Can The Moving Image Influence the Experience of Music?

An Experimental Investigation of the Impact of The Moving Image on Musical Emotions

Elias Gram-Nilsen

Master's Thesis in Musicology

Department of Musicology

University of Oslo

Spring 2021

Abstract

The combination of music and the moving image is prevalent in our society, occurring in visual art forms and medias like movies, music videos and commercials. The relationship between the two has received much interest in existing research, but the investigation has mainly been one sided. While the influence of music on the moving image is well documented, the reverse relationship has not been given much attention. The present study is devoted to the latter approach and examines how a visual context and extramusical information from the moving image may influence the experience of music.

Sixty-six participants were presented with three music excerpts ("ambient track", "string quartet", and a "pop tune") that were either presented alone or paired with one of two videos expected to elicit strong emotions ("surfing video", expected to evoke awe; "Christian the lion video", expected to evoke feelings of being moved). The participants were randomly assigned to one of three experimental conditions, where the different music-video (or no video) combinations were balanced across conditions. After a short filler task, the three music excerpts were presented again (without any visual accompaniment) to see if a possible influence of previously presented visual information was persistent. Participants were requested to provide ratings of enjoyment, congruency between music and visuals, the occurrence of physiological responses, the experience of awe and being moved, - and perceived valence and arousal.

During the first presentation, the moving image (in particular the "surfing video") had a significant positive effect on the experience of awe when compared to music alone. This was true for the "ambient track" and the "pop tune". The "Christian the lion video" also had a significant positive effect on being moved when paired with the "pop tune". When the "string quartet" was paired with "Christian the lion video", there was a significant negative effect on both enjoyment and valence (compared to music alone) and this pairing was also perceived as the least congruent by the participants. This negative effect was sustained and present also during the second presentation of the "string quartet", being the only significant carry-over effect.

The current study suggests that the moving image can have an intensifying effect on emotions induced while listening to music, especially when the music and visuals are congruent. Interestingly a mismatch between visuals and music seemed to have the most persistent effect, observable also during the second presentation of music alone.

Acknowledgments

Firstly, I would like to give special thanks to my supervisors Jonna Vuoskoski and Laura Bishop for providing me with indispensable support while working on this thesis. Thank you Jonna for helping me with the design of the study, finding relevant literature, recruiting participants, and also for showing great patience when teaching me how to analyze and interpret the results. Thank you Laura for giving me valuable feedback in the writing process and many great suggestions in relation to the structure of the thesis. The writing has really improved with your help. I would also like to thank my fellow students for many great discussions, both concerning my thesis and other music related subjects. Mika deserves a special thanks for helping me with recruiting participants from Denmark, and I would also like to thank Hanne and Sander for testing the internet survey for me before it was published. Last (but not least), thank you to friends and family for showing great moral support and helping me with sharing the internet survey. My mother, Ninna, needs an extra thanks for proofreading in the final stage of the writing process.

Elias Gram-Nilsen

Oslo, April 2021

Table of Contents

1 Introduction

The combination of music and visuals has in the course of history been prevalent. Whereas music listening in these days can be liberated from specific situations and activities, the situation was quite different before the development of sound recording. The necessity of performance in real time made music not only an auditory but also a visual act. The visual accompaniment was not limited to a simple performance of the music but could be of a more sophisticated nature. «Music is rarely alone. Throughout civilization, it has typically been performed in conjunction with dance, poetic text, theatre, and, more recently, film and television» (Cohen, 2008: 441).

This tendency to combine music with other art forms is particularly prevalent in the tradition of motion pictures. Used originally to drown the noise from the film projector, music continued to be a fundamental part of the film experience even though the noisy film projector quickly disappeared (Cohen, 2001: 250). How these two artforms connect with each other has been a subject of interest in the scientific community. With a foundation in human perception this relationship has been given much attention in the field of psychology. The perspective has been rather one-sided with the majority of the existing literature looking at how music might influence the perception of film. Several studies have investigated how music might account for the interpretation of and emotional reactions to film (Cohen, 2001; Boltz, 2013). The opposite claim, how film might influence music, has on the other hand not been given much attention (Boltz, 2013: 217-18).

Surroundings and the context of music listening is undoubtedly an integral part of the musical experience. In the book *Music Emotions Explained*, Patrik Juslin describes the context of the music as determining for causing certain emotional responses (Juslin, 2019: 384-85). In the course of history there exist many examples of how extramusical information has been applied to enhance the musical experience. In the 19th century composers of program music attempted to compose music that depicted an extramusical narrative (Lavy, 2001: 80-81) and the audience was often given program notes that contained a written down narrative. The Russian composer Alexander Scriabin is known for applying multi-sensory elements to his music. His symphonic work *Prometheus* from 1910 was written for the color organ (Cohen et al., 2013: 3), and the unfinished and highly ambitious work *Mysterium* was intended to be performed in the foothills of the Himalayas and to include all of the arts and

trigger all of the senses, even smell and touch (Garcia, 2005: 9). The example that probably demonstrates to the fullest the impact extramusical information might have is a composition by the Polish composer Krzysztof Penderecki from 1960. Originally given the name 8´37, referring to the length of it, Penderecki decided to change the name of the piece to *Threnody to the Victims of Hiroshima* after hearing a performance of it. With this political title the composition gained much attention with the dissonance in the strings being compared to screams of pain (Guldbrandsen, 2012: 298). In addition to much recognition Penderecki also won the UNESCO award for the composition. All these cases support the notion of a reverse influence from film on music being possible.

Despite not being a topic of much interest, some research has looked into the influence of visual mediums on the experience of music. The popularity of music videos demonstrates how visual information has been used to enhance the popularity of the music. Although few studies have investigated how visual aspects of music videos might influence the experience of the music, Goldberg and colleagues found that music videos could extend the number of times participants could listen to songs without growing tired of them in a study conducted in 1993 (Goldberg, 1993). Music videos that were ambiguous and open for multiple interpretations were most effective in postponing so called «wear outs» (Boltz et al., 2009: 44) and these findings demonstrate how videos might have an influence on our interaction with music. Two studies from 1996 and 1997 looked more specifically at the affective response to music when combined with visuals. Participants in these studies were presented with music extracts either alone or paired with excerpts from the film *Fantasia.* The combination of video and music led to some increase in involvement, emotional reactions as well as preference, but the effect was not large (Geringer et al., 1996; Geringer et al., 1997). Research has also found musical parameters like tempo, rhythm, amplitude to be perceived differently when music is accompanied by videos. Boltz, Ebendorf, and Field found music in combination with visual information to be perceived as faster, louder and more rhythmical than music played alone. The affective quality of the music was also influenced by the video accompaniment in a mood congruent manner (Boltz et al., 2009).

Since visual information is a fundamental accompaniment to music listening, either arbitrary in the form of our surroundings or more intentional with a live performance or video/film accompaniment, the influence it may have on music is an important topic to study. Due to the amount of information that is present in film/video, this medium is a useful stimulus for investigating the impact of both visual information and the context surrounding the music. The studies mentioned all look into this but are all lacking in certain central

features. Particularly the influence on emotions is an area that should be investigated more systematically. The studies conducted by Geringer and colleagues looked into affective responses to music through ratings of emotions experienced as well as degree of liking and involvement, but the characteristics of emotions induced was not studied, and the occurrence of physical responses like chills were not measured. In the paper from 2009, the focus was primarily on how visuals could alter the perception of music. Only perceived emotions (affect, and activity) were investigated and the relationship between visual accompaniment and induced emotions was not looked into.

The current study offers more comprehensive look at emotional reactions to music, covering enjoyment, specific emotions (awe and being moved) that may be experienced, physiological responses, as well as emotions perceived, including both valence and arousal. One of the aims is to find out if emotions that are expected to be perceived and induced by selected videos will carry over to the music, when played in combination, and thus enhance the experience of it. Whether or not this influence is persistent and can change how music is experienced when played alone subsequent to a first exposure, will also be examined.

2 Theory

2.1 Terminology: The Moving Image

Film, motion picture, moving picture and *moving image* are different terms used to describe the visual simulation of experiences. Strictly speaking, this medium is the result of still images played in a rapid succession that captures real life, staged, or even animated experiences. Which of the terms then is most suitable for discussing this phenomenon?

Film is probably the most common and the term used most frequently both in and outside of academic literature. One disadvantage with the term is that it is often associated with cinematography and full-length feature films and tends to exclude other formats like for instance video. In the book *Theorizing the Moving Image* Nöel Carrol deliberately substitutes the term film with moving image due to its limited scope.

