Temporal Analysis of Emotion Perception in Film Music: Insights from the FME-24 Dataset

Ruby O. Crocker[1] and **György Fazekas**[1]

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

ABSTRACT

Understanding how composers evoke emotions through music remains limited due to the lack of diversity in existing datasets used for computational analysis of film music. To address this gap, the Film Music Emotion Dataset (FME-24) was created to explore emotion perception in modern and contemporary film music, covering various genres and compositional styles. FME-24 introduces precise temporal control through comprehensive time stamps, enabling nuanced analyses often overlooked in previous studies. Over 185 participants contributed annotations for the dataset. While there is consensus on arousal, there's significant valence variation across samples, possibly due to temporal aspects and participant differences. On average, each participant completed 17 annotations across 10 randomly generated samples, totaling 1580 annotations. These annotations will help uncover parallels between emotional states and musical transitions, enriching our understanding of music's impact on audience immersion. Initial findings indicate potential for meaningful insights, future work including extracting musical features to compare with valence-arousal values and expanding the sample size with additional participants will enhance the reliability of the results, and uncover patterns between emotion perception and musical elements.

1. INTRODUCTION

Composers and filmmakers use music to manipulate emotion to aid storytelling. Changes in music based on characteristics or elements by composers can alter the way in which visual scenes are interpreted, film music can differ from mainstream or commercial music as there can be alternative motivations for the direction of the composition. Juslin and Sloboda [1], propose that film music stands out as one of the most powerful tools for conveying emotion in cinematic experiences.

Juslin and Laukka [2] proposed that one of the most prevalent reasons for listening to music is emotion. It is suggested that understanding the human listening response to music in everyday context provides a better understanding of how emotion in music is perceived. There is evidence indicating distinctions between emotions evoked by

music and those experienced in everyday life [3].

The unique ability of film music to replicate or evoke emotions through immersion makes it an intriguing subject for exploration, especially considering the distinctions observed between music-induced emotions and those experienced in everyday life [4]. There is recent research based around composition and film music being the carrier of emotion in film [5], however there is minimal research focusing on how composers manipulate emotion based on musical attributes and techniques.

The distinction between emotion expressed by a piece of music and the emotion felt by a listener is an essential part of research for emotion in music research [6]. Both perceived and felt emotional responses are relevant for studies on emotional responses to music [7]. There is a disparity between the emotional response perceived by listeners and the emotions they feel. These discrepancies may be attributed to various factors, including contextual information, memory, physiological arousal, age, and musical features such as tempo and suggested emotions [6].

2. RELATED WORK

Wei et al. [8] explored the impact of music on perceived emotions in film and stated that the impact of film music on viewers' emotions is still not fully understood. An experiment was conducted to analyse how music influences emotion perception in film, they compared viewers responses with video only, audio only, and both video and audio, they had 16 clips that covered 4 film genres. The focus was on the influence of both image and sound. There is limited research in how mood is measured over-time in film music and how these emotions change and develop over-time.

Russell [9] proposed a valence-arousal model of emotion, Figure 1, offering a dimensional framework for understanding emotional experiences. This model, widely adopted in emotion research, provides detailed insights into emotional nuances compared to categorical approaches [10]. Arousal in this model signifies the energy level of an emotion, while valence indicates its positivity or negativity.

Existing film music datasets predominantly focus on classical and mainstream music, lacking representation of modern genres and arthouse cinema [11, 12]. There is a scarcity of film music datasets from the past two decades, particularly those reflecting diverse cultural backgrounds and cinematic styles [13].

Walus [14] explored approaches to composing film music, emphasizing the dynamic nature of scoring during pre-
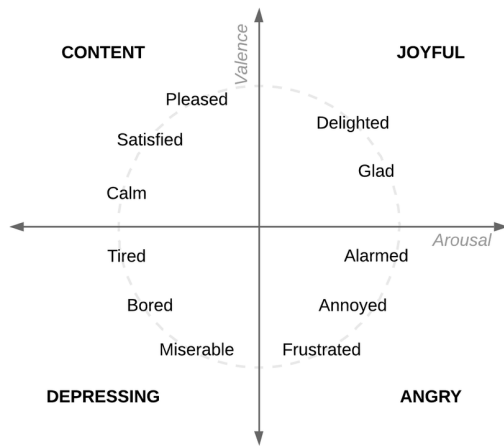
Figure 1: Russell's Arousal-Valence Emotion Model

production. Changes in the score's layers can significantly impact narrative, character development, and mood in the film.

