

The Effect of Audio-Visual Cues in Understanding Musical Performance

Julie Kim¹ and Nicholas Pellegrino^{1#}

Northern Valley Regional High School at Old Tappan

#Advisor

ABSTRACT

This study investigated the effects of adding visual aesthetics to a music performance. Two modes of visual performance are explored: multimedia and performance cues. Multimedia and performer cues serve different purposes in expressing music in a performance and allowing the audience to interpret the music. Multimedia refers to the use of more than one medium of expression or communication in performance, while performance cues refer to the implementation of gestures and facial expressions to guide the direction of the music. In this research study, two experiments were performed with two separate samples of 24 subjects. These experiments observe the two modes' effects on engagement, understanding, and overall enjoyment of the performance by their audience. Participants were required to listen to the audio performance of an excerpt of music and listen to the same excerpt with the visuals as well. The subjects answered a series of questions rating the listening and watching experience. The results overall showed that having the visual experience while listening to a musical performance enhances the subjects' engagement, enjoyment and understanding of the music. A Wilcoxon ranked sum test was performed on the two different parts of this experiment and the p values were 0.0486 and 0.00148, which showed that there was enough statistical evidence in the experiment that showed engagement levels were higher when both audio and visual modes were provided to the audience.

Introduction

Music performance is often overlooked as just a simple playback of a tune. One could simply listen to a recording of a piece instead of going to a live performance to listen to the same piece. However, a live concert provides a unique element to the music that recordings do not offer; the visual aspect of performance significantly increases the appreciation and engagement of music (Platz & Kopiez, 2012). At the same time, music performance does not consist only of a performer and the music. To make a performance more effective in moving the audience, stage aesthetics are implemented and modified according to the piece. Stage aesthetics may include an accompanying piece of art or performer cues. Oftentimes, visuals tend to communicate the first attraction of a song or performance (Walker, 2017). The audience may not understand the meaning behind a song solely based on the sounds or lyrics. In fact, visuals may be used individually to understand the meaning of a song (Walker, 2017). Additionally, visual components of a song highlight certain parts of the message of the song.

Multimedia

Music and art are interconnected in many ways. Many musicians/composers may be inspired by certain paintings to write music in a certain way (*Debussy Quartet*, 2008), and vice versa. A new trend in music performance explores performing accompanied by a medium such as paintings or poems. Recent studies have observed that there is a correspondence between the two modes, also known as the audio-visual cross-modal correspondence. Audio-visual cross-

modal correspondence refers to the association between auditory and visual features (Liu, 2022). The auditory stimulus is the music while the visual stimuli are the art that accompanies the music. In its most basic form, music in major tonality corresponded with light colors and music in minor tonality corresponded with the opposite (Liu, 2022). Correspondences between colors and sound primarily seen in synesthesia, a neurological condition in which information meant to stimulate one of your senses stimulates other senses (Reymore, 2019). Synesthetes tend to associate high pitches with light colors (Sun et al., 2018). Non-synesthete musicians have strong color-pitch associations (Reymore, 2019). Sun et al. (2018) observed in a study investigating the cross-modal correspondences between music and art that gray and blue colors are associated with sadder music. Certain color properties were found to be easier for people to associate with a certain sound (Sun et al., 2018). Tan and Antony (2020) conducted a study observing the effects of generating music from images and found that there was a correlation between the emotional ratings of the image and the generated music from the image. However, one danger is that evaluators may express fondness for the music instead of the emotion it triggers (Tan & Antony, 2020). Crossmodal pairing preferences may come from structural or functional overlaps in the brain (Reymore, 2019). There is a possibility that people make pairings because of what occurs most often in the environment (Reymore, 2019).

An implication of audio-visual cross-modal correspondences is that as people listen to the music and observe the accompanying media, they have emotional responses or recognize the emotional content of the music (Liu, 2022). The audience also observes the painting features that may reflect the structure of the music (Liu, 2022).