> I prefer the idiom moving images rather than film because I predict that what we call film and, for that matter, film history will, in generations to come, be seen as a larger continuous history that will not be restricted to things made only in the so-called medium of film but, as well, will apply to things made in the media of video, TV, computer-generated imagery, and we know not what (Carrol, 1996: xiii).

The terms *motion picture, moving picture,* and *moving image* are on the other hand wider and encompassing in their meaning, simply describing the process of certain types of visual mediums. What separates these terms from each other in relation to their meaning is a question difficult to answer. *Moving image* is chosen by Carrol because he regards the term *image* as more «nonrepresentational and abstract» than *picture,* which only implies «recognizable representations» (Carrol, 1996: xiii). Whereas the term picture is often used to describe concrete objects like paintings and photography's, the term image tends to also be associated with abstract phenomena like mental images.

Despite being present in the introduction, the term film will not be used further in this thesis. This is because it is believed to be too narrow in its meaning and not concerned with medias like video, which is the stimulus relevant for the current study. Another disadvantage with the term film is that it is not only concerned with visual information, but also audio of dietetic sounds like speech. Terms such as the moving image are, on the other hand, only concerned with visual information from a film or a video, as the name implies. The use of the

term film in relation to music may also signal that the tradition of film music is the object of study. This is not the case, and this master thesis is more concerned with music in general and how it might be influenced by visual information. Choosing the term moving image over other terms like moving picture and motion picture, is mostly an intuitive choice, and the other terms just highlighted are believed too also be suitable. That being said, the more abstract nature of the term image may be a bit more relevant for the current thesis, due to a focus on the perception and experience, rather than the objective nature of visual information. Although the term film is most applied in the existing literature, moving image has, for instance, been applied by Annabel Cohen in a chapter in *The Oxford Handbook of Music Psychology* (Cohen, 2008), further supporting the relevance of it.

2.2 Music and Emotions

One of the reasons why music is prevalent all over the world may be its ability to evoke emotions in listeners «wherever there are human beings, there is *music*; and wherever there is music, there is *emotion*» (Juslin, 2019: 3). It has from an evolutionary perspective been suggested that the emotional quality of music may be grounded in melodic and emotional aspects of vocal expression, and music may in that regard evoke emotions through the imitation of the human voice (Juslin and Laukka, 2003). Expression of emotions is in spite of this not necessarily a primary aim in all musical activities, and in some cultures the link between music and emotions seems to be quite weak. The Mafa people in Cameroon for instance do not seem to systematically connect music with emotions, but rather with certain rituals. They have nevertheless been found to be able to recognize emotions conveyed in western music, suggesting that music may have a universal emotional aspect (Fritz et al., 2009, supplementary data).

In the field of psychology, it is in the literature on music and emotion common to distinguish between what is known as *perceived emotion* and *felt emotion.* Whereas the former is related to emotions expressed and recognized in the music, the latter relates to music's ability to induce emotions and physical responses in a listener (Gabrielsson, 2001). One might for instance perceive a song as happy without experiencing any changes in the emotional state, and in more extreme cases the perceived valence may be conflicting with felt emotions. Listening to a song characterized as happy can for instance trigger negative

emotions if the song is linked to sad life events, like being the favorite song of a deceased family member. Although there exists a distinction between felt and perceived emotions, they are quite often interacting with each other and the line is to some extent believed to be blurred (Eerola and Vuoskoski, 2011:19).

In the understanding of music's ability to evoke feelings in listeners, the BRECVEMA model created by Patrik Juslin and Daniel Västfjäll is a central framework worth mentioning. This model, which is grounded in an evolutionary perspective, gives explanations for the underlying mechanisms behind our emotional responses to music. The model includes the following mechanisms: *brain stem reflex, rhythmic entrainment, evaluative conditioning, contagion, visual imagery, episodic memory, musical expectancy,* and *aesthetic judgments* (the name of the model is an acronym for each mechanism). Many of these mechanisms are influenced by cultural upbringing and personal experiences (evaluative conditioning, episodic memory), while some are more independent and similar across cultures (brain stem reflex, contagion) (Juslin and Västfjäll, 2008).

Just as the BRECVEMA model is used to comprehend the mechanisms behind emotional responses, there also exist models for mapping and understanding different music specific emotions. A two-dimensional model containing *valence,* which is related to positivity and negativity, and *arousal* (sometimes labeled activity), which is more centered on level of energy, is commonly used in the organization of music related emotions (Eerola and Vuoskoski, 2011: 20). This model treats a specific emotion as a result of certain combinations of the two dimensions. Happiness contains both a high valence (positivity) and a high level of arousal, while sadness contains a low level of both dimensions. Fear and aggression, on the other hand, are the result of a high level of arousal but a low level of valence. Compared to an alternative organizing of emotions into distinct categories, nuances are better captured with this model. It is nevertheless worth noting that the two-dimensional model, although operating with scales, still utilizes certain categories such as positivity and negativity in the valence dimension (Juslin, 2019: 51).

The emotions mentioned above (anger, fear happiness and sadness) are some examples of basic emotions expected to be universally perceived across culture (Ekman, 1992; Ekman, 1999). While these emotions seem to be pretty stable in terms of both valence and energy, there exist other emotions that are more complex and conflicting in nature. Some noteworthy examples of more complex emotions, relevant for the present research, are *being moved* (also known as *Kama muta*) and *awe* (these emotions are sometimes also referred to as *aesthetic emotions*).

The former emotions *being moved* is characterized as a deep and sincere emotion which contains both elements of sadness and joy. The emotion is believed to be highly social in its nature and, according to Menninghaus and colleagues, serves the purpose of promoting social bonds (Menninghaus et al., 2015: 26). Events that are particularly predicted to trigger this feeling are crucial life events like birth and death, and relationship events like marriage and reunion. Being moved has also a tendency to occur in situations where the observer has little impact on events or the cause of them, «those who are moved by the event do not, as a rule, cause it but rather only witness it, and they neither need nor wish to change it» (Menninghaus et al. 2015: 5/8-10).

The latter emotion *awe* is hard to give an adequate description of but similar to being moved, this emotion contains both elements of positivity, in the form of wonder and admiration, in addition to negativity represented by fear. The two core ingredients in this emotion are believed to be *vastness*, as well as *accommodation*. The former is related to «anything that is experienced as being much larger than the self» (Keltner and Haidt, 2003: 303) and can be connected to both physical objects of great size like mountains and hurricanes, but also mental concepts such as fame and prestige. The latter ingredient can be defined as an adjustment of existing mental schemas in the encounter with new extraordinary information. Objects of great physical or mental size, like a large mountain or a charismatic leader are, in addition to vastness, also associated with accommodation. In addition to these main elements, *threat*, *beauty*, *ability*, *virtue*, and the *supernatural* are themes that might flavor the experience of awe (Keltner and Haidt, 2003: 303-06). Vastness may for instance be altered by threat as well as beauty, while certain skills and traits of virtue can lead to accommodation. Similar to being moved, a social component is also expected to be present in awe. In a study from 2007, participants reported experiencing a feeling of being part of something bigger minimizing the individual self (Shiota et al., 2007). A paper from 2015 found the experience of awe to have an impact on prosocial behavior (Piff et al., 2015), and one of the few studies that have looked into the relationship between awe and physiological responses found triggers to be social in 46 % of times awe was evoked (Schurtz et al., 2012: 210). From an evolutionary perspective, this emotion has also been linked with submission to a powerful leader (Keltner and Haidt, 2003: 306).

Even though there seem to be clear differences between awe and being moved, some scientists do not fully separate them from each other. Konecni, for instance, describes awe as a peak experience that is always accompanied by the state of being moved (Konecni, 2005: 31-32). This view is however not shared by many scientists. While being moved is believed to

occur in relationships of a close and egalitarian nature, awe is related to admiration and an uneven relationship in relation to authority (Menninghaus et al., 2015: 25). They will be regarded as two separate emotional states containing different components in this thesis.

