American Psychological Association

Psychomusicology: Music, Mind, and Brain

© 2019 American Psychological Association
0275-3987/19/$12.00

2019, Vol. 29, No. 4, 200–208
http://dx.doi.org/10.1037/pmu0000242

Effects of Audiovisual Congruency on Perceived Emotions in Film

Ninett Rosenfeld and Jochen Steffens
Technical University of Berlin

This study examined the influence of film music on the emotional perception of unambiguous semantically (in)congruent audiovisual film scenes. We predicted that the visual stimulus would dominate the emotional perception of the combined audiovisual stimuli and that a sarcastic or melancholic effect would be conveyed when the visual and musical stimuli are semantically incongruent. Therefore, one visual stimulus and one musical stimulus clearly representing one of the four emotions (anger, fear, happiness, and sadness) were chosen and combined in a congruent and incongruent way with each other. Utilizing an online experimental methodology, participants watched 16 stimulus combinations, rating them in terms of the perceived emotions (anger, fear, happiness, and sadness) and the emotional effects (melancholy and sarcasm) on a unipolar 5-point Likert scale. In addition, participants rated the perceived congruence between image and music for each stimulus. The results showed that the perceived audiovisual emotion was determined by the visual emotion, and semantically incongruent music was able to decrease the visually perceived emotions, compared to the congruent condition. Furthermore, a sarcastic effect was perceived when happy music accompanied negatively valent visual content. The results also showed that the perception of a melancholic effect was influenced by sad film music. The study provides further empirical foundations for the interaction of visual and auditory sensory channels, and its effect on the emotional perception of film scenes. Thus, it contributes to the understanding of audiovisual perception and highlights the importance of both film content and film music on the interpretation of a scene.

Keywords: audiovisual perception, film music, emotion, audiovisual interaction, semantic congruence

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Music can convey different emotions, such as happiness or sadness, and can influence how visual scenes, for example, in feature films, are perceived. Filmmakers are familiar with the emotional effects of music and have used it to shape the cinematic experience for decades. Different studies have already shown the influence of music on the perception of film characters (Hoeckner, Wyatt, Decety, & Nusbaum, 2011; Marshall & Cohen, 1988), the recall of details of a film scene (Boltz, 2001), and the effect of incongruent music on the visually perceived semantic content (Bolivar, Cohen, & Fentress, 1994).

We always couple visual actions and auditory sounds from our environment and therefore develop steady representations of common cross-modal actions that help associating one modality with the other (Prinz, 1990). For example, the unmistakable sound of using a lighter elicits a mental imagery of a lighter and the visual action. We will also have an idea of how the lighter sounds like if we only perceive its use visually. In the course of the "ventriloquist effect" by Howard and Templeton (1966, p. 361), different locations of semantically congruent visual and auditory sources led to a conflict in the perception of the position of both sources. To adapt this to a common localization of both – the visual and auditory sources – the auditory impression is "discarded" in favor of the visual stimulus, and the auditory stimulus is perceived in the same place as the image. This can be observed while watching TV and especially in cinema. Here, the sound system and the screen are physically separated from each other, but the words of the person can be perceived as coming directly from the actor and not from the speakers, unless there is a glitch in the sound system (Howard & Templeton, 1966). This research demonstrates how the visual channel can dominate the auditory one, even influencing its spatial perception. Another example where the visual stimulus alters the perception of the auditory channel is known as the McGurk effect (McGurk & MacDonald, 1976). This refers to the auditory information of the syllable (ba) being heard as (da) when it is combined with the visually presented lip movement of the syllable (ga).

This effect, however, might also work in the opposite direction. For instance, a lightning flash combined with an acoustical click is perceived earlier if the click precedes the flash (Fendrich & Corballis, 2001). Sekuler, Sekuler, and Lau (1997) observed that two identical shapes meeting in the center of the image and moving back to the edges are perceived to be colliding and drifting apart when combined with an auditory stimulus. If the objects were shown without this auditory stimulus, objects were perceived to be in a flowing movement to the other side of the image. These

This article was published Online First July 22, 2019.

Ninett Rosenfeld and Jochen Steffens, Audio Communication Group, Technical University of Berlin.

This article was written in the course of a master thesis at the Technische Universität (TU) Berlin. This master's thesis provides preliminary results of the article and was published on the homepage of the Audio Communication Group of TU Berlin: https://www2.ak.tu-berlin.de/~akgroup/ak_pub/abschlussarbeiten/2017/RosenfeldMasA.pdf.

Correspondence concerning this article should be addressed to Ninett Rosenfeld, Audio Communication Group, Technical University of Berlin, Einsteinufer 17, 10587 Berlin, Germany. E-mail: ninettrosenfeld@web.de

studies illustrate how auditory stimuli can influence the perception of ambiguous visual scenes and show how we associate specific auditory stimuli, like a bang or a click, with a vision of colliding or fast visual actions.

If the content of both the visual and auditory channels is clear and congruent, both channels can be considered to have an additive effect (Boltz, Schulkind, & Kantra, 1991). For example, the perceived intensity of an LED can be increased by a short auditory stimulus (Stein, London, Wilkinson, & Price, 1996). However, in cases where the conveyed content of both channels is clear but different, the interpretation of audiovisual perception seems to depend on the content of the two channels. For example, the music stimulus of a cello played with a bow is influenced by the visually presented plucked play of the cello, whereas the music stimulus of a plucked cello seems to be less affected by the visually presented bow-playing cellist (Saldaña & Rosenblum, 1993). Thus, depending on the content, the visual channel influences the perception of the auditory channel.