Performance Cues

A musical performance typically includes the performers sitting or standing in a way that the audience can hear the music and stay engaged with the performance. The format of the stage is also modified by the performers for the purpose of communication. Several performance cues are leveraged, which include gestures and facial expressions. Performers also internally create mental imagery to guide the direction of the music they produce (Keller, 2012). Audience members, as a result, may empathize with the performers in order to understand the performers' intentions (Brinck, 2017).

Gestures in a music performance are physical movements. Regardless of the number of performers on stage, many implement gestures in their performances as a way of communicating their own interpretation of the music they are performing (Aroso, 2021). Dotov et al. (2021) has conducted a study observing the energy and coordination between performers of a wide range of music and found that the complexity of gestures are leveraged to match the context of the music (Dotov et al., 2021). Gestures also help performers match to tempos better and maintain coordination and energy between each other (Dotov et al., 2021). As a result, gestures cause higher emotional responses from the audience (Goodchild et al., 2017). Audiences reacted to gestures highlighted by texture, spectral centroid, dynamics, and tempo, according to the study conducted by Goodchild et al. However, there are instances when visual components of performances contribute negatively to the performance. A study conducted by Huang & Krumhansl (2011) demonstrated that seeing the performer evoked lower ratings for the performance than simply listening to the music. Huang & Krumhansl found that non-musicians may not have been familiar with the performer's stage behavior, or there may have been a lack of expression from the performers that negatively contributed to the audience's ratings. On the other hand, According to Waddel & Williamon (2017), appropriate stage entrance correlated with a more positive first impression of the performance. However a performance error and an inappropriate facial reaction between the performers led to a lower performance rating. For example, sad facial expressions referred to lower trait dominance while happy/surprised facial expressions referred to higher dominance (Waddel & Williamon, 2017). The difference is that a negative first impression was quickly forgiven based on the performance quality that followed (Waddel & Williamon, 2017). Therefore, musical content is deemed more effective in changing the audience's opinions. The level of musical expertise may also have an effect on an audience member's perception of the musical performance.

Musicians may use mental imagery to play together in a performance. Mental imagery refers to pre-planning the performance by creating imagery of the music beforehand. Musicians use mental imagery and, as a result, play the right notes, play fast, and play accurately (Keller, 2012). The process is also referred to notational audiation, which is imagining what the music sounds like before hearing it in reality. This process involves the interplay of brain regions implicated in auditory and motor processing (Keller, 2012). One negative impact is when musicians in an ensemble each use different imagery, there is variation in expression and coordination (Keller, 2012).

An additional inquiry explored is how audience members understand the intentions of the performers. A study conducted by Brinck (2017) investigated the emotional responses of audience members while watching a performance. Brinck observed that body movements cue emotions and the audience to bodily and emotionally move with the performers. It was also observed that the performance can cause feelings of insight or awe that is unfamiliar to the audience. Additionally, Reason & Reynolds (2010) observed that audience members had admiration for virtuous gestures and appreciation for the effort and emotional intention in the movements. As a result, audience members with more empathy were more accurate in recognizing the musicians' intentions (Wöllner, 2012). Therefore, empathy has an impact on the appreciation of music. Additionally, the ability to recognize intentions was not related to the musical experience of an audience member (Wöllner, 2012). Observing musicians' body movements elicits automatic internal response mechanisms.

Experimental Design

The study contained two experiments of two random samples of 24 subjects each. The two experiments explore two aspects of audio-visual cross-modal correspondence with multimedia cues and performance cues. For ease of data collection and recording, the questionnaire was administered via Google Forms. Before the experiment, demographic participant demographic information was collected, such as age, ethnicity, education level, and musical experience. Musical experience included any experience involving a musical instrument or a musical group. Each participant now had to listen to just the audio of an excerpt of music and listen to the audio with visual aids. The participants were asked to answer a series of questions about engagement and overall enjoyment. The audio part asks the participants to listen to the audio of a piece of chosen music before answering its designated questions. The same is asked again in the video part, except the participants watch the video performance of the music from the same audio.

