*Short Report*

PERCEPTION

# The Auditory Kuleshov Effect: Multisensory Integration in Movie Editing

***Perception***
2017, Vol. 46(5) 624–631

DOI: 10.1177/0301006616682754
journals.sagepub.com/home/pec
SAGE

**Andreas M. Baranowski and H. Hecht**
Department of Psychology, Johannes Gutenberg-University Mainz, Germany

**Abstract**
Almost a hundred years ago, the Russian filmmaker Lev Kuleshov conducted his now famous editing experiment in which different objects were added to a given film scene featuring a neutral face. It is said that the audience interpreted the unchanged facial expression as a function of the added object (e.g., an added soup made the face express hunger). This interaction effect has been dubbed "Kuleshov effect." In the current study, we explored the role of sound in the evaluation of facial expressions in films. Thirty participants watched different clips of faces that were intercut with neutral scenes, featuring either happy music, sad music, or no music at all. This was crossed with the facial expressions of happy, sad, or neutral. We found that the music significantly influenced participants' emotional judgments of facial expression. Thus, the intersensory effects of music are more specific than previously thought. They alter the evaluation of film scenes and can give meaning to ambiguous situations.

**Keywords**
Kuleshov effect, film editing, facial expressions, emotional context, multisensory integration

In the wake of the 20th century, the Soviet filmmaker Lev Kuleshov (1899–1970) conducted a series of experiments. In one experiment, he filmed then-famous actor Ivan Mozhukin looking into the camera with a neutral expression and intercut the face with a bowl of soup, a female in a coffin, a girl playing, or a beautiful woman lying on a couch. Unfortunately, the original footage was lost and reports about the experiment are somewhat conflicting (Hill, 1967; Pudovkin, 1970). All reports agree on that the audience who watched the short film interpreted the unchanged facial expression of Mozhukin differently depending on the object. The face was interpreted as sad in the presence of the coffin but hungry when the soup was added. This technique of intercutting faces with emotional context to alter the viewers' interpretation of the faces is nowadays cited in almost every introductory film study book and has been dubbed Kuleshov effect.

**Corresponding author:**
Andreas M. Baranowski, Department of Psychology, Johannes Gutenberg-University Mainz, Germany.
Email: baranowski@uni-mainz.de

Only in recent years have researchers begun to study the Kuleshov effect, and only three studies on it have been published so far. A first attempt to replicate the original study by Prince and Hensley (1992) did not find any evidence for the existence of the Kuleshov effect. They presented a close-up of a face to 137 students, edited together with a shot of a little girl playing, a woman lying in a coffin, and a bowl of soup. The film was professionally made and edited and stayed close to the alleged design of the original experiment. However, this was a single-trial between-subject experiment, which is prone to noise in the data.

Mobbs et al. (2006) presented very short (4 seconds) film sequences consisting of negative, positive, or neutral valence. After this context manipulation, 14 participants had to rate the emotional expression of mostly neutral faces, some of which were edited to appear slightly happy or sad. Film sequences and faces were separated by a jittered interstimulus interval of between 4 and 8 seconds. Participants rated the faces for emotional expression and mental state. Additionally, Mobbs et al. used fMRI to detect neuronal activity changes in the brain. They found that positive context leads the participants to rate the faces as happier, whereas negative context led to higher ratings of sadness. The fMRI data also suggested that the emotional context had been particularly salient and activated regions related to emotion processing such as the amygdala. While the study supports the Kuleshov effect, the design is somewhat problematic. In particular, object shots were shown before the faces, reversing the order of the original Kuleshov effect. To collect fMRI data, the authors were also forced to leave a time gap between the presentation of the context scene and the face, thus removing the direct transition from scene to face and reducing the chance that they are perceived as temporally connected.

More recently, Barratt, Rédei, Innes-Ker, and van de Weijer (2016) tested 36 participants using 24 film sequences across five emotional conditions (happiness, sadness, hunger, fear, and desire) and a neutral control condition. The stimuli were presented by first showing the still image of a neutral face, followed by a short video clip ending with the same neutral face. Participants were asked to judge the emotional expression of the faces, as well as valence and arousal. Barratt et al. (2016) found that participants tended to choose the appropriate category more frequently than the alternative options, while the answers to the valence and arousal questions also went in the expected directions (but were only significant for the sadness and desire conditions).

