*Cognitive Dissonance: An Experiment through Video and Audio*

Ayla Favati, [redacted]

Psychology of Music

Professor Pablo Ripollés

Fall 2020

**New York University, Department of Psychology**

## Abstract:

Festinger (1962) defines cognitive dissonance as the notion that a person can understand that various items are not necessarily consistent in a psychological manner with another and that it is human nature to strive to make consistency a priority. We set out to better understand the factors that contribute to cognitive dissonance between auditory and visual stimuli. We have identified three overarching categories that are fundamental for the study of cognitive dissonance in the context of audio-visual stimuli : Visual Imagery, Auditory Stimuli, and Emotional Engagement. We created a set of stimuli that varied over two features: familiarity and dissonance modification. Specifically, we selected popular and unpopular video clips, for which we created three modifications by carefully swapping the audio: comedic, dark and popular. We asked participants to rate how well the video and audio matched, as well as how familiar and pleasurable they were. Lastly, we asked participants to describe their rating and the emotion they experienced while watching the clip. We hypothesized that, for unfamiliar clips, participants would give a poorer rating to audio modifications. We found in our results evidence supporting previous research showing that participants were adept at matching original audio to video, even for the unfamiliar videos. We also found that the opposite of our hypothesis was true: for familiar clips participants actually rated specific audio modifications as better matched. It may be interesting to conduct further studies exploring if this is indeed true and why. Interestingly, familiarity and unfamiliarity ratings were not distinct enough to categorize clips as familiar and unfamiliar. Future experiments may curate a more individualized set of video clips to account for these discrepancies.

## Introduction:

Cognitive dissonance is the notion that a person can understand that various items are not necessarily consistent in a psychological manner with another and that it is human nature to strive to make consistency a priority (Festinger, 1962). By way of defining cognitive dissonance, the author proves that this psychological process can happen both consciously and subconsciously, making it not only complicated to research, but also constantly expanding (Festinger, 1962). Using Festinger's (1962) definition of cognitive dissonance, an interesting venue to study this phenomenon is to use audiovisual material. This study aims to understand the effects of cognitive dissonance on audiovisual material: video clips and original or dissonant audio. The modality-appropriateness hypothesis (Guttman et. al., 2005) based theorizes that we use our senses methodically, defaulting to vision for spatial awareness and audition for temporal awareness. They asked whether we can use the audition/temporal processes for visual changes and the visual/spatial processes for auditory changes (Guttman et. al, 2005). Indeed, there is evidence that the visual and auditory processes influence each other and that visual temporal stimuli may be cognitively transformed into auditory information (Guttman et. al, 2005). There is also evidence that suggests music can affect visual perception and that some aspects of visual stimuli, like the overall emotion perceived, can have a strong correlation with dissonance (Bravo, 2013). These findings from Bravo (2013) and Guttman et. all (2005) suggest a deeply ingrained process by which we experience interacting visual and auditory information and that one may influence the other, even if that information is contradictory to expectation. Most importantly, an article by Wu et al. (2016) tested how well humans and machines were able to determine the consonance or dissonance of audiovisual stimuli. Humans were used to determine if the 45,000 audiovisual stimuli, between music and one image selected from music videos, were well-matched or not. Afterwards, through the use of cross-modal ranking analysis the researchers were able to achieve a 91.5% match rate, using the computer created music-image pairs with human labelers. In the experiment the subjects were asked to select between two images and determine which fit the given piece better. The subjects were able to accurately select the image from the original music video at an average percentage of 86.78% (Wu et al. 2016). This suggested that participants could, regardless of familiarity with stimuli, accurately match audio to its video counterpart.

The literature thus points to three categories that are fundamental for the study of cognitive dissonance in an audio-visual context: visual imagery, auditory stimuli, and emotional/engagement responses. We hypothesize that through the understanding of the individual variables, visual and auditory stimuli, and emotional or engagement responses to those variables, one can further investigate the variables that affect cognitive dissonance in the context of audio-visual stimuli.