Part 1: Multimedia as Visual Mode

The chosen music for the mode was titled *On the Blue Shore of Silence*, performed and curated by violinist Francesca DePasquale. *On the Blue Shore of Silence* is a multimedia commissioning project inspired by Pablo Neruda's collection of poems "On the Blue Shore of Silence: Poems of the Sea." Works for solo violin, piano, and electronics were created to respond to the poems. Visual art pieces were created as well to depict the image of water as one of the principal forces and elements of life. According to DePasquale, the music was created carefully in order to draw upon the colors, textures, and emotions of the poems (*On the Blue Shore*, n.d.). Participants were asked to listen to the audio of a segment (1:27) of the chosen music. The designated questions to be answered after listening to the audio were in the form of Likert scale. For example, participants were asked, "What was your overall performance rating?"; the participants would submit a rating from 1 to 5. Along with Likert scale questions, participants were asked to answer qualitative questions, requiring participants to provide open-ended responses. One such question was, "What emotions do you feel were provoked by the music?" which allowed participants to expand on their answers. After answering the audio-listening segment of questions, participants were asked to watch the video performance of the same music. Participants were not told of the fact that the audio-listening part and the video performance played the same music. As the music was being performed, one of Neruda's poems and a complimentary art piece were displayed on a screen

in the background. The video performance's designated questions included the same Likert scale and qualitative questions as the audio-listening part. Participants were also asked questions that allowed a comparison between the audio listening and the video performance. Accordingly, the participants were asked "Do you think watching the performers perform helped you enjoy the performance better?"

Part 2: Performance Cues as Visual Mode

The second survey, with a total of 24 participants, tested the visual cues of visual performance. The same demographic questions were asked in the beginning of the survey as the first survey. Again, the survey consists of an audio performance part and a video performance part, each including designated questions very similar to the first survey.

The chosen music for the mode was titled *String Quartet in G minor, op. 10* by Claude Debussy. Although Debussy wrote the quartet without a specific message to convey, he emphasized the French style that he wrote with. Important to note is that the colors of the paintings of Turner, one of Debussy's favorite painters, are strongly elicited throughout his quartet (*Debussy Quartet*, 2008). The performers in the recording include two violinists, one violist, and one cellist, who are all members of the FaMa Quartet. Like the first survey, participants were first asked to listen to the audio of a segment (1:52) of the chosen music. The designated questions to be answered after listening to the audio were in the form of Likert scale. The questions are very similar to those of the first survey but are modified to the mode of performance cues. Again, participants were asked qualitative questions. After answering the audio-listening segment of questions, participants were asked to watch the video performance of the same music. Participants were not told of the fact that the audio-listening part and the video performance played the same music. The video-performance's designated questions included the same Likert scale and qualitative questions as the audio-listening part. Participants were also asked questions that allowed a comparison between the audio-listening and the video performance like those of the first survey.

Results

To analyze and depict the data I used cross tabulation across different demographic information such as musical experience, education level as well as cross tabulated across the self-evaluation of the subjects on how well they believed the different modes were helpful in understanding the theme and emotions. Additionally, I ran a Wilcoxon ranked sum test to see if there was a difference in performance rating and engagement rating by the subjects with the different modes.

Part 1: Multimedia Cues as Visual Mode

The first survey, testing the use of multimedia in music performance, had a total of 24 participants. Ages varied from 16 to 59 years, with an average age of 27 years. The participant body included 11 Asians/Pacific Islanders, 7 White/Caucasians, 2 Middle Eastern/North Africans, and 1 Multiracial/Biracial. 3 participants have achieved a bachelor's degree, 2 have achieved a Doctorate, 3 have achieved a high school diploma, 12 have not achieved a high school diploma, 3 have achieved a master's degree, and 1 has achieved a professional degree. Musical experience ranged from 0 to 30 years, with an average of 9.938 years.