The rationale behind the Kuleshov effect is that a given facial expression is always perceived relative to the entire context. The facial close-up provides an emotional range, which is further interpreted. A laughing person is most likely not sad, but the context is used to determine the degree of happiness perceived by the viewer. In particular, if the facial expression stays neutral, the context will provide information that is used to evaluate the emotional state of the actor (Carroll, 1996). Studies of contextual influence on affective face processing have shown similar effects outside the realm of films. Contextual cues are used to assess facial expressions (Carroll & Russell, 1996; Wallbott, 1988; Wieser & Brosch, 2012; Zhang, Fu, Chen, & Fu, 2014).

Do similar effects exist for nondiegetic sound? Nondiegetic effects of sound are almost universally employed in modern day film. For instance, Bouhuys, Bloem, and Groothuis (1995) demonstrated that music can be used as an affective primer to influence a viewer's evaluations about a character's emotions. Likewise, Tan, Spackman, and Bezdek (2007) found that music is an effective primer for emotions and additionally for expected intentions. Eldar, Ganor, Admon, Bleich, and Hendler (2007) presented participants with neutral film clips containing emotional music. They found that the emotional evaluation of the clips changed in accordance with the music. A further study showed that empathy-related judgments of film characters were modulated by music. The protagonist was rated as more likable when accompanied by melodramatic music as compared with tense music (Hoeckner, Wyatt, Decety, & Nusbaum, 2011). Jolij and Meurs (2011) manipulated the mood of the participants with music. In a stimulus detection task that followed, they found that not only were faces congruent with the participant's mood easier detected. Participants produced a high number of false alarms, seeing faces congruent with their mood, which were not there.

In addition, neurophysiological studies have confirmed the enhanced affective effect of combined congruent auditory and visual stimuli. Baumgartner, Esslen, and Jäncke (2005), for example, found that psychophysiological measures (heart rate, skin conductance, respiration, and skin temperature) as well as alpha power density as measured by electroencephalogram, for sad, happy, and fearful faces, in combination with congruent music, increased, compared with pictures or music alone. In an fMRI study with a similar design, Baumgartner, Lutz, Schmidt, and Jäncke (2006) found increased activation of brain areas associated with emotion processing, when presenting visual and auditory stimuli simultaneously. Most importantly, the fusiform gyrus (FG) and areas involved in the auditory processing like the superior temporal gyrus showed increased activation. In agreement with Baumgartner et al. (2006), Jeong et al. (2011) found increased activation in the FG and the superior temporal gyrus in congruent conditions for sad and happy faces, while they found activation to be greatest in FG for incongruent conditions. The authors suggested that there is an increased activation in areas related to auditory stimuli in congruent conditions, while incongruent conditions result in a higher activation of regions involved with the processing of facial information.

Music elicitation of emotions can be explained by several factors including association or memory, empathic responses, cognitive evaluation, and signaling (Hanser & Marks, 2013). Film music often works by association and memory. Sounds that have strong associations with particular emotions are used to activate these associations in the viewer. The advantage of the use of sound to elicit emotions is that they directly and quickly activate associated emotions in the perceiver (Kamiyama, Abla, Iwanaga, & Okanoya, 2013; Logeswaran & Bhattacharya, 2009). Combining the sound with congruent visual information usually leads to an enhanced intensity of the emotional reaction while incongruent pairing can produce paradox results (Pavlović & Marković, 2011).

In the current study, we tested whether next to these nondiegetic sound effects, the emotions induced by a soundtrack are able to change the perceived expression of a face seen in the film. In other words, is there an auditory Kuleshov effect? If this is so, changing the emotional tone of a neutral film sequence by playing emotional music (Eldar et al., 2007) should alter the perception of a facial expression. We chose happy and sad music, because these are the most distinctive musical categories, and crossed them with happy and sad facial expressions. Our main hypothesis was that the evaluation of the facial expression changes according to the emotional category of the music.

We also aimed at improving the design of previous studies on the visual Kuleshov effect. Barratt et al. (2016) rightfully pointed out that the Kuleshov effect relies on point-of-view editing. This means that the scene of the onlooker and the film scene must be temporally and spatially connected; the onlooker is actually looking at the emotional scene. If this is not the case, the Kuleshov effect cannot work because the facial expression of the observer is not linked to the emotional scene.

