HAL open science

Influence of Music on Perceived Emotions in Film

Chaoyang Wei, Thomas Kronland-Martinet, Yann Frachi, Mathieu Barthet

To cite this version:

Chaoyang Wei, Thomas Kronland-Martinet, Yann Frachi, Mathieu Barthet. Influence of Music on Perceived Emotions in Film. AES - Journal of the Audio Engineering Society Audio-Accoustics-Application, 2022. hal-04114230v2

HAL Id: hal-04114230

https://hal.science/hal-04114230v2

Submitted on 2 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Influence of Music on Perceived Emotions in Film

Chaoyang Wei[1], Thomas Kronland-Martinet[1,2,3], Yann Frachi[1], and Mathieu Barthet[1,*]

[1] *Centre for Digital Music, Queen Mary University of London, United Kingdom*
[2] *Aix Marseille Univ, CNRS, ISM, Marseille, France*
[3] *Aix Marseille Univ, CNRS, IM2NP, Marseille, France*
[*] *Corresponding author: Mathieu Barthet (m.barthet@qmul.ac.uk)*

Abstract

Film music plays a core role in film production and reception as it not only contributes to the film aesthetics and creativity, but it also affects viewers' experience and enjoyment. Film music composers often aim to serve the film narrative, immerse viewers into the setting and story, convey clues, and importantly, act on their emotions. Yet, how film music influences viewers is still misunderstood. We conducted a perceptual study to analyse the impact of music on the perception of emotions in film. We developed an online interface for time-based emotion annotation of audio/video media clips based on the Valence/Arousal (VA) two-dimensional model. Participants reported their perceived emotions over time in the VA space for three media conditions: film scene presented without sound (video only), film music presented without video (audio only), and film scene with accompanying music and sound effects (both video and audio modalities). 16 film clips were selected covering four clips for each of four genres (action & drama, romance, comedy, and horror). 38 participants completed the study (12 females and 26 males from many countries, average age: 28.9). Density scatter plots are used to visualise the spread of emotion ratings in the VA space and differences across media conditions and film clips. Results from linear mixed effect models show significant effects of the audiovisual media condition and film genre on VA ratings, in line with previous results by Parke et al. [32]. Perceived VA ratings across media conditions follow an almost linear relationship with an increase in strength in the following order: film alone, film with music/sound, music alone. We illustrate this effect by plotting the VA rating centre of mass across conditions. VA ratings for the film-alone condition are closer to the origin of the space compared to the two other media conditions, indicating that the inclusion of music yields stronger emotions (higher VA ratings). Certain individual factors (musical ability, familiarity, preference) also seem to impact the perception of arousal and valence while viewing films. Our online emotion annotation interface was on overall well received and suggestions to improve the display of reference emotion tags are discussed.

Keywords: film music, emotions, music perception, valence/arousal, online time-based emotion annotation, audiovisual production

© Wei, C., Kronland-Martinet, T., Frachi, Y., & Barthet, M. Licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). **Attribution:** Wei, C., Kronland-Martinet, T., Frachi, Y., & Barthet, M. (2022, October). Influence of Music on Perceived Emotions in Film. In Audio Engineering Society Convention 153. Audio Engineering Society.
Permalink: http://www.aes.org/e-lib/browse.cfm?elib=21959

1 Introduction

Film is a medium in which music plays an essential creative role. Juslin and Sloboda [23] suggest that film music is one of the most potent sources of emotion in film. Depending on film directors and composers' intentions, film music can follow, anticipate, or contrast the narrative, acting on the emotions and associations that are evoked. Film music is integral to the cinematic experience as it offers significant clues about the film characters and situations and heightens a film's emotional impact [5, 42]. The connection between film music and emotions is linked to the ability that music has to communicate emotional expression to listeners, certain authors describing it as the "language of emotions" [33]. Depending on context and individual factors, listeners can either perceive emotional expression in music (*perceived emotions*) without a change of their emotional state, or have their own emotions affected by music (*felt emotions*) [16, 28, 35]. Even though the importance of music in viewing films is acknowledged [13], there is a limited number of perceptual studies investigating the relationships between music and film, and how film music affects the perception (rather than induction) of emotions while viewing film, which is the focus of this work.

2 Related work

2.1 Link between music and emotion

Affect scientists have studied emotions at least since Darwin as emotion and affective states are pervasive in all forms of communication [36]. A common experimental method to collect emotional response is to present participants with linguistic labels or pictures of expressive faces. Music features complex acoustic and temporal structures that affect emotional response. People often listen to and engage with music because it can induce emotion and regulate mood [27]. Empirical investigations demonstrated that music stimuli could induce basic emotions across adults and young children, such as sadness, happiness, anger, and fear [9, 10]. Other works investigated the recognition of emotional expression in sound/music (perceived emotions). For example, [30] compared young and old adults' emotion recognition ability from acted speech, synthesized speech, and short electric guitar melodies with different emotion intensity. The results show that for both speech and music stimuli, old adults yielded lower recognition rates for negative emotions. Previous research showed that basic emotional expressions can be recognised across widely different cultures, although recognition performance increases when the musical style is congruent to participants' culture [15]. For example, Fritz et al. [15] report that Cameroon's Mafa tribe members were able to identify emotional expressions of happiness, sadness, or fear above chance level in Western music, but Western listeners have largely higher recognition rates for these expressions. Studies advanced that such universal emotional expressions in music are linked to a common structure shared between music and movement evident within and across cultures [39].

2.2 Music emotion models

Several general or domain-specific emotion models were created in scientific areas such as psychology, musicology, and neuroscience [1]. The most common models can be divided into two categories, categorical and dimensional [2]. Categorical models present emotions in a discrete way (e.g. clusters), while dimensional models organise emotions along continuous dimensions. The categorical emotion model is linked to the assumption that several emotion types can be universally recognised and understood by the human brain [11]. Ekman and Friesen [12] proposed six fundamental emotion components: happiness, anger, disgust, fear, surprise, and sadness. However, there are some shortcomings with the categorical model; emotions that do not fall into predetermined categories cannot be represented; the relationship between two different emotions cannot be analysed; the degree of an emotion cannot be clearly reflected. The dimension model was proposed to circumvent these disadvantages [17, 18]. One of the most widely used dimensional model is the Valence/Arousal (VA) model derived from the work by Russell [36] (see Figure 1). The VA model, which we have used in this study, is more precise and may simplify emotion annotations compared to categorical models. The valence dimension is related to the degree of pleasantness, and the arousal dimension is related to the degree of excitation. The VA model also has limitations considering music: it is not specific to music and fails to capture all the possible emotions expressed by music [4, 7]. Scherer argues that the aesthetic emotions caused by artistic works are often more subtle than and inconsistent with utilitarian emotions [38]. For example, guilt and shame are rarely evoked by music. Moreover, music emotions can be contradictory which is hard to reflect in the VA plane [22].

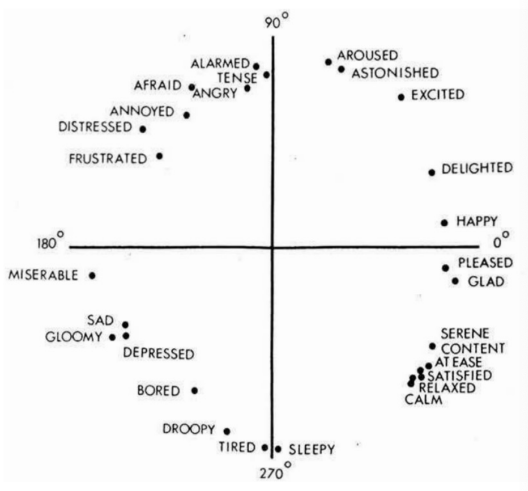


Figure 1: Valence-Arousal model from *'A circumplex model of affect'* by Russell [36]