The complexity and ambivalence in awe and being moved makes it harder to map them onto a two-dimensional model, and the two emotions do not seem to have a very clear connection to any side of the spectrum of the two dimensions. In studies conducted by Menninghaus and colleagues four terms related to being moved (*moving, touching, stirring, and deeply moving)* were mapped onto a valence and arousal diagram. The concepts were shown to have a « slightly positive valence and a low to mid- affective arousal score» (Menninghaus et al., 2015:17), with terms like moving and touching having a higher arousal scale than stirring and deeply moving. The ratio between positivity and negativity is nevertheless likely to vary in different eliciting situations. Birth and reunion are for instance likely to lead to significantly more positive emotions than the death of a loved one. Awe is also a challenging emotion to define in relation to valence and arousal and has been described as ambiguous and dependent on the specific situation (Shiota et al., 2007: 945). Keltner and Haidt lists both *threat* and *beauty* as some of the themes that may alter the experience of awe, demonstrating the conflicting nature of this emotion (Keltner and Haidt, 2003: 304/311). In an article from 2013, Elvira Brattico and Marcus Pearce stress the difficulty of measuring mixed emotion with dimensional models: «dimensional models are fundamentally unable to accommodate the possibility of mixed emotions combining the extremes of a single dimension, such as feeling happy and sad at the same time» (Brattico and Pearce, 2013: 50). This conflicting nature may also explain why these emotions are not able to be perceived in the music in the same manner as many basic emotions, as they are mostly related to induced emotions (Gabrielsson, 2001: 136).

Despite this ambivalence, these complex emotions may have a capacity for inducing stronger responses than other basic emotions in an art related context. Patrik Juslin suggests that the mixed nature of emotions such as being moved may come from several mechanisms being at play at the same time, and the result may be «a sense of being overwhelmed by emotion (..) which may tend to occur in peak experiences» (Juslin, 2019: 491-93). Peak experiences have also been connected to awe by the American psychologist Abraham Maslow. Wonder, reverence, humility and «fear of being overwhelmed by more than what one can bear» (Gabrielsson, 2010), central ingredients in awe, are also highlighted as salient features of peak experiences. A relationship between being moved and awe, and the occurrence of chills and goosebumps has also been highlighted in several papers (Konecni,

2005; Schurtz et al., 2012; Wassiliwizky et al., 2015; Schubert et al., 2018). One possible explanation for this may be an overlap between mechanisms in the brain responsible for chills and social processes (Panksepp, 1995: 199). As mentioned, particularly being moved but also awe seems to contain social aspects.

Other physiological responses have also been found to accompany both feelings of being moved and awe. In their paper on being moved, Menninghaus and colleagues lists tears and warmness as common responses, in addition to chills (Menninghaus et al., 2015: 26). This is further supported by a study from 2018 that found tears and warm sensations to be connected to the experience of being moved (Schubert et al., 2018). When it comes to awe, little research seems to have investigated bodily responses, but tears have in addition to goosebumps been connected to the experience of awe in a paper from 2001 (Braud, 2001). David Huron also suggests that awe is likely to be accompanied by gasps, which is assumed to come from a *freeze* response that may be triggered by surprise in certain situations (Huron, 2006: 32).

2.3 Four Theoretical Approaches to the Influence of the Moving Image on Music

Music and visual information go hand in hand in many ways. As mentioned, visuals are present in the majority of our listening experiences but the connection between the two stimuli seems to be even more fundamental. Studies have found areas in the occipital lobe, which is connected to visual processing, to be activated under exposure to harmonies in music (Satoh et al., 2001; Schmithorst, 2005). The processing of both musical emotions and visual emotions seems to be connected to many of the same brain areas (Logeswaran et al., 2009), which may explain why the pairing of music and visuals often works so well and is effective in eliciting emotions in listeners. Studies have also found the combination of music and visuals to result in increased activity in brain structures related to emotions, like the amygdala and the hippocampus. This pattern was found both in a study where sad or fearful pictures were either presented alone or combined with congruent music (Baumgartner et al., 2006), and a study in which neutral films and either joyful, neutral, or fearful music excerpts were either presented alone or paired with each other. In the last study both the music presented alone, and the films presented alone resulted in far less activity in emotional centers in the

brain, compared to the combination of the two stimuli. Not only did music enhance the experience of the films but the neutral films had a strong impact on the experience of the music, assumed to be caused to the amount of concrete information present in the films (Eldar et al., 2007). The relationship between music and visuals may, based on these results, be characterized as a symbiotic one in which each part is able to induce far less emotions than the pairing of the two.

These findings from brain studies, in addition to the empirical studies mentioned in the introduction, give substantial support for a reverse investigation, focusing on the influence the moving image can have on music. Given that the literature that addresses this theme explicitly is relatively sparse, a more in-depth investigation of this topic will demand a look into different theories that may give insight into a possible influence from the moving image. Four theories related to visual imagery, narrative processes, cross-modal interaction, and emotional contagion will be presented in the following section, giving different angles to the possible influence of the moving image on music.

2.3.1 Visual Imagery

In our perception of music, we tend to connect sounds we hear with certain visual images. A high-pitched sound is for instance often characterized as bright while one with a low pitch is often associated with darkness. These descriptions of music through metaphors are particularly prevalent in our speech about music and are intuitive rather than conscious choices. A tone gradually increasing in pitch is for instance described as rising in pitch, giving it an upward motion. This notion of up and down is one common *image schema* used to describe the location of pitch in our speech about music (Snyder, 2001: 110 -111).

It is not only in the language of music that elements of visuals manifest themselves, but visual information can also be a central part of our listening experience. *Visual imagery* is in relation to the experience of music how one might conjure up images and let the imagination fly while listening to music. Imagination and fantasy are not dependent on the presence of music, but a piece of music may trigger the creation of mental images. Just as musical features are connected to visual concepts in our language, mental images can also be caused by certain structures in the music. For example, «if the music features a melody with rising pitch, a listener might imagine that he or she is flying higher and higher in the air» (Juslin, 2019: 331).

Research on *visual imagery* as a response to music can be traced back to the beginning of the 1980s. In a study from 1980 John Osborne found mental images to be the most common response compared to thoughts, emotions and bodily sensations. Natural scenes and out-of-body experiences were among the most common images that participants produced (Juslin, 2019: 336-67). After this study, several other scientists have examined the connection between music and visual imagery. Alexandra Quittner found the use of music to be a better predictor of the occurrence of mental images than the use of relaxing techniques (relaxing condition) or no stimulus (control condition) (Quittner, 1983). In the SEM (Strong Experiences with Music) project which contains reports of peak experiences with music from almost a thousand people, around 10% of the reports included the experience of inner images (Gabrielsson, 2011: 118). In an experience sampling study from 2008 visual imagery was found to be present in 7% of all music experiences of an emotional character (Juslin et al., 2008) and Küssner and Eerola found visual imagery to have been experienced by 77% of the responders in a study from 2019. In the last study, images of landscapes, scenes from the nature, life events, and the performance of the music were among the images that were reported. Some participants even imagined themselves performing the music or being a part of the audience. (Küssner and Eerola, 2019).

The accompaniment of visual imagery to music has been suggested to be a source of emotion by several scientists investigating the phenomenon. Mental images are in general expected to be a direct cause of emotions with different forms of images resulting in different reactions. Juslin and Västfjäll have pointed to visual imagery as one of the underlying mechanisms from the BRECVEMA model that cause emotional responses to music (Juslin and Västfjäll, 2008). Mental images are believed to be able to evoke emotion due to the simulation of real events of an emotional nature. Images of important life events, like meeting your first love or losing a close relative may for instance lead to strong emotions due to the impact of the event. Although mental images like these mentioned are unlikely to lead to the exact same emotional reaction as in real life events, they are believed to be unconsciously experienced as real by us. «Mental imagery is a form of mental simulation, and it is so good that it may fool your body» (Juslin, 2019: 334). Visual imagery can lead to a wide variety of emotions and complex emotional reactions (mixed emotions) such as being moved and awe. The former (being moved) is for instance expected to be triggered by autobiographical scenes that are connected to crucial life events, while awe might be the result of mental images of vast landscapes such as mountains. In the research on strong experiences with music, the content of the visual images was often related to earlier life experiences and the mental

images also contained religious aspects in several reports (Gabrielsson, 2011: 110-18). The phenomenon of visual imagery from interactions with music may therefore be a source of strong emotional responses.

The emotional impact mental images from music may have is demonstrated through the application of imagery in a therapeutic setting. In the 70s the *Bonny Method of Guided Imagery and music* (GIM) was developed by Helen Bonny. In this therapy form the patient or "traveler" listens to a pre-programmed playlist that is expected to be of an emotional nature, offer focus to the traveler, and make visual imagery more likely to occur. While listening to music the therapist serves as a guide who helps the traveler to process emotional images that he or she may conjure up during the session (Juslin, 2019: 333). This method, which is expected to be helpful in the treatment of anxiety, stress, and depression, has gained a foothold in music therapy (Küssner and Eerola, 2019).