Table 1. Cross Tabulation of Average Ratings based on Musical Experience

Music experience	Average Performance Rating - Audio	Average Performance Rating - Video	Average Engagement Rating - Audio	Average Engagement Rating - Video	Number of Participants
> 5 years	4.4	4.467	3.933	4.2	15
<= 5 years	3.889	4.11	3.33	3.667	9

Table 1 above shows the average performance ratings and level of engagement ratings for both audio and video performance on a scale from 1 to 5, comparing participants with greater than 5 years of musical experience and participants with less than or equal to 5 years of musical experience. More participants had greater musical experience than those with lower musical experience. In both groups, the performance and engagement ratings increased to some extent between watching the audio and the video performance. Participants with lower experience demonstrated a more distinct difference between the ratings of audio and video than those with higher experience.

Table 2. Cross Tabulation based on Effectiveness of Audio Modality of Music in Understanding Emotions

Level of how much listening contributed to understanding performance's emotions	Emotions provoked by performance	Average Performance Rating	Average Engagement Rating	Number of Participants
0-3	Longing, loss, sadness, eeriness	4	3.167	12
4-5	Confusion, mystery, anxiety, sadness, loneliness	4.416	4.25	12

Table 2 above shows the emotions perceived by the audio performance and average performance ratings and level of engagement ratings for the audio performance, comparing participants who believed that listening to the performance contributed to understanding the emotions of the performance (rated 4 or 5) and those who believed that listening did not contribute much (rated 0-3). The higher rating group had a significantly higher performance and engagement rating than the lower rating group. The lower rating group tended to provide less explanation for their qualitative answers.

Table 3. Cross Tabulation based on Effectiveness of Visual Modality of Music in Understanding Emotions

Level of how much watching contributed to understanding performance's emotions	Emotions provoked by performance	Average Performance Rating	Average Engagement Rating	Number of Participants
0-3	suspense, confusion, wondering	4.333	3.333	6
4-5	calmness, unease, sadness, longing, confusion	4.333	4.222	18

Table 3 shows emotions perceived by the video performance and average performance ratings and level of engagement ratings for the video performance, comparing participants who believed that watching the performance contributed to understanding the emotions of the performance (rated 4 or 5) and those who believed that watching did not contribute much (rated 0-3). The performance ratings for both groups were the same, while the higher rating group had a significantly higher engagement rating than the lower rating group. The number of participants increased in the higher rating group from the audio performance. The emotions provoked by the performance were more similar between the groups.

Table 4. Cross Tabulation based on Effectiveness of Audio Modality of Music in Understanding Themes

Level of how much listening contributed to understanding performance's themes	Themes provoked by performance	Average Performance Rating	Average Engagement Rating	Number of Participants
0-3	death, tragedy, loneliness, longing, loss	4.222	3.333	15
4-5	challenge, curiosity, wonder, ambition, discovery, amazement	4.2	3.933	9

Table 4 above shows themes perceived by the audio performance and average performance ratings and level of engagement ratings for the audio performance, comparing participants who believed that listening to the performance contributed to understanding the emotions of the performance (rated 4 or 5) and those who believed that listening did not contribute much (rated 0-3). Both groups had a similar average performance rating, while the higher rating group were significantly more engaged in the performance than the lower rating group. However, there were more participants who were in the lower rating group than the higher rating group.