2.3 Film music and emotion

Raksin suggests that the avowed purpose of music in a film is to help realize its meaning [34]. Film music can combine a series of images that are disconnected on their own and impart a rhythm to their unfolding [24]. It encourages audiences' absorption into the film by distracting them from its technological basis. The origins of film music can be traced back to Paris in the early 1890s [8]: Reynaud's animated Pantomimes lumineuses were presented in 1892 with piano music specially composed by Paulin. One of the most critical connections between film and music is how the combination of the audio and visual modalities influence the emotions of audiences. Parke et al. [32] studied

how film scenes with and without music impact perceived emotions using quantitative and visual analyses. 47 participants used a web interface to rate stress, activity, and dominance for film excerpts without sound (silent), with various soundtracks, and for the soundtracks alone. The results show that film music significantly affects the perceived emotions of viewers. The authors also highlight that the centre of mass of perceived emotion ratings for film clips with music is consistently located between the centres of mass of the film-alone and music-alone conditions. Lipscomb and Kendall investigated the relationship between film scenes and musical soundtracks [31]. Five scenes were selected from *Star Trek IV: The Voyage Home* and the video and composer-intended music were separated. Each visual excerpt was combined with each soundtrack resulting in 25 composites. Participants had to identify the most suitable musical score and rate all the composites on semantic differential scales. Results show a high accuracy in finding the composer-intended musical score and the correct pairs have higher mean scores on the evaluative dimension. However, they also found that the accuracy drops fast when there is no human in the scene. It may be because participants do not have the actor gestures and expressions to build a frame of reference. This study concluded that music has a strong and continuous influence on participants' responses, whatever the visual stimulus was. Bullerjahn and Güldenring studied how film music impacts the interpretation and perception of emotions in a scene and its outcome [6]. They created a 10-min long experimental film with five soundtracks in various genres, styles, orchestration, motifs, lengths, and time positions. Feedback from 412 participants was collected using standardized rating scales and open-ended questions related to their interpretation of the film clips. The results show that film music profoundly influences the perception of the film's plot, polarizes the emotional atmosphere and anticipated outcomes. For the same film scene, each musical soundtrack created a particular type of narrative and plot. For example, when the soundtrack was composed to support a crime film, the viewers believed that a crime or violent confrontation would happen. Tan et al. [41] analysed how the loudness of film music altered viewers' perceptions. The study found that adjusting the volume of the same piece of film music could bring a more significant difference in the interpretation of a film scene than changing the soundtrack. Film music can also shape viewers' understanding of characters. Tan et al. [40] studied interpretations of film characters' emotions. They invited 177 undergraduates to watch film excerpts with music being presented before or after a scene with a single character. They found that participants always tended to interpret the character's emotions based on the emotions expressed by the film music. Hoeckner et al. [21] analysed whether film music can influence the connection between viewers and the movie characters. Participants watched film clips showing a character's neutral or ambiguous reactions to some events or person, and rated the character likability and their certainty about the character's thoughts in three conditions: melodrama music, thriller music, and no music. The results showed that the ratings of character likability and certainty about the character's thoughts were significantly lower for the thriller music compared to the

melodrama music. This indicates that film music can influence viewers' perception of movie characters.

3 Time-based emotion annotations for film/music clips

We conducted an online perceptual study to collect emotion annotations from participants while they experienced film/music excerpts in conditions where the video and music modalities were either combined or separated. The study received ethics approval from the Queen Mary Ethics of Research Committee.

3.1 Film/music clip selection and experimental conditions

Film/music excerpts were selected from highly-rated films recognised for their soundtrack quality, most of which are discussed in Karlin and Wright's book [25]. The films belonged to four different genres (Comedy, Horror, Romance, and Action & Drama) as described in Table 1. 16 one-minute long scenes with non-diegetic music were chosen (no diegetic music), four for each genre (scene start timings are provided in Appendix 7.1). The scenes were selected as they contained limited or no dialogues and the sound effects were diegetic sounds (e.g. footsteps). In scenes with dialogues, English subtitles were displayed so that the narrative could be followed even without sound. Three media conditions were prepared for each excerpt: (i) film scene without music and sound effects (video only), (ii) film music without video (audio only), and (iii) film scene with music and sound effects (both video and audio). For the condition combining film with music and sound effects, the clip corresponded to the original film excerpt with sound. For the film-only condition, the original audio soundtrack was removed. The music-only condition was obtained by synchronising the commercial soundtrack to the film clip and slicing it to match the length of the scene. This way the music-only condition features the music but not the dialogues nor sound effects. The audio stimuli were all loudness equalised. In total, 48 media clips were gathered, all encoded in the MP4 format.

Table 1: Title and genre of selected film clips

Film clips title	**Genre**
Airplane!	Comedy
The Pink Panther	Comedy
Bean The Ultimate Disaster Movie	Comedy
Beverly Hills Cop	Comedy
A Nightmare On Elm Street (1984)	Horror
Psycho	Horror
Jaws	Horror
The Shining	Horror
Braveheart	Romance
Manhattan	Romance
Out of Africa	Romance
Sleepless in Seattle	Romance
Gladiator	Action & Drama
Speed	Action & Drama
The Fugitive	Action & Drama
Vertical Limit	Action & Drama

3.2 Online emotion annotation interface

Emotion self reports can provide information on the otherwise inaccessible cognitive part underlying emotions [23]. We developed an online emotion annotation interface based on the Mood Conductor [14] and

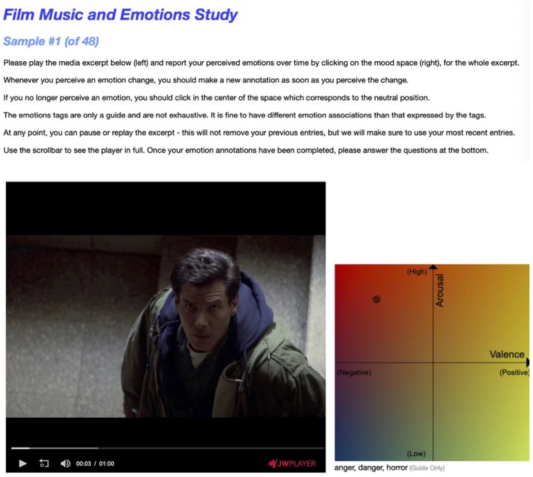


Figure 2: Online emotion rating interface synchronised to film/music clips

Mood Rater [43] frameworks using an interactive two-dimensional Valence/Arousal space for self reports, synchronised with film/music clips. The interface enables participants to rate emotions by clicking with the mouse in the VA space at a given time while watching/listening to a media clip. Illustrative mood tags are displayed under the VA space to facilitate the understanding of the VA space. Figure 2 displays a screenshot of the interface with an example of emotional annotation corresponding to the tags "anger, danger, horror". The timestamp and VA data are stored in a Google Firebase database. Dedicated survey questionnaires are also integrated into the interface. The interface was developed using node.js with the Glitch framework. The media clips are stored online using JWPlayer.

3.3 Procedure

The study was composed of three main parts, training, emotion rating task, and questionnaire survey. The training stage provided background information on the Valence/Arousal model with music examples for the four quadrants of the VA space (happy, relaxing, angry, sad). Participants could try the interactive VA space while watching film clips (not used in the main task). They were also instructed to adjust the sound level during this stage and to avoid changing it during the task. The study took about 1h30 mins to complete and participants were compensated for their time.

Participants had to make an emotion annotation whenever and as soon as they perceived an emotion change in the media clip. During the main task, participants annotated 48 media clips (16 x 3 media conditions) which were presented sequentially. To minimise any order effect, the films were randomised across participants and for each film, the order to the media conditions was also randomised.

For each clip, participants had to answer the open-ended question *"What aspects of the film and/or music contributed to the perceived emotions?"*. As familiarity with the film/music and enjoyment may influence emotion ratings, we also asked participants how familiar they were with the media clip and how much they liked it, using five-point Likert items.

After completion of the main task, participants had to complete a short survey. Feedback on the emotion annotation interface was collected also using five-point Likert items. The following demographics information were collected: age, gender, occupation, nationality, country of residence. Participants also had to report the model or earphone/headphone used for the study, their musical proficiency, and their English

language proficiency (which may affect understanding of instructions, mood tags, dialogue subtitles). They also had to report results of an Internet-speed test integrated to the interface (to control for potential streaming issues), and which web browser they used.

3.4 Participants

Participants over the age of 18 were recruited using the Prolific platform[1] as well as departmental and student mailing lists. Participants were instructed that they should sit in a quiet environment, and use a computer with headphones or earphones.

Fully completed surveys were collected for 30 participants. We also used results from another 8 participants who had missing emotion rating data for a maximum of three clips, making a total of 38 participants (12 females and 26 males with an average age of 28.9 years). Participants were from many different nationalities: British (8), Portuguese (5), French (4), Italian (3), Polish (3), other participants are Turkish, Russian, Greek, Egyptian, Mexican, Hungarian, American, Romanian, Norwegian, Moroccan and Chinese. 23 participants were students, others were working in different areas such as education. No participants had hearing nor significant visual impairments. A large number of participants had a good command of English. Regarding musical ability, participants included non musicians (14), beginners (8), intermediate (7), advanced (5), and professionals (4).

