

## Bidirectional Cross-Modal Influence on Emotion Ratings of Auditory and Visual Stimuli

**Abstract:** Previous research concerning cross-modal influences on emotional perception has focused primarily on how auditory stimuli affect emotional responses to visual stimuli. The present study examines whether such effects are bidirectional. Different participants were tested in one of these two directions of influence, using a slider-bar rating task to judge the emotionality (sadness/happiness) of stimuli in an attended modality (auditory or visual). Stimuli were presented in auditory and visual pairs, with instructions to ignore stimuli from the irrelevant modality. All stimuli (auditory and visual) had been previously categorized as sad, ambiguous, or happy. Results showed that ratings depended primarily on the emotional categories of stimuli in the attended modality (auditory or visual). In addition, participants were subject to smaller cross-modal influences from the unattended modality in both the auditory and visual attentional conditions. Bidirectional influences were thus observed, showing that perceptual influence is not limited to a single cross-modal direction.

BY HARRISON JAMES RAMSAY

### Introduction

Auditory and visual information are frequently combined to produce unique effects in many types of entertainment and performance. The simultaneous presentation of music and film, in particular, is a very powerful combination that is utilized as a way to elicit emotion in cinema. The ability of music to influence emotion has been well documented in the literature.<sup>1</sup> Similar evidence documents the effectiveness of film segments to evoke emotions within participants and across participants.<sup>2</sup> While these studies individually highlight the capabilities of music and film to manipulate emotion, it is important to consider specific effects that may result when auditory and visual stimuli are experienced together.

Multiple studies have combined auditory and visual information in order to assess the resulting effects of such cross-modality pairings on various parameters of emotion and cognition.<sup>3-9</sup> For present purposes, it will be necessary to recognize the distinction between experienced emotion and the perception of stimulus emotionality. With regards to the perception of emotionality in stimuli, fewer studies<sup>11, 12</sup> have illustrated the possibility of a cross-modal influence. While research on this topic is less plentiful, it does begin to elucidate the phenomenon of interest. Until recently, research on the topic of cross-modal influence has focused primarily on the auditory-to-visual direction. This is not surprising, as combinations of auditory and visual information are frequently found in television, movies, and theatrical performances as mentioned earlier, where the primary focus is on visually oc-

curing activity. Therefore, it naturally appeals to questions about whether and why the addition of music matters in these examples.

The present study addresses whether these sorts of phenomena can also occur in the opposite direction. The handful of studies that have investigated the visual-to-auditory direction of influence seem to focus less on the topics of experienced affect and stimuli emotionality that have been studied in the auditory-to-visual direction, showing instead, a greater emphasis on the perception of other stimuli characteristics.<sup>13-17</sup>

In summary, while much of the research in the auditory-to-visual direction appears to focus mainly on the experience of affect, some studies have instead chosen to explore the perception of stimulus emotionality. When comparing the two directions of influence, it is apparent that the visual-to-auditory direction as a whole, has been less explored. From reviewing available research in this direction, the lack of emphasis on the topic of affect in general is clear. There are few studies of experienced affect, and even fewer studies suggesting cross-modal effects on the perception of stimulus emotionality. While some of these studies do explore participant ratings of auditory stimuli, few ask whether the emotionality of auditory stimuli can be altered by the presence of emotional visual information.

Given this review of the current literature, we hypothesized that emotionality ratings of both auditory (musical excerpts) and visual (images) stimuli will be influenced by their respective cross-modal stimuli. We also suggest that these effects will be greater when stimuli in the attended modality are ambiguous and those in the unattended modality are unambiguous (happy or sad), than from other possible combinations. We identified “ambiguous” stimuli by finding those that maximized the value of a unique difference score calculation, reasoning that such stimuli could be judged as either somewhat happy or somewhat sad. To the best of our knowledge, there are no studies to date that have investigated the bidirectional cross-modal influence on emotionality ratings of both auditory and visual stimuli, with ambiguous stimuli operationalized in this way.