Table 5. Cross Tabulation based on Effectiveness of Visual Modality of Music in Understanding Themes

Level of how much watching contributed to understanding performance's themes	Themes provoked by performance	Average Performance Rating	Average Engagement Rating	Number of Participants
0-3	curiosity, overcoming something	4.375	3.5	8
4-5	growth, challenge, hope, nature	4.312	4.25	16

Table 5 above shows themes perceived by the video performance and average performance ratings and level of engagement ratings for the video performance, comparing participants who believed that watching the performance contributed to understanding the emotions of the performance (rated 4 or 5) and those who believed that watching did not contribute much (rated 0-3). Both groups had a similar average performance rating, while the higher rating group were significantly more engaged in the performance than the lower rating group. There were more participants who were in the higher rating group than the lower rating group. Additionally, I ran two Wilcoxon ranked sum test between the paired samples of audio and visual responses of two different questions (overall rating and engagement rating) from the subjects the table below depicts the results of each test.

Table 6. Wilcoxon Ranked Sum Test on Paired Sample of Performance Rating and Engagement Rating of Two Treatments (Audio Listening vs Audio-Visual Listening)

Test Sample	Hypothesis	Test Statistic	P-value
Overall Performance Rating	Null Hypothesis: The two modes have a mean difference rating of 0. Alternative Hypothesis: The two modes have a mean difference rating that is not 0.	$z = -1.0142$	$p = 0.155$ (not significant)
Engagement Rating	Null Hypothesis: The two modes have a mean difference rating of 0. Alternative Hypothesis: The two modes have a mean difference rating that is not 0.	$z = -1.680$	$p = 0.046$ (significant)

Towards the end of the survey, participants were asked if the video performance including the poetry distracted them from the music. There were nearly the same number of participants who believed the poetry did distract them from the music (13 participants) and the participants who believed the poetry did not distract them (11 participants). Those who believed that the poetry distracted them from the music have stated that reading the poems did not line up with the phrases, the words of the poem kept them from understanding the music, and there was a disconnection between the poetry and the music. On the other hand, those who believed that the poetry did not distract them from the music have stated that the poetry added another layer of understanding to the music and helped narrow their own interpretations of the music.

Participants were also asked if the video performance helped them enjoy the performance better. 18 participants responded that the video performance did help them enjoy the performance better while only 6 participants responded with the opposite. Those who responded with the former explained that they could understand what the performer was feeling and what the piece was about. Those who responded with the latter explained that they could not see the performer for the entirety of the excerpt.

Part 2: Performance Cues as Visual Mode

The second survey, testing the use of performance cues, had a total of 24 participants. Ages varied from 16 to 60 years, with an average age of 27 years. The participant body included 11 Asians/Pacific Islanders, 8 White/Caucasians, 1 Middle Eastern/North Africans, 2 Hispanics/Latinos, and 2 Multiracial/Biracial. 3 participants have achieved a bachelor's degree, 2 have achieved a Doctorate, 4 have achieved a high school diploma, 9 have not achieved a high school diploma, 3 have achieved a master's degree, 1 has achieved a professional degree, 1 has achieved an Associate's Degree, and 1 has attended some college. Musical experience ranged from 0 to 16 years, with an average of 4.958 years.

Table 7. Cross Tabulation of Average Ratings based on Musical Experience.

Music experience	Average Performance Rating - Audio	Average Performance Rating - Video	Average Engagement Rating - Audio	Average Engagement Rating - Video	Number of Participants
> 5 years	4.556	4.778	4.111	4.667	10
<= 5 years	4.222	4.333	3.444	4.222	14

Table 7 above shows the average performance ratings and level of engagement ratings for both audio and video performance on a scale from 1 to 5, comparing participants with greater than 5 years of musical experience and participants with less than or equal to 5 years of musical experience. More participants had less musical experience than those with greater musical experience. In both groups, the performance and engagement ratings increased to some extent between watching the audio and the video performance. Participants with lower experience demonstrated a more distinct difference between the ratings of audio and video than those with higher experience.