[1] https://www.prolific.co/

Krippendorff's Alpha Reliability Estimate for Valence						
	Alpha	LL95%CI	UL95%CI	Units	Observrs	Pairs
Interval	.5665	.5593	.5720	48.0000	38.0000	33304.0000
Krippendorff's Alpha Reliability Estimate for Arousal						
	Alpha	LL95%CI	UL95%CI	Units	Observrs	Pairs
Interval	.3418	.3306	.3532	48.0000	38.0000	33304.0000

Figure 3: Krippendorff's alpha reliability estimate results

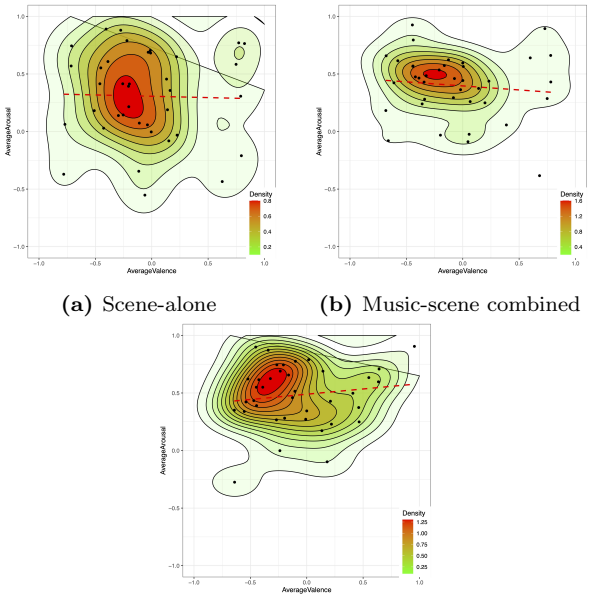
3.5 Emotion annotation resampling

Participants made several emotion annotations for each media clip at discrete times (raw rating data). In order to be able to compare annotations across participants and take into account the time span for each reported emotion, the VA raw rating data were resampled. A sampling period of 1 s was chosen. In the resampled signal, an emotion annotation gets repeated until the following new emotion annotation. The average perceived emotion computed over the resampled VA ratings take into account how long each emotion lasted. This is a difference in our study compared to [32] where only one summative emotion rating was collected for a whole clip.

4 Effects of film and music modalities on perceived emotions

4.1 Inter-rater reliability (IRR)

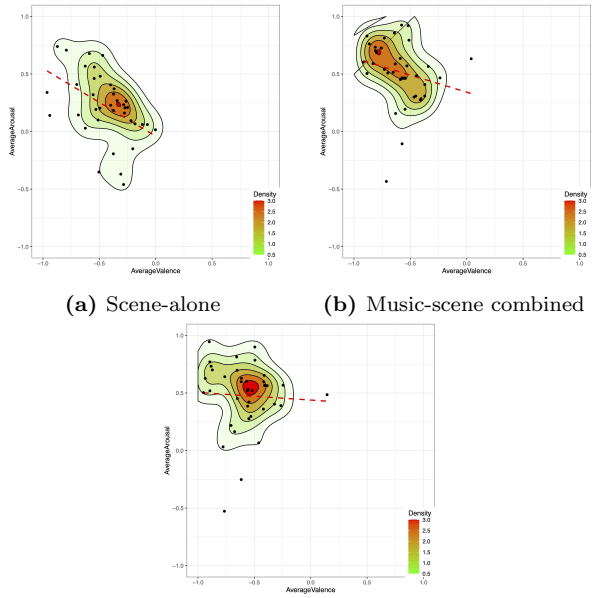
For each participant and media clip, we computed the mean valence and arousal across time using the resampled data. The means were subjected to inter-rater reliability (IRR) analysis. We chose Krippendorff's alpha [20] as IRR measure given that it handles missing data [19]. Results are reported in Figure 3. A Krippendorff alpha of 0.57 was obtained for valence and 0.34 for arousal.



(a) Scene-alone **(b)** Music-scene combined

(c) Music-alone

Figure 4: Emotion distribution for the three media conditions of the *AirPlane!* scene



(a) Scene-alone **(b)** Music-scene combined

(c) Music-alone

Figure 5: Emotion distribution for the three media conditions of the *Shining* scene

Although agreement between participants is well above chance indicating the presence of a general perceived emotion trend, there is some disagreement between participants, meaning different participants may provide different VA ratings. This result supports the findings in Kuppens et al. [29] and Yang et al. [43] which report significant variations in valence and arousal between individuals.

4.2 Emotion distribution analysis

In order to visually analyse the distribution of VA ratings across the 38 participants for each clip, we generated density scatter plots for the average valence and arousal based on the resampled ratings. Figures for all the media clips are provided in Appendix 7.2.

Figure 4 shows the density scatter plots of the three media conditions for the comedy film *AirPlane!*. The most frequent perceived emotions are represented by the darkest colour zones. For example, the most frequently perceived emotion in the film-only condition for *AirPlane!* has a valence of about -0.2 and an arousal of about 0.25. For the *AirPlane!* clips, the emotion distribution in the VA space spreads across two or more quadrants which indicates a fairly large variance across participants (in line with the relatively low Krippendorff alpha reported above). This spread is often observed for the other film clips (see Appendix 7.2). However, as can be seen in Figure 5, the variance for the three media conditions of the *Shining* excerpt is much smaller. For *AirPlane!*, the spread may be due to the contrasting emotions expressed by the scene, some participants focusing on the comic aspects and others on the tense aspects exacerbated by the music as the captain looses control of the

plane. In contrast, the emotions expressed through the *Shining* scene are more homogeneous and mostly located in the top left quadrant (low valence, high arousal). For both film clips, the location of the highest density point slightly shifts across the three media conditions; this occurs for most film clips. These observations suggest that there are effects of the media conditions and genre on perceived valence and arousal which we assess more closely in the next section.

Information Criteria[a]

-2 Restricted Log Likelihood	842.280
Akaike's Information Criterion (AIC)	856.280
Hurvich and Tsai's Criterion (AICC)	856.342
Bozdogan's Criterion (CAIC)	901.818
Schwarz's Bayesian Criterion (BIC)	894.818

The information criteria are displayed in smaller-is-better form.

a. Dependent Variable: AverageValence.

Type III Tests of Fixed Effects[a]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	18.557	3.069	.096
Genre	3	13.837	27.749	.000
Media_condition	2	33.686	6.751	.003

a. Dependent Variable: AverageValence.

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval Lower Bound	Upper Bound
Intercept	.441965	.075788	19.577	5.832	.000	.283655	.600275
Genre[Comedy]	-.237881	.098011	13.837	-2.427	.029	-.448326	-.027435
Genre[Horror]	-.810307	.098011	13.837	-8.268	.000	-1.020753	-.599862
Genre[Romance]	-.614949	.098011	13.837	-6.274	.000	-.825395	-.404504
Genre[Action&Drama]	0[b]	0	.	.	.	.	.
Media condition[Music-scene combined]	-.105226	.046858	33.686	-2.246	.031	-.200486	-.009967
Media condition[Scene-alone]	-.170646	.046858	33.686	-3.642	.001	-.265905	-.075386
Media condition[Music-alone]	0[b]	0	.	.	.	.	.

a. Dependent Variable: AverageValence.

b. This parameter is set to zero because it is redundant.

Figure 6: Media condition and genre linear mixed effect model results for valence

4.3 Effects of media condition and genre: linear mixed effects model

We used linear mixed effects (LME) models to assess whether the media condition (three levels: music-scene combined, music-alone, scene-alone) and genre have a significant influence on perceived emotions. A Type I error of 0.05 is considered in the following analyses. Two models were established with valence and arousal as dependent variables. The media condition and genre were set as fixed effects. Four random effects were considered: intercept per participant (participants may rate emotions differently), intercept per film clip (film clips may express different emotions), slope per participant considering media condition (media condition effects may not be the same for all participants), slope per film clip considering media condition (media condition effects may not be the same for all clips). Figures 6 and 7 show the results of the LME models for valence and arousal respectively. The information criteria show how the LME models fit the data, the smaller the number, the best the model fit. The media condition yields a significant effect on both valence (p=.003) and arousal (p=.005). The film genre also significantly affects valence (p¡.001) and arousal (p=.001).