To draw meaningful conclusions about perceptions of film music, it is crucial to analyze a diverse array of compositions spanning different genres, styles, cultural backgrounds, and cinematic contexts [15].

The paper proceeds as follows: Section 3 introduces the Film Music Emotion Dataset (FME-24), outlining its creation process and characteristics. Section 4 analyzes the demographics, inter-rater agreement, and initial evaluation of FME-24. Initial findings are discussed in Section 5, followed by future research directions in Section 6.

3. FME-24 DATASET

FME-24 dataset was developed to address the gaps in existing research explained in Section 1. FME-24 was created to study composers' techniques for manipulating emotions in film music across different styles. The dataset is populated with annotations from various participants to provide valence-arousal annotations with corresponding timestamps within film compositions. The initial purpose of creating the dataset was to identify music features that coincide with participants' emotional perceptions, aiming to pinpoint specific characteristics during moments of emotional change.

The dataset consists of 300 compositions from 21 different film genres with over 156 different composers. Many films in the dataset are classified under a blend of multiple genres. The predominant film genre label in the dataset, occurring 229 times, is 'Drama.' This predominance stems from the frequent combination of 'Drama' with other genre labels within many film genres. The other most common film genres in the dataset consist of; Action: 51, Adventure: 64, Biography: 38, Comedy: 67, Romance: 42, Crime: 38, Mystery: 39, Sci-Fi: 26 and Thriller: 27. A combination of multiple genres are used to describe each film. All film metadata, including film genre classifications, release dates, directors, composers, and other relevant details, were sourced from the IMDB database web-

site [1].

The audio files selected were from a combination of art-house cinema and mainstream cinema from the last twenty years. While the dataset primarily concentrates on music produced within the last two decades, a slight emphasis is placed on compositions from more recent scores due to the increased variety observed in genres and styles during this period. The compositions were selected from award-winning films, regardless of whether the film's soundtrack itself won an award. This decision was made because the focus is on examining the immersive mood created within the film. Award-winning films often undergo rigorous evaluation processes by industry professionals, ensuring a high standard of emotional impact and artistic merit. Selection criteria also included critical acclaim and the capacity to evoke a wide spectrum of emotions, ensuring a diverse range of compositions for comprehensive emotional analysis. In many cases, the emotional impact of a film extends beyond its original soundtrack. It's worth noting that not all films have original scores; some utilize music from older films or mainstream music to enhance mood and atmosphere. In the selection process for creating the dataset, music from sources other than original scores was included, enabling a broader exploration of musical influences on perceived emotions in film music. While the film's soundtrack may not always be recognized with awards, it remains pivotal in evoking emotion within the film. Consequently, if a film has received awards or nominations, it is deemed valuable for inclusion in the dataset. During the selection of compositions for the dataset, discussions were conducted with both film and music students to identify impressive and dynamic scores. When compiling the audio files for the dataset, consideration was given to acclaimed and highly rated films from various prestigious film festivals. The aim was to incorporate a diverse selection of films beyond Western cinema, ensuring a rich representation of cultural influences and styles.

3.1 Dataset Creation and Process

Annotations from participants were collected using an online interactive survey. The interactive space consisted of a playable waveform and an interactive valence-arousal space, shown in Figure 2. Participants were presented with a 15-second excerpt from a film soundtrack, devoid of any visual accompaniment, and solely consisting of the audio waveform. It was explicitly communicated to the participants at the start of the study that the investigation solely related to the perception of emotion in music. Unlike, Wei et al. [8], this paper focuses solely on sound, not including image or visual stimuli. The decision to focus on emotion perception rather than induced emotions is based on the understanding that participants may find it easier to discern and articulate their own perceived emotions. In large-scale data collection, involving participants in self-analysis offers valuable insights and is more efficient, thus encouraging greater participation. By focusing on perceived emotions, it becomes easier to systematically observe and interpret emotional responses, ideal for scientific and analyt-