To increase the likelihood of the viewer to infer such a relation, Persson (2003) proposes eight conditions that should be met: (a) the actor does not look directly into the camera; (b) the object shot is presented from the perspective of either the actor or an observer; (c) the object shot is preceded and followed by a glance shot; (d) the environment of the actor close-up matches the environment of the object; (e) the actor changes their behavior just before the cut in the first glance shot; (f) the actor shows some form of reaction in the second glance shot; (g) the soundtrack for all three shots is continuous; and (h) the spatial relation between the actor and the object is established beforehand. In our stimuli, we managed to realize all but the fourth and fifth rule (see Methods section for details). Reactions (if present) always followed the object shot.

We also crossed the music with facial expression, making our study a $3 \times 3$ (facial expression: neutral, happy, sad $\times$ music: neutral, happy, sad) design. There were two reasons to include different emotional expressions in our study. First, we wanted to explore the intersensory integration of music and facial expression. Would the visual modality carry more weight such that a smiling face with sad music is still perceived as happy? Second, the fully crossed design is preferable to merely change the soundtrack, as it reduces demand characteristics (see Barratt et al., 2016).

We hypothesized that the intersensory cross-talk is mutual, that is, we expected to find that soundtrack and facial expression both contribute to the perceived emotionality of the film clip.

## Methods

### Participants

Thirty psychology students (female = 21, male = 9) participated in the experiment on a voluntary basis. Age ranged from 20 to 31 years ($M = 23.40$, $SD = 3.60$). All participants had normal or corrected-to-normal vision.

### Materials

Four actors were filmed separately with identical camera settings. The actors were sitting at a 30° angle with respect to the camera and a computer screen was visible on the left side of the frame. In accordance with Mobbs et al. (2006), the actors looked at the screen, which allowed for the use of multiple neutral scenes and settings that would not have been possible if we tried to create the illusion that the actor was actually present in the neutral context scene. This is a departure from the original Kuleshov design, where the illusion was created that the actor was actually present in the viewed scene. The camera was placed on a tripod 1 m away from the actors, and a head and shoulder close-up was produced. In postproduction, we produced nine versions of each clip. In the beginning, the actor was always looking with a neutral expression at the computer screen for 3 seconds, followed by a neutral scene of random events (e.g., a man walking down a hallway, a woman grocery shopping) that lasted 15 seconds, taken from the video platform YouTube. The sound of these neutral scenes was muted. Afterward, we cut back to the actor for another 3 seconds, amounting to a total of 21 seconds for each clip. In the last shot, the actor had a neutral, happy, or sad facial expression. To ensure authentic reactions, we asked the actors to watch congruent self-selected emotional stimuli during filming. Six researchers then validated each facial expression independently. The neutral expression was in accordance with Kuleshov's original experiment, while the happy or sad facial expressions were a departure from the original design. These emotional expressions were included to explore the intersensory integration of music and facial expression. In addition, crossing music with facial expression reduced demand characteristics of the stimulus material. The clips were edited in accordance with Persson's (2003) eight rules for creating a point-of-view structure (p. 6), increasing the likelihood for a Kuleshov effect to appear (see Barratt et al., 2016, for a discussion of these rules and their relation to the Kuleshov effect). Additionally, we added happy or sad music to each clip or kept it without music as control. We considered using white or pink noise as a control, but while some people find it neutral, others find it calming or annoying. The music was taken from various online film music databases, and only unknown music was chosen. All music was emotionally categorized by the composer and was only instrumental. Six researchers additionally validated the music with high concordance.

### *Procedure*

Each participant was tested individually. Upon arrival in the laboratory, participants were seated 3 m from the screen (size 115 cm × 65 cm; horizontal visual angle 22°). The room was darkened, and sound was provided via speakers. After signing a consent form, we asked participants to watch 27 film clips, 3 for each of the 9 conditions (3 music × 3 facial expressions). Every clip was followed by a pause during which participants had time to fill in a short questionnaire. Participants were instructed to rate the perceived emotional state of the actor in each clip on the six basic emotions anger, disgust, fear, happiness, sadness, and surprise (Ekman & Friesen, 1971) on scales ranging from 0 (*emotion not present*) to 5 (*emotion strongly present*). Target emotions were happiness and sadness, whereas the others served as distractors. This was to ensure that participants reported all emotions they perceived and avoid demand characteristics that would have been present if only one emotion was allowed to select. Additionally, this way, paradox emotions that could result from incongruent music and facial expressions were recorded. The average ratings of the target emotions for the three films per condition were then used for further analysis. The film clips were shown in random order and the whole experiment lasted about an hour. After they finished, participants were asked whether they had previously seen any of the short video clips and were debriefed. None of the participants had seen any of the short video clips before.