## Methods

### I. Participants

Fifty student participants (32 females, 17 males, and 1 participant who declined to report gender) were selected from the Research Participation Program (RPP) through the Psychology Department at the University of California, Berkeley. Students received one credit for participating. All participants were between eighteen and sixty-one years of age, with the average age being in the early twenties ( $M = 22.88$ ,  $SD = 6.95$ ). One participant declined to report their age, and was not calculated into the age range. Participants were primarily Asian, with thirty- three identifying as such. Ten participants iden-

tified as Caucasian, three as Hispanic/Latino, two as both Caucasian and Asian, one as both African-American and Hispanic/Latino, and one as an ethnicity not specified. Four participants reported neurological conditions, with three of those four participants also reporting the type of medication prescribed for the listed condition. Although English was not the only native language reported, all of the fifty participants spoke English fluently, and had no difficulty understanding the instructions for the experiment.

### II. Materials

The auditory and visual stimuli used in this experiment were chosen through the review of numerous university stimuli sets that had been made publicly available online. The auditory stimuli were selected from a set obtained from the University of Jyväskylä,<sup>23</sup> and the visual stimuli from a set obtained from Cornell University.<sup>24</sup> The auditory stimuli consisted of film score segments spanning a wide range of genres such as romance, horror, comedy, and drama, that had been rated on levels of valence, energy, and several emotions. The visual stimuli consisted of a broad range of static photographic images, including, but not limited to, people, animals, nature, inanimate objects, and various social situations. The stimuli in both the auditory and visual sets, had all been previously rated on numerous emotions including sadness and happiness. The same selection criteria were used for the auditory and visual stimuli sets to define three conditions: sad, ambiguous, and happy. For the sad and happy stimuli, this involved computing a difference score between the value listed for the intended emotion (either sadness or happiness) and the value for the unintended emotion (the opposite emotion). For ambiguous stimuli, the absolute value of the difference score was subtracted from the addition of the happy and sad values. This simple calculation yielded an ambiguity score for each stimulus. Stimuli belonging to the highest 30 difference scores in each emotionality category of both modalities were identified. However, due to high content similarity between the top 30 stimuli in the visual sad emotionality category, the highest 95 difference scores were used instead. The 15 stimuli in each category that were judged to be the most different from each other were selected by the experimenter as final stimuli. The final stimuli set used in the experiment thus consisted of 90 stimuli; 45 auditory stimuli and 45 visual stimuli, with each category of emotion in both modalities containing 15 stimuli. The top three stimuli in both the happy and sad emotion categories from each of the two stimuli modalities were used in an instructional anchoring task designed to teach participants how to navigate through the experiment. All auditory stimuli were shortened to 10 seconds, with the first and last 2 seconds of each excerpt fading in and out, respectively. All images were cropped to the same size (8x8 inches, 72 pixels/inch). It is important to note that, although the same method for stimulus selection was used for both the auditory and visual stimuli sets, different numerical scales had been used in the original stimuli sets, as they had not been collected from the same source.



Figure 1. The images from each emotion category shown above were used in combination with short audio segments for each individual trial.

Figure 2. The 45 trials that each participant experienced consisted of 5 random pairings in each of the 9 combinations of Auditory and Visual stimuli shown to the right.

### III. Design

Three independent variables were included in the experimental design; the modality of the to-be-rated attended stimuli (auditory or visual) which participants were assigned, the emotionality category of the auditory stimulus (sad, ambiguous, or happy), and the emotionality category of the visual stimulus (sad, ambiguous, or happy). The first variable (attended and rated modality) was a between-subjects variable, and the second and third variables were within-subjects variables. The dependent variable consisted of the ratings that participants made to stimuli in their attended modality, which depended on the rating condition to which the participant was assigned. Figure 2. depicts the categories of emotionality pairings presented to participants and the number of each such pairing trials. The pairings of auditory and visual stimuli were chosen at random (without replacement) for each participant on each trial. This design, and randomization was used for both rating conditions.