### Visual Imagery

Visual Imagery proves to be a leading variable within the study of cognitive dissonance and music. As shown in the Chetan and Iancu study (2020), visual stimuli can increase musical interest, as well as influence musical perception. Visual imagery not only can stimulate musical interest, but it can influence a specific performance, as shown in the Bishop, Bailes and Dean

(2014) and the Boltz and Field (2009) studies. Researchers within this study express how both auditory feedback and musical imagery work together to make a musical experience more pleasant, proving that imagery stands to be a leading factor in cognitive perceptions (Boltz and Field, 2009). Progress in the experimentation of visual imagery continues to grow, as researchers use different techniques to experiment on this subject. In the study case of Evans (2005), the scientists expand on the idea that visual music not only stands as a fundamental of art, but that it enhances the idea of visual consonance. By conducting a series of experiments, they present data that suggests the idea of visual “rightness”, or consonance, is built (Evans, 2005). Studies also present that visual imagery and music connections are also found within video and music, as found in the Oliver and Swarbrick (2019) study. Within this experimental brief, theoretical foundations were discussed by the scientists about the integration of cinematographic theory and musicology (Oliver and Swarbrick, 2019). This article expands on the idea that the connection between visual mediums and music have strong data that further proves the notion of cognitive dissonance and consonance (Oliver and Swarbrick, 2019). Musical engagement and moving image relations are also examined in the Afra (2015) article, where researchers use Einstein’s audiovisual theory, as well as synesthesia and synthesis, to prove a relationship between image and music. This article outlines how the level of movement and musical phrasing contribute to the cohesion of image and music (Afra, 2015).

**Auditory Stimuli**

The importance of auditory stimuli in affecting cognitive dissonance cannot be understated. Music has the ability to symbolize and invoke different forms of visual imagery within the mind of the listener. For Nagari (2015), music can be seen as a Jungian symbol in that it has the capacity to trigger an imaging process, work on different psychic levels, and stir emotive dynamics within the psyche. Due to the symbolic nature of music, an experiment can be constructed to determine whether a particular piece of music’s effect is at odds with or complementary to a visual stimulus. A test by Boltz et al. (2009) looked to assess how musical stimuli affects a viewer’s perception of the accompanying visual stimulus. In doing so, the researchers created two research scenarios to determine the influence visual stimuli had on the emotional impact, interpretation, and memory of the given piece of music. The two experiments were constructed with similar methodologies but contrasted in which of the two stimuli (audial or visual) were set at odds with the other’s “ambiguous” effect. Other sections of the experiment, however, were shown without visual stimulus as a control. Many elements of music including rhythm and emotional affect can affect what is seen from a given visual stimulus. In an article by Miller et al. (2013), over the course of multiple experiments, it was found that there are modes of musical understanding that influence the processing of audiovisual stimuli: rhythmic and continuous modes. Within the experiments, it was found that there was an influence of auditory rhythm “on the temporal allocation of visual attention” (Miller et al., 2013). Through a series of dot tests, the researchers were able to support their hypothesis that what is heard can influence what is seen.

**Emotion and Engagement Responses to Stimuli**

Emotional responses and engagement to stimuli can convey a lot of information about our cognitive processes. Presicce and Bailes’ study (2019), for example, shows association between musical engagement and visual stimuli. Their research indicates a strong effect of visual stimuli on auditory stimuli, though an individual's engagement would vary by piece and musical experience (Presicce & Bailes, 2019). For Thompson et. al (2008), auditory and visual stimuli are paired and played back in ways that both meet and upend listener's expectations. The researchers present significant data that suggests visual performance and auditory cues are automatically processed together to form one emotional response and that visual aspects of performance may significantly alter musical interpretation (Thompson et. al, 2008). Baltes et. al (2014), however, introduces some important aspects of empathy and mood on an individual level. When accounting for multiple explanations of their results, Baltes et. al (2014) discusses the possibility of a higher tolerance for negative emotions and varying levels of focused attention influencing results of tension perception in music. Their findings also suggest a pattern of emotional connection by genders and age. The audience members that identified as women and those of greater age reported more emotional reactivity in the study (Baltes et. al, 2014). Bourbakis et. al (2016) explains that their findings suggest a more complicated relationship between auditory and visual stimuli. In their experiment, cognitive dissonance between positive visual stimuli and negative auditory stimuli affect participants in a less linear way than negative visual stimuli with positive auditory stimuli (Bourbakis et. al, 2016). These studies show that there are many contributing factors to the ways in which we perceive and process various stimuli and it is becoming increasingly clear that emotion and engagement play a significant role in understanding these processes.