4.4 Centre of mass plots

Centre of mass plots were analysed to illustrate the effects of media conditions on average valence and arousal. Figure 8 displays the centre of mass for the three media conditions for the *AirPlane!* excerpt. Each media condition is represented by a different colour (green: scene only, red: film with music/sound, blue: music only). It can be seen

Information Criteria[a]

-2 Restricted Log Likelihood	588.404
Akaike's Information Criterion (AIC)	602.404
Hurvich and Tsai's Criterion (AICC)	602.466
Bozdogan's Criterion (CAIC)	647.942
Schwarz's Bayesian Criterion (BIC)	640.942

The information criteria are displayed in smaller-is-better form.

a. Dependent Variable: AverageArousal.

Type III Tests of Fixed Effects[a]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	16.797	55.823	.000
Genre	3	14.623	8.889	.001
Media_condition	2	48.249	5.832	.005

a. Dependent Variable: AverageArousal.

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Intercept	.009100	.070000	17.602	.130	.898	-.138204	.156404
Genre[Comedy]	.230232	.094463	14.622	2.437	.028	.028434	.432029
Genre[Horror]	.318646	.094463	14.622	3.373	.004	.116848	.520443
Genre[Romance]	.477037	.094463	14.622	5.050	.000	.275239	.678834
Genre[Action&Drama]	0[b]	0	.	.	.	.	.
Media condition[Music-scene combined]	.045573	.032382	48.249	1.407	.166	-.019528	.110673
Media condition[Scene-alone]	-.064477	.032382	48.249	-1.991	.052	-.129577	.000624
Media condition[Music-alone]	0[b]	0	.	.	.	.	.

a. Dependent Variable: AverageArousal.

b. This parameter is set to zero because it is redundant.

Figure 7: Media condition and genre linear mixed effect model results for arousal

that the film with music and sound effects condition tends to be located between the scene-only and music-only conditions on an almost linear trajectory. For some of the film clips, we obtain results in line with the following hypothesis proposed in [32]: perceived VA of a film scene with music occurs most nearly on a linear trajectory between the VA of the scene- and music-only versions. However, as can be seen in Appendix 7.2, there are film clips for which the centres of mass of the three conditions are related in nonlinear ways.

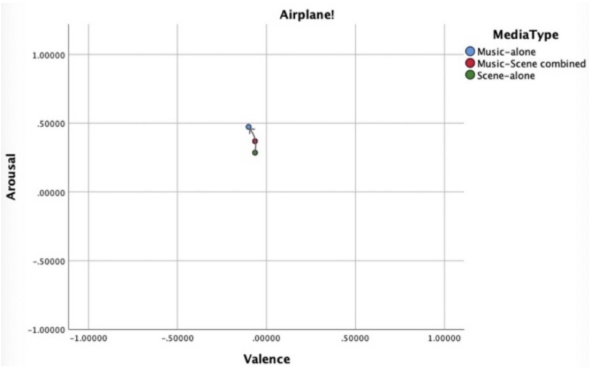


Figure 8: VA center of mass in three media conditions for *AirPlane!*

4.5 Emotion strength

We computed the distances between the average emotion and the origin of the VA space for the different media conditions. Results show that the perceived emotion for the scene-only clip is always much closer to the origin compared to the two other media conditions (average distance to VA space origin - film without music and sound effects: 0.387, film with music and sound effects: 0.509, music alone: 0.515). This finding suggests that a film scene expresses less strong emotions without music. It is in line with the hypothesis reported in Baumgartner et al. [3] that music can remarkably enhance the perceived emotion in a scene.

4.6 Effects of individual factors: linear mixed effects model

We used linear mixed effect models to assess the effect of gender, age, musical ability, occupation, media familiarity and preference (fixed effects), on average valence and arousal. Clip items were considered as a random factor (film clips and media conditions may affect emotions). To have

Type III Tests of Fixed Effects[a]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	48.524	3.220	.079
Familiarity	4	205.527	6.084	.000
Preference	4	1764.107	28.181	.000
Age	1	1712.438	.139	.709
Gender	1	1703.993	1.365	.243
MusicalAbility	1	83.597	.036	.851
Occupation	1	1708.555	1.043	.307

a. Dependent Variable: AverageValence.

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Intercept	.069241	.055338	84.748	1.251	.214	-.040790	.179272
Strongly disagree familiar with the media [Familiarity=1]	-.016161	.031342	330.704	-.516	.606	-.077816	.045494
Disagree familiar with the media [Familiarity=2]	.058636	.029258	283.852	2.004	.046	.001045	.116226
Neutral[Familiarity=3]	.053423	.032493	373.804	1.644	.101	-.010469	.117315
Agree familiar with the media[Familiarity=4]	-.034419	.027361	211.004	-1.258	.210	-.088356	.019517
Strongly agree familiar with the media [Familiarity=5]	0[b]	0	.	.	.	.	.
Strongly disagree enjoyed the media [Preference=1]	-.248263	.041080	1767.708	-6.043	.000	-.328834	-.167692
Disagree enjoyed the media[Preference=2]	-.299850	.029306	1772.038	-10.232	.000	-.357328	-.242371
Neutral[Preference=3]	-.169359	.026088	1773.873	-6.492	.000	-.220525	-.118193
Agree enjoyed the media[Preference=4]	-.112852	.022630	1761.174	-4.987	.000	-.157236	-.068468
Strongly agree enjoyed the media [Preference=5]	0[b]	0	.	.	.	.	.
Participants younger than 26 years old including 26[Age=0]	-.007570	.020316	1712.438	-.373	.709	-.047417	.032277
Participants older than 26 years old not including 26[Age=1]	0[b]	0	.	.	.	.	.
Female[Gender=0]	.019998	.017117	1703.993	1.168	.243	-.013575	.053571
Male[Gender=1]	0[b]	0	.	.	.	.	.
Low musical ability [MusicalAbility=0]	.003397	.017982	83.597	.189	.851	-.032363	.039158
High musical ability [MusicalAbility=1]	0[b]	0	.	.	.	.	.
Not student [Occupation=0]	-.019197	.018794	1708.555	-1.021	.307	-.056058	.017665
Student[Occupation=1]	0[b]	0	.	.	.	.	.

a. Dependent Variable: AverageValence.

b. This parameter is set to zero because it is redundant.

Figure 9: Individual factor linear mixed effect model results for valence

Type III Tests of Fixed Effects[a]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	50.836	63.323	.000
Familiarity	4	178.220	1.012	.403
Preference	4	1772.074	3.259	.011
Age	1	1755.299	.196	.658
Gender	1	1730.446	.111	.739
MusicalAbility	1	1741.426	9.267	.002
Occupation	1	1743.396	.303	.582

a. Dependent Variable: AverageArousal.

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Intercept	.365491	.042432	133.263	8.614	.000	.281564	.449417
Strongly disagree familiar with the media [Familiarity=1]	-.047541	.030693	287.410	-1.549	.122	-.107951	.012870
Disagree familiar with the media [Familiarity=2]	-.055452	.028697	246.170	-1.932	.054	-.111974	.001070
Neutral[Familiarity=3]	-.044200	.031802	324.668	-1.390	.166	-.106763	.018363
Agree familiar with the media[Familiarity=4]	-.030096	.026925	184.091	-1.118	.265	-.083218	.023026
Strongly agree familiar with the media [Familiarity=5]	0[b]	0	.	.	.	.	.
Strongly disagree enjoyed the media [Preference=1]	-.055473	.039690	1774.739	-1.398	.162	-.133316	.022370
Disagree enjoyed the media[Preference=2]	-.079197	.028315	1781.186	-2.797	.005	-.134732	-.023663
Neutral[Preference=3]	-.070833	.025206	1782.588	-2.810	.005	-.120269	-.021396
Agree enjoyed the media[Preference=4]	-.021732	.021901	1766.211	-.992	.321	-.064686	.021223
Strongly agree enjoyed the media [Preference=5]	0[b]	0	.	.	.	.	.
Participants youger than 26 years old including 26[Age=0]	-.008729	.019711	1755.299	-.443	.658	-.047388	.029930
Participants older than 26 years old not including 26[Age=1]	0[b]	0	.	.	.	.	.
Female[Gender=0]	.005526	.016597	1730.446	.333	.739	-.027025	.038077
Male[Gender=1]	0[b]	0	.	.	.	.	.
Low musical ability [MusicalAbility=0]	-.050084	.016453	1741.426	-3.044	.002	-.082353	-.017815
High musical ability [MusicalAbility=1]	0[b]	0	.	.	.	.	.
Not student [Occupation=0]	.010026	.018225	1743.396	.550	.582	-.025718	.045771
Student[Occupation=1]	0[b]	0	.	.	.	.	.

a. Dependent Variable: AverageArousal.

b. This parameter is set to zero because it is redundant.