### IV. Procedure

Participants were assigned to either the auditory or visual rating condition in an alternating pattern. Instructions had been tailored appropriately for each of the two rating conditions, and were read aloud to the participant by the experimenter. Participants were told that they would be presented with a series of stimuli. The modality of these stimuli depended on the rating condition of the participant (auditory or visual). A rating scale was shown below the instructions that consisted of a single horizontal line, positioned between two smaller vertical tick marks on either side of the scale, which indicated the left and right extremities. The left end of the scale was labeled “Sad”, and the right end of the scale was labeled “Happy”. Participants were told that this rating scale would appear on each trial, ten seconds after presentation of the stimulus, and that this scale was to be used to record their judgment about the emotionality of the attended stimulus in that trial. Participants were also told that in each trial, a stimulus from another modality (the modality of the other rating condition) would accompany the stimulus that they had been asked to rate. In each rating condition, participants were asked to ignore these additional stimuli, but to do so in a way that would not prevent them from experiencing the stimuli (i.e., without covering their ears or closing their eyes). Attended and unattended stimuli modalities were consistent across trials for all participants in a given rating condition. Participants were asked to use the full scale when making their ratings throughout the course of the experiment.

The experiment itself consisted of 45 trials with a single auditory-visual pairing in each trial. All participants were exposed to the entire collection of stimuli in both modalities,

Total Trials = 45	Visual <u>Sad</u> (15)	Visual <u>Ambiguous</u> (15)	Visual <u>Happy</u> (15)
Auditory <u>Sad</u> (15)	SS (5)	SA (5)	SH (5)
Auditory <u>Ambiguous</u> (15)	AS (5)	AA (5)	AH (5)
Auditory <u>Happy</u> (15)	HS (5)	HA (5)	HH (5)

however, the order of stimulus presentation was randomized for each participant. This resulted in participants experiencing sequences of unique auditory and visual combinations.

### Results

Average ratings of the sadness/happiness of attended stimuli are plotted in Figure 3. for the auditory and visual modalities. A mixed factorial design (2x3x3) ANOVA revealed significant main effects for emotionality conditions in both auditory ( $F(2, 423) = 89.0, p < 0.0001$ ) and visual rating conditions ( $F(2, 423) = 92.6, p < 0.0001$ ). A significant interaction between rating condition and auditory stimuli emotionality was found ( $F(2, 423) = 261.90, p < 0.0001$ ). Similarly, a significant interaction between rating condition and visual stimuli emotionality was found ( $F(2, 423) = 276.16, p < 0.0001$ ). An interaction between auditory stimuli emotionality and visual stimuli emotionality, was not found to be significant ( $F(4, 423) = 0.33, p = 0.855$ ). An interaction between rating condition, auditory stimuli emotionality, and visual stimuli emotionality, was not found to be significant ( $F(4, 423) = 0.32, p = 0.87$ ).

As expected, when participants were in the auditory rating condition, the emotional category of the musical selections had a significant effect on their ratings ( $F(2, 207) = 130.77, p < 0.0001$ ). Happy music was rated as reliably happier than ambiguous music, and ambiguous music was rated as reliably happier than sad music. Similarly, in the visual rating condition, the emotional category of the photographic images had a significant effect on the ratings that participants assigned to them ( $F(2, 207) = 185.43, p < 0.0001$ ). Again, happy images were rated as reliably happier than ambiguous images, and ambiguous images were rated as reliably happier than sad images.