Through exploration of past research, we wanted to further understand the factors that contribute to cognitive dissonance by presenting familiar and unfamiliar audiovisual material and using auditory modifications. We hypothesized that for unfamiliar clips participants would give a more dissonant rating to audio modifications. Previous research outlined above suggested that participants could, regardless of familiarity with stimuli, accurately match audio to its video counterpart. So we also expected to see that participants would largely be able to match original audio to its video clip. Also we hypothesized that a difference in rhythm (of the video and audio) and affect could lower the matched rating between video and modified audio.

## Methods:

**Participants**

There was a mix of both musicians and non-musicians that participated in this experiment. There were 9 female participants and 7 male participants with an average age of 30 years and a standard deviation of 10.702. The participants that were musically trained had a range of musical history, as well as length of training: classically trained, trained through jazz, or by ear/self-trained with an average age of 7 for the beginning of their instruction, with a standard deviation of 0.205. The participants were tested on both musical reward and attention during the experiment with the BMRQ test, which resulted in a mean of 88.509 and a standard deviation of 7.213 and the Gold-MSI resulted in a mean of 0.790 with a standard deviation of 0.138. Based on these tests, all participants have received an education at the university level with a degree and are continuing so through NYU. Though the experiment is completely voluntary, some of the participants were required to complete our experiment as fulfillment of a grade in an NYU class entitled the Psychology of Music.

**Materials**

The stimuli for the experiment consisted of four separate video clips from two well-known movies and two lesser-known movies. For each video there were four possible iterations in which the audio was altered. One iteration was the original audio which for our purposes we considered to be well-matched with the video. Each video clip was also shown with three audio modifications which were meant not to match with the video but have distinct characteristics or moods. The modifications fit into three separate categories; comedic, dark, and popular.

The first well-known video clip was Hyacinth Hippo and her servants from Fantasia which includes the original audio "Danza delle ore" by Amilcare Ponchielli, the comedic modification "Come Di" by Paolo Conte, the popular modification "Dead on Time" by Queen, and the dark modification "Track A-Solo Dancer" by Charles Mingus. The second well-known video clip was from La La Land which includes the original audio "Epilogue" from the original score by Justin Hurwitz, the comedic modification "Barbie Girl" by Aqua, the popular modification "Them Changes" composed by Thudercat and performed by Ariana Grande, and the dark modification "Rated X" by Miles Davis.

The first lesser-known video clip was from Pride and Prejudice which includes the original audio "Dawn" from the original score by Dario Marianelli, the comedic modification "Yakety Sax" by Boot Randolph and James Rich, the popular modification "Forever and For Always" by Shania Twain, and the dark modification "Scarbo" by Maurice Ravel. The second lesser-known video clip, and final clip overall, was from Kiki's Delivery Service which includes the original audio "Embraced in Softness" by Arai Yumi, the comedic modification "Rossini's Cat Duet" by Kiri Te Kanawa and Norma Burrows, the popular modification "I'll Be Seeing You"written by Sammy Fain and performed by Billie Holiday, and the dark modification "The Banshee" by Henry Cowell.

**Procedure**

Participants received a link to the experiment and were instructed to complete it at their own time. Because this experiment was completely remote, participants were able to use their own equipment and were instructed that only a computer is necessary. They were told to complete the experiment with speakers or headphones if they wish, but it was not required. Only a computer, browser, and connection to the internet was necessary.

Participants were asked to complete a questionnaire in which they completed both the Goldsmiths Musical Sophistication Index and the Barcelona Musical Reward Questionnaire to measure both musical proficiency and musical reward. Participants were presented with either thirty second or one minute video clips with randomized, accompanying audio. They were able to watch once and repeat if necessary. During each of the video/audio clips and in real time, participants used a sliding scale to rate how well the video and audio match in order to assess that the participants were on task. Later found in the results via the “dprime” calculations, we were able to see if the participants were completing the task in a timely manner. After each set of clips, participants were asked to answer a short questionnaire (See Appendix A).