In the following text, these findings of audiovisual perception will be discussed in terms of the effect of film music on a film scene. As suggested by the studies mentioned earlier (Boltz et al., 1991; Saldaña & Rosenblum, 1993; Stein et al., 1996), the perceived congruence of the auditory/musical and visual channels is crucial for this effect. In the congruence–associationist model (Cohen, 2010; Marshall & Cohen, 1988), congruence describes formal or structural audiovisual similarities, whereas incongruent stimuli are formal and structurally different. As a result of audiovisual similarities, the music directs attention to parts of the film with the same formal or structural information, and the musical meaning is connected with the visual object (Marshall & Cohen, 1988). The perceived information is then classified based on previous knowledge and associated with mental images of experiences and emotions. In detail, mental imagery describes mental association of a certain sensory impression, triggered by a perceived sensory impression that does not have to correspond to the imagined one (Nanay, 2018). As perception is influenced by mental imagery (Nanay, 2018), the mental images may influence if the film scene and film music are perceived as matching. For example, a sad film scene accompanied by happy film music may not correspond to the mental association of the character of a sad film scene (e.g., sad-looking people accompanied by calm sad music) and, therefore, influence how intensely the emotion is perceived. In this case, the film music is prevented from connecting to the film scene, and as Ellis and Simons (2005) stated, the music is "merely adjunctive rather than integral to the subsequent understanding and coloring of the overall emotional response" (p. 35). In contrast to Marshall and Cohen's (1988) statement, the music would not direct attention to similar parts in the film scene due to its incongruity and the lack of audiovisual similarities, and therefore, the musical meaning does not have a straightforward connection to the visual action.

This is corroborated by the findings of Boltz (2001), who stated that film items that are semantically congruent with the perceived valence of the music are more likely to be remembered than those that are semantically incongruent. Boltz (2001) also showed that more objects were associated with the music due to its semantically congruent valence and, therefore, falsely identified, as compared with semantically incongruent objects. In a study by Bolivar et al. (1994), scenes of aggressive or friendly interactions of wolves were combined with aggressive or friendly music in a semantically congruent and a semantically incongruent way. Here, it was found that the semantic content of the film scene was amplified by the semantically congruent music, compared with the incongruent combinations. This shows the positive effect of emotional congruence of audiovisual similarities. If, according to Boltz et al. (1991), attention is focused on visually relevant details in congruent audiovisual stimuli, we might assume that audiovisual incongruence would reduce attention enhancement and reduce the visual impact. The results obtained by Bolivar et al. (1994) showed that the rating of the perceived semantic content for the friendly scene was reduced by the aggressive music, and the rating for the aggressive scene was reduced by the friendly music. However, in this study, aggressive film scenes were less effected by the friendly music than friendly film scenes by aggressive music. Aggression, whether it is conveyed visually or musically, dominated the perception of the audiovisual stimuli. Thus, the differential impact of music on visual events seems to depend on the characteristics of the music, such as valence (Bolivar et al., 1994).

Combined emotionally incongruent visual and musical material in films can be used further as an intended element. For example, in comedies, brutal scenes are often combined with cheerful music. In the film *Shaun of the Dead* (2004) by Edgar Wright, the zombie fight scene in the bar is accompanied by the song *Don't Stop Me Now* by Queen. The positive music with the lyrics "Cause I'm having a good time" is in contrast to the threatening situation and the fear of the protagonists. This can be described by the so-called "ironic contrast" (Boltz et al., 1991, p. 594), in which the effect of incongruous background music leads to an emotional neutralization of the film scene and partly to a sarcastic effect. With regard to ironic insults used in communication, Dews and Winner (1995) stated, "the positive literal meaning tinges the negative intended meaning, resulting in a less critical evaluation" (p. 4). The effect of "ironic contrast" in films mainly arises in the course of the combination of emotionally negative scenes (sadness, fear, and anger) with emotionally positive music (happiness; Boltz et al., 1991). In communication, sarcastic meaning is also indicated by the use of a positive cue. Tepperman, Traum, and Narayanan (2006) identified laughter as one contextual cue for the sarcastic effect. Furthermore, Lee and Katz (1998) highlighted the importance of ridicule of a specific victim to create a sarcastic effect. As a basis for sarcastic mockery, they speak of a person's failure to fulfill a certain expectation specific for that person. Sarcasm highlights the person's failure and recalls the actual expectation.

Melancholic flashbacks are another example for emotionally incongruent combined audiovisual material. Here, emotionally incongruent visual and musical materials are combined like happy film scenes with calm or sad music to emphasize a past incident, or a melancholic flashback. The nature of melancholy is not easy to describe, as it resembles the emotion sadness. According to Brady and Haapala (2003), melancholy has a dual nature, because it consists of positive and negative aspects. For example, in the film *Her* (2013) by Spike Jonze, the protagonist thinks of his wife, and happy and serious scenes of them are shown accompanied by music. The character of the music is slow and uplifting, which means that the calm melody is perceived as sad but also as slightly positive. Therefore, the visual channel and the auditory channel in this film scene contain both sadness and a slight form of happiness. As described by Brady and Haapala (2003), melancholy is not as

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

debilitating as sadness and "involves the pleasure of reflection and contemplation of things we love and long for" (p. 2) or "a memory of a person, place, event, or state of affairs" (p. 3). Therefore, melancholy has a kind of comforting and exhilarating character like in the film scene from *Her*. According to Burton and Jackson (1978), melancholic mood is associated with solitude. Commonly, reflection or contemplation happens in a situation of solitude or silence. Such a situation may also be caused by music that contains only the emotion sadness, as this emotion may be associated with a situation of solitude and silence. Hence, by combining the sad music with an emotionally incongruent visual scene, a melancholic effect may be created.