Table 8. Cross Tabulation based on Effectiveness of Listening of Music in Evoking Emotions

Level of how much listening contributed to understanding performance's emotions	Emotions provoked by performance	Average Performance Rating	Average Engagement Rating	Number of Participants
0-3	calmness, excitement, uneasiness, magical	4.307	3.231	13
4-5	calmness, excitement, curiosity, happiness	4.636	4.182	11

Table 8 above shows the emotions perceived by the audio performance and average performance ratings and level of engagement ratings for the audio performance, comparing participants who believed that listening to the performance contributed to understanding the emotions of the performance (rated 4 or 5) and those who believed that listening did not contribute much (rated 0-3). The higher rating group had a significantly higher performance and engagement rating than the lower rating group. The lower rating group noted that they were confused about what the emotions were.

Table 9. Cross Tabulation based on Effectiveness of Performance Cues in Evoking Emotions

Level of how much watching contributed to understanding performance's emotions	Emotions provoked by performance	Average Performance Rating	Average Engagement Rating	Number of Participants
0-3	amazement, frustration, satisfaction, gloominess	4	3.5	4
4-5	excitement, calmness, joy, deep connection	4.75	4.5	21

Table 9 above shows emotions perceived by the video performance and average performance ratings and level of engagement ratings for the video performance, comparing participants who believed that watching the performance contributed to understanding the emotions of the performance (rated 4 or 5) and those who believed that watching did not contribute much (rated 0-3). The performance rating and engagement rating were significantly higher in the higher rating group. The number of participants increased in the higher rating group from the audio performance. However, the emotions provoked by the music were very distinct.

Table 10. Cross Tabulation based on Effectiveness of Visual Synchrony in Engagement

How effective do you think the visual synchrony (if any) between the performers contributed to your engagement with the music?	Average Performance Rating	Average Engagement Rating	Number of Participants
0-3	3.5	3	2
4-5	4.727	4.454	22

Table 10 above shows the average performance ratings and level of engagement ratings for the video performance, comparing participants who believed that the visual synchrony between the performers (if any) contributed significantly to their engagement with the music (rated 4 or 5) and participants who believed the opposite (rated 0-3). There were significantly more participants in the higher rating group than the lower rating group. The former group also had a much higher average performance and engagement rating than the latter group.

Table 11. Cross Tabulation based on Performers Facial Expressions and Empathy for Music

How strongly did you empathize with the performers' facial expressions (if any) during the performance?	Average Performance Rating	Average Engagement Rating	Number of Participants
0-3	4.615	4.154	13
4-5	4.636	4.545	11

Table 11 above shows the average performance ratings and level of engagement ratings for the video performance, comparing participants who empathized strongly with the performers' facial expressions (rated 4 or 5) with participants who did not empathize as much (rated 0-3). The two groups had very similar performance ratings while the engagement ratings were higher for the higher rating group.

Similarly, to part 1, I ran two Wilcoxon ranked sum test between the paired samples of audio and visual responses of two different questions (overall rating and engagement rating) from the subjects. With the overall performance rating as the paired sample the data had too many similar values therefore a Wilcoxon ranked test was invalid. However, the results for the Wilcoxon ranked sum test proved to be significant. The results are shown in Table 12.

Table 12. Wilcoxon Ranked Sum Test on Paired Sample of Performance Rating and Engagement Rating of Two Treatments (Audio Listening vs Audio-Visual Listening)

Test Sample	Hypothesis	Test Statistic	p-value
Overall Performance Rating	Null Hypothesis: The two modes have a mean difference rating of 0. Alternative Hypothesis: The two modes have a mean difference rating that is not 0.	<i>Invalid study the sample pairs had too many similar responses.</i>	<i>Invalid study the sample pairs had too many similar responses.</i>
Engagement Rating	Null Hypothesis: The two modes have a mean difference rating of 0. Alternative Hypothesis: The two modes have a mean difference rating that is not 0.	$z = -3.1798$	$p = 0.00148$ (significant)

Towards the end of the survey, participants were asked what aspects of the performance moved them. 22 participants indicated that they were moved by some aspect of the performance while 2 participants did not believe that the performance moved them. Many participants noted that the body movements and coordination significantly moved them. Others noted that the dynamics, phrasing, style, timing, and facial expressions moved them.