**Analysis**

For data analysis, we used the open-source software JASP. We averaged participant’s responses over each category (familiar original, familiar dark, familiar comedy, familiar pop, and the same for unfamiliar). We then submitted our data to a 2 (Familiarity: Familiar/Unfamiliar) x 4 (Modification: Original, Dark, Comedy, Pop) Repeated Measures ANOVA. Post-hoc t-tests with hold correction were planned for any significant main effects or interactions. We also analyzed the qualitative responses to find similarities in the participants’ reasoning behind rating certain clips as more or less well-matched.

Results:

**Quantitative**

In the 2x4 Repeated Measure ANOVA, participants rated the original audio as significantly more well-matched than the modifications regardless of their familiarity level. Indeed, there was a significant effect of Modification within-subjects effects that the original audio (for both familiar and unfamiliar clips) was rated as significantly more matched than the dark, comedic, and popular modifications (See Figures 1 and 2). This reinforces our original assertion that the modifications would cause cognitive dissonance in the participants. However, there was not an effect of Familiarity on participants' ranking of the original audio. Another interesting finding is that for the dark and Comedic modifications familiarity seems to decrease the level of cognitive dissonance significantly. This is reflected by a significant Familiarity x Modification interaction and by post-hoc tests showing significant differences when comparing the Familiar, Dark and Unfamiliar, Dark modifications and the Familiar, Comedic and Unfamiliar, Comedic modifications (See Figures 2 + 3). This runs counter to a portion of our original hypothesis that familiarity would cause the participants to rate the modifications as less well-matched.

# **Matched**

Within Subjects Effects

| Cases | Sum of Squares | df | Mean Square | F | p | $\eta^2$ |
|---|---|---|---|---|---|---|
| Familiarity | 573.758 | 1 | 573.758 | 4.601 | 0.049 | 0.009 |
| Residuals | 1870.430 | 15 | 124.695 | | | |
| Modification | 44984.031 | 3 | 14994.677 | 69.353 | < .001 | 0.670 |
| Residuals | 9729.406 | 45 | 216.209 | | | |
| Familiarity ∗ Modification | 4168.211[a] | 3[a] | 1389.404[a] | 10.819[a] | < .001[a] | 0.062 |
| Residuals | 5778.852 | 45 | 128.419 | | | |

*Note*. Type III Sum of Squares

[a] Mauchly's test of sphericity indicates that the assumption of sphericity is violated ($p < .05$).

*Figure 1: Within subjects effects from repeated measure ANOVA showing that the modifications had a significant impact on the matched rating*

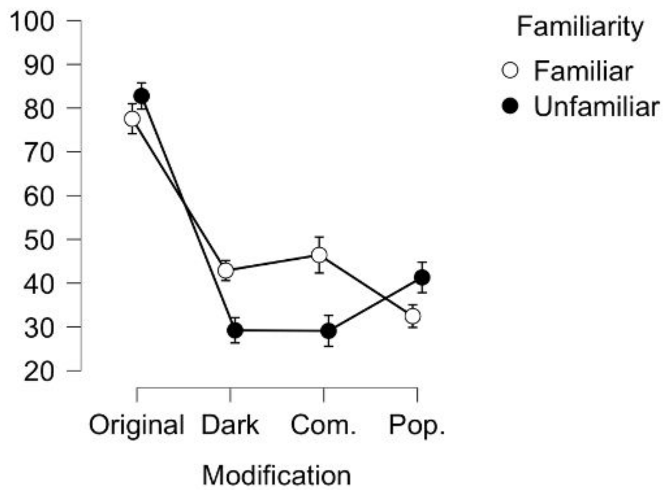


*Figure 2: Descriptives plot from repeated measure ANOVA comparing the matched rating (y-axis) with the modifications (x-axis) between the familiar and unfamiliar video clips*