These examples illustrate that music can fulfill different functions in a film, depending on the semantic content and congruence to the visual channel. In ambiguous film scenes, it decisively determines and dominates the effect of the visual channel (Boltz, 2001; Boltz et al., 1991; Hoeckner et al., 2011). The semantically congruent combination of distinctive film scenes and music results in the reinforcing effect of the visual scene (Bolivar et al., 1994; Marshall & Cohen, 1988). In semantically incongruent scenes, the musical meaning conflicts with that of the visual channel. As in ambiguous scenes, music affects attention and leads to a distortion of the emotional effect in this way, such that the influence of the visually presented emotion is reduced (Bolivar et al., 1994) or new elements are added to the scene relative to the musical meaning (Boltz, 2001).

The following study aimed to further clarify how film music influences the emotional perception of the visual channel depending on the semantic (in)congruence. Therefore, visual and musical stimuli were used to convey clear basic emotions and were combined in a congruent and an incongruent manner. In detail, we investigated to what degree visual and musical stimuli contribute to the overall emotional effects of audiovisual film scenes and whether it is possible to create a sarcastic or melancholic effect by pairing semantically incongruent stimuli. In contrast to the study by Bolivar et al. (1994), our study used not only two emotional states but also four different visually and auditorily perceived emotions.

Due to the abstractness of music, the visual content often has a greater emotional clarity (Cohen, 2010). Hypothesis 1, therefore, predicted that the visual stimulus would dominate the emotional perception of the combined audiovisual stimuli over the musical emotion. Sarcastic and melancholic effects in a film are interesting, because they are created by combining different emotions to achieve this result. According to Boltz et al. (1991) and Tepperman et al. (2006), a sarcastic effect occurs when negative content and positive content are combined. Brady and Haapala (2003) stated that melancholy consists of positive and negative aspects. Hypothesis 2, therefore, predicted that additional emotional effects such as sarcasm and melancholy would be conveyed when the film scene and music have opposing perceived emotional valence. Thus, the study aimed to provide further empirical evidence to understand the interaction of the visual and musical channels when watching feature films.

Method

Participants

The study included a sample of 61 people in total (27 women, 32 men, and two other) aged 20 to 58 years (M = 29.7 years, SD = 7.8). The sample consisted predominantly of bachelor's (N = 23) and master's/diploma students (N = 27). Eight participants indicated that they had a high school diploma, and three participants reported a doctorate as their highest degree. Participation was incentivized, and all participants in the study had the opportunity to enter a lottery to win a €30 Eventim voucher. Participants were able to state if they need to be credited with 30 min of participation time to pass a statistics course at the master's program Audio Communication and Technology at the Technische Universität Berlin. The survey took between 20 and 30 min.

Materials

The perceived emotions of the chosen visual and musical stimuli were aimed to be semantically unambiguous. For our experiment, we chose basic emotions that can be clearly conveyed both visually and musically, namely: happiness, sadness, fear, and anger. Basic emotions, in general, can be found in every culture and are essential for everyday life (Ekman, 1992). With the different arousal states of basic emotions, these form the basis for other emotional states such as "happiness" for "ecstasy" or "sorrow" for "pensiveness" (Plutchik, 1991, pp. 112–115). No stimuli were chosen that contained the emotional effects melancholy and sarcasm. To compare their respective contribution to the perceived emotions of the audiovisual combination, the aim was to use the same emotional terms (happiness, sadness, fear, and anger) for both the visual and auditory channels, to reduce individual differences in the understanding and use of terminology.

Video stimuli. Twenty film scenes with five excerpts for each of the four emotions (fear, happiness, sadness, and anger) were selected. To avoid the recognition of individual examples, scenes from mainly unknown films from the years between 1969 and 2006 were chosen, as we assumed that more recent films would have a greater potential to be recognized by the rather young sample. The period of the release years for the film selection was based on the selected music examples. To achieve variation in the material, films from different genres and countries were selected.

The selected scenes had an average length of 14 s (range: 8–25 s). The most important criteria for the selection of the scenes were the unambiguity of the expressed emotion and the duration of the film scene. The emotion should be expressed superficially by gestures and facial expressions of the actor and should give no indication of the plot of the film. In addition, the selected scenes should contain few cuts and no or only little dialogue. Two of the selected films were taken from other studies (*Star Trek IV: The Voyage Home* by Leonard Nimoy [1986], study by Lipscomb and Kendall, 1994, and *Butch Cassidy and the Sundance Kid* by George R. Hill [1969], study by Boltz et al., 1991). Several examples were taken from the films *Fear Dot Com* by William Malone (2002), *When a Stranger Calls* by Simon West (2006), *Oldboy* by Park Chan-wook (2003), and *Star Trek IV*. The selected scenes showed close-ups and medium and long shots of one or more actors.

Music stimuli. The selected music examples were taken from the study by Eerola and Vuoskoski (2011). Among others, the authors evaluated a broad and structured pool of musical stimuli according to the basic emotions anger, fear, happiness, sadness, and tenderness. From this pool, film music was used that, accord-

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

ing to Eerola and Vuoskoski (2011), conveyed the emotions fear, happiness, sadness, and anger.

A total of 10 music excerpts (two stimuli each for fear and sadness, and three stimuli each for happiness and anger) from the years between 1989 and 2005 were selected. These excerpts had an average duration of 20 s (range: 16–30 s).

The visual and music stimuli were extracted from the original DVDs or CDs. The music samples were adjusted in their volume channel using Cubase (Steinberg Media Technology GmbH, Hamburg, Germany) music software. The original music and sound of the film scenes were removed, and the scenes were faded in and out using the video software Adobe Premiere Pro (Adobe Inc., San Jose, California).