We now turn to the contextual effects of the emotional condition of the unattended stimuli on ratings of the

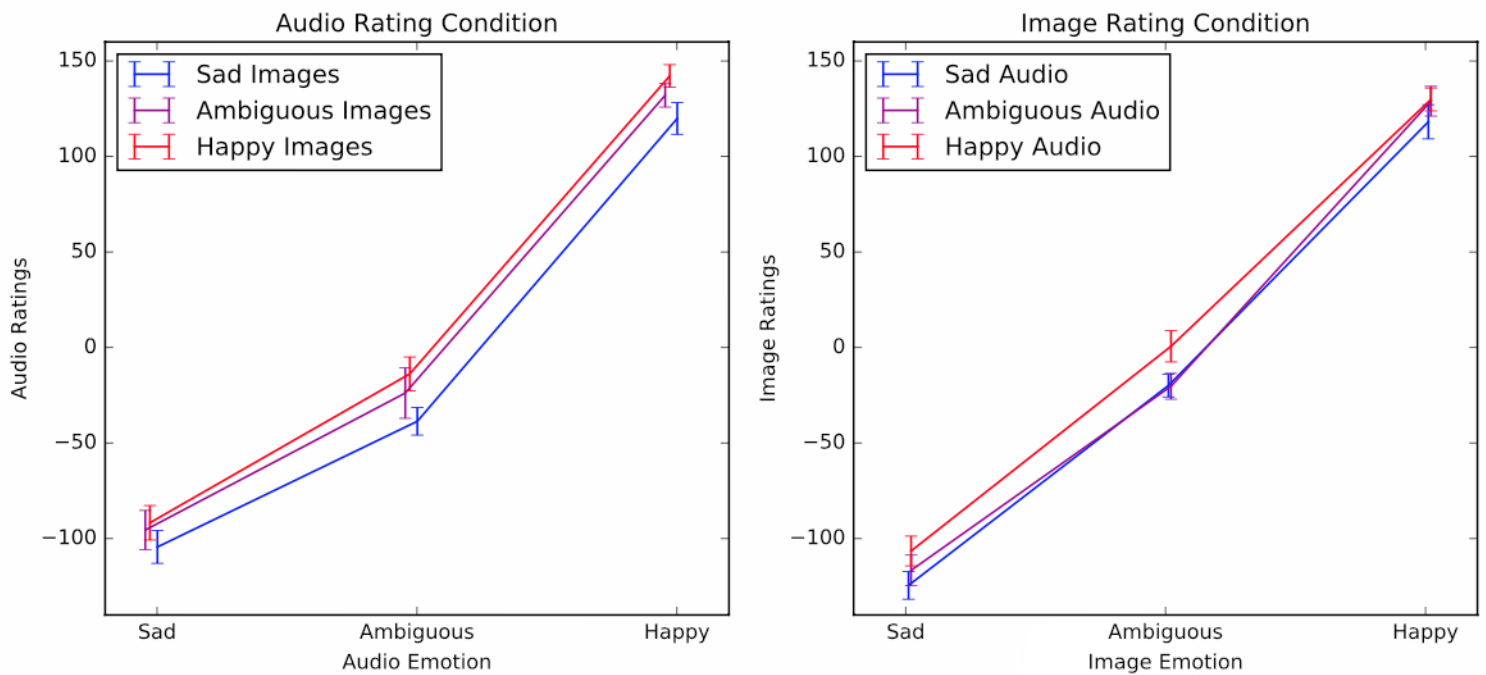


Figure 3. The graphs above display the means of each possible combination of auditory and visual emotionality in both the Auditory and Visual rating conditions.

attended stimuli. A pairwise comparison with Bonferroni correction between the effects of happy versus sad images on ratings of the happiness of the musical selections revealed a small, but significant, difference ( $t(49) = 3.18, p = 0.003$ ), in that happy images increased happiness ratings of the music and sad images decreased them. An analogous pairwise comparison with Bonferroni correction between happy and sad musical excerpts on ratings of the happiness of photographic images emotionality also revealed significant results ( $t(49) = 2.67, p = 0.01$ ), in that happy music increased happiness ratings of the images and sad music decreased them. These effects show that unattended cross-modal stimuli do indeed influence people's judgments of the emotionality of auditory and visual stimuli in both directions, as predicted. Somewhat surprisingly, there was no indication that ratings of the ambiguous stimuli were more strongly affected than the happy and sad stimuli, since the curves plotted in Figure 3. are largely parallel.