---

[1] https://www.imdb.com/

ical contexts, ensuring comprehensive and impartial analysis. Participants were instructed to click a button upon perceiving an emotion change while listening to the excerpt. The instruction was to: "add markers to the waveform as soon as you feel a change in emotion/mood. Add as many markers as you'd like". Simultaneously, a dot appeared on the valence-arousal space in a color matching the marker, which participants could move. Valence and Arousal were labelled 'Positivity' and 'Energy' respectively to ensure understanding for participants. Participants were able to manipulate markers on the waveform and adjust the position of dots on the valence-arousal space. They were then prompted to add a sentence describing their feelings and familiarity rating before submitting their response. Participants were instructed to complete the entire survey, which comprised of 10 excerpts, each lasting 15 seconds. However, they had the option to stop at any time. The 10 excerpts were randomly loaded from the collection of 300 audio files for each participant as they initiated the survey. Additionally, the 15-second samples were randomised for each participant. Participants were encouraged to look at short video before starting the survey to help direct how to navigate the interface, additionally there was a pop-up diagram for further help with interpreting the valence-arousal space. Participants were recruited by distributing the survey as a website link through GitHub Pages [2] and sharing it via targeted mailing lists within musicology, film, and emotion-related fields. Academic mailing lists were also utilized to ensure a broad participation base. While the study focused on film and music, it did not necessitate participants to be musicians or experts in the field. Hence, social media platforms were employed to reach a wider audience. This approach aimed to gather a diverse range of responses from participants with varied backgrounds, including those in psychology, film studies, music, production and related academic disciplines. Ethical approval for this study was obtained, ensuring participant safety and data confidentiality [3].

4. ANNOTATION PROCESS

4.1 Demographics

The predominant age bracket among participants fell within the range of 25 to 34 years old. There were 17 different nationalities and 22 countries given in the demographics form. The majority of participants were from the UK and Europe. There were 62 participants who identified as female, 67 participants who identified as male, 13 provided a different identification or left the field blank and 43 had errors. In regard to employment status, the majority of participants were either employed full-time or enrolled in university. The distribution of age groups and employment status among the 130 participants who took part in the study, after removing errors, is illustrated in Figure 3
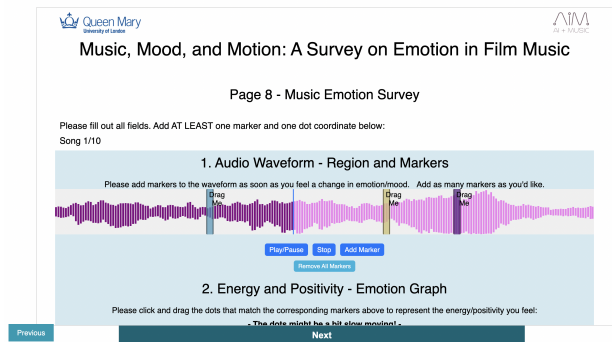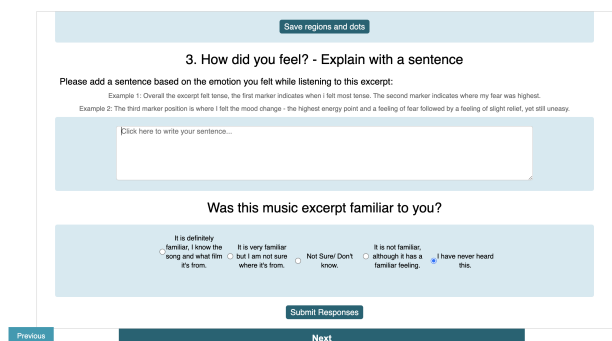
(a) Online study - Visual waveform with interactive markers
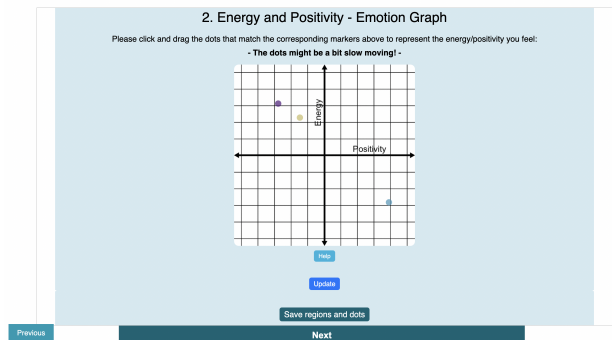
(b) Online study - Valence-Arousal interactive graph

(c) Online study - Emotion sentence and familiarity rating

Figure 2: Screenshot of online annotation collection study with interactive interface

and Table 1, respectively. It is clear from the results there is a gap in age range representation, more participants are needed.

4.2 Annotations

FME-24 consists of 300 compositions, with each composition featuring two randomly selected excerpts, thus totaling 600 samples in the dataset. These samples were annotated by 185 different annotators, each annotator provided annotations for 10 audio samples. Overall, there are 1580 individual annotations across the entire dataset. After eliminating errors, the average number of annotations per sample ranges from 4 to 5.