The 20 film scenes and 10 music excerpts were narrowed down to four of each through pretesting (to be detailed in the pretest section) and were combined in both a congruent and an incongruent manner (Table 1). Sixteen stimulus combinations with an average length of 14 s (range: 12–16 s) were created and embedded into an online questionnaire via YouTube links. The stimuli were presented in a randomized order to avoid order effects. Each participant watched the 16 pairings in a different order.

Pretest

For the experiment, the least ambiguous emotional film scenes and music were selected in a preliminary test from a pool of 30 stimuli. Fifty-three people (21 women, 32 men, aged between 24 and 59 years, $M = 31.1$ years, $SD = 2.9$) participated in this pretest, and none of them participated in the main experiment. They rated each stimulus (film scene or film music) on a 5-point scale with regard to the perceived emotions sadness, happiness, fear, and anger. In addition, participants indicated whether they knew the film scene or the film music using a bipolar scale ("yes," "no," "I don't know"). Furthermore, they were asked to report the film title in a text box, if applicable.

Those stimuli that most clearly expressed the desired emotion were chosen for the main experiment (see Appendix, Table A1). The mean value for each emotion rating for each stimulus was calculated and compared against the other stimuli in an emotion group. The choice fell on the stimulus with the highest mean value for the respective emotion.

Table 1
Overview of the Audiovisual Congruent and Incongruent Stimuli Combinations

Visual emotion	Musical emotion
Anger	Anger
	Fear
	Happiness
	Sadness
Fear	Anger
	Fear
	Happiness
	Sadness
Happiness	Anger
	Fear
	Happiness
	Sadness
Sadness	Anger
	Fear
	Happiness
	Sadness

For the emotion fear, the title *Dear Clarice* ($M = 4.47$, $SD = 0.84$, recognized by <5%) by Hans Zimmer from the soundtrack for the film *Hannibal* (2001) and a scene ($M = 4.98$, $SD = 0.15$, recognized by <5%) from the film *Fear Dot Com* by William Malone (2002) were selected. The music was a calm disharmonic piece with no specific melody or rhythm consisting of synthetic tones and bow-played violin. The film scene showed a frightened woman looking around, startled. The song *Strip the Willow* ($M = 4.84$, $SD = 0.52$, recognized by <5%) by Simon Boswell for the film *Shallow Grave* (1995) and the scene ($M = 4.98$, $SD = 0.15$, recognized by <5%) from the western *Butch Cassidy and the Sundance Kid* by George R. Hill (1969) reached the highest mean for the emotion happiness. The song is a medium-fast, folklore piece with a simple melody played by an accordion accompanied by bass and percussion. In the film scene, a man shows bike tricks to a woman while joking with her. The highest average for the emotion sadness was reached by the music title *Ask Your Saint Who He's Killed* ($M = 4.29$, $SD = 0.99$, recognized by <5%) by Gabriel Yared for the film *The English Patient* (1996) and the film scene ($M = 4.98$, $SD = 0.15$, recognized by <5%) from *The Fall* by Tarsem Singh (2006). The music piece has a slow, plaintive melody, consisting of long notes played by a violin. In the scene, the crying protagonist is shown in close-up. For the emotion anger, the title *Futile Escape* ($M = 3.98$, $SD = 0.92$, recognized by <5%) by Cliff Eidelman from *The Alien Trilogy* soundtrack (1996) was chosen. The music piece is a fast, energetic melody that is mostly played by timpani, deep wind instruments, and strings. As video stimulus, the scene ($M = 4.76$, $SD = 0.44$, recognized by <5%) from *Star Trek IV: The Voyage Home* by Leonard Nimoy (1986) was chosen. It reached only the third-highest average of all film samples for the emotion anger. However, unlike the other stimuli, it was unknown and emotionally explicit. The scene showed a woman arguing with a man and slapping him.

Design and Procedure

The online study was advertised via university e-mailing lists and Facebook. It was publicly available for 4 weeks and was designed to take 20 to 30 min to complete.

Participants carried out the study on their own on a desktop computer with speakers or headphones. They were advised to not use smartphones to complete the study, to ensure a similar atmosphere such as in a home TV situation without much distracting surrounding noise. A within-subject design was chosen, to use the participants for every stimulus combination. Therefore, we needed only one group of subjects with fewer participants than in a between-subjects design. To lessen the effects associated with the repeated presentation of the stimuli, the stimuli were presented in a random order for each participant.

The task of the participants was to watch and rate 4 × 4 combinations of emotionally congruent and incongruent film and music samples, in terms of the perceived emotions happiness, sadness, fear, and anger and the melancholic and sarcastic effects (labelled: melancholy and sarcasm), on a unipolar 5-point Likert scale. In addition, they indicated whether they had known the film scene before the experiment ("yes," "no," "I do not know") and, if

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

possible, reported the title of the film in a text box. Finally, participants rated the perceived congruence between image and sound on another 5-point scale ranging from 1 ("do not match at all") to 5 ("match very well"). After evaluating all stimuli, participants finally reported on sociodemographic information (such as sex, age, and highest educational degree).

Results

Table 2 displays the mean values and standard deviations of the audiovisual stimulus combinations, depending on the emotion conveyed by the visual and the musical modality.

Hypothesis 1 predicted that the visual channel dominates the emotional perception of the combined audiovisual stimuli. Table 2 reveals that the ratings of the four emotions (anger, fear, happiness, and sadness) were always highest for the emotion suggested by the visual stimulus (bold ratings), except for the audiovisual stimulus combination of visual happiness and musical sadness. For all four dependent variables, repeated-measures (RM) analyses of variance (ANOVAs)–with the two four-level factors visual emotion and musical emotion as main effects–were calculated. The results for anger, fear, happiness, sadness, sarcasm, and melancholy are shown in Table 3.