Despite visual imagery not being centered on how the accompaniment of moving image might influence our experience of music, it demonstrates our tendency to connect sounds we hear with visual information. Moving image may therefore influence our emotional experience of music. In the experience sampling study conducted by Patrik Juslin and colleagues 10% of emotional episodes with music occurred during the activity of watching TV or Movies (Juslin et al., 2008). The fact that we respond to mental images in the same manner as to real images, (Juslin, 2019: 333) makes is likely to assume that real visual images can have the same and maybe an even stronger impact on our emotional response to music. A study in which participants either saw drawings or were asked to visualize specific pictures found many of the same brain regions to be activated in both conditions (Ganis et al., 2004). The underlying mechanisms that elicit emotions under visual imagery is therefore expected to be present when music is paired with the moving image.

One may in spite of this come across challenges when studying the influence of real visual images and it may in many ways be more challenging than letting participants conjure up their own images. For the moving image to have a positive influence on the emotional experience of music, participants need to be emotionally moved by the visuals. Finding visual information that triggers emotions might be difficult due to it depending on the personal taste of participants. Whereas participants during the occurrence of visual imagery are in charge of the visual information and can create mental images that are earmarked for triggering emotions, that is not the case with the accompaniment of real visual images. Mental images connected to personal memories are therefore likely to have a stronger impact than real images chosen by an experimenter, even when they are of an emotional character. That being

said, the phenomenon of visual imagery illustrates how extramusical information and visual images may have a strong impact on our experience of music. In that matter, it supports the theory of the moving image being able to influence the experience of music.

2.3.2 Narrative Comprehension

The experience of the moving image is not only a visual but often a narrative one as well. Although the narrative aspects of the moving image may in some cases be of little relevance, like documentary footage depicting a specific landscape, it is often a central part of the experience. The narrative content of the moving image can be as complex as that in literature, which is often the case with movies made for cinema. The unfolding events in the moving image may in the same manner as literature be analyzed through the use of a dramaturgical model that is centered on the structure of the story and important elements such as plot and turning point. The impact narratives may have on music should therefore not be neglected when the influence of the moving image on music is investigated.

Narratives are not only a central part of the moving image but also expected to be essential in our structure of experiences. «Narrative is the fundamental scheme for linking individual human actions and events into interrelated aspects of an understandable composite» (Polkinhorne, 1988: 13). In the book *Consciousness Reconsidered* the neurobiologist Owen Flanagan describes how a narrative conception is a prerequisite for the experience of the personal identity (Flanagan, 1992: 197- 99). The findings from a famous empirical study from 1944 conducted by Heider and Simmel illustrates our disposition for understanding the world in terms of narratives. Participants were found to have a tendency to ascribe human qualities and agency to geometrical figures in a short film. The largest geometrical figure was described as more dominant than the smaller ones, and participants also found it to be bullying and chasing the smaller geometrical figures (Cohen, 2013: 20; Margulis, 2017: 235).

This inclination to categorize our experiences in terms of narratives is also reflected in our interaction with music. Lyrics have throughout history been a central part of the music experience, and the words that accompany the music are often telling a story like in an opera, a pop song, or a musical. Even instrumental music such as program music and film music contains narrative elements often manifested through musical imitation of extramusical phenomena. Several empirical findings also support this connection between music and narratives. A study from 2004 in which participants were presented with short pieces of orchestral music found a tendency among non-musicians to create stories that were related to

the music (Tan and Kelly, 2004). A brain imaging study from 2001 found musical syntax to be comprehended in a similar manner to language syntax, with both resulting in activity in the Broca's area (Maess et al., 2001). This area has also been found to be active in the comprehension of stories (Schmithorst et al., 2006; Schmithorst et al., 2007), further strengthening the link between music perception and narrative processes.

A narrative understanding of music is particularly evident in the terminology used to describe music. In the Western tradition tension and release are often ascribed to musical features, with the dominant chord representing tension and instability, while the tonic chord is perceived as stable. Matthew Lavy has suggested that suspense is as common in our understanding of music as in more narratively oriented stimuli such as books or films.

> there exists a striking similarity between the evocation of suspense in narrative (…) and a musical phenomenon that need not be connected at all with an external text or story, namely the phenomenon of expectation. (Levy, 2001: 88).

Expectations and uncertainty about the evolution of the story in for instance thriller movies are therefore not expected to be dissimilar from our expectations of the direction the music will take. The role of expectations when listening to music has been highlighted as a central part of the listening experience by other scientists. In the BRECVEMA model *musical expectancy* is named as one underlying mechanism of emotional responses to music. This is described as a process in which emotions are induced through violations, delays or confirmation of expectancies (Juslin and Västfjäll, 2008).

Whereas music is commonly understood in terms of narratives in the field of philosophy as well as music analysis, narrative responses to music have not been investigated much in the field of music psychology (Lavy, 2001: 75-77; Margulis et al., 2019). Partially in response to this lack of focus, a model on the emotional response to music, in which narrative processes are incorporated, has been created by Matthew Lavy. In this model the narrative comprehension is defined as a mechanism that combines context, sound, and utterances into «a dynamic coherent experience» (Levy, 2001: 77). The relevance of the model is supported by experiments conducted by Lavy in which the role of narratives during music listening has been investigated. Due to the hypothesis that we tend to integrate the extramusical context into the music experience, it was expected that extramusical narratives could have an influence on the music experience. In one study, participants were asked to read short stories while listening to music and then set to describe their experience of both stimuli. Although the results were somewhat ambiguous, narratives were found to have an influence on the

emotional response to music and particularly the valence was found to be affected by the short story (Lavy, 2001:150).

The assertion that extramusical narratives might have an impact on the emotional response to music has been supported by more recent research. In a study from 2013, the impact of extramusical information on emotions induced by music was investigated. Before listening to an excerpt of music expected to be experienced as sad, participants were given descriptions of either a sad or a neutral narrative. The sad narrative was about a shocking encounter between the US Army and starving prisoners in a holocaust camp, while the neutral narrative was related to nature in Yellowstone. These conditions were compared with two additional conditions from a previous study: One where participants were exposed to the same music but not given any narrative descriptions, and another, for comparison, where a different piece of music described as neutral was played. Narrative descriptions from the sad narrative were found to help increase emotions obtained from the music (Vuoskoski and Eerola, 2015).

The role of narratives in music listening has also been investigated by the researcher Elizabeth Hellmuth Margulis. In a study from 2017, music that was familiar and high in contrast was found to increase the number of narratives produced by participants when listening to music (Margulis, 2017). The creation of narratives has also been found to be correlated with level of enjoyment to music in a paper from 2019. This suggests that a narrative listening mode can be the result of engagement and emotional involvement with the music. In this experiment narrativization was also found to be a phenomenon independent of cultural background but dependent on familiarity with the music (Margulis et al., 2019). Although narrativization was not prevalent in all participants, these findings support the hypothesis of narratives being a central part of the music experience.

Based on Lavy's theory of how we tend to combine context and sounds into a unified experience one might expect the context from the moving image to influence and be integrated into the experience of the music in the same manner as a short story or specific descriptions of the background of the music might. Lavy actually suggests that silent films may be a more suitable stimulus in the measurement of how external narratives can influence the experience of music (Lavy, 2001:152). Given that narrativization seems to be related to certain musical features (Margulis, 2017), it is (similar to emotionally congruent descriptions: Vuoskoski and Eerola, 2015) also possible that congruency between the structure of external narratives and the structure of the music may help to enhance the overall experience. In particular elements of suspense in an external narrative may increase felt emotions in a similar manner as musical emotions are induced through *musical expectancy*. Narratives may also

make it easier to remember the musical experiences due to an established connection between the music and concrete information from a narrative.

2.3.3 Cross-Modal Perception

Influence from vision on sound is, in spite of a lack of focus on the influence of the moving image on music, a topic given much attention in the fields of both psychology and neuroscience. Researchers have over the years found interactions between different senses to be a common phenomenon. In our perception we tend not only to rely on one sense but rather combine different senses to better orient ourselves. A result of this organization is a more unified and less ambiguous perception of the world (Lalanne and Lorenceau, 2004). This integration of sensory inputs, known as *cross-modal perception*, makes it harder to separate different forms of sensory information and in some cases an influence from one modality on another can occur. The *McGurk effect* is one famous example worth mentioning of how visual information can alter, and, in some cases, override auditory information in speech perception (Tiippana, 2014). Interactions between vision and sound is also relevant in our perception of music. In a study from 2007, in which audio of notes played on a marimba was paired with videos of different stroke types, participants consistently rated notes paired with long visual gestures as having a longer duration, compared to notes paired with short visual gestures (Schutz and Lipscomb, 2007). Several studies have also found visual information from a musical performance to have an influence on the emotional responses to music. (Vines et al., 2006; Thompson et al., 2008; Vines et al., 2011; Krahé et al., 2015). Movements and visual expressions from a performer can therefore be a direct influence on our perception of the music.