**Post Hoc Comparisons - Familiarity ✻ Modification**

| | | Mean Difference | SE | t | $p_{holm}$ |
|---|---|---|---|---|---|
| Familiar, Dark | Unfamiliar, Dark | 13.656 | 3.992 | 3.421 | 0.016 |
| | Familiar, Com. | -3.563 | 4.641 | -0.768 | 1.000 |
| | Unfamiliar, Com. | 13.781 | 4.628 | 2.977 | 0.045 |
| | Familiar, Pop. | 10.406 | 4.641 | 2.242 | 0.248 |
| | Unfamiliar, Pop. | 1.562 | 4.628 | 0.338 | 1.000 |
| Familiar, Com. | Unfamiliar, Com. | 17.344 | 3.992 | 4.345 | < .001 |
| | Familiar, Pop. | 13.969 | 4.641 | 3.010 | 0.045 |
| | Unfamiliar, Pop. | 5.125 | 4.628 | 1.107 | 1.000 |

*Note.* P-value adjusted for comparing a family of 28

*Figure 3: Post-hoc comparisons illustrating a significant P-holm values when comparing familiar and unfamiliar dark and comedic modifications as it relates to how well-matched the participants rated the stimuli*

The experiment also cultivated an interesting finding in the relationship between the effects of cognitive dissonance and pleasure. Figure 4 shows that in the 2 x 4 Repeated Measures ANOVA for Pleasure, the modifications had a significant effect (P-value <.001) on the subject's ability to feel pleasure from the audiovisual stimulus. This is underscored by comparing it to the effect of familiarity on pleasure, which is not significant (P-value of .952). By evaluating the descriptives plot from the Repeated Measure ANOVA (See Figure 5) and the post-hoc for pleasure (See Figure 6), it can be deduced that pleasure is rated significantly higher for the original audio than for all of the modifications. This implies that better-matched stimuli were more pleasurable. This finding is also reflected in Figure 8 and 9 by noting the significant P-value of .004 when comparing the level of pleasure with the matched rating using Pearson's correlation. The scatter plot in Figure 8 further illuminates this by showing that as cognitive dissonance was decreased or the matched rating increased, the level of pleasure increased.

## Pleasure

**Within Subjects Effects**

| Cases | Sum of Squares | df | Mean Square | F | p | η² |
|---|---|---|---|---|---|---|
| Familiarity | 1.320 | 1 | 1.320 | 0.004 | 0.952 | 4.190e -5 |
| Residuals | 5385.805 | 15 | 359.054 | | | |
| Modification | 12631.141 | 3 | 4210.380 | 22.373 | < .001 | 0.401 |
| Residuals | 8468.734 | 45 | 188.194 | | | |
| Familiarity ∗ Modification | 713.570 | 3 | 237.857 | 2.483 | 0.073 | 0.023 |
| Residuals | 4310.305 | 45 | 95.785 | | | |

*Note.* Type III Sum of Squares

[a] Mauchly's test of sphericity indicates that the assumption of sphericity is violated (p < .05).

*Figure 4: Within subjects effects from repeated measure ANOVA showing that the modifications had a significant impact on the pleasure rating*

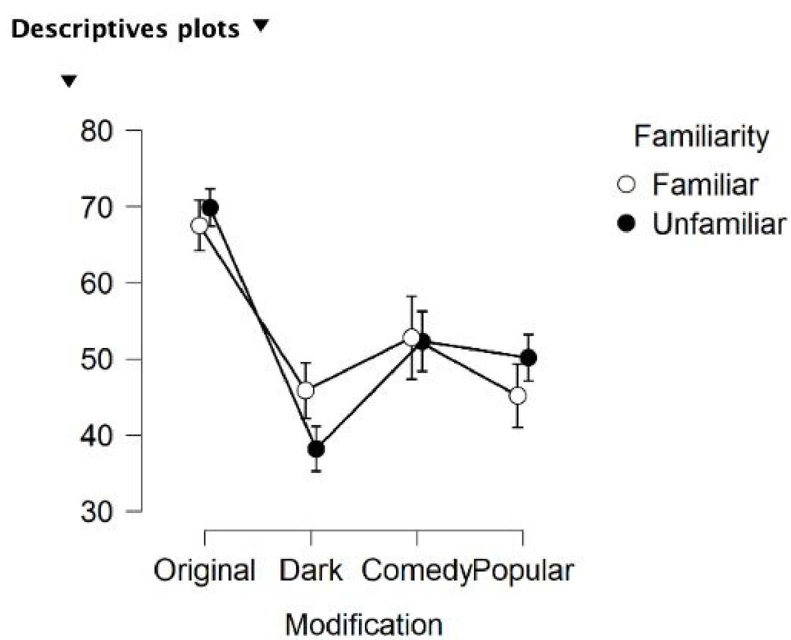


*Figure 5: Descriptives plot from repeated measure ANOVA comparing the pleasure rating (y-axis) with the modifications (x-axis) between the familiar and unfamiliar video clips*