When looking at the graph in Figure 3., it seems that there might be an effect in the auditory rating condition, such that sad images are more easily able to decrease ratings of ambiguous auditory stimuli than are ambiguous images. However, a significant difference was not found to exist between these two emotionality categories in affecting participant ratings of ambiguous auditory stimuli ( $t(24) = 1.14, p = 0.27$ ). Interestingly, there seemed to be an opposite effect in the visual rating condition. The one case in which the curves deviate from the parallelism arises in the visual rating condition, where the ambiguous musical excerpts seemed to produce somewhat lower ratings of the ambiguous images than would be expected if the effects were purely additive (i.e., if the curves were fully parallel) and also somewhat higher ratings of the happy images than would be expected if the ef-

fects were purely additive. However, the results revealed no significant differences between happy and ambiguous musical excerpts on emotionality ratings given to ambiguous visual stimuli ( $t(24) = 1.95, p = 0.06$ ). However, both of these p-values were approaching significance. In the visual rating condition, there appeared to be a ceiling effect between happy and ambiguous musical excerpts on the emotionality ratings of happy images. However, there was no significant differences between the happy and ambiguous audio ( $t(24) = 0.13, p = 0.90$ ).

### Discussion

Emotionality ratings of stimuli in both the auditory (musical) and visual (image) conditions primarily depended on the emotionality of stimuli in the attended modality. That is, participants in the auditory rating condition, tended to make their ratings based largely on the emotion in the attended auditory modality (i.e., the musical excerpts), and participants in the visual rating condition, tended to make their ratings based largely on the emotion in the attended visual modality (i.e., the images). This was expected to occur, as unambiguous stimuli from both modalities produced high ratings for their respective emotionality. If this effect had not been present, it would have possibly indicated a discrepancy of the original stimulus ratings and their assignment to emotionality categories in the context of this experiment.

Also consistent with our predictions, the perception of emotionality in stimuli of both the auditory and visual modalities can also be influenced, though to a much lesser extent, by stimuli in another modality. More specifically, the emotionality ratings that participants assigned to attended stimuli in the auditory rating condition were influenced by

stimuli from the unattended visual modality. Similarly, the emotionality ratings that participants assigned to attended stimuli in the visual rating condition were influenced by stimuli in the unattended auditory modality. However, this finding was only shown when looking at the difference between happy and sad unattended stimuli on attended stimulus ratings.

Our findings replicate previous research in the sense that an auditory-to-visual direction of influence on emotionality ratings of visual information has already been shown. However, our findings also extend previous research by showing that this effect can occur in the auditory-to-visual direction through the use of visual stimuli containing general types of information such as scenes, nature, animals, etc. Previous studies examining visual emotionality have presumably used more limited types of stimuli, as the majority of research in this direction has focused on film and how it is influenced by music, where human characters serve as foundational elements in the construction of plot lines. However, human characters are not the only elements in film, or life in general, and it is important to investigate whether the perceived emotionality of other visual components in film and images (such as those mentioned), can also be influenced by stimuli from another modality. By using images that were not confined to a single form (e.g., human faces), we feel that we were able to more accurately represent influences that may occur in every day life.

Our cross-modal finding also expands the more limited amount of research showing a visual-to-auditory direction of influence. While previous studies have shown influences from visual stimuli to auditory stimuli, few have focused on the perceived emotionality of the auditory stimuli themselves. A single study<sup>22</sup> did find evidence for bidirectional cross-modal influences on stimulus emotionality ratings, but it used images of faces and single sentence vocal recordings as visual and auditory stimuli, respectively. It remains a unique finding, as research in the visual-to-auditory direction related to stimuli emotionality has been extremely limited. However, their use of single sentence audio recordings seems insufficient to represent visual-to-auditory effects in general. It is therefore important to study how other types of visual information can shift perception of emotionality in more general types of auditory information, such as music. We aim to determine whether such bidirectional cross-modal influences on ratings of emotionality can be generalized to musical and pictorial stimuli. The present results establish such effects.