The phenomenon of cross-modal perception substantiates the notion of an influence from the moving image on music, but the effect is not expected to be present in all forms of visual information. Cross-modal perception is assumed to rely on the perceiver believing that the senses interacting with each other have the same origin. This phenomenon is known as the *unity assumption* and is believed to be the reason why we tend to experience several modalities as one unified experience.

> The unity assumption certainly serves as one of the key mechanisms by which the human brain solves the crossmodal binding problem; that is, how signals from the different senses are encoded into a unified object/event representation (Chen and Spence, 2017).

When the moving image is combined with music, a unity between visuals and music is quite often absent. *Non-diegetic music*, music that does not originate from events in the moving image, is in the form of film scores or a soundtrack a central part of the movie experience. With the use of this form of music, the *unity assumption* will not automatically occur.

Although cross-modal interactions rely on a shared origin between several modalities, this is not determining. As the name implies, the unity does only need to be an assumption and one might be manipulated into believing that input from several modalities comes from the same event when this is in fact false. This is evident with the *McGurk effect* which solely relies on manipulation of the visual accompaniment to speech. What prevents the unity assumption from being rejected is, to a large extent, a perceived congruency between several modalities. «The unity effect can also be induced *implicitly* by stimuli with highly congruent properties» (Chen and Spence, 2017). This might be explained by the fact that we in our natural world tend to experience congruency between several modalities. When one is walking the movement of the legs and the sound coming from the feet hitting the ground would for instance be correlated. If the movements of the legs increase in tempo, then the sound produced by this movement will also increase in the same tempo. Congruent patterns that do not have the same origin might therefore create the illusion of modalities being united.

In the existing research perceived congruency between the moving image and music tends to be organized into *formal congruency* and *semantic congruency* (Iwamiya, 2013). The former is related to the time in which events occur and the matching between the two stimuli, like for instance having a person walk in the same tempo as music is played. *Semantic congruency*, on the other, hand is more concerned with the correspondence between two mediums in terms of meaning, like for instance the image of a cat and a meowing sound. The tempo of certain objects and the tonality of music accompanying the objects might for instance be perceived as either in synch or out of synch depending on the combination. In one study, fast visual objects in combination with music in the major key were perceived as congruent and the results were the same when minor music was combined with slow visual objects (Iwamiya, 2013: 149). Congruency is therefore not only concerned with the matching of temporal structures, but also aspects related to the affective characters of stimuli. Sad music is for instance predicted to be congruent with sad faces. Both formal congruency and semantic congruency is assumed to lead to an assumption of unity (Chen and Spence, 2017).

Achieving congruency is, in spite of non-diegetic music not being rooted in actions in the moving image, an important part of the creation of visual experiences. «A vital element in making visual media presentations more impressive is ‘perceived congruence’ between sound

and moving pictures» (Iwamiya, 2013: 141). In the tradition of cinema, the technique of having music temporally synchronized with visual movements is known as *mickey mousing.* The name comes from the extensive use of music mimicking the animated motions in Disney films like *Fantasia*. Although most associated with animated films, this technique has also been utilized frequently in non-animated films as well. Theories on cross-modal perception may therefore be relevant in the context of more visually advanced and non-dietetic mediums like film.

Several researchers who have investigated interactions between the perception of the moving image and music have a point of departure in theories on congruence and cross-modal perception. The psychologists Annabel Cohen and Sandra Marshall have developed a model for investigating the function of music and film called the *congruency association model* (CAM). This model describe how influence from music is the result of two processes. The first process revolves around how the «shared (congruent) temporal structure of the music and film» (Cohen, 2013: 20) might lead our attention to specific parts of the moving image. The other process is concerned with how meaning from music, which is the result of associations evoked by the music, might be ascribed to the formal congruent part of the moving image. Although Cohen focuses on how formal congruency might lead our attention to specific parts of the moving image, research has also found semantic congruence to direct attention to formal features in the moving image (Boltz et al., 1991). This congruency, either formal or semantic, is something we intuitively seek and if one modality is ambiguous then this is assumed to be adjusted to obtain unification between the modalities. «An ambiguous event (music or visual scene) accompanied by another event with a clear, unambiguous meaning is even more likely to result in unification» (Boltz, 2013: 228).

The phenomenon of how one dominant modality can override another subservient modality has been supported by empirical studies. In a study from 1993, conducted by Annabel Cohen, participants were shown an ambiguous scene of a man and a woman who were engaging in an activity of either a loving or a violent nature. This scene was paired with one of two excerpts of music given the titles “Say Hello to Love” and “Conflict”, judged by the participants to be highly appropriate titles for the pieces. Results showed that participants who saw the clip accompanied by the romantic excerpt judged the activity between the two people to be of a loving nature while participants who saw the scene in combination with “Conflict” judged it to be of a violent nature. The clear affective character of the music was shown to have an impact on the scene when it was ambiguous. In this study the music was also paired with an unambiguous scene of people fighting but the romantic music except did

not have a significant influence on the perception of the film (Cohen, 2013: 27). This might be explained by the fact that both modalities were dissimilar in terms of semantic congruency and none of them contained ambiguous elements.

Since influence can be the result of pairing an unambiguous modality with an ambiguous one, an influence from the moving image on music is therefore possible. In one of the few studies, conducted by Marilyn Boltz and colleagues, that have found the moving image to influence the perception of music, ambiguous music was paired with videos with a clear affect (Boltz et al., 2009). The CAM model developed by Annabel Cohen is also relevant for the influence of visuals on music. Just as associations from music may be ascribed to visual patterns, a reverse phenomenon in which meaning from the moving image is imposed on the music may also occur (Boltz, 2013: 228).

The reason why we in our perception tend to seek for congruency and unification, and erase ambiguity is assumed to be a limited mental capacity. «The mind cannot attend to everything equally. If a choice must be made, it makes sense to direct attention to information that is coordinated between the two modalities» (Cohen, 2013: 23). Congruence is more easily achieved when ambiguous stimulus is overruled by unambiguous. This influence one modality may have on another can be an advantage in certain real-life situations. In a noisy environment where sounds are unclear, visual aspects of speech (movements of the mouth) may for instance help improve communication. Unification and congruency between senses is therefore a necessary tool in our perception: «unity (…) ensures a more efficient use of cognitive resources» (Boltz, 2013: 228). Because of this strong need for unification, visuals can be expected to have an influence on the perception of sound even when the sensory information does not appear to have the same origin. It is therefore important to take cross-modal interaction into account when investigating the influence of the moving image on music.

2.3.4 Contagion

Cross-modal interactions seems to demonstrate how the moving image may change our perception of music, but this is not the only way in which visual patterns may have an impact on our music experience. The influence visuals from a live performance may have on the music can also be caused by what is known as *emotional contagion.* Contagion is a process in which emotions from an external source triggers an internal emotional experience through mimicry. This phenomenon can express itself in many ways. Watching a person smile may

induce positive emotions and cause the observer to smile, and one may become sad by seeing someone suffering. Contagion effects have been found in brain imaging studies on the experience of pain, disgusting odors, and touch (Molnar-Szakacs et al., 2011: 321). In addition to experiencing the same emotions, emotional contagion may also result in the observer mimicking the emotions from an external source. In a study from 2000, participants who were shown pictures of happy, neutral, and angry faces responded by mimicking the facial expression in the pictures (Dimberg et al., 2000). In our perception of the world, we tend to mirror and catch emotions of others and the phenomenon of contagion is closely linked with empathy (Juslin, 2019: 288).

One of the causes of emotional contagion may be what is known as *mirror neurons.* In a study from 1992 done on *Macaca menestrina* monkeys, researchers found neurons that were activated during certain actions, such as grasping food, to also be activated when the monkeys only watched an experimenter grasp the same objects (Di Pellegrino et al., 1992). This was the first experiment that found evidence of certain neurons that were firing both when an action was executed and when observed, mirroring the emotions of the executor. Further studies have investigated and supported the existence of mirror neurons in monkeys (Gallese et al., 1996; Rizzolatti et al., 1996). The tendency to imitate perceived stimuli may therefore be explained by the activity of certain neurons. Although present in monkeys, few studies have been able to give direct evidence of the existence of *mirror neurons,* but these neurons are nevertheless believed to exist in humans. Data from fMRI have found certain brain areas to be activated for both observed and executed actions (Molenberghs et al., 2009), and equivalent brain areas to the ones that activate mirror neurons in monkeys also exist in the human brain (Molnar-Szakacs et al., 2011: 319).