## Results

As illustrated in Figure 1, music and facial expression mutually influenced the emotional ratings. We conducted a 3 × 3 (facial expression: neutral, happy, sad × music: neutral, happy, sad) two-way rmMANOVA with the dependent variables happiness and sadness. Using Pillai's trace, we found main effects for music, $F(4, 26) = 5.42$, $p = .003$, $\eta_p^2 = .46$, facial expression, $F(4, 16) = 22.02$, $p < .001$, $\eta_p^2 = .77$, and a significant interaction between the two, $F(8, 22) = 3.06$, $p = .018$, $\eta_p^2 = .53$. Music significantly influenced happiness, $F(2, 58) = 9.52$, $p < .001$, $\eta_p^2 = .25$, and sadness ratings, $F(2, 58) = 3.29$, $p = .044$, $\eta_p^2 = .10$, as did the facial expression, $F(4, 16) = 34.42$, $p < .001$, $\eta_p^2 = .54$ and $F(4, 16) = 9.00$, $p < .001$, $\eta_p^2 = .24$, respectively. An interaction effect existed only for sadness ratings, $F(4, 116) = 4.54$, $p = .002$, $\eta_p^2 = .14$, but not for happiness ratings, $F(4, 116) = 0.63$, $p = .590$, $\eta_p^2 = .02$.

A contrast analysis revealed that our actors were rated as significantly less happy when accompanied by sad music ($M = 1.19$, $SD = 0.62$), as compared with happy music ($M = 1.74$, $SD = 0.64$) or no music at all ($M = 1.47$, $SD = 0.65$), with $F(1, 29) = 15.42$, $p < .001$, $\eta_p^2 = .35$ and $F(1, 29) = 9.54$, $p = .004$, $\eta_p^2 = .25$, respectively. Happy music had the reverse effect on sadness ratings, making actors look less sad ($M = 0.70$, $SD = 0.47$) compared with sad music ($M = 1.00$, $SD = 0.77$), with $F(1, 29) = 3.12$, $p = .037$, $\eta_p^2 = .14$. The no music condition received intermediate ratings ($M = 0.91$, $SD = 0.64$) but did not differ significantly from either of the two groups.