Participants were also asked if the video performance helped them enjoy the performance better. Many participants noted that the visual component positively contributed to the performance, but it did not change their impression of the piece. The participants were much more engaged in the performance as well.

Conclusion

After analyzing the results of the two surveys, several conclusions can be made. The overall performance and engagement ratings increased after including visuals (playing the video performance) to the performance regardless of the type of visual performance (multimedia/visual cues), or musical experience. Visual aesthetics, therefore, allow the audience to enjoy the performance better and stay more engaged with the performance. However, participants with greater musical experience (greater than 5 years) had higher performance and engagement ratings than those with less musical experience (less than or equal to 5 years). Visual aesthetics contribute greatly to understanding the emotions and themes elicited by the music. Recognizing a piece of music has a significant effect on one's rating of the music. Body movements, such as gestures and facial expressions have a positive effect on the audience's ratings of the music. Visual synchrony has a positive effect on the audience's ratings of the music. Visual aesthetics allow the audience to empathize strongly with the emotions in the performance. However, the use of multimedia has the potential of distraction.

In addition to the cross-tabulations, the Wilcoxon ranked sum tests resulted in significant performance engagement of the subjects. The p-values of the part 1 Wilcoxon ranked sum test shown in table 6 was 0.0486 and for part 2 shown in table 12 was 0.00148. This showed that there was a significant difference in engagement when listening to just the audio and finally listening to the video. Since the z score was negative, this meant that overall the engagement rating was higher when the visual aspect of the experiment was shown as well.

Other studies that have preceded this one has promoted the notion that visual aesthetics allow audiences to enjoy and engage with a musical performance better than a performance without visuals. More specifically, older studies have proven that there is a cross-modal correspondence between visual and auditory stimuli, when there is another medium involved in a performance. As a result, audience members are able to understand the performers' musical intentions better compared to a performance without visual aesthetics. The data collected in this study has proven to be statistically significant, showing that audience members enjoyed a musical performance better and became more engaged to the performance when visual components were added to the performance. Additionally, the musical experience has a significant effect on an audience member's likeness and engagement to a performance, where audience members with more musical experience enjoyed and were more engaged to a performance than audience members with less experience. The visual component of a performance allows audience members to understand the meaning of a performance better.

When examining the performance cues of a musical performance, older studies have proven that gestures and certain facial expressions significantly impact an audience member's liking for a performance. Previous studies have also shown that performers utilize mental imagery to play synchronously. Additionally, audience members can empathize with the performers' facial expressions and movements to understand what the performers' intentions for the music are. The data collected in this study has proven to be statistically significant, showing that audience members enjoyed a musical performance better and became more engaged in the performance when visual components (performance cues) were added to the performance, similar to the performance, including multimedia. The musical experience has a similar effect on an audience member's liking and engagement to a performance. When performers play synchronously the audience members tend to stay more engaged to the performance. Audience members are able to empathize with the performers' emotions when visual components are added to a performance. Recognition of a musical piece contributes to the liking and engagement to a musical performance.