Figure 10: Individual factor linear mixed effect model results for arousal

a balanced number of participants across age groups, we divided participants into a younger group (age ≤ 26) and an older group (age > 26). We also grouped participants according to their music ability, a lower musical ability group (not a musician, beginner) and a higher musical ability group (intermediate, advanced, professional). We divided participants based on their occupation (student or not). After testing the restricted log-likelihood, we found that the model with random effects influencing both gradient and intercept fitted the data better. The results for this model are reported in Figures 9 and 10.

No effects were found for the age, occupation and gender groups. However, the participants' preference were found to significantly influence both valence (p¡.001) and arousal (p=0.011). Results also show that familiarity with the film significantly influences valence (p¡.001) but not arousal. Another finding is that musical ability can influence arousal significantly but not valence. It could be because the perception of arousal is less pronounced or important for participants with lower musical ability.

Figure 11: Evaluation of emotion annotation interface for film/music clips

4.7 Evaluation of emotion annotation interface

The interface was assessed by measuring agreement to Likert items between 1 (strongly disagree) to 5 (strongly agree). The results are shown in Figure 11.

Most participants agreed that it was easy to rate the emotions with the interface (Mean=3.37, SD=1.15). However, two out of 38 participants strongly disagreed that the interface made emotion rating easy. Most participants disagreed that it was complex to rate the emotions while listening/watching the media (Mean=2.70, SD=0.95). However, one participant found it difficult to self report emotions given the task. Most participants disagreed that the emotion tags at the bottom of the interface distracted them (Mean=2.30, SD=1.20). Mood tags were found helpful by most participants when rating perceived emotions (Mean=3.85, SD=1.06). Participants also gave some feedback to improve the interface. Three participants mentioned that certain emotions cannot be selected with the interface (limitation of the VA model), and that some mood tags were misplaced. Four participants also reported that the emotion tags may influence their perceived emotion judgement. Two participants suggested that it could be helpful to display emojis along tags when selecting emotion annotations.

4.8 Discussion and limitations

Most of the selected scenes only contain common and expected diegetic sound effects used for realism. However, in addition to music, the presence of prominent sound effects may also influence emotion ratings. Five out of 16 clips include at times some pronounced sound effects which may have influenced the participants' emotion judgements (see Appendix 7.1). For example, the scene from "A Nightmare on Elm Street" features a sound used to increase emotional duress (a surprisingly loud door slam). The analysis of the relative influence of music and sound effects (whether diegetic or non diegetic) would require to separate them from each other. This could be achieved using original soundtrack production multitracks (assuming availability), or an adequate source separation model. This would for instance enable to produce a version of a film scene with sound effects and dialogues but without music. However, even if this process was technically achievable, it could pose challenging questions since the boundary between music and sound effects can sometimes be voluntarily blurry for aesthetic reasons [26]. The emotional expression of characters communicated auditorily in dialogues could also influence emotion perception. Investigating the sonic influence of dialogues on emotions separately from music

and sound effects would also deserve further investigations. Relatively to the overall duration of the clips, the presence of dialogues and pronounced sound effects remains limited in the scenes selected for this study, which gives music more weight in explaining observed differences.

In this study, we used the two-dimensional Valence/Arousal model to represent perceived emotions. This model has some limitations as it is not sufficient to distinguish certain emotions (e.g. fear from anger) and more elaborated models have been proposed, such as Russell and Mehrabian's three-dimensional Valence/Arousal/Dominance (VAD) model [37]. However, we wanted to keep the emotion rating task simple for participants by limiting the cognitive load for time-based emotional annotations. Feedback on the interface shows that most participants found it easy to rate emotions using the proposed VA-based interface that also displays reference tags (see Section 4.7). Some participants however noted that some emotions could not be selected in the proposed interface, and that some of the reference emotion tags were misplaced, indicating that there is scope for further improvement in the Mood Rater interface (e.g. adding emojis along with tags when selecting a location in the VA space).

To better understand individual differences (see moderate Krippendorff alpha reported in Section 4.1) and represent a wider range of compositional techniques and strategies in film music, a wider pool of participants and film excerpts should be used in follow-up studies.

The linear mixed effects model results indicate that there is scope for predicting perceived emotions for a film scene with music from ratings obtained from conditions with isolated audio and visual modalities (film alone and music alone), which is in line with [32]. However, qualitative observations indicate that nonlinear models would be better suited (see Figure 8).

A more in-depth analysis of the relationship between the film music, film genre, and emotion ratings would require to take into account the dramaturgical approach that has been used by the composer for the film scene (e.g. whether the music is in line or contrasting the narrative). However, obtaining knowledge about the director and composer's creative intentions is challenging and techniques such as interviews are prone to participants' consent and availability.

5 Conclusion

We conducted an online perceptual experiment to assess the influence of film music on perceived emotions. We collected time-based valence and arousal from 38 participants for 16 different film clips in three media conditions (film, music, film+music). A moderate inter-rater reliability was obtained for valence and arousal which is in line with previous works by Kuppens et al. [29] and Yang et al. [43] which reported significant variance across participants. Linear mixed effect model results suggest that there are significant effects due to the presence or absence of music (media condition). Film genre also appears to influence perceived valence and arousal significantly, as can be expected. By modeling the strength of a perceived emotion by the distance to the origin of the VA space, results indicate that the emotions expressed by a film tend to be stronger with than without music. Sur-

prisingly, when participants listened to film music alone without viewing the scenes, valence tended to be perceived more positively (no significant difference for arousal). Certain individual factors (musical ability, familiarity, and preference) were shown to significantly affect perceived emotions, while others (age, occupation, and gender) did not. This probes for further research to elicit relationships between individual factors and perception of emotions in film music. Further extension of this work could also focus on analyzing variations of perceived emotions over time and how these related to acoustic and film attributes.

6 Acknowledgements

This study was partly supported by industry partner Ovomind. We would like to thank the reviewers for their useful comments.

References

[1] Aljanaki, A., Wiering, F., and Veltkamp, R. C. (2016). Studying emotion induced by music through a crowdsourcing game. *Information Processing & Management*, 52(1):115–128.

[2] Barthet, M., Fazekas, G., and Sandler, M. (2012). Music emotion recognition: From content-to context-based models. In *Int. Sympos. on Computer Music Modeling and Retrieval*, pages 228–252. Springer.

[3] Baumgartner, T., Esslen, M., and Jäncke, L. (2006). From emotion perception to emotion experience: Emotions evoked by pictures and classical music. *Int. J. of Psychophysiology*, 60(1):34–43.

[4] Bigand, E., Vieillard, S., Madurell, F., Marozeau, J., and Dacquet, A. (2005). Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts. *Cognition & Emotion*, 19(8):1113–1139.

[5] Brown, R. S. (1994). *Overtones and undertones: Reading film music*. Univ of California Press.

[6] Bullerjahn, C. and Güldenring, M. (1994). An empirical investigation of effects of film music using qualitative content analysis. *Psychomusicology*, 13(1-2):99.

[7] Collier, G. L. (2007). Beyond valence and activity in the emotional connotations of music. *Psychology of Music*, 35(1):110–131.

[8] Cooke, M. (2008). *A history of film music*. Cambridge University Press.

[9] Cunningham, J. G. and Sterling, R. S. (1988). Developmental change in the understanding of affective meaning in music. *Motivation and emotion*, 12(4):399–413.

[10] Dolgin, K. G. and Adelson, E. H. (1990). Age changes in the ability to interpret affect in sung and instrumentally-presented melodies. *Psychology of Music*, 18(1):87–98.

[11] Ekman, P. and Friesen, W. V. (1971). Constants across cultures in the face and emotion. *J. of Personality and Social Psychology*, 17(2).

[12] Ekman, P. and Friesen, W. V. (2003). *Unmasking the face: A guide to recogniz-*

ing emotions from facial clues, volume 10. Ishk.

[13] Ellis, R. J. and Simons, R. F. (2005). The impact of music on subjective and physiological indices of emotion while viewing films. *Psychomusicology*, 19(1):15.

[14] Fazekas, G., Barthet, M., and Sandler, M. B. (2013). Novel methods in facilitating audience and performer interaction using the mood conductor framework. In *Int. Sympos. on Computer Music Multidisciplinary Research*, pages 122–147. Springer.