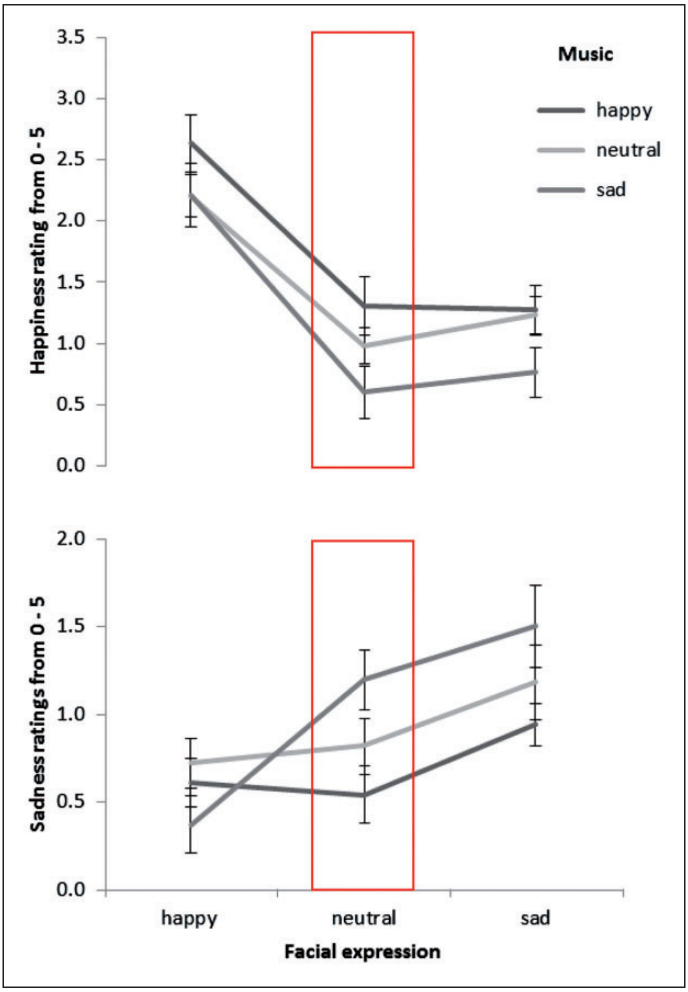


**Figure 1.** Happiness and sadness ratings for music crossed with facial expression. The expected effect was strongest in the neutral facial condition (red boxes).

### *Error Bars Indicate SEM*

The facial expression of the actors had a strong effect on perceived emotions. Happy faces were rated as happier ($M = 2.35$, $SD = 0.80$) than neutral ($M = 0.96$ $SD = 0.59$) or sad ($M = 1.09$ $SD = 0.97$) faces, with $F(1, 29) = 93.54$, $p < .001$, $\eta_p^2 = .76$ and $F(1, 29) = 30.25$, $p < .001$, $\eta_p^2 = .51$, respectively. Likewise, sad facial expressions lead to significantly higher sadness ratings ($M = 1.21$, $SD = 0.79$) than neutral ($M = 0.85$, $SD = 0.75$) or happy ($M = 0.57$, $SD = 0.44$) faces, $F(1, 29) = 93.54$, $p < .001$, $\eta_p^2 = .76$ and $F(1, 29) = 30.25$, $p < .001$, $\eta_p^2 = .51$.

## Discussion

Faces were rated as happier when they watched scenes accompanied by happy music. And they were rated as less happy when they watched scenes accompanied by sad music.

This intervention worked well and in accordance with our hypothesis. Participants rated faces that watched a scene accompanied by happy music as less sad than when accompanied by sad music. Sad music did not lead to significantly higher sadness ratings than no music; however, when looking only at the neutral faces, the music had the expected effect for all categories. Happy music made neutral faces seem significantly happier and sad music made neutral faces significant sadder.

The interaction effect seems to be carried by the happy faces in the sad music condition. They did not reduce the sadness ratings as expected, but rather seemed to increase them. Happy faces were rated least sad when accompanied by sad music and sadder when accompanied by happy music or no music at all. An explanation for this trend would be that happy faces are generally perceived as more arousing than sad faces (Barratt et al., 2016) or that the sad music is so out of character that it has a reversed effect. Note that no such reversal occurred for happiness ratings.

The facial expressions of our actors worked very well. The intended emotions were correctly categorized in almost all cases in the neutral condition. Other emotions than happiness and sadness were rarely chosen and produced negligible mean ratings.

The effect sizes for music ($\eta_p^2 = .46$) and facial expression ($\eta_p^2 = .77$) were large, but the ratings were skewed toward the bottom of the scale, indicating a floor effect. It turned out that the scale ranging from *emotion not present* to *emotion strongly present* was not fully used. Further studies on the Kuleshov effect should take this into account. Emotions are clearly perceived in the face of the actor, but their verbal coding was weak.

To explore the interaction of the auditory and visual information further, it would be interesting to use a $3 \times 3 \times 3$ design, where music, facial expressions, and context are varied. In addition, the questionnaire could include more nuanced emotions to detect more subtle changes. Ratings of valence and arousal might further be added to help explain variance.

In sum, we could demonstrate an auditory Kuleshov effect. The emotional timbre of nondiegetic music influences how viewers judge the film clip. Facial expression and music both contribute to the resulting impression, thus there is multisensory integration. However, it is still not clear how exactly the Kuleshov effect works. There are two general explanatory models for the Kuleshov effect (Barratt et al., 2016; Mobbs et al., 2006). First, it is possible that the visual (or in our case auditory) stimuli of the object shot induce emotions in the participants, who project their own emotional state onto the actor. Second, a more popular explanation suggests that we set the observed face in the context of the object shot and cognitively adapt our perception. To pit these hypotheses against each other will be a challenge for further studies.

## Acknowledgements

Tha authors thank specially to Mike Kirschke, Theresa Reul, Wubamlak Endale, Zeeb-U-Nisah Ahmad, and Marzia Ablakulova for preparing the stimuli and conducting the testing.

## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## References


Barratt, D., Rédei, A. C., Innes-Ker, A., & van de Weijer, J. (2016). Does the Kuleshov effect really exist? Revisiting a classic film experiment on facial expressions and emotional contexts. *Perception*, *45*, 847–874.

Baumgartner, T., Esslen, M., & Jäncke, L. (2005). From emotion perception to emotion experience: Emotions evoked by pictures and classical music. *International Journal of Psychophysiology*, *60*, 34–43.

Baumgartner, T., Lutz, K., Schmidt, C. F., & Jäncke, L. (2006). The emotional power of music: How music enhances the feeling of affective pictures. *Brain Research*, *1075*, 151–164.

Bouhuys, A. L., Bloem, G. M., & Groothuis, T. G. (1995). Induction of depressed and elated mood by music influences the perception of facial emotional expressions in healthy subjects. *Journal of Affective Disorders*, *33*, 215–226.

Carroll, N. (1996). Toward a theory of point-of-view editing: Communication, emotion, and the movies. In N. Carroll (Ed.), *Theorizing the moving image* (pp. 125–138). Cambridge, England: Cambridge University Press.

Carroll, J. M., & Russell, J. A. (1996). Do facial expressions signal specific emotions? Judging emotion from the face in context. *Journal of Personality and Social Psychology*, *70*, 205–218.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, *17*, 124–129.

Eldar, E., Ganor, O., Admon, R., Bleich, A., & Hendler, T. (2007). Feeling the real world: Limbic response to music depends on related content. *Cerebral Cortex*, *17*, 2828–2840.

Hill, S. P. (1967). Kuleshov—Prophet without honor? *Film Culture*, *44*, 6–9.

Hanser, W. E., & Mark, R. E. (2013). Music influences ratings of the affect of visual stimuli. *Psychological Topics*, *22*, 305–324.

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.

Jeong, J., Diwadkar, V. A., Chugani, C. D., Sinsoongsud, P., Muzik, O., Behen, M. E., . . . Chugani, D. C. (2011). Congruence of happy and sad emotion in music and faces modifies cortical audiovisual activation. *NeuroImage*, *54*, 2973–2982.

Jolij, J., & Meurs, M. (2011). Music alters visual perception. *PLoS One, 6*, doi:10.1371/journal.pone.0018861

Kamiyama, K. S., Abla, D., Iwanaga, K., & Okanoya, K. (2013). Interaction between musical emotion and facial expression as measured by event-related potentials. *Neuropsychologia*, *51*, 500–505.

Logeswaran, N., & Bhattacharya, J. (2009). Crossmodal transfer of emotion by music. *Neuroscience Letters*, *455*, 129–133.

Mobbs, D., Weiskopf, N., Lau, H., Featherstone, E., Dolan, R., & Frith, C. (2006). The Kuleshov Effect: The influence of contextual framing on emotional attributions. *Social Cognitive and Affective Neuroscience*, *1*, 95–106.

Persson, P. (2003). *Understanding cinema: A psychological theory of moving imagery*. Cambridge, England: Cambridge University Press.

Prince, S., & Hensley, W. E. (1992). The Kuleshov effect: Recreating the classic experiment. *Cinema Journal*, *31*, 59–75.

Pavlović, I., & Marković, S. (2011). The effect of music background on the emotional appraisal of film sequences. *Psihologija*, *44*, 71–91.

Pudovkin, V. I. (1970). *Film technique and film acting* (I. Montagu, Ed. & Trans.). New York, NY: Grove Press, Inc.

Tan, S. L., Spackman, M. P., & Bezdek, M. A. (2007). Viewer's interpretations of film characters' emotions: Effects of film music before or after a character is shown. *Music Perception*, *25*, 135–152.

Wallbott, H. (1988). In and out of context: Influences of facial expression and context information on emotion attributions. *British Journal of Social Psychology*, *27*, 327–369.

Wieser, M., & Brosch, T. (2012). Faces in context: A review and systematization of contextual influence on affective face processing. *Frontiers in Psychology*, *3*, 471. doi:10.3389/fpsyg.2012.00471

Zhang, M., Fu, Q., Chen, Y. H., & Fu, X. (2014). Emotional context influences micro-expression recognition. *PLoS One*, *9*, 0095018.