Results revealed that in all cases, except for the sarcastic effect, both visual and musical channels significantly contributed to the perceived emotion of the audiovisual combinations. Comparing the overall effect size η^2 of both channels from Table 3 across the four basic emotions anger, fear, happiness, and sadness reveals that, on average, the impact of the visual channel (mean $\eta^2 = .84$) was 23% larger than the impact of the musical channel (mean $\eta^2 = .61$), confirming Hypothesis 1. Table 2 further suggests that semantically incongruent music is able to lessen the perceived emotion implied by the visual channel of the audiovisual combinations, compared with the congruent condition. This is corroborated by post hoc (Bonferroni-corrected) pairwise comparisons across the four musical emotions, calculated for each of the four different film excerpts. As displayed in Table 4, the congruent condition significantly differed from all noncongruent conditions (all ps $<$ .01) for the excerpts conveying visual fear, happiness, and sadness. Regarding the excerpt visually conveying anger, comparisons showed that happy ($p < .001$) but not sad and fearful music (*n*s) significantly reduced perceived audiovisual anger.

Table 3
Statistics of the RM ANOVAs With the Perceived Emotions of the Audiovisual Stimulus Combinations as Dependent Variables

AV combinations	Factor	Wilks' Λ	F	df	p	η^2
Anger	Visual	.18	89.5	3	<.001	.82
	Musical	.55	16.1	3	<.001	.45
Fear	Visual	.15	108.7	3	<.001	.85
	Musical	.32	41.7	3	<.001	.68
Happiness	Visual	.15	109.0	3	<.001	.85
	Musical	.43	25.8	3	<.001	.57
Sadness	Visual	.15	107.4	3	<.001	.85
	Musical	.27	51.1	3	<.001	.73
Melancholy	Visual	.34	37.5	3	<.001	.66
	Musical	.24	62.4	3	<.001	.76
Sarcasm	Visual	.90	2.1	3	.11	.10
	Musical	.42	26.5	3	<.001	.58

Note. RM = repeated-measures; ANOVAs = analyses of variance; AV = audiovisual.

In addition, congruency between film scene and music, as rated by the participants, significantly affected perceived emotions of the audiovisual combination. The numeric differences in averaged perceived emotions of the audiovisual combinations between the four congruent conditions (e.g., visual fear and musical fear) and the respective incongruent conditions (e.g., visual fear and musical sadness) can be directly predicted by the averaged rated congruency of video and music, as shown by a simple linear regression, $F(1, 14) = 12.37$, $R^2 = .47$, $b = -0.38$, $p < .01$. This means that the higher the incongruence between video and music was rated by the participants, the more that music reduced the emotion conveyed by the video.

Hypothesis 2 predicted that a sarcastic and melancholic effect would be conveyed when the film scene and music had opposing

Table 2
Mean Perceived Emotions of the Audiovisual Stimulus Combinations and Congruency Ratings

Visual emotion	Musical emotion	Anger	Fear	Happiness	Sadness	Melancholy	Sarcasm	Congruency
Anger	Anger	**4.02 (1.34)**	2.08 (1.17)	1.16 (0.55)	1.69 (0.98)	1.39 (0.76)	1.39 (0.84)	3.02 (1.22)
	Fear	**3.75 (1.45)**	2.67 (1.31)	1.05 (0.28)	1.98 (1.18)	1.61 (1.04)	1.21 (0.55)	2.41 (1.05)
	Happiness	**3.41 (1.38)**	1.48 (0.81)	1.92 (1.08)	1.51 (0.89)	1.23 (0.56)	2.92 (1.54)	2.07 (1.03)
	Sadness	**4.02 (1.19)**	2.00 (1.08)	1.11 (0.37)	3.21 (1.31)	2.80 (1.36)	1.25 (0.67)	2.95 (1.06)
Fear	Anger	1.95 (1.07)	**4.21 (1.18)**	1.05 (0.22)	1.61 (0.76)	1.20 (0.51)	1.34 (0.85)	3.18 (1.06)
	Fear	1.54 (0.67)	**4.75 (0.60)**	1.05 (0.38)	1.69 (0.99)	1.43 (0.87)	1.10 (0.40)	4.61 (0.69)
	Happiness	1.59 (0.84)	**3.70 (1.46)**	1.54 (0.94)	1.52 (0.85)	1.18 (0.50)	2.77 (1.74)	1.38 (0.82)
	Sadness	1.61 (0.86)	**4.13 (1.09)**	1.08 (0.33)	2.70 (1.39)	2.54 (1.41)	1.41 (0.80)	2.59 (1.04)
Happiness	Anger	1.44 (0.83)	1.66 (0.79)	**3.33 (1.38)**	1.20 (0.63)	1.38 (0.82)	1.97 (1.33)	1.54 (0.87)
	Fear	1.21 (0.58)	2.67 (1.36)	**3.08 (1.41)**	1.43 (0.87)	1.85 (1.19)	1.51 (0.83)	2.16 (1.11)
	Happiness	1.07 (0.31)	1.05 (0.22)	**4.74 (0.66)**	1.08 (0.33)	1.41 (0.82)	1.51 (0.94)	4.46 (0.79)
	Sadness	1.21 (0.61)	1.64 (0.95)	3.28 (1.33)	2.84 (1.40)	**3.66 (1.44)**	1.23 (0.59)	3.18 (1.01)
Sadness	Anger	2.46 (1.19)	2.33 (1.19)	1.10 (0.40)	**3.89 (1.27)**	2.15 (1.24)	1.48 (1.07)	1.95 (1.02)
	Fear	2.13 (1.15)	3.28 (1.25)	1.08 (0.42)	**4.18 (1.06)**	2.43 (1.45)	1.13 (0.39)	3.15 (1.08)
	Happiness	1.51 (0.89)	1.98 (1.12)	1.67 (1.04)	**3.67 (1.38)**	2.39 (1.31)	2.67 (1.71)	1.46 (0.77)
	Sadness	1.77 (0.97)	2.43 (1.31)	1.08 (0.42)	**4.87 (0.34)**	3.72 (1.42)	1.03 (0.18)	4.43 (0.85)