[15] Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., Friederici, A. D., and Koelsch, S. (2009). Universal recognition of three basic emotions in music. *Current biology*, 19(7):573–576.

[16] Gabrielsson, A. (2001). Emotion perceived and emotion felt: Same or different? *Musicae scientiae*, 5(1_suppl):123–147.

[17] Gunes, H. and Pantic, M. (2010). Automatic, dimensional and continuous emotion recognition. *Int. J. of Synthetic Emotions (IJSE)*, 1(1):68–99.

[18] Gunes, H. and Schuller, B. (2013). Categorical and dimensional affect analysis in continuous input: Current trends and future directions. *Image and Vision Computing*, 31(2):120–136.

[19] Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in quantitative methods for psychology*, 8(1):23.

[20] Hayes, A. F. and Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication methods and measures*, 1(1):77–89.

[21] Hoeckner, B., Wyatt, E. W., Decety, J., and Nusbaum, H. (2011). Film music influences how viewers relate to movie characters. *Psychology of Aesthetics, Creativity, and the Arts*, 5(2):146.

[22] Hunter, P. G., Schellenberg, E. G., and Schimmack, U. (2008). Mixed affective responses to music with conflicting cues. *Cognition & Emotion*, 22(2):327–352.

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

[24] Kalinak, K. (2010). *Film music: A very short introduction*. Oxford University Press.

[25] Karlin, F. and Wright, R. (2013). *On the track: A guide to contemporary film scoring*. Routledge.

[26] Knight-Hill, A. (2019). Sonic diegesis: reality and the expressive potential of sound in narrative film. *Quarterly Review of Film and Video*, 36(8):643–665.

[27] Knobloch, S. and Zillmann, D. (2002). Mood management via the digital jukebox. *J. of Communication*, 52(2):351–366.

[28] Konečni, V. J. (2008). Does music induce emotion? a theoretical and methodological analysis. *Psychology of Aesthetics, Creativity, and the Arts*, 2(2):115.

[29] Kuppens, P., Tuerlinckx, F., Russell, J. A., and Barrett, L. F. (2013). The relation between valence and arousal in subjective experience. *Psychological bulletin*, 139(4):917.

[30] Laukka, P. and Juslin, P. N. (2007). Similar patterns of age-related differences in emotion recognition from speech and music. *Motivation and Emotion*, 31(3):182–191.

[31] Lipscomb, S. D. and Kendall, R. A. (1994). Perceptual judgement of the relationship between musical and visual components in film. *Psychomusicology*, 13(1-2):60.

[32] Parke, R., Chew, E., and Kyriakakis, C. (2007). Quantitative and visual analysis of the impact of music on perceived emotion of film. *Computers in Entertainment (CIE)*, 5(3):5.

[33] Pratt, C. C. (1948). Music as a language of emotion. *Bulletin of the American Musicological Society*, 11:67–68.

[34] Prendergast, R. M. (1992). *Film music: a neglected art- A critical study of music in films*. WW Norton & Company.

[35] Rickard, N. S. (2004). Intense emotional responses to music: a test of the physiological arousal hypothesis. *Psychology of music*, 32(4):371–388.

[36] Russell, J. A. (1980). A circumplex model of affect. *J. of Personality and Social Psychology*, 39(6):1161.

[37] Russell, J. A. and Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *J. of Research in Personality*, 11(3):273–294.

[38] Scherer, K. R. (2004). Which emotions can be induced by music? what are the underlying mechanisms? and how can we measure them? *J. of New Music Research*, 33(3):239–251.

[39] Sievers, B., Polansky, L., Casey, M., and Wheatley, T. (2013). Music and movement share a dynamic structure that supports universal expressions of emotion. *Proc. of the National Academy of Sciences*, 110(1):70–75.

[40] Tan, S.-L., Spackman, M. P., and Bezdek, M. A. (2007). Viewers' interpretations of film characters' emotions: Effects of presenting film music before or after a character is shown. *Music Perception*, 25(2):135–152.

[41] Tan, S.-L., Spackman, M. P., and Wakefield, E. M. (2017). The effects of diegetic and nondiegetic music on viewers' interpretations of a film scene. *Music Perception*, 34(5):605–623.

[42] Wierzbicki, J. (2009). *Film music: A history*. Routledge.

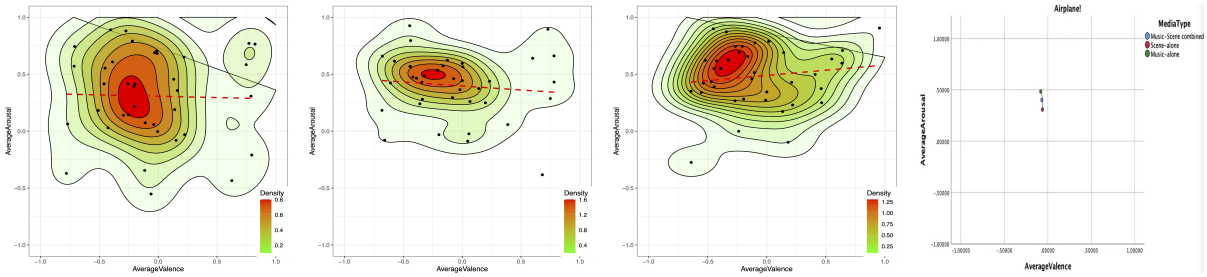
[43] Yang, S., Reed, C. N., Chew, E., and Barthet, M. (2021). Examining emotion perception agreement in live music performance. *IEEE Transactions on Affective Computing*.

7 Appendix

7.1 Film clip characteristics

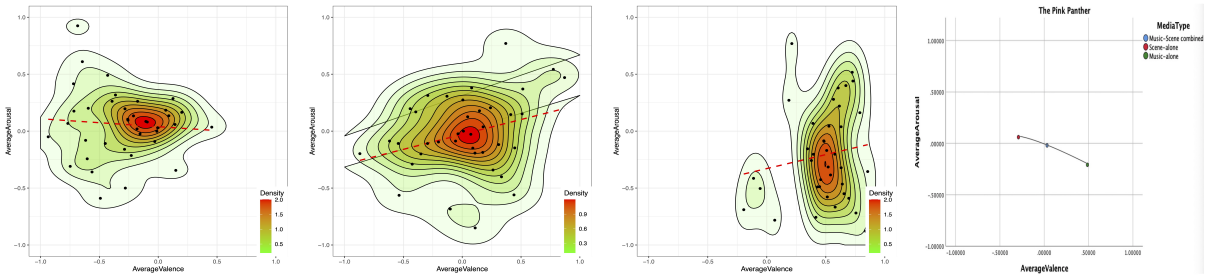
Film clips title	**Timing scene beginning**	**Dialogues & subtitles**	**Pronounced sound effects**
Airplane!	01:17:27	Yes	Siren, shouts
The Pink Panther	00:44:16	Yes	-
Bean The Ultimate Disaster Movie	00:13:35	Yes	-
Beverly Hills Cop	00:39:45	Yes	-
A Nightmare On Elm Street (1984)	01:25:45	Yes	Door slam
Psycho	00:52:11	No	-
Jaws	01:58:50	Yes	-
The Shining	00:36:19	Yes	Processed voices
Braveheart	02:16:44	No	-
Manhattan	01:28:35	Yes	-
Out of Africa	01:44:40	No	-
Sleepless in Seattle	01:37:47	Yes	-
Gladiator	00:08:44	Yes	War battle sounds
Speed	01:43:35	Yes	Fight sounds
The Fugitive	01:19:54	Yes	-
Vertical Limit	00:38:46	Yes	Avalanche sounds

7.2 Emotion distribution and center of mass



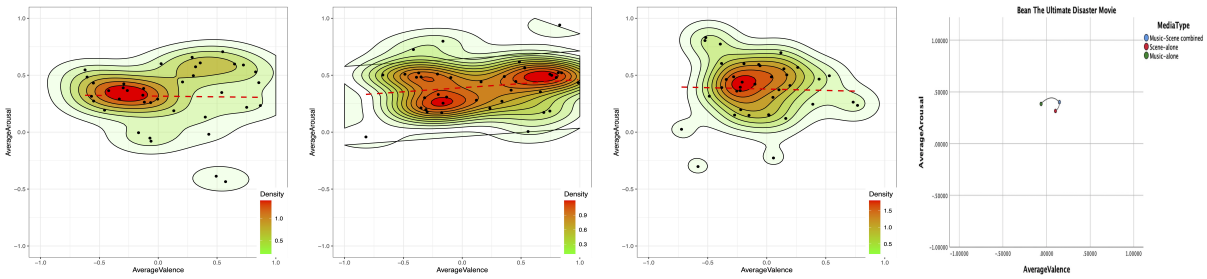
(a) Scene-alone **(b)** Music-scene combined **(c)** Music-alone **(d)** Center of mass

