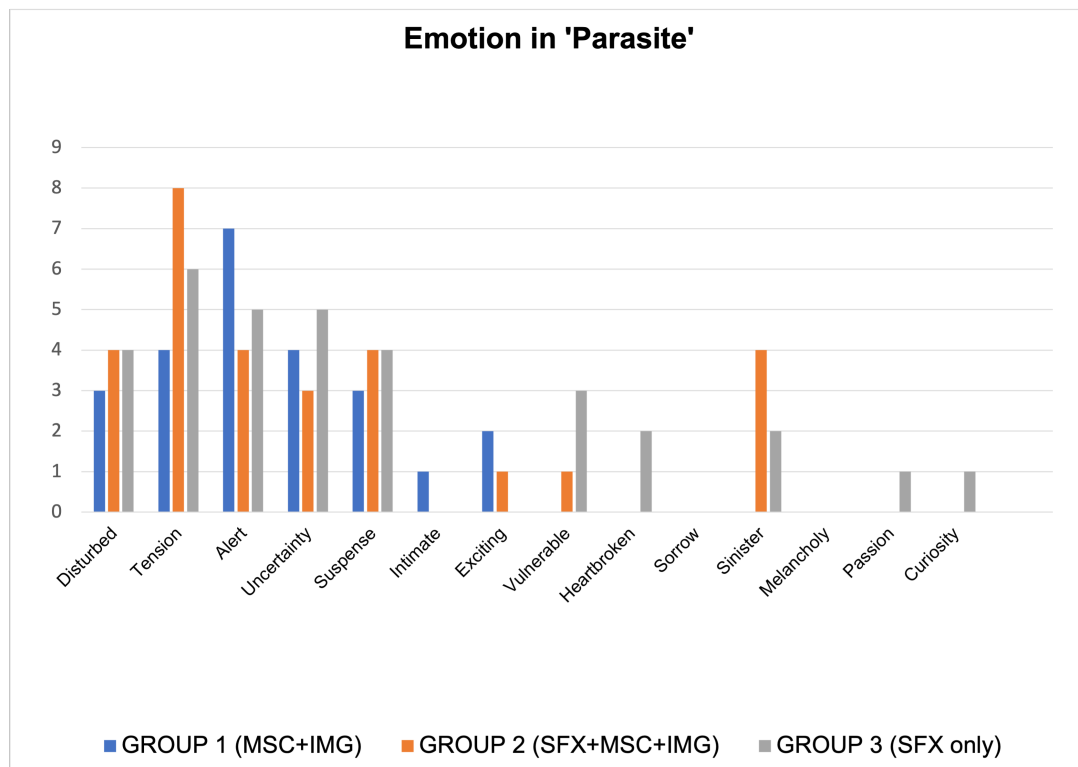
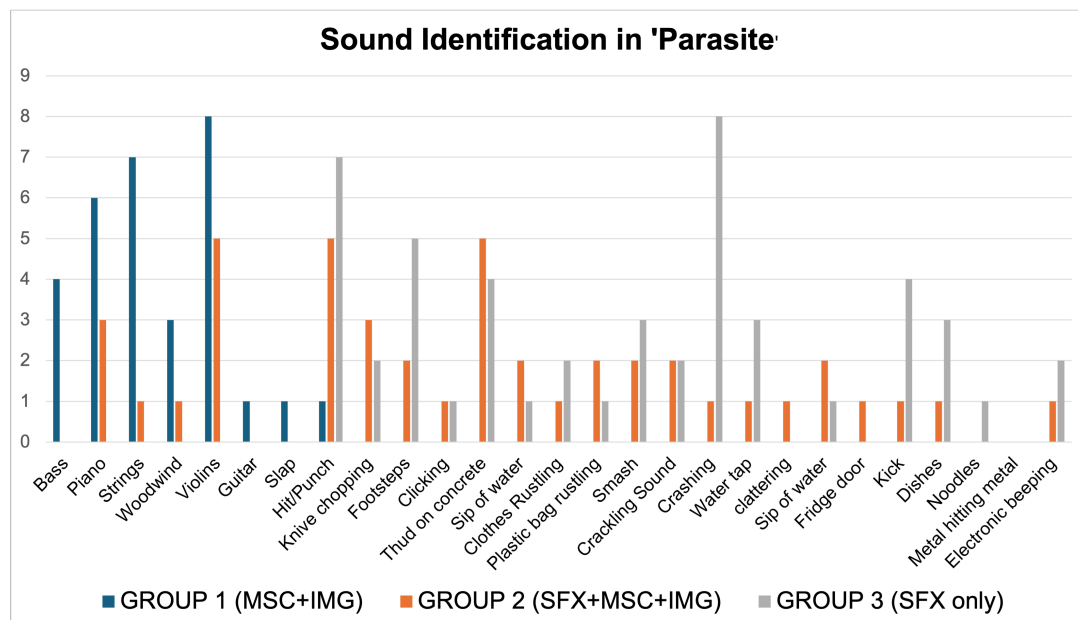**Fig. 1:** Identified Emotions for the Three Groups for 'Parasite'

Fig. 2: Identified Sounds for the Three Groups for 'Parasite'

Group	Std Deviation	Kurtosis	Skewness
Group 1	2.300	2.107	1.746
Group 2	1.739	0.026	0.973
Group 3	2.333	0.575	1.148

Table 5: Standard deviation, kurtosis and skewness of the different groups in sound identification in 'Parasite'

compositions could further refine audience responses. Future research should be extended to different film genres, which may reveal different patterns. Obtaining individual audio tracks would allow for more flexible sound presentations, and the inclusion of participants with theatre or film experience is recommended.