Note. Boldface indicates the emotion with the highest mean for the audiovisual combination.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Table 4
Pairwise Comparisons (Bonferroni-Corrected) Between Congruent and Incongruent Musical Emotions for the Different Visual Emotions

Musical emotions	Mean difference (congruent − incongruent)	*SE*	*p*	95% CI
		Visual anger		
Musical anger				
Musical fear	0.262	0.148	0.485	[−0.141, 0.665]
Musical happiness	0.607	0.169	0.004	[0.145, 1.068]
Musical sadness	0.000	0.124	>.999	[−0.338, 0.338]
		Visual fear		
Musical fear				
Musical anger	0.541	0.139	<.001	[0.161, 0.921]
Musical happiness	1.049	0.172	<.0001	[0.579, 1.520]
Musical sadness	0.623	0.108	<.0001	[0.330, 0.916]
		Visual happiness		
Musical happiness				
Musical anger	1.410	0.169	<.0001	[0.948, 1.871]
Musical fear	1.656	0.181	<.0001	[1.162, 2.149]
Musical sadness	1.459	0.156	<.0001	[1.033, 1.885]
		Visual sadness		
Musical sadness				
Musical anger	0.984	0.156	<.0001	[0.558, 1.409]
Musical fear	0.689	0.129	<.0001	[0.336, 1.041]
Musical happiness	1.197	0.171	<.0001	[0.729, 1.665]

Note. CI = confidence interval.

perceived emotional valence. As expected and as indicated by Table 2, the sarcastic effect was rated highest when negatively valenced film content conveying fear, anger, and sadness was presented together with positively valenced music conveying happiness. An RM ANOVA–with the sarcastic effect ratings as the dependent variable–revealed a significant difference between the 16 stimuli, depending on the musical emotion (see Table 3). Here, the factor visual emotion did not have a significant effect on the sarcastic effect ratings.

Furthermore, results of a linear mixed-effects model, including a random intercept for each participant, and predicting sarcasm by visual emotion, musical emotion, and the perceived congruency as independent variables (fixed effects), confirmed and extended the results of the RM ANOVA. Here, Type III tests of fixed effects revealed that musical emotion and congruency (but not visual emotion) significantly contributed to the prediction of the sarcastic effect ratings of the audiovisual combination (musical emotion: $F(3, 908.5) = 76.1$, $p < .001$; congruency: $F(1, 936.7) = 67.2$, $p < .001$; and visual emotion: $F(3, 907.5) = 1.1$, $p = .32$). Taken together, the results suggest that (rather) independent from the actual visual context, the sarcastic effect is conveyed when happy music is played, which does not fit semantically to the visual content.

In addition, with regard to the perception of a melancholic effect of the audiovisual combinations, results of the RM ANOVA (Table 3) revealed that the melancholic effect was conveyed to a different extent across the 16 stimuli. As can be seen from Table 2, melancholic effect ratings were highest in all four cases, where sad film music was presented compared with the combination with happy, fearful, and angry music. Comparing the effect size η^2 of both the visual and musical channels in Table 3, results reveal that the musical channel ($\eta^2 = .76$) can be considered more important than the visual one ($\eta^2 = .66$) to convey a melancholic effect. However, the rating of sadness and melancholy showed the difficulty of telling the difference between both emotions. Except for the combination of visual happiness with musical sadness where melancholy was rated highest (Table 2), the rating of the perceived sadness for the other combinations with sad music was higher than the rating for melancholy.

To investigate the effect of congruency, another linear mixed-effects model was computed predicting a melancholic effect of the audiovisual combination with the visual emotion, the musical emotion, and the perceived congruency as independent variables. The results from this model confirmed and extended the results of the repeated ANOVA. Here, all three independent variables significantly contributed to the prediction of an audiovisual melancholic effect (visual emotion: $F(3, 907.4) = 56.9$, $p < .001$; musical emotion: $F(3, 908.8) = 132.2$, $p < .001$; and congruency: $F(1, 948.0) = 8.9$, $p < .01$). The combination of visual happiness and musical sadness achieved the highest rating of congruency compared with the other incongruent combinations with sad music. Thus, the results suggest that the melancholic effect is conveyed when sad music fits the visual content.

Discussion

The primary aim of this article was to investigate the influence of semantically congruent and incongruent film music on the perceived emotional expression of combined audiovisual stimuli. Results of the study confirmed the hypothesis that the visual content of film scenes dominates the emotional perception of combined audiovisual stimuli, compared with the underlying film

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

music. This is in line with Cohen (2010), who stated that due to the abstractness of music, the visual content often has a greater emotional clarity. As the used scenes, in contrast to ambiguous scenes, already contain all the clues that are relevant for understanding the emotion, the visual domain mainly appears to determine the detection of perceived emotion. Marshall and Cohen (1988) also stated in their study that distinct actions can already be interpreted based on the visual stimulus, and therefore, music is not needed to understand the plot.

Furthermore, the fact that semantically incongruent music lessens the emotion conveyed by the visual channel compared with a congruent combination is consistent with findings by Boltz et al. (1991) and Bolivar et al. (1994). Similar to the study by Boltz (2001), incongruent music might distract attention from visually relevant characteristics, and thus reduce the emotional impact of the film scene.