Figure 12: Airplane! - Comedy



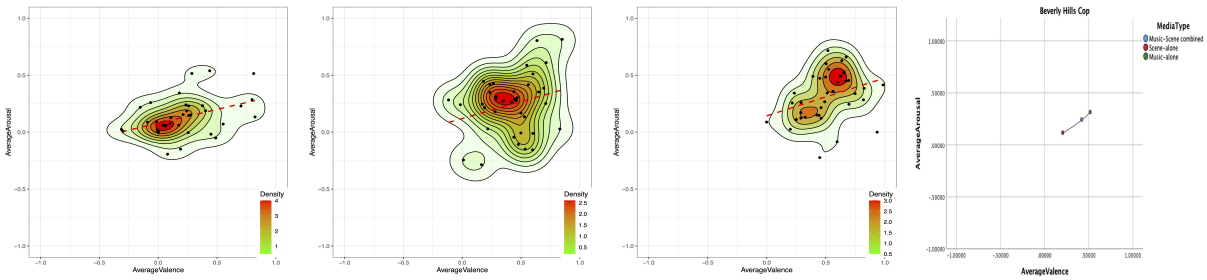
(a) Scene-alone **(b)** Music-scene combined **(c)** Music-alone **(d)** Center of mass

Figure 13: The Pink Panther - Comedy



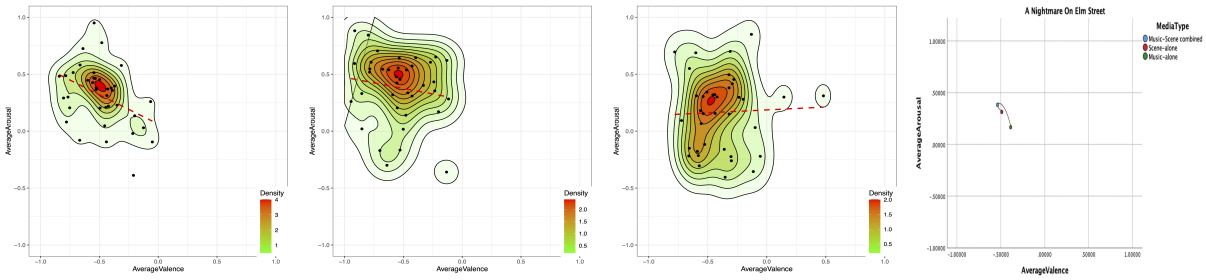
(a) Scene-alone **(b)** Music-scene combined **(c)** Music-alone **(d)** Center of mass

Figure 14: Bean The Ultimate Disaster Movie - Comedy



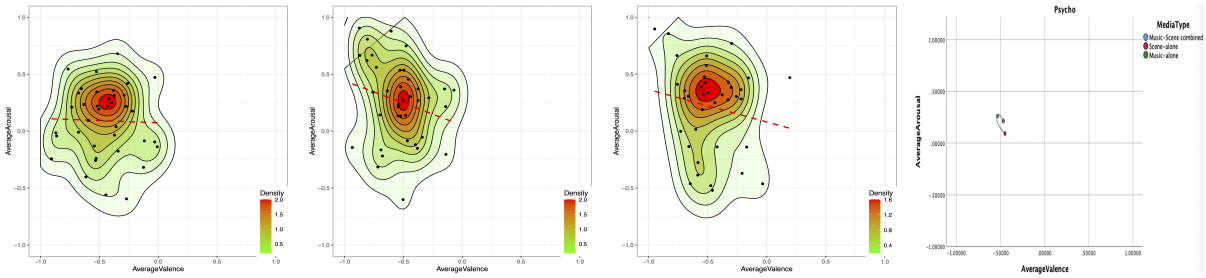
(a) Scene-alone **(b)** Music-scene combined **(c)** Music-alone **(d)** Center of mass

Figure 15: Beverly Hills Cop - Comedy



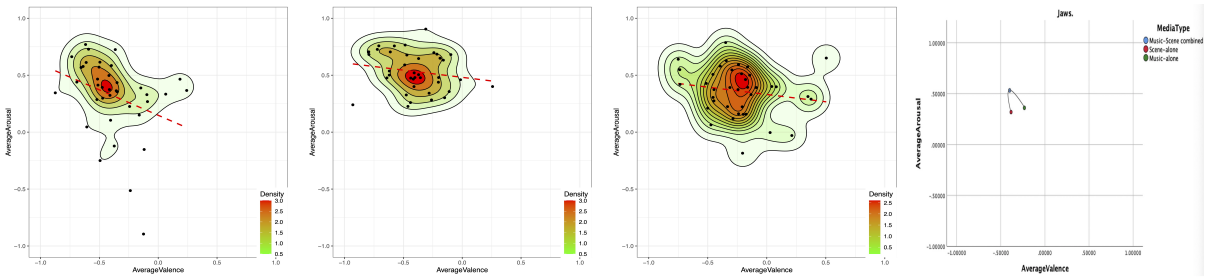
(a) Scene-alone **(b)** Music-scene combined **(c)** Music-alone **(d)** Center of mass

Figure 16: A Nightmare On Elm Street - Horror



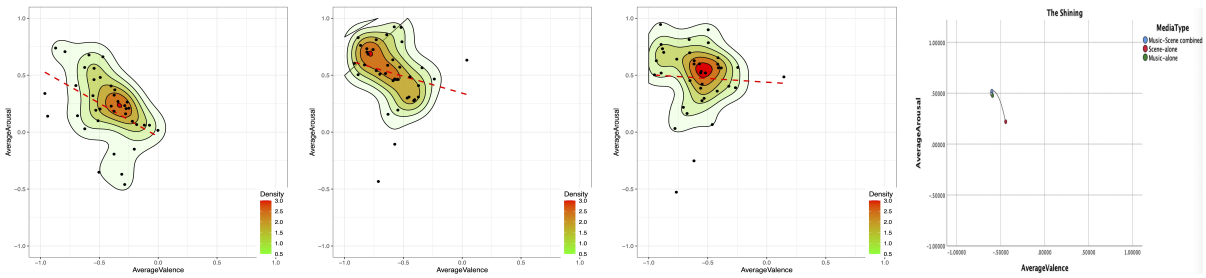
(a) Scene-alone **(b)** Music-scene combined **(c)** Music-alone **(d)** Center of mass

Figure 17: Psycho - Horror



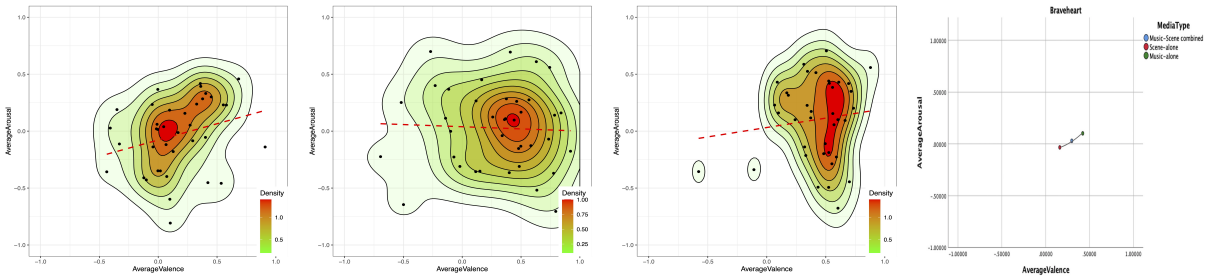
(a) Scene-alone **(b)** Music-scene combined **(c)** Music-alone **(d)** Center of mass

Figure 18: Jaws - Horror



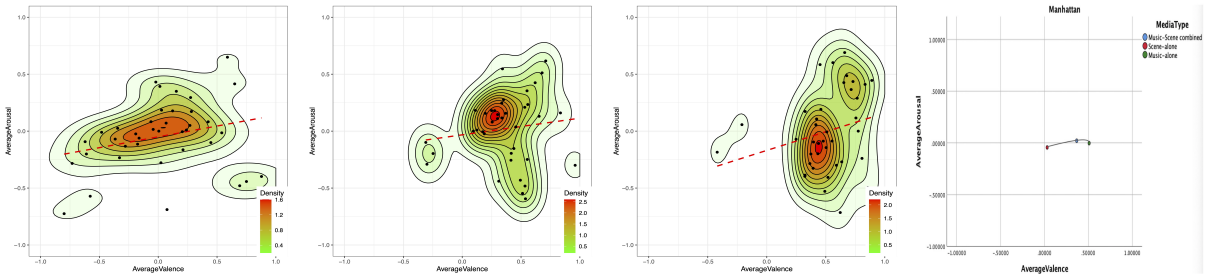
(a) Scene-alone **(b)** Music-scene combined **(c)** Music-alone **(d)** Center of mass