An interesting effect was noted across rating conditions. It appears as if the emotionality ratings of ambiguous auditory stimuli in the auditory rating condition, were more heavily influenced by sad images, than they were by ambiguous images. In the visual rating condition, the ratings that participants assigned to ambiguous visual stimuli, seemed to be altered to a great extent by happy musical excerpts, but

not nearly as much when presented with ambiguous or sad musical excerpts. While both of these pairwise comparisons did not reveal a significant result, the p-values were trending towards significance in both cases. Most likely, a larger sample size would have yielded significant results in both of these tests. An interesting observation is the fact that bidirectional cross-modal effects were still shown, even though the experimental design explicitly asked participants in both rating conditions to ignore stimuli from the unattended modality. This may indicate that when individuals are not putting effort into ignoring these stimuli, as in the case of viewing a film or performance, that these effects would be even larger than shown here.

### **Limitations**

The sample of participants used in this experiment consisted entirely of UC Berkeley undergraduate students who were elicited through the Psychology department's RPP system. As a result, most of these students were Psychology majors. It is likely that there are characteristics of this sample that are not representative of the general population. In the future, a larger and more accurately representative sample might be collected through an online recruitment process such as Amazon Mechanical Turk.

A second limitation of our experiment involves the high degree of similarity that existed between images in the sad emotional category for stimuli in the visual modality. For instance, many of the images in the sad emotional category portrayed individuals covering their face with their hands, crying, looking down, or in other poses that are evocative of sadness. This theme dominated approximately the upper third of the sad images. Such an observation would be unworthy of mention, if it were not for the wide variability of image content in the other emotional categories. Upon evaluation of difference scores in this category, we found it necessary to search more possible stimuli to find a set that we felt were of a more general nature and suitable for use in this experiment. In future studies, it would be ideal to incorporate images that reflect sadness in more varied representations. However, we suspect that this is not easy to accomplish.

### **Implications**

The phenomenon of bidirectional cross-modal influence in emotional responses supports the idea that music and film have a unique effect when paired, an effect that has most likely been understood by film makers for decades. Such effects could be useful in many other sub-areas of entertainment, not necessarily limited to the realm of film and visual performances, as implied by prior findings. Acknowledging the lack of research in the visual-to-auditory direction of influence, the findings of this study may be of particular importance to musically based forms of entertainment, especially those with simultaneous visual components, such as music videos.

The use and popularity of the music video has increased steadily in recent years. Essentially, such videos are the building blocks for creating more powerful productions of auditory and visual material. Knowledge of bidirectional cross-modal influence could possibly be used in assisting artists to construct film sequences with the intention of changing the ways in which listeners hear their music. Similarly, these findings could be put to use in developing a music visualizer that is far superior to what is currently available in altering perceived and experienced emotion.

Knowledge of bidirectional cross-modal influence may not be entirely restricted to forms of entertainment either. While our study focused specifically on the perceived emotionality of auditory and visual stimuli, participants most likely experienced some level of emotional change throughout the course of the experiment. It is possible that the results of this study could help clinicians in incorporating more effective emotional stimuli into various forms of therapy.

### Acknowledgments

We would like to thank the UC Berkeley Psychology department, and all of our RPP participants for their time and effort. We thank research assistants Liang Hao, and Sai Ting Chu for their help in running participants, and Liang Hao for his additional help in data analysis.