Emotional contagion is expected to play an important role in our experience of music. With a point of departure in research on mirror neurons, the model *Shared Affective Motion Experience* (SAME) focuses on how emotions may be communicated through musical activities, creating a shared experience between people and «making possible empathy without the need for verbal explanation» (Molnar-Szakacs et al., 2011: 319). Emotional contagion is also included as a mechanism in the BRECVEMA model by Patrik Juslin and Daniel Västfjäll (Juslin and Västfjäll, 2008). Although they focus more on how emotional cues from the voice can lead to contagion, this process may also be triggered by visual information accompanying the music. «The most obvious way in which musical events can produce contagion effects is through the non-verbal expressions (face, body) shown by performers during a live concert» (Juslin, 2019: 290). In relation to the SAME model,

physical gestures are mentioned as one example of how communication of emotions may occur (Molnar-Szakacs et al., 2011: 316).

Several studies have found facial expressions from performances to enhance emotions perceived from the music (Thomson et al., 2005; Thomson et al., 2008; Livingstone et al., 2015). In one of the experiments conducted in 2005 by Thompson and colleagues, participants were presented with audio-visual performances of B.B King playing guitar. In excerpts where Kings facial expressions were of a dissonant character, the music was described as more dissonant than excerpts where his expressions were more neutral. Facial mimicry has in addition to experienced emotions also been found to occur during music listening. Chan and colleagues found participants to respond to an audio-visual recording of a vocalist's performance by mimicking the same emotions as the singer was conveying (Chan et al., 2012). This clearly demonstrates the presence of emotional contagion under exposure to music.

In research done on how visual information may influence music through emotional contagion, musical performance seems to be the topic given most interest. Contrary to cross-modal interactions, which rely on an assumption of unity, emotional contagion may occur in the absence of congruency. It would therefore be highly possible that visual information that is not united with the music may also have an influence on the experience of music. A video depicting the facial expressions of people smiling or crying paired with music may in that sense lead to an enhanced emotional experience, regardless of the structural relationship between visuals and music. That being said, it is highly possible that facial expressions conveying emotions that are conflicting with those of the music, a happy face combined with angry loud music for instance, will reduce the emotional experience due to the mismatch between visuals and music. Emotional contagion is nevertheless expected to be less dependent on the relation between the music and the visual than cross-modal interactions.

2.3.5 Discussion of Theories

The four theories mentioned all have different angles in relation to the question of how the moving image may influence the experience of music. The phenomenon of conjuring up visual images to music demonstrates our tendency to combine music and visuals. It is also an example of how emotions from music may be caused by visual information. Our inclination to understand the world, and thereby also music, in terms of narratives and Lavy's theory of how we combine sounds with context offer an explanation of how the moving image may be

integrated into the musical context and influence the experience of music. Theories on cross-modal perception probably reveals most directly how sounds may be perceived differently in combination with visuals, like with the McGurk effect. Despite it being grounded in the unity assumption, several researchers have demonstrated how this assumption may be sustained through the presence of different forms of congruency. The last example of emotional contagion, although restricted to emotions conveyed by someone observed, demonstrates how our emphatic nature can cause visuals to alter the experience of music.

The tendency to create narratives while listening to music strikes many resemblances with the mechanism of visual imagery. Given that the content of visual imagery may be biographical events, the creation of narratives is in addition to the visual images a central part of imagery as a response to music. Although visual imagery is not mentioned in the experiments conducted by Margulis, narrative engagement was measured through imagined stories which suggest that visuals are a central part of the narrative listening experience. In the study conducted by Vuoskoski and Eerola, visual imagery related to the narrative description was reported by 80% of the participants. It was further suggested that emotions triggered by narrative descriptions might be evoked through the visual imagery mechanism (Vuoskoski and Eerola, 2015).

Isolating the mechanisms at play may also be challenging when it comes to cross-modal perception and emotional contagion. Despite these theories having different points of departure, the effect of both phenomena may occur in similar situations. As mentioned, most of the research on emotional contagion in relation to music is concerned with facial expressions during live performance of music, but facial gestures are also relevant to the phenomenon of cross-modal interactions. The *McGurk effect* is for instance the result of manipulation of mouth gestures. This mixture of mechanisms is to some extent also reflected in the existing research. In one experiment conducted by Thomson and colleagues in 2008, facial expressions from a singer were found to have an influence on the perceived affect of the music (Thomson et al., 2008). The theoretical foundation for this experiment is mostly connected with cross-modality and emotional contagion is not mentioned in the article. Although cross-modal interactions may explain some of the change of the perceived valence of the music, emotional contagion from a happy or sad face may also explain the influence from facial expressions on the music.

2.4 Theoretical Accounts for the Possibility of a Persistent Carry-Over Effect

With these different theoretical approaches, a foundation for the possibility of an influence from the moving image on music has been made. Although these theories give insight into mechanisms that may be at play, they do not dive into the possibility of a persistent carry-over effect that may be present in later encounters with the music. Visuals may alter the experience of music when the two are played in combination, but this does not necessarily mean that lasting changes in the experience of the music have been established. Since this aspect is a central part of the research question, a separate investigation into the possibility of this phenomenon needs to be made and the following part of this thesis will examine this question with a point of departure in theories on memory related processes, conditioning, and priming.

2.4.1 Memory-Related Mechanisms

Similar to our perception of the world, the recollection of past events consists of interactions from many systems like vision, audition, spatial imagery, olfaction (smell), taste, language, narrative, emotions, etc. (Rubin, 2006). Different sensory input is not memorized separately but rather as a multimodal experience and this way of organizing memories is highly relevant in a musical context. A melody and a text are for instance memorized more easily when the two are connected to each other, as opposed to being learned separately. Visual aspects of hand movements related to playing an instrument can serve an important role in learning a music piece. A change in the layout of a musical score is likely to have a negative impact on the memorization of a piece of music, due to changes in the spatial location of notes on the page (Chaffin et al., 2009: 354- 356).

This tendency to connect several inputs in our recollection of past experiences can explain how episodic memories may be trigged by an associated stimulus. A particular smell may bring back memories from the past that co-occurred with the smell. Similar incidents may for instance also be triggered by food, as in the novel *In search of lost time* by Marcel Proust where a Madeleine cake brings the author back to his early childhood. Music is no exception to this and even people with severe memory loss due to dementia may recollect earlier experiences when presented with music with a personal meaning (Juslin, 2019: 326). Episodic memories are in a musical context also expected to be a source of emotions and the phenomenon is included as one of the mechanics in the BRECVEMA model. Emotions that

are associated with a particular memory are expected to be induced when the memory is triggered by music that has co- occurred with the memory earlier. Particularly autobiographical memories of a social nature tend to be recollected in encounters with music, but other memories such as those of a music concert or a movie may also unfold in encounters with associated music (Juslin and Västfjäll, 2008). The recollection of music-associated events is not entirely arbitrary in nature. Emotional experiences are expected to be remembered more easily than non-emotional experiences (Talmi et al., 2007), and songs that have a strong emotional impact on listeners have been found to more frequently evoke episodic memories (Juslin, 2019: 318).

The integration of information from many modalities in our memory may, in addition to enhance emotions induced, also have an influence on how a particular piece of music is remembered. In the study from 2009 conducted by Marilyn Boltz and colleagues, which investigated the impact of visual information on music, the recollection of music pieces after an initial pairing was also investigated in a second experiment. After being presented with music in combination with visuals with either a positive or negative affect, participants were presented with the pieces of music without visuals in addition to two manipulated versions of each music piece. The manipulated versions were either more positive in character (faster tempo, higher pitch etc.) or more negative (slower tempo, lower pitch etc.) and participants were set to judge which one of the versions that they had heard in combination with the video. Results showed that video accompaniment distorted the recognition of the melodies in a mood congruent fashion. Music excerpts paired with a video with a positive affect tended to be remembered as higher in pitch while the opposite pattern was found for music paired with videos with a negative affect. To what extent this influence is the result of a conscious recollection of the videos however, is not clear, as the study did not go into underlying mechanism for changes in memory.

2.4.2 Conditioning

Associations made between stimuli co-occurring in time are in addition to memory related processes also relevant for the phenomenon of conditioning. First discovered by Ian Pavlov by accident in a study on salivation in dogs, Pavlov realized that repeatedly pairing food with different stimuli, like for instance the sound of a bell, would result in the dogs drooling whenever the paired stimulus was presented (Pavlov, 2010). With these experiments Pavlov would lay the foundation for *classical conditioning* which is a process in which repeatedly

pairing a potent unconditioned stimulus like for instance food with a neutral stimulus like a bell can change the response to the latter, resulting in the neutral stimulus being conditioned. Conditioning, as opposed to episodic memory, does not result in a conscious recollection of past experiences, but is only concerned with changes in responses (Juslin, 2019: 319).