The number of annotations per sample varied among participants, ranging from a minimum of 1 annotation to a maximum of 18. The annotations consist of time stamps where participants felt a change in mood, correspond-
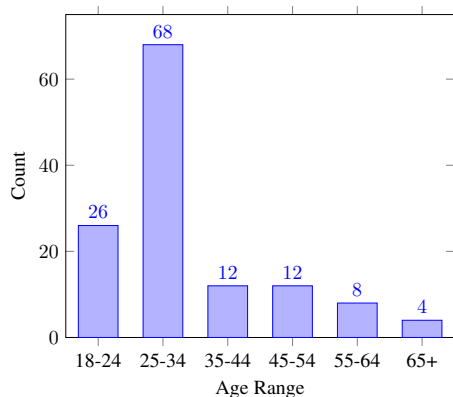
Figure 3: Distribution of Age Groups

Employment Status	Count
Full-time employment	**42**
Full-time parent	2
Other	6
Part-time employment	24
Retired	4
Self-employment	22
University	**44**

Table 1: Distribution of Employment Status

ing coordinates that represent the energy and positively felt at that time point in valence-arousal space, an emotion sentence from the participant about the excerpt, and a familiarity rating each. After filtering out entries containing errors and refining the dataset by eliminating entries with only one annotator, as well as excluding participants self-describing limited music interest or experience from the Goldsmiths Musical Sophistication Index (Gold-MSI) [4] assessments, the validation process confirmed 192 samples. While the study initially involved 185 annotators, subsequent error adjustments resulted in a final count of 130 participants. Increasing the number of participants and annotations would enhance the reliability of the dataset, as the numbers are significant but considerably insufficient for the scope of the data. Table 2 shows the summary of statistics in the dataset before and after refining the dataset and removing errors. Although this refinement reduces the size of annotations, participants and samples it enhances the overall quality and reliability of the data, ensuring more accurate and meaningful analysis.

	Original	Refined
Participants	185	130
Soundtracks	278	161
Samples	439	192
Total Annotations	1580	922
Annotations per Sample	3-4	4-5

Table 2: Summary statistics of the FME-24 dataset

4.3 Inter-rater Agreement

	Arousal	Valence
Range	0.418103	0.435423
IQR	0.208950	0.218257

Table 3: Range and Inter-Quartile Range (IQR) for Arousal and Valence

Table 3 shows the range and Inter-Quartile Range (IQR) for Arousal and Valence. The larger range in valence compared to arousal indicates greater variability in valence perceptions. This suggests that individuals typically encounter greater difficulty in perceiving valence compared to arousal. This observation is consistent with findings from previous research studies [16–18]. The dataset exhibits considerable variation in inter-rater agreement, suggesting that some songs are more consistently perceived in agreement than others. This could be due to many factors including; cultural differences, subjective nature of emotions, music and film, familiarity of the stimuli, personal preferences, music ability, contextual factors - the ambiguity of the story with it being sound only. During the online survey Gold-MSI was used to analyse music ability and preferences of the participants. This approach facilitates the identification of participants with higher levels of musical ability and facilitates further analysis of their responses.

5. DISCUSSION

Plotting the distribution of participant annotations reveals a notable consensus in arousal levels, see Figure 4, indicating a general alignment in the perceived energy across responses. While there might be minor discrepancies in valence interpretations, the overall trend suggests a shared understanding of the emotional intensity conveyed by the music. This observation is interesting in terms of film music as it highlights composers' ability to effectively evoke a consistent energetic atmosphere, even if participants have subjective responses to the positivity and negativity in music.