### ABOUT THE AUTHOR



Harrison James Ramsay  
Senior  
Psychology Major  
Palmer Lab

Completing my senior honors thesis helped to strengthen my interest in scientific research. However, I soon realized that it would be difficult to answer the types of questions that I found myself asking, if I were to continue my studies within the field of Psychology. I then made the necessary decision to reroute my path towards the direction of Cellular and Molecular Neuroscience. I am currently in my third and final year at UC Berkeley, enrolled in additional science courses that will help me to succeed in the Neuroscience Ph.D. programs that I will be applying to next Fall. I am highly interested in studying learning and memory at the synaptic level, and I would like to secure an industry job after I complete my Ph.D. I also plan on starting a business that specializes in brain health supplements. In my free time, I produce hip-hop and electronic music, promoting myself through the persona, "The Neuroscientist".

### References

1. Krumhansl, C. L. (1997). An exploratory study of musical emotions and psychophysiology. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 51(4), 336.
2. Philippot, P. (1993). Inducing and assessing differentiated emotion-feeling states in the laboratory. *Cognition & Emotion*, 7(2), 171-193.
3. Thayer, J. F., & Levenson, R. W. (1983). Effects of music on psychophysiological responses to a stressful film. *Psychomusicology: A Journal of Research in Music Cognition*, 3(1), 44.
4. Baumgartner, T., Esslen, M., & Jäncke, L. (2006a). From emotion perception to emotion experience: Emotions evoked by pictures and classical music. *International Journal of Psychophysiology*, 60(1), 34-43.
5. Baumgartner, T., Lutz, K., Schmidt, C. F., & Jäncke, L. (2006b). The emotional power of music: how music enhances the feeling of affective pictures. *Brain research*, 1075(1), 151-164.
6. Boltz, M. G. (2001). Musical soundtracks as a schematic influence on the cognitive processing of filmed events. *Music Perception: An Interdisciplinary Journal*, 18(4), 427-454.
7. Bullerjahn, C., & Guldenring, M. (1994). AN EMPIRICAL INVESTIGATION OF EFFECTS OF FILM MUSIC USING QUALITATIVE CONTENT ANALYSIS. *Psychomusicology*, 13, 99-118.
8. Vitouch, O. (2001). When your ear sets the stage: Musical context effects in film perception. *Psychology of Music*, 29(1), 70-83.
9. Tan, S. L., Spackman, M. P., & Bezdek, M. A. (2007). Viewers' interpretations of film characters' emotions: Effects of presenting film music before or after a character is shown. *Music Perception: An Interdisciplinary Journal*, 25(2), 135-152.
11. Jeong, J. W., Diwadkar, V. A., Chugani, C. D., Sinsongsud, 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(4), 2973-2982.
12. Logeswaran, N., & Bhattacharya, J. (2009). Crossmodal transfer of emotion by music. *Neuroscience letters*, 455(2), 129-133.
13. Geringer, J. M., Cassidy, J. W., & Byo, J. L. (1996). Effects of music with video on responses of nonmusic majors: An exploratory study. *Journal of Research in Music Education*, 44(3), 240-251.
14. Boltz, M. G. (2004). The cognitive processing of film and musical soundtracks. *Memory & Cognition*, 32(7), 1194-1205.
15. Schutz, M., & Lipscomb, S. (2007). Hearing gestures, seeing music: Vision influences perceived tone duration. *Perception*, 36(6), 888-897.
16. Saldaña, H. M., & Rosenblum, L. D. (1993). Visual influences on auditory pluck and bow judgments. *Perception & Psychophysics*, 54(3), 406-416.
17. Boltz, M. G., Ebdendorf, B., & Field, B. (2009). Audiovisual interactions: The impact of visual information on music perception and memory. *Music Perception: An Interdisciplinary Journal*, 27(1), 43-59.
22. De Gelder, B., & Vroomen, J. (2000). The perception of emotions by ear and by eye. *Cognition & Emotion*, 14(3), 289-311.
23. Eerola, T. & Vuoskoski, J. K. (2011). A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music*, 29 (1), 18-49.
24. Peng, K. C., Chen, T., Sadovnik, A., & Gallagher, A. A (n.d.) *Mixed Bag of Emotions: Model, Predict, and Transfer Emotion Distributions.*