### Post Hoc Comparisons - Modification

| | | Mean Difference | SE | t | $p_{holm}$ |
|---|---|---|---|---|---|
| Original | Dark | 26.656 | 3.430 | 7.772 | < .001 |
| | Comedy | 16.141 | 3.430 | 4.706 | < .001 |
| | Popular | 21.016 | 3.430 | 6.128 | < .001 |
| Dark | Comedy | -10.516 | 3.430 | -3.066 | 0.011 |
| | Popular | -5.641 | 3.430 | -1.645 | 0.214 |
| Comedy | Popular | 4.875 | 3.430 | 1.421 | 0.214 |

*Note.* P-value adjusted for comparing a family of 6

*Note.* Results are averaged over the levels of: Familiarity

*Figure 6: Post-hoc comparisons illustrating a significant P-holm values when comparing familiar and unfamiliar dark and comedic modifications as it relates to how pleasurable the participants rated the stimuli*

### Post Hoc Comparisons - Familiarity

| | | Mean Difference | SE | t | $p_{holm}$ |
|---|---|---|---|---|---|
| Familiar | Unfamiliar | 0.203 | 3.350 | 0.061 | 0.952 |

*Note.* Results are averaged over the levels of: Modification

*Figure 7: Post-hoc comparisons showing that there is no significant effect of familiarity on pleasure*

## Correlations

### Pearson's Correlations

| | | | Pearson's r | p |
|---|---|---|---|---|
| FamOriginalPleasure | - | FamOriginalMatch | 0.679 ** | 0.004 |

* p < .05, ** p < .01, *** p < .001

*Figure 8: Pearson's correlation showing there is a significant correlation between the matched rating and the level of pleasure felt*

**Scatter plot – Familiar Original Pleasure vs. Familiar Original Match**

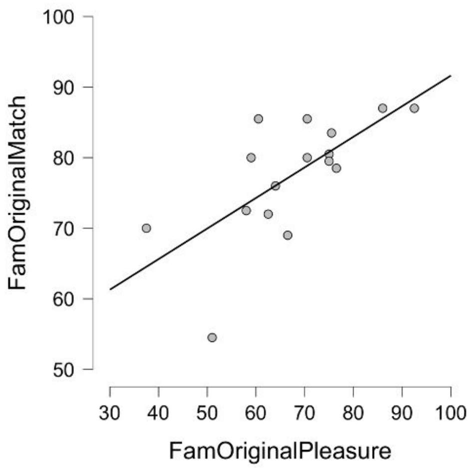


*Figure 9: Scatter plot depicting the correction between the matched rating and level of pleasure felt*

**Qualitative**

By evaluating the qualitative data, a better understanding of why this experiment produced these findings can be attained. As per our original hypothesis, many of the participants pointed to the difference in rhythm between the video and audio. A direct quote from participant 2 on the Hyacinth Hippos (familiar) with the Dark modification ("Track A- Solo Dancer") shows this most clearly, saying "The rhythm of the other hippos walking made it match less." This is not an atypical response although different participants used different verbiage to describe the difference in rhythm saying that the music felt "off-beat" or "not lined-up" with the video. Participants often connected the rhythm of the music with the "movements" in the video. Also per our original hypothesis, many participants noted a disconnect between the affect of the video and the affect of the audio as a prime reason for the cognitive dissonance created by the audiovisual stimulus. An example from participant 10 about the Kiki's Delivery Service (unfamiliar) clip with the Dark modification ("The Banshee") best exemplifies this, saying "The dark, unsettling music seems too scary for the video." One common response that we did not foresee was that the music did not seem to fit the period in which the video was set. This anachronistic nature of some of the modifications is best exemplified by participant 1's response to the La La Land (familiar) clip with the Dark modification ("Rated X"), saying "The music was from the wrong era." Again, in evaluating qualitative data it was important to connect differences in the language used to describe the same phenomena. Other ways in which the anachronism was described was as "vintage vs. contemporary" or "from a different period." These responses were often inverted for the stimuli in which the participant considered more well-matched (often the original audio). This is reflected in participant 12's response to Kiki's Delivery Service (unfamiliar) with the original audio ("Embraced in Softness"), saying "The cat movement and moon scene totally matched the vibe and rhythm of the music."