Figure 4 illustrates the emotion perception over time in the valence-arousal space of two audio files assessed by four different annotators. Upon visual inspection, it becomes apparent that there was a degree of consensus among these raters. While the plotted points are not densely clustered, they are not excessively sparse either, suggesting a moderate level of agreement. However, it is worth noting that the initial impression based solely on the first plotted point may lead to a different interpretation. The temporal aspect of the plotting may give the appearance of sparsity, but this could be attributed to annotators emphasizing perceptual changes rather than a lack of diversity. The black marker represents the first time point added by a participant, while the dotted lines show the progression to subsequent change points. Despite these variations, a consistent trend emerges, suggesting relative agreement on arousal or valence. Most plots exhibit clustering in either the top or bottom half of the valence-arousal space,
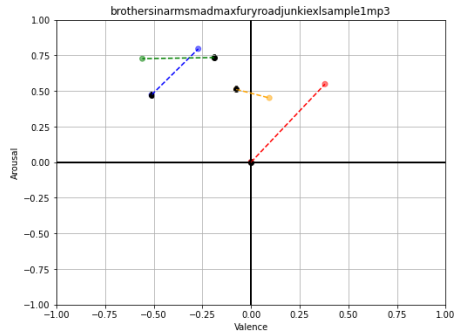
(a) AV participant annotations over-time for *'Brothers In Arms'*
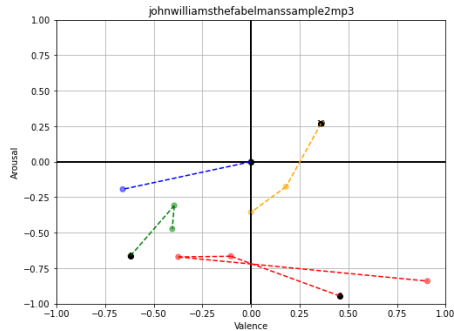
(b) AV participant annotations over-time for *'The Fabelmans'*

Figure 4: AV participant annotations over-time

(a) Arousal Over time for 2 participants for *'Remains - Volker Bertelmann'*

(b) Valence Over time for 2 participants for *'Remains - Volker Bertelmann'*

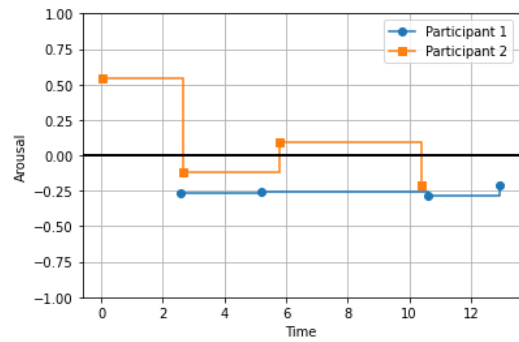Figure 5: AV participant annotations over-time for *'Remains - Volker Bertelmann'*

suggesting a consensus on arousal perception, even if disagreement in valence perception, this is also shown in Figure 4.

Figure 5 shows how 2 participants perceive arousal and valence over-time in the composition: *'Remains - Volker Bertelmann'* from *All Quiet on the Western Front*. The two participants seem to mostly agree on the time points where they perceive a change, in arousal there appears to be a lot of agreement but for valence the perceptions are almost opposite. This is interesting because the time-points are still relatively close together but complete disagreement in positivity and negativity but agreement in energy levels in the excerpt.
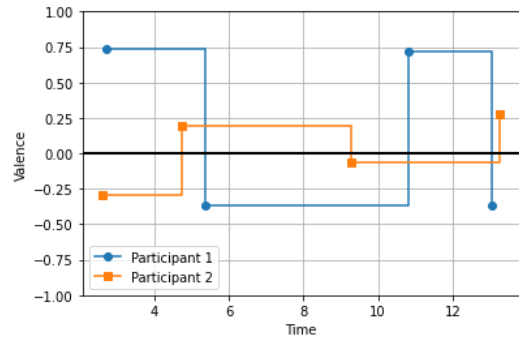
Looking at the distribution of annotations across genres within the dataset it is clear that participants seem to plot significantly more emotion ratings in the top half of the arousal-valence graph. This is shown in Table 4. These values are calculated from the refined dataset. Each column represents a quadrant of the arousal-valence plain, many of these genres are combined with other genres. The positive arousal areas have the highest number of annotations.

6. CONCLUSIONS AND FUTURE WORK

In summary, the FME-24 dataset is the first film composition dataset with valence-arousal annotations and timestamps. It will help provide insights on how individuals perceive arousal and valence in music. This comprehensive dataset, curated from diverse cultural sources and film festivals, reveals a remarkable consensus in the perceived

energy levels of compositions. These findings underscore the dataset's significance in advancing our understanding of music's emotional impact and its potential for future research. The FME-24 dataset serves as a valuable resource for exploring the intricate relationship between music and emotion for various contexts.

While the dataset comprises a substantial number of data points and participants, it's important to acknowledge that the available information per track is relatively limited at this stage. The current participant count is significant but insufficient for the study's scope. Increasing the sample size would enhance reliability. Further exploration and analysis may lead to deeper insights into the richness of the data and its potential for drawing robust conclusions. Examining further into the dataset, the aim is to uncover patterns in the distribution of arousal and valence over time. The objective is to identify correlations between shifts in emotional states, regardless of matching arousal and valence values. This involves closely analyzing the temporal aspect of data collection, scrutinizing the proximity of timestamps across different participants using various window intervals.