Conditioning is relevant in a musical context and may explain many of the attitudes we have toward specific forms of music. Patrik Juslin uses the term *associative coding* to describe how the perception of music can be altered as a result of being repeatedly associated with a stimulus or specific events with a clear affect. This is further highlighted as a central aspect of how we perceive emotions in music (Juslin, 2019: 173-75). The heroic quality ascribed to the horn, or the perception of organ music as spiritual is likely to be the result of associative coding. Even the concept of major chords and modes being perceived as happy and minor ones as sad, may possibly be the result of learned associations (Athanasopoulos et al., 2021).

Conditioning may not only alter the perception of music but can also have an influence on emotions induced in a listener. *Evaluative conditioning* is a form of conditioning in which emotions from a potent stimulus are carried over to a conditioned stimulus after repeated pairings. In the BRECVEMA model by Juslin and Vastfjall, evaluative conditioning is included as one of the mechanisms and strong emotions may be induced in a listener as direct consequence of this process. Compared to classical conditioning, the process related to evaluative conditioning seems to differ in several ways. Studies have found the latter to be less dependent on awareness and attention, and it is also expected to be more resistant to extinction (Baeyens et al., 2005; Vansteenwegen et al., 2006). This is a process in which the conditioning effects dissipates due to the lack of followed up pairings between an unconditioned and a conditioned stimulus. This form of conditioning is also expected to be established faster than classical conditioning. A simple tone has been found to be conditioned fairly quickly in fear conditioning experiments in which the presentation of the tone is followed by an unpleasant experience such as a shock. Even one pairing is expected to be sufficient for conditioning to occur if the potent stimulus is of a certain intensity (LeDoux, 1995: 211).

Although much of the music-related research on evaluative conditioning has used music as an unconditioned stimulus (Gorn, 1982; Eifert et al., 1988; Dickson and Schubert, 2020), there exists some research that have looked into how music may be conditioned by being repeatedly paired with affective stimuli. A study on dogs found a musical piece that had repeatedly been presented when dogs were with their owner to have a calming effect on the

dogs when they were alone. This effect was not present in an unconditioned classical piece of music (Bernardini and Niccolini, 2015). Studies on humans have also found evaluative conditioning to be a source of influence on people's relationship with music. One experiment found the attitudes toward music, paintings and photographs to be altered after six free lunches (Razran, 1954). In a study from 1992, participants who were exposed to a piece of music in an unpleasant environment tended to dislike the music and hold less favorable attitudes toward a brand name the music was paired with, compared to a control group in which participants were not familiar with the music (Blair and Shimp, 1992). In a study conducted by Klas Hellström and Petri Laukka from 2012, participants were presented with previously rated neutral music that was repeatedly paired with either positive or negative pictures. In a succeeding task, the participants were asked to rate the valence of words that were presented after one of the conditioned music excerpts. The study found the reaction time to be faster for positive words that were presented after music excerpts previously paired with positive pictures compared to positive words that succeeded incongruent music excerpts, but no effect was found on the negative words (Juslin, 2019: 311-12).

2.4.3 Priming

It is not only through conscious memory processes and conditioning that a carry-over effect from one stimulus on another may occur. *Priming* refers to how the perception of one stimulus (target) may be influenced by a preceding stimulus (prime) and compared to conditioning this influence is not dependent on a pairing of the two stimuli. One central aspect of this effect is its presence regardless of conscious awareness and attention to the prime. Priming has even been found to occur in people with brain damage in areas connected to explicit memory, clearly demonstrating that the mechanisms related to this effect relies on implicit memory (Schacter and Bruckner, 1998). Much of the research on priming is related to the processing of words and how target words carrying certain similarities with a prime is interpreted faster than dissimilar primes. One classic example of this is how the word nurse is processed faster when following the word "doctor" compared to a word not related to nurse like "bread" (Tillmann and Bigand, 2002). In addition to the processing of words, research from social psychology have found priming to have an influence on participants' behavior and attitudes, suggesting a more comprehensive influence from the prime, but it is important to mention that several of the studies are surrounded by controversies due a lack of replicability (Doyen et al., 2012; Pashler et al., 2012).

Priming is also relevant in terms of the evaluation of emotions and primes with a strong affective value can influence the perception of targets (affective priming). Similar to other forms of priming, affective priming tends to be measured in terms of how quickly a target stimulus is evaluated, but in the case of affective priming, the prime carries a specific emotion. Studies on affective priming have found photographs, color slides, and even odors either positive or negative to work as primes, demonstrating that the concept of priming is relevant for several types of stimuli (Musch and Clauer, 2003: 9-12). Some studies have also looked at how priming may work in a musical context. Although much of this research is either concerned with the priming effect within a musical chord progression, or how music may work as a prime on target words (Tillmann and Bigand, 2002; Sollberge et al., 2003; Koelsch et al., 2004), the effect of non-musical stimulus on music have also been investigated. In a study conducted by Katharina Sophia Goerlich and colleagues, music was treated as a target while visual words were treated as primes. Results from one experiment revealed that participants evaluated the affective character of the music significantly faster when the music was preceded by affectively congruent words, compared to incongruent words (Goerlich et al., 2012). Although not concerned with the response time in the evaluation of a target stimuli, the study conducted by Vuoskoski and Eerola (2015), which found extramusical information to intensify induced emotions, may possibly also be regarded as a form of priming with the descriptions working as primes and the music as a target.

2.4.4 Relevance of Theories in relation to a Persistent Carry-Over Effect

The theories related to memory processes, conditioning and priming support the possibility of a carry-over effect from the moving image on music when music is present subsequent to an initial pairing. Primarily the phenomenon of episodic memories may explain how emotions may be induced through the recollection of the video when the music is repeated. Although these memory processes are primarily associated with autobiographical events, the information from the moving image may possibly be recollected in a similar manner, particularly if the overall experience is of a certain emotional intensity. Given that the recollection of the affect of music has been found to be influenced by a previously paired visual stimulus (Boltz et al., 2009), it is not unlikely that music can be perceived differently when repeated in a similar mood-congruent manner.

The mechanisms related to conditioning may also account for a possible influence, with the moving image working as an unconditioned stimulus and music as a conditioned stimulus. A possible influence nevertheless seems to rely heavily on repeated pairings between the two stimuli. It is therefore possible that one paring between music and the moving image is not enough for conditioning to occur. An influence after one pairing, particularly in relation to induced emotions, is nevertheless not unlikely to occur in cases where the unconditioned stimulus is of a particular strength.

Priming on the other hand is not dependent on repeated pairings and further supports the possibility of an influence after one repetition. Music in combination with the moving image may in that sense be regarded as a prime and the same music repeated alone may be viewed as a target. The problem with priming is that it seems not to be very versatile. In most of the research the target is presented right after a prime and research have found the effect of affective priming to be greater when a target is presented quickly after a prime (Musch and Clauer, 2003: 17). A priming effect may therefore not necessarily come from the moving image in the current experiment, but a preceding excerpt may unintentionally also work as a prime on a succeeding excerpt. Another problem with priming is that many studies mainly tend to measure the response time. Investigations of felt and perceived emotion in terms of self-report and measurements of physiological responses do not seem to have been given much attention in the existing research. There is therefore some uncertainty related to the magnitude of the priming effect. To what extent priming may lead to persistent changes is also something that can be discussed, but evidence for a long-term effect have been found in some forms of priming (Moutsopoulou et al., 2019; Was et al., 2019).

An important consideration that needs to be taken into account in relation to the repetition of the music is the fact that a stimulus is expected to be preferred as it gets more familiar. This is known as a *mere exposure effect* and several studies have found exposure to be an effective way of increasing the preference for music (Krugman, 1943; Mull, 1957; Peretz, 1998). Although repetition is mainly associated with increased preference, it may in some cases have a negative effect on preference and this is known as *the boredom effect* (Brattico and Pearce, 2013: 55). The extent to which music is likely to be preferred after exposure seems to be dependent on the complexity of it. A complex piece of music is expected to be preferred more as a result of repetition while a piece quite low in complexity may be preferred less after repeated exposure (Juslin, 2019: 415-16). The second experience of the music is regardless of visual accompaniment therefore expected to be different than the first, either in a positive or negative way due to the familiarity of it.