## Discussion:

In this experiment, we sought out to understand the factors that contribute to cognitive dissonance by presenting familiar and unfamiliar audiovisual material and using auditory modifications. During research from previous experiments, we found that similar data was presented in which researchers sought to understand relative cognition through video/auditory stimuli. We hypothesized that participants would be more likely to give a lower rating for familiar clips but found the opposite was true in the data: participants were more likely to rate only certain modifications higher if they were familiar with the clip.

Our data is in line with Wu and colleagues 2016. In their experiment, it was found that people were adept at picking out the original audio to its video and that subjects were able to accurately select the image from the original music video at an average percentage of 86.78% (Wu 2016). This aligns with our findings of our own experiment in the notion that participants have relative cognitive understanding to associate the original audio with the original clip based on each participant. Similar data was found in the Guttman et. al (2005) article in that we use our senses methodically, defaulting to vision for spatial awareness and audition for temporal awareness. In this experiment, the researchers asked whether we can use the audition/temporal processes for visual changes and the visual/spatial processes for auditory changes (Guttman et. al, 2005). Their experiment strived to focus on how visual and auditory processes influence each other. They ultimately found that there is a deeply ingrained process by which we experience interacting visual and auditory information and that one may influence the other, even if that information is contradictory to expectation (Guttman 2005). Through these previous experiments, we can see that not only has data been found similar to our own, but that the question of visual and auditory perception is a subject that has been questioned over time.

Ultimately the data shows the opposite of a portion of our original hypothesis is true: we believed that participants were more likely to rate lower if they were familiar with the clip but the opposite was true; participants were more likely to rate the dark and comedic modifications higher if they were familiar with the clip. This may be caused by a sort of "familiarity spillover" into the modifications in which the listeners familiarity with the video clips causes them to connect different modifications even though they know it does not connect as much as the original audio. The nuance of this finding is that even though the familiar dark and comedic modifications were rated higher than their unfamiliar counterparts, they still were rated less than .5 in the matched rating which is still significantly lower than the original audio (reflected in the *P-holm* values in Figure 2). This subversion of our original assumption has potential to illuminate future experimentation. An interesting note is that this is not the case for the popular modification, this may be because the participant is aware of both to the point that it becomes obvious that the two do not fit together. The data also shows that the difference in video's affect caused lower ratings in how well the video and audio 'matched.' However, familiarity and unfamiliarity ratings were not distinct enough to categorize them as familiar and unfamiliar. This could have affected our results in any number of ways. For example, the disproving of our hypothesis could have been affected by the categorization of familiar and unfamiliar clips.

Therefore, in future experiments it would be useful to reconsider the differences between familiar and unfamiliar video clips. Because participants had varying levels of familiarity with each set of clips, individualizing the familiar and unfamiliar categories would be useful in designing related experiments in the future. $These results may be useful for those in the film, advertisement, or other related industries, in which audiovisual materials and dissonance between them could affect the viewer's perception of the clip.

Bibliography:

Afra, K. (2015). “Vertical Montage” and Synaesthesia. *Music, Sound & the Moving Image*, Vol. 9, No. 1, 33–61, doi: https://doi-org.proxy.library.nyu.edu/10.3828/msmi.2015.2

Baltes, Felicia Rodica., & Miu, Andrei C. (2014). Emotions During Live Music Performance: Links With Individual Differences in Empathy, VIsual Imagery, and Mood. *Psychomusicology: Music, Mind, and Brain*, Vol. 24 No. 1, pp. 58-65. doi: 10.1037/pmu0000030

Bishop, Laura., Bailes, Freya., and Dean, Roger T. (2014). Performing Musical Dynamics: How Crucial are Musical Imagery and Auditory Feedback for Expert and Novice Musicians? *Music Perception: An Interdisciplinary Journal*, Vol. 32, No. 1, pp. 51-66. doi: https://www.jstor.org/stable/10.1525/mp.2014.32.1.51?seq=1