Increasing the number and diversity of participants (age, gender, cultural background) would increase the robustness and generalizability of the data. Overall, the drama genre suggests that participants are more in tune with their emotions in the absence of visuals, while sound effects broaden the emotional range and music sets the tone.

References

[1] Juslin, P. N. e. a., *Handbook of music and emotion: Theory, research, applications*, Oxford University Press, 2011.

[2] Rudolph, P., "The musical idea work group: production and reception of pre-existing music in film," *Twentieth-Century Music*, 20(2), pp. 136–156, 2023.

[3] Marshall, K., "The Long-standing Relationship between Composer Alan Silvestri and Director Robert Zemeckis.", " *A Critical Companion to Robert Zemeckis, ritstỳrt af Adam Barkman & Antonio Sanna*, pp. 75–88, 2020.

[4] Wei, C. e. a., "Influence of Music on Perceived Emotions in Film," *Journal of the Audio Engineering Society.*, 2022.

[5] Carlsson, R., "Non-diegetic film music as a narrative agency in Christopher Nolan's Interstellar (2014)," 2019.

[6] Wu, K. and Wu, Y., "The Cinematic Symphony of Sound: An Analysis of the Music and Sound Effects in" Tenet"," *International Journal of Education and Humanities*, 11(2), pp. 176–183, 2023.

[7] Evans, P. e. a., "Relationships between expressed and felt emotions in music," *Musicae Scientiae*, 12(1), pp. 75–99, 2008.

[8] Song, Y. e. a., "Perceived and induced emotion responses to popular music: Categorical and dimensional models," *Music Perception: An Interdisciplinary Journal*, 33(4), pp. 472–492, 2016.

[9] Wang, J.-C. e. a., "Exploring the relationship between categorical and dimensional emotion semantics of music," in *Proceedings of the second international ACM workshop on Music information retrieval with user-centered and multimodal strategies*, pp. 63–68, 2012.

[10] Russell, J. A., "A circumplex model of affect." *Journal of personality and social psychology*, 39(6), p. 1161, 1980.

[11] Schubert, E., "Update of the Hevner adjective checklist," *Perceptual and motor skills*, 96(3_suppl), pp. 1117–1122, 2003.

[12] Ardakani, M. S. e. a., "A Categorical Approach for Recognizing Emotional Effects of Music," 2017.

[13] Thiparpakul, P., Mokekhaow, S., and Supabanpot, K., "How Can Video Game Atmosphere Affect Audience Emotion with Sound," pp. 480–484, 2021.

[14] Chion, M., *Audio-Vision: Sound on Screen*, Columbia University Press, New York, 1994.

[15] Selbo, J., *Film genre for the screenwriter*, Routledge, New York, USA, 2014.

[16] Gabrielsson, A., "Emotion Perceived and Emotion Felt: Same or Different?" *Musicae Scientiae*, 5, pp. 123–147, 2002.

[17] Hunter, P. and Schellenberg, E., *Music and Emotion*, volume 10, pp. 129–164, 2010.

[18] Richardson, D. e. a., "Engagement in video and audio narratives: contrasting self-report and physiological measures," *Scientific Reports*, 10, 2020.

[19] Ronan, D., Reiss, J., and Gunes, H., "An empirical approach to the relationship between emotion and music production quality," 2018, doi:10.13140/RG.2.2.34288.76800.

[20] Shteynberg, G. e. a., "Feeling more together: group attention intensifies emotion." *Emotion*, 14(6), p. 1102, 2014.

[21] Cohen, A. J., "Music as a Source of Emotion in Film," in P. N. Juslin and J. A. Sloboda, editors, *Handbook of Music and Emotion: Theory, Research, Applications*, pp. 879–908, Oxford University Press, Oxford, 2010.

[22] Herget, A.-K., "Well-known and unknown music as an emotionalizing carrier of meaning in film," *Media Psychology*, 24(3), pp. 385–412, 2021.

[23] Martin, D., *Within Worlds: Immersion and Subjectivity in Film Music*, Ph.D. thesis, University of Huddersfield, 2017.

[24] Fang, C.-W., Chen, L.-C., and Chen, C. S., "An Analysis of Film Voice-Over and Sound Interpretation Techniques from a Cross-Cultural Perspective," in *The Asian Conference on Psychology & the Behavioral Sciences 2013 Official Conference Proceedings*, p. 0281, Osaka, Japan, 2013.