Figure 19: The Shining - Horror



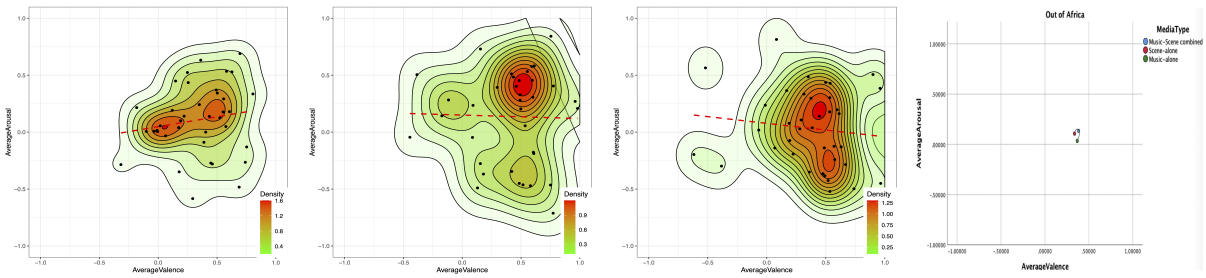
(a) Scene-alone **(b)** Music-scene combined **(c)** Music-alone **(d)** Center of mass

Figure 20: Braveheart - Romance



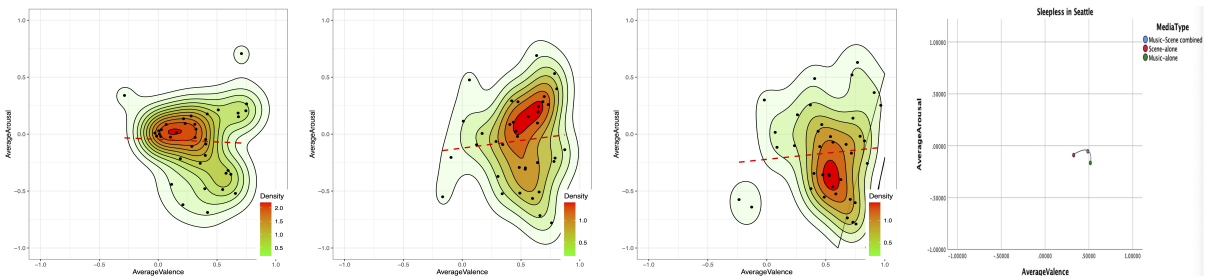
(a) Scene-alone **(b)** Music-scene combined **(c)** Music-alone **(d)** Center of mass

Figure 21: Manhattan - Romance



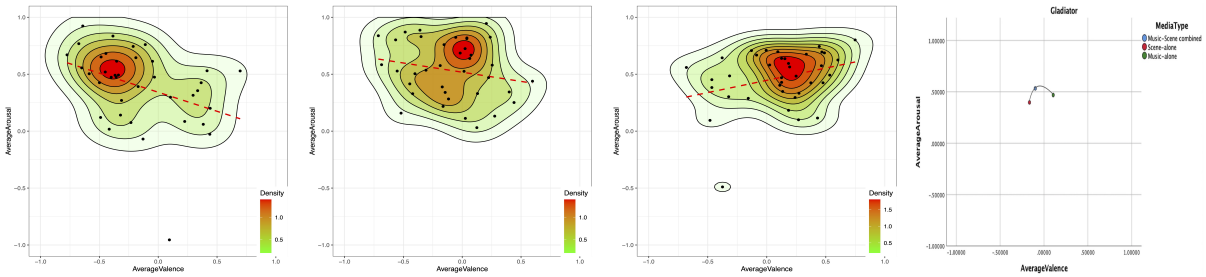
(a) Scene-alone **(b)** Music-scene combined **(c)** Music-alone **(d)** Center of mass

Figure 22: Out of Africa - Romance



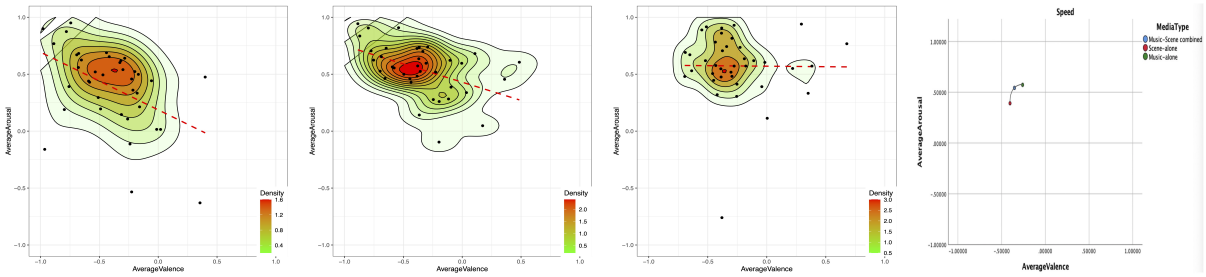
(a) Scene-alone **(b)** Music-scene combined **(c)** Music-alone **(d)** Center of mass

Figure 23: Sleepless in Seattle - Romance



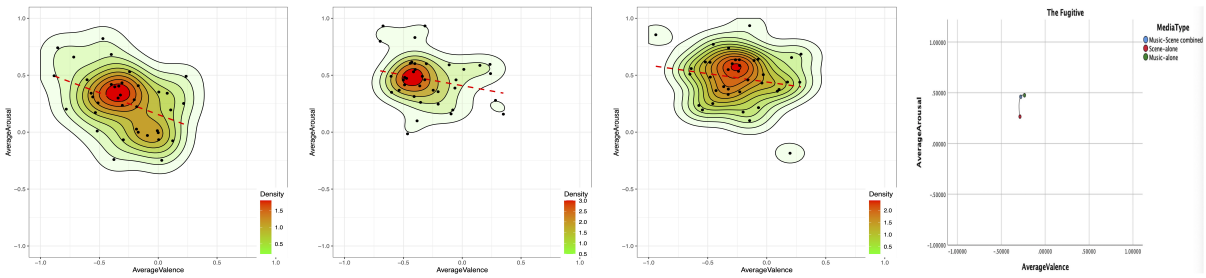
(a) Scene-alone **(b)** Music-scene combined **(c)** Music-alone **(d)** Center of mass

Figure 24: Gladiator - Action & Drama



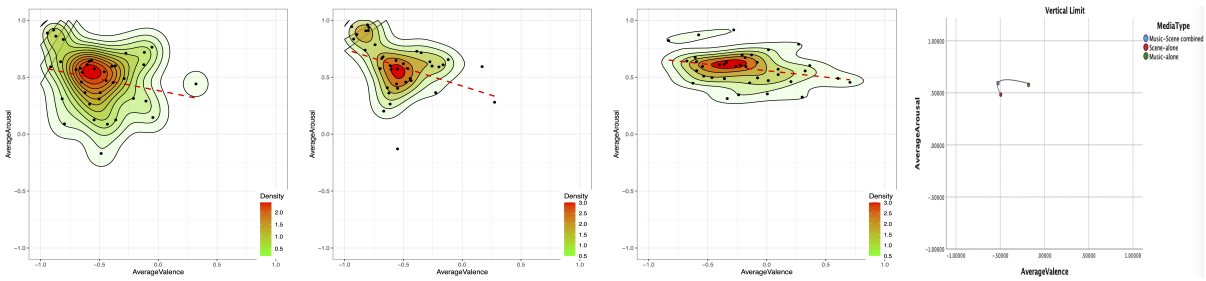
(a) Scene-alone **(b)** Music-scene combined **(c)** Music-alone **(d)** Center of mass

Figure 25: Speed - Action & Drama



(a) Scene-alone **(b)** Music-scene combined **(c)** Music-alone **(d)** Center of mass

Figure 26: The Fugitive - Action & Drama



(a) Scene-alone **(b)** Music-scene combined **(c)** Music-alone **(d)** Center of mass

Figure 27: Vertical Limit - Action & Drama