Regarding Hypothesis 2, it was demonstrated that a sarcastic effect is conveyed when negative valent visual content is combined with incongruent, happy music. Analogous to findings by Tepperman et al. (2006), who investigated the use of laughter in communication, happy music can thus be seen as an indicator for perceiving sarcasm. In this study, the three stimulus combinations conveying sarcasm showed a person in the role of victim. By using happy music, which does not match the scene, it seems that the visual action is ridiculed. This is in line with findings by Lee and Katz (1998), who highlighted the importance of ridicule of a specific victim for sarcastic effect. In this context, it is necessary to point out that sarcasm is portrayed not only by combining visual action with auditory happiness but also through visual content in isolation; however, this was not the goal of the present study.

Furthermore, the results showed that the film scene (e.g. visual fear) is perceived as less negative with happy music than with the semantically congruent music, although the intended visual emotion is still clearly perceived. Here, the explanation offered by Dews and Winner (1995) of a less critical evaluation of an ironic insult, due to the positive literal meaning, is adaptable to the incongruent combination with happy music. The music decreased the perception of the negative visual channel compared with the congruent combinations.

Using positive music, the perception of the negative visual content can be mitigated. In some cases, such as films like *Bowling for Columbine* by Michael Moore (2002), the combination with positive music can also appear morbid, and therefore perceived more terrifying than the visual alone. In this film, the song *What a Wonderful World* by Louis Armstrong accompanied scenes of wars and their victims as a strong contrast to the violent actions. In connection with the film's previous scene, in which an agent of a large weapons company claims that weapons are not employed without good reason, the above scene underlines the failure of this statement and is an example of sarcastic mockery according to Lee and Katz (1998).

Conversely, our study showed that sad music has a significant effect on the rating of the melancholic effect and that sad music increases the perception of melancholy regardless of the visual channel. The combination of the visual emotion with sad music had always the highest melancholy rating compared with the combination with happy, fearful, or angry music. As our results showed, the perceived congruency between the visual and auditory channels influences the melancholic effect of the audiovisual scene. In the combination of visual happiness and auditive sadness where the sad music fits best the film scene, melancholy was rated highest compared with the other emotions (even higher than happiness). Based on the data, it is a logical assumption that the music was able to evoke the emotion because it was perceived as connected to the narrative of the film scene (Ellis & Simons, 2005). In the case of the incongruent combination of visual happiness and auditory sadness, we may find evidence in the rating that melancholy is not just a form of sadness and consists of positive and negative aspects (Brady & Haapala, 2003); here, in the form of a happy action accompanied by sad music. Brady and Haapala (2003) stated that melancholy involves reflection and memories, and Burton and Jackson (1978) said that melancholy is a situation of solitude where reflection and contemplation happen; thus, the sad music may have led to a mental imagery of such a situation and the association of perceiving memories from somebody else. However, further studies are needed to investigate in detail the nature of melancholy in film scenes and its difference from sadness.

By using original motion picture material, the results provide an empirical basis for the design of emotional effects in films. However, some limitations raised with the design of the study have to be addressed. First, in the pretest, video stimuli were rated as slightly stronger with regard to their intended emotion, compared with the musical stimuli. This might have affected the relative influence of both channels on the perceived emotions of the audiovisual combinations. Further research should therefore replicate these findings, using stimuli with emotional "clarity" more balanced across the visual and auditory channels. In addition, the use of only one visual and musical excerpt as a representative for a specific emotion presented several times in the experiment might limit the generalizability of the findings. To address this limitation, further studies should use multiple film music excerpts specifically composed for a film scene, and the use of more and longer extracts to increase ecological validity of the findings.

In this context, the limitation of the within-subject design regarding the effects associated with the repeated presentation of the stimuli should be addressed. A between-subjects design might be advantageous to lessen the effect of a repeated stimulus presentation.

Nevertheless, the present study adds to the existing research on the emotional effects of semantically congruent and incongruent audiovisual stimuli and provides further empirical evidence to understand the interaction of the visual and musical channels in feature films. The results further highlight the importance of the visual channel for the overall emotional perception of the audiovisual scene and suggest that the semantic (in)congruency of film music affects the visually perceived emotion, potentially leading to a sarcastic or melancholic perception of the whole scene.

References

Bolivar, V. J., Cohen, A. J., & Fentress, J. C. (1994). Semantic and formal congruency in music and motion pictures: Effects on the interpretation of visual action. *Psychomusicology: A Journal of Research in Music Cognition, 13,* 28–59. http://dx.doi.org/10.1037/h0094102

Boltz, M. G. (2001). Musical soundtracks as a schematic influence on the cognitive processing of filmed events. *Music Perception, 18,* 427–454. http://dx.doi.org/10.1525/mp.2001.18.4.427

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Boltz, M., Schulkind, M., & Kantra, S. (1991). Effects of background music on the remembering of filmed events. *Memory and Cognition, 19,* 593–606. http://dx.doi.org/10.3758/BF03197154

Brady, E., & Haapala, A. (2003). Melancholy as an aesthetic emotion. *Contemporary Aesthetics, 1*. Retrieved from http://hdl.handle.net/2027/spo.7523862.0001.006

Burton, R., & Jackson, H. (Eds.). (1978). *The anatomy of melancholy* (Reprinted). London, United Kingdom: Dent.

Cohen, A. J. (2010). Music as a source of emotion in film. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, applications* (pp. 878–908). Oxford, United Kingdom: Oxford University Press.