References

- Acosta, C. M. F. (2018, November 30). *OJSL 2018 - 01 - 02 - 04*.
<http://centerprode.com/ojsl/ojsl0102/coas.ojsl.0102.04057a.html>
- Aroso, N. (2021). Role of Movement and Gesture in Communicating Music Expressiveness to an Audienc. *Convergences - Journal of Research and Arts Education*, 13(25), 83–94.
<https://doi.org/10.53681/c1514225187514391s.25.5>
- Brinck, I. (2017). Empathy, engagement, entrainment: the interaction dynamics of aesthetic experience. *Cognitive Processing*, 19(2), 201–213. <https://doi.org/10.1007/s10339-017-0805-x>
- Castillo Acosta, M. F. (2018). A Multimodal Discourse Analysis of Childish Gambino's Music Video "This is America." *Open Journal for Studies in Linguistics*, 1(2), 57–72.
<https://doi.org/10.32591/coas.ojsl.0102.04057a>
- Debussy Quartet*. (2008, October 21). Brentano Quartet. <https://www.brentanoquartet.com/notes/debussy-quartet/>
- Dotov, D., Bosnyak, D., & Trainor, L. J. (2021). Collective music listening: Movement energy is enhanced by groove and visual social cues. *Quarterly Journal of Experimental Psychology*, 74(6), 1037–1053.
<https://doi.org/10.1177/1747021821991793>
- Goodchild, M., Wild, J., & McAdams, S. (2017). Exploring emotional responses to orchestral gestures. *Musicae Scientiae*, 23(1), 25–49. <https://doi.org/10.1177/1029864917704033>
- Huang, J., & Krumhansl, C. L. (2011). What does seeing the performer add? It depends on musical style, amount of stage behavior, and audience expertise. *Musicae Scientiae*, 15(3), 343–364.
<https://doi.org/10.1177/1029864911414172>
- Keller, P. E. (2012). Mental imagery in music performance: underlying mechanisms and potential benefits. *Annals of the New York Academy of Sciences*, 1252(1), 206–213. <https://doi.org/10.1111/j.1749-6632.2011.06439.x>

- Liu, M. (2022). Paintings of Music. *The Journal of Aesthetics and Art Criticism*, 80(2), 151–163.
<https://doi.org/10.1093/jaac/kpac003>
- Nakra, T. M., & BuSha, B. F. (2014). Synchronous Sympathy at the Symphony. *Music Perception*, 32(2), 109–124.
<https://doi.org/10.1525/mp.2014.32.2.109>
- On the Blue Shore of Silence*. (n.d.). Francesca dePasquale. Retrieved November 9, 2022, from
<https://www.francescadepasquale.com/blueshore>
- Platz, F., & Kopiez, R. (2012). When the Eye Listens: A Meta-analysis of How Audio-visual Presentation Enhances the Appreciation of Music Performance. *Music Perception*, 30(1), 71–83.
<https://doi.org/10.1525/mp.2012.30.1.71>
- Reason, M., & Reynolds, D. (2010). Kinesthesia, Empathy, and Related Pleasures: An Inquiry into Audience Experiences of Watching Dance. *Dance Research Journal*, 42(2), 49–75.
<https://doi.org/10.1017/s0149767700001030>
- Reymore, L. (2019). Making the rainbow connection: audio-visual crossmodal correspondences between colors and musical instrument timbres.
- Sun, X., Li, X., Ji, L., Han, F., Wang, H., Liu, Y., Chen, Y., Lou, Z., & Li, Z. (2018). An extended research of crossmodal correspondence between color and sound in psychology and cognitive ergonomics. *PeerJ*, 6, e4443. <https://doi.org/10.7717/peerj.4443>
- Tan, X., Antony, M. (2020). Automated Music Generation for Visual Art through Emotion.
<http://computationalcreativity.net/iccc20/papers/137-iccc20.pdf>
- Touizrar, M., Knoll, A. L., & Siedenburg, K. (2021). Repetition and Aesthetic Judgment in Post-tonal Music for Large Ensemble and Orchestra. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.673706>
- Waddell, G., & Williamon, A. (2017). Eye of the Beholder: Stage Entrance Behavior and Facial Expression Affect Continuous Quality Ratings in Music Performance. *Frontiers in Psychology*, 8.
<https://doi.org/10.3389/fpsyg.2017.00513>
- Walker, R. J. (2017). Multimodality and the song: Exploiting popular song in the university classroom. *The Pansig Journal*.
- Wöllner, C. (2012). Is empathy related to the perception of emotional expression in music? A multimodal time-series analysis. *Psychology of Aesthetics, Creativity, and the Arts*, 6(3), 214–223.
<https://doi.org/10.1037/a0027392>