In future research, musical features will be extracted and compared with valence-arousal values using machine learning techniques. Further statistical analysis will be undertaken to understand how participants perceive changes in emotion over time. Various approaches, such as comparing the distance between arousal-valence points and analyzing emotion sentence data, will be employed to uncover

Genre	+A/+V	-A/-V	+A/-V	-A/+V
Biography	**39**	14	30	11
Comedy	**112**	12	44	27
Crime	28	16	**37**	10
Drama	**291**	122	167	123
History	**14**	0	10	0
Musical	**9**	0	7	0
Animation	**44**	6	14	7
Adventure	**96**	11	56	21
Sci-Fi	28	12	**30**	12
Thriller	26	10	**29**	11
Action	53	17	**64**	12
War	**16**	6	10	5
Romance	**63**	19	16	31
Mystery	48	18	**56**	20
Horror	23	12	**35**	9
Documentary	**9**	2	5	0
Music	**22**	4	9	2
Fantasy	**38**	7	15	11
Family	**10**	1	3	0
Western	**6**	0	0	0
Sport	**6**	4	3	0

Table 4: Distribution of Annotations by Genre Across Valence-Arousal Quadrants

additional insights from the collected data.

In conclusion, the FME-24 dataset is well-positioned to help improve our understanding of the intricate relationship between music and emotion in film, providing a valuable resource for future research and enhancing our comprehension of the profound influence of music on our emotional responses.

Acknowledgments

7. REFERENCES

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

[2] P. N. Juslin and P. Laukka, "Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening," *Journal of new music research*, vol. 33, no. 3, pp. 217–238, 2004.

[3] P. N. Juslin, *Musical emotions explained: Unlocking the secrets of musical affect*. Oxford University Press, USA, 2019.

[4] S. K. Jensen and N. Salmose, "Media and modalities: film," 2022.

[5] A.-K. Herget, "Well-known and unknown music as an emotionalizing carrier of meaning in film," *Media Psychology*, vol. 24, no. 3, pp. 385–412, 2021.

[6] E. Schubert, "Emotion felt by the listener and expressed by the music: literature review and theoretical perspectives," *Frontiers in psychology*, vol. 4, p. 837, 2013.

[7] P. N. Juslin, G. Barradas, and T. Eerola, "From sound to significance: Exploring the mechanisms underlying emotional reactions to music," *The American journal of psychology*, vol. 128, no. 3, pp. 281–304, 2015.

[8] C. Wei, T. Kronland-Martinet, Y. Frachi, and M. Barthet, "Influence of music on perceived emotions in film," in *Audio Engineering Society Convention 153*. Audio Engineering Society, 2022.

[9] J. A. Russell, "A circumplex model of affect." *Journal of personality and social psychology*, vol. 39, no. 6, p. 1161, 1980.

[10] P. Evans and E. Schubert, "Relationships between expressed and felt emotions in music," *Musicae Scientiae*, vol. 12, no. 1, pp. 75–99, 2008.

[11] M. Richards, "Film music themes: Analysis and corpus study." *Music Theory Online*, vol. 22, no. 1, 2016.

[12] H. D. Jones, "Euro-million arthouse films: Diverse and insightful stories, niche audiences," in *Transnational European Cinema: Representation, Audiences, Identity*. Springer, 2024, pp. 107–140.

[13] F. Morgan and J. Mooney, "'the same trade as mozart': Convincing the sceptics of electronic music's value," *The Journal of Popular Television*, vol. 9, no. 1, pp. 55–77, 2021.

[14] B. P. Walus, "A new modular approach to the composition of film music." Ph.D. dissertation, 2013.

[15] B. Ma, T. Greer, D. Knox, and S. Narayanan, "A computational lens into how music characterizes genre in film," *PloS one*, vol. 16, no. 4, p. e0249957, 2021.

[16] J. Grekow, "Audio features dedicated to the detection and tracking of arousal and valence in musical compositions," *Journal of Information and Telecommunication*, vol. 2, no. 3, pp. 322–333, 2018.

[17] B. Schuller, J. Dorfner, and G. Rigoll, "Determination of nonprototypical valence and arousal in popular music: features and performances," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2010, pp. 1–19, 2010.

[18] H. Egermann, N. Fernando, L. Chuen, and S. McAdams, "Music induces universal emotion-related psychophysiological responses: comparing canadian listeners to congolese pygmies," *Frontiers in psychology*, vol. 5, p. 1341, 2015.