Dews, S., & Winner, E. (1995). Muting the meaning a social function of irony. *Metaphor and Symbolic Activity, 10,* 3–19. http://dx.doi.org/10.1207/s15327868ms1001_2

Eerola, T., & Vuoskoski, J. K. (2011). A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music, 39,* 18–49. http://dx.doi.org/10.1177/0305735610362821

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion, 6,* 169–200. http://dx.doi.org/10.1080/02699939208411068

Ellis, R. J., & Simons, R. F. (2005). The impact of music on subjective and physiological indices of emotion while viewing films. *Psychomusicology: A Journal of Research in Music Cognition, 19,* 15–40. http://dx.doi.org/10.1037/h0094042

Fendrich, R., & Corballis, P. M. (2001). The temporal cross-capture of audition and vision. *Perception and Psychophysics, 63,* 719–725. http://dx.doi.org/10.3758/BF03194432

Hoeckner, B., Wyatt, E. W., Decety, J., & Nusbaum, H. (2011). Film music influences how viewers relate to movie characters. *Psychology of Aesthetics, Creativity, and the Arts, 5,* 146–153. http://dx.doi.org/10.1037/a0021544

Howard, I. P., & Templeton, W. B. (1966). *Human spatial orientation*. London, United Kingdom: Wiley.

Lee, C. J., & Katz, A. N. (1998). The differential role of ridicule in sarcasm and irony. *Metaphor and Symbol, 13,* 1–15. http://dx.doi.org/10.1207/s15327868ms1301_1

Lipscomb, S. D., & Kendall, R. A. (1994). Perceptual judgment of the relationship between musical and visual components in film. *Psychomusicology: A Journal of Research in Music Cognition, 13,* 60–98. http://dx.doi.org/10.1037/h0094101

Marshall, S. K., & Cohen, A. J. (1988). Effects of musical soundtracks on attitudes toward animated geometric figures. *Music Perception, 6,* 95–112. http://dx.doi.org/10.2307/40285417

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264,* 746–748. http://dx.doi.org/10.1038/264746a0

Nanay, B. (2018). Multimodal mental imagery. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior, 105,* 125–134. http://dx.doi.org/10.1016/j.cortex.2017.07.006

Plutchik, R. (1991). *The emotions* (rev. ed.). Lanham, MD: University Press of America.

Prinz, W. (1990). A common coding approach to perception and action. In O. Neumann & W. Prinz (Eds.), *Relationships between perception and action: current approaches* (pp. 167–201). Berlin, Heidelberg: Springer. http://dx.doi.org/10.1007/978-3-642-75348-0_7

Saldaña, H. M., & Rosenblum, L. D. (1993). Visual influences on auditory pluck and bow judgments. *Perception and Psychophysics, 54,* 406–416. http://dx.doi.org/10.3758/BF03205276

Sekuler, R., Sekuler, A. B., & Lau, R. (1997). Sound alters visual motion perception. *Nature, 385,* 308. http://dx.doi.org/10.1038/385308a0

Stein, B. E., London, N., Wilkinson, L. K., & Price, D. D. (1996). Enhancement of perceived visual intensity by auditory stimuli: A psychophysical analysis. *Journal of Cognitive Neuroscience, 8,* 497–506. http://dx.doi.org/10.1162/jocn.1996.8.6.497

Tepperman, J., Traum, D., & Narayanan, S. (Eds.). (2006). *Yeah Right. Sarcasm Recognition for Spoken Dialogue Systems*. Proceedings of InterSpeech ICSLP, Pittsburgh, PA.

(*Appendix follows*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Appendix

Table A1
Description of the Used Visual and Musical Stimuli

Stimuli	Film/music	Description
Visual fear	Malone, William (2002): *Fear Dot Com* [DVD], United Kingdom, Germany, Luxembourg and United States: Sony Pictures Home Entertainment, 00:43:54–00:44:10.	Terrified woman in a room looking around, startled
Visual happiness	Hill, George R. (1969): *Butch Cassidy and the Sundance Kid* [DVD], United States: Twentieth Century Fox, 00:29:37–00:29:49.	Man showing bike tricks to a woman while joking with her
Visual sadness	Singh, Tarsem (2006): *The Fall* [DVD], United States and India: Alive – Vertrieb und Marketing/DVD, 01:40:32–01:40:42.	Crying man
Visual anger	Nimoy, Leonard (1986): *Original: Star Trek IV: The Voyage Home* [DVD], United States: Paramount (Universal Pictures), 01:08:30–01:08:42.	Woman arguing with a man and slapping him
Musical fear	Hans Zimmer (2001): *Dear Clarice*, on: *Hannibal* – Original Motion Picture Soundtrack [CD], Decca, Nr. 1, 00:40–00:54.	Long, single notes, no melody, no rhythm, synthetic tones combined with bow-played violin
Musical happiness	Boswell, Simon (1995): *Strip the Willow*, on: *Shallow Grave* – Original Soundtrack [CD], EMI, Nr. 6, 02:02–02:17.	Folk music piece with an accordion playing a simple melody, simple bass line and percussions, medium tempo
Musical sadness	Yared, Gabriel (1996): *Ask Your Saint Who He's Killed*, on: *The English Patient* – Soundtrack [CD], Fantasy Records, Nr. 18, 00:14–00:32.	Calm, classical tune, bow-played violin, playing a slow melody with long notes
Musical anger	Eidelman, Cliff (1996): *Futile Escape*, on: *The Alien Trilogy* – Soundtrack [CD], Colosseum (Alive), Nr. 9, 00:03–00:18.	Played by an orchestra dominated by the timpani, deep wind instruments and strings, fast rhythm short notes

Received June 21, 2018
Revision received June 4, 2019
Accepted June 11, 2019 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.