Boltz, M., Ebendorf, B., & Field, B. (2009). Audiovisual Interactions: The Impact of Visual Information on Music Perception and Memory. *Music Perception: An Interdisciplinary Journal,* Vol. *27*, No. 1, 43-59. doi:10.1525/mp.2009.27.1.43

Bourbakis, N., Esposito, A., Esposito, A. M., Esposito M., Esposito, M. T., & Vinciarelli, A. (2016). Effects of Emotional Visual Scenes on the Ability to Decode Emotional Melodies. *IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI)*. doi: 10.1109/ICTAI.2016.0124

Bravo, F. (2013). [In Aramaki, M., Barthet, M., & Ystad, S. (Eds.)], From sounds to music and emotions*. 9th International Symposium CMMR,* pp 367-376. doi: https://ebookcentral-proquest-com.proxy.library.nyu.edu/lib/nyulibrary-ebooks/detail.action?docID=3093056

Chetan, A., & Iancu, I. (2020). Influencing Music Perception: A Systematic Theoretical Review on Visual Stimuli Applied in Branding and Communication. *Styles of Communication*, Vol. 12, No. 1, pp. 104-120. doi: https://doaj-org.proxy.library.nyu.edu/article/bee485087b3b4409b89d7395a67d3dd0

Evans, Brian (2005). Foundations of Visual Music. *Computer Music Journal*, Vol. 29, No. 4, pp. 11-24. doi: https://www.jstor.org/stable/3681478?seq=1

Festinger, Leon (1962). Cognitive Dissonance. *Scientific American*, Vol. 207, No. 4, pp. 93-106. doi: https://www.jstor.org/stable/24936719?seq=1

Guttman, Sharon E., Gilroy, Lee A., and Blake, Randolph (2005). Hearing What the Eyes See: Auditory Encoding of Visual Temporal Sequences. *Psychological Science*, Vol. 16, No. 3, pp. 228-235. doi: https://www.jstor.org/stable/40064206?seq=1

Mas-Herrero, E., Marco-Pallares, J., Lorenzo-Seva, U., Zatorre, R. J., & Rodriguez-Fornells, A. (2013). Barcelona Music Reward Questionnaire (BMRQ) [Database record]. *APA PsycTests*. Doi: https://doi.org/10.1037/t31533-000

Miller, Jared E., Carlson, Laura A., and McAuley, J. Devin (2013). When What You Hear Influences When You See: Listening to an Auditory Rhythm Influences the Temporal Allocation of VIsual Attention. *Psychological Science*, Vol. 24, No. 1, pp 11-18. doi: https://www.jstor.org/stable/23408585?seq=1

Müllensiefen, Gingras, Musil, & Stewart, (2014) Goldsmiths Musical Sophistication Index (GOLD-MSI) [Database record]. *APA PsycTests*, doi: https://www.gold.ac.uk/music-mind-brain/gold-msi/

Nagari, B. (2015). *Music as image : Analytical psychology and music in film*. ProQuest Ebook Central, doi: https://ebookcentral-proquest-com.proxy.library.nyu.edu

Oliver, Josep Torelló and Swarbrick, Josephine (2019). The Musicalized Image: A Joint Aesthetic of Music and Image in Film. *Aisthesis,* Vol. 12, No. 2. https://doiorg.proxy.library.nyu.edu/10.13128/Aisthesis10733

Presicce, G., & Bailes, F. (2019). Engagement and Visual Imagery in Music Listening: An Exploratory Study. *Psychomusicology: Music, Mind, and Brain*, Vol. 29, No. 2-3, pp. 136-155. doi: 10.1037/pmu0000243

Thompson, W. F.., Russo, F. A., & Quinto, L (2008). Audio-visual Integration of Emotional Cues in Song. *Cognition and Emotion*, Vol. 22, No. 8, pp. 1457-1470. doi: 10.1080/02699930701813974

Wu, X., Qiao, Y., Wang, X., and Tang, X. (2016). "Bridging Music and Image via Cross-Modal Ranking Analysis” *IEEE Transactions on Multimedia*, Vol. 18, No. 7, pp. 1305-1318. doi: 10.1109/TMM.2016.2557722.