3 Method

3.1 Aim and Hypothesis

The aim of the current study was to examine the possible influence the moving image could have on emotions perceived and induced in a musical context. The findings from the study conducted by Boltz et al. (2009) gives valuable insights into how cross-modal interactions may occur when music is of an ambiguous character and visuals are unambiguous. It therefore seems like the moving image may have an influence on the perception of music. The present research set out to further examine the magnitude of this type of influence to see if unambiguous music, in addition to ambiguous music, could be influenced by selected videos predicted to be perceived as unambiguous. A verification of this hypothesis would further suggest that the moving image can have an even stronger impact on perceived emotions in music than what the present research has documented.

An investigation of induced emotions was also an important aspect of the current study since literature particularly related to visual imagery and narrative processes indicates that the moving image may be a source of felt emotions. The selection of videos was based on their ability to induce strong emotions and the presence of narrative elements. One video was predicted to induce awe, while another was hypothesized to evoke feelings of being moved. Videos associated with these two emotions were selected since they (awe and being moved) were expected to induce stronger responses than basic emotions, in addition to be connected to the occurrence of chills. The videos wer therefore hypothesized to have a positive influence on the experience of awe and being moved, as well as the occurrence of chills. Narrative elements from the videos were, based on the highlighted research, also expected to have an influence on induced emotions and to result in a more unified experience (Lavy, 2001; Vuoskoski and Eerola, 2015).

Felt emotions tend to be associated with the enjoyment aspect of music listening (Thompson, 2006; Vuoskoski et al., 2011), and the latter was therefore also investigated in the present research. Feelings of being moved have for instance been found to mediate the enjoyment of sad music, resulting in what is often characterized as pleasurable sadness (Jonna and Vuoskoski, 2017). In two studies conducted by Geringer and colleagues, the moving image had a positive, but small, effect on both emotional involvement and liking (Geringer et al., 1996; Geringer et al., 1997). It was therefore not unlikely that a positive influence from

the videos on induced emotions, such as being moved and awe, would also lead to a positive effect on the enjoyment of the music.

Emotions induced by the videos were further hypothesized to have an influence on the paired music when it was repeated without visuals. Since research has found emotional events to be remembered more easily than neutral events (Talmi et al., 2007), strong emotions predicted to be induced by the videos in the first presentation were expected to be recollected in the second presentation of the music. Mechanisms related to episodic memory were expected to account for a sustained influence from the moving image, but an influence through mechanisms related to evaluative conditioning was not excluded as a possible source of an influence. Since the recollection of perceived emotions in music has been found to be influenced by visuals (Boltz et al., 2009), a lasting influence in relation to perceived emotions was also hypothesized to occur.

The present research also set out to investigate if an influence from the moving image could occur in the absence of congruency. This choice was mainly motivated by the lack of attention given to this topic in the small number of papers that have been occupied with the impact of the moving image on music. A possible effect from the moving image was in general expected to be negative when a pairing was perceived as incongruent.

3.2 Experiment Design

For my empirical investigation on the influence of the moving image on music, it was important to have control over both the music and the moving image variable. Methods based on interviews or observations in a natural environment were therefore not chosen. A controlled experiment was in that regard believed to be the only sufficient method for separating these variables from each other. For this controlled experiment, participants were exposed to different stimuli, either music alone, video alone, or music and video in combination. After the presentation of each stimulus, participants were asked to answer questions relating to their experience of the stimulus (questions were mainly related to enjoyment, felt emotions, and perceived emotions). An *independent group design* (also known as a between subject design), in which participants were randomly assigned to one of three experimental conditions, was applied to the experiment. Since the main aim of this study was to look at the influence of the moving image on music, the former (moving image) was

manipulated and treated as an independent variable, while the latter (music) was treated as a dependent variable and held constant. In the current experiment this meant that the music presented was the same in each condition while the video accompaniment would vary amongst the different groups. All the music excerpts were repeated without visuals after the first presentation.

A *repeated measures design* in which all conditions consisted of the same participants was omitted for several reasons. Participants would have been presented with three times the amount of stimuli as in an independent group design. This would further make completion of the experiment more demanding and, a *fatigue effect* (Cozby and Bates, 2015: 170) would be more likely to be experienced by several participants. More importantly, the experience of music in one condition would have an impact on the experiences of the remaining conditions, known as a *carry-over effec*t (Cozby and Bates, 2015: 170). Although an independent group design is believed to be the most suitable option, there are some disadvantages with this design. *Confounding variables* related to the specific situation, personality traits, musical preference, and musical background are likely to have had an impact on responses, and variations found between different groups may, due to the independent group design, be partly explained by these variables.

3.2.1 Distribution of Stimuli

A total number of three music excerpts and two videos were presented in all the groups. A music piece was in one group presented alone and in the two remaining ones paired with one of the two videos. The experiment more specifically contained three different stimulus categories for the music: music alone category (control stimulus), music and video 1 category, and music and video 2 category (experimental stimuli). Participants were not assigned to conditions containing music from only one of these three categories and a design with control groups and experimental groups was avoided. Rather, the different stimulus categories were balanced across the 3 different conditions. More specifically, each condition contained music excerpt from all the stimulus categories (one from each category). The stimulus categories the music excerpts were allocated to were further varying between the different conditions (see table on page 32 for a more detailed illustration).

Despite making the analysis of the data less demanding, a control group and two experimental groups would have resulted in several disadvantages. Whereas the response to for instance music alone now has been gathered from all the participants, the experience

would otherwise have been determined by one third of the total population with the application of control and experimental conditions. The lack of variation in stimuli, which control and experimental groups would have resulted in, could also have caused participants to get in a bad mood. Participants who for instance were shown one video paired with three music excerpts would probably not be as emotionally influenced by the content of the video after the third presentation of it. A control group might also have found the experience to be less enjoyable due to the presentation of the same music without any visual accompaniment twice, making the repetition more evident. This could have had a negative impact on the result.

	Music alone	**Music and Video 1**	**Music and Video 2**
Music 1	Condition 3	Condition 1	Condition 2
Music 2	Condition 2	Condition 3	Condition 1
Music 3	Condition 1	Condition 2	Condition 3

Table depicting which conditions the music stimuli for the first presentation were allocated to.

In addition to the different music excepts presented either alone or in combination with visuals, all participants were also randomly presented with one of the two videos alone in the beginning of the study. Randomization was conducted within all the conditions with one half of a group being presented with video 1 and the other half video 2. The presentation of the visuals alone was intended to capture reactions the moving image alone could account for, and responses from this stimulus will be compared with those from both music alone and music in combination with the specific video. The video which was presented alone might not have had the same emotional influence when paired with music as the video participants saw for the first time. This is because familiarity with this video might have diluted emotions elicited by video when presented in combination with music. The surprise element, particularly from the breaking of the wave in the “surfing video” might for instance be weaker after the second presentation. By having an evenly distribution of these videos to participants, the negative impact from the video played twice will have been more or less the same in each group.

3.2.2 Repetition of Music

The three music excerpts were repeated in all the groups without any video accompaniment. The aim of this repetition was to see if music excerpts combined with visuals during the first presentation, would be experienced differently during the second presentation from those that had not been presented with visuals. A verification of this hypothesis would support a lasting influence from the moving image on music. It would also more strongly indicate that emotions elicited by the moving image can carry over to the music and change the experience of it. In particular, in relation to felt emotions is it quite difficult to measure a *carry-over effect* from music over to video when the two are played simultaneously. Emotions may simply be induced by the videos without having an impact on the experience of the music. There are nevertheless some challenges with the repetition of the music. The second presentation of the music stimuli is, regardless of the moving image, likely to be experienced differently during the second presentation. The *mere exposure effect* may lead to a stronger preference for the music stimuli, while the *boredom effect,* working in the opposite direction, may result in a dislike for a music piece that is repeated. Although factors like complexity may predict the preference of music, it would be impossible to fully separate exposure effects from the influence of the moving image. Different exposure effects might therefore operate as confounding variables.

3.2.3 Procedure of the Experiment

The first stimulus presented was, as mentioned, one of the two videos. This was succeeded by the presentation of the tree music excerpts, one alone and two with video accompaniment, and finally, the music excerpts were repeated without video accompaniment. Before the repetition of the pieces of music, participants were asked to solve a small filler task. This task was a question taken from an IQ test and the purpose of it was to take focus away from the stimuli participants recently had been exposed to and prevent the repetition of the music from being experienced too soon after the first presentation.

Details	Condition 1	Condition 2	Condition 3
One of two videos randomly presented	Video 1 or Video 2	Video 1 or Video 2	Video 1 or Video 2
Stimuli presented in a randomized order	Music 1/ Video 1	Music 1/ Video 2	Music 1/ No Video
	Music 2 / Video 2	Music 2/ No Video	Music 2 / Video 1
	Music 3 / No Video	Music 3/ Video 1	Music 3 / Video 2
IQ question	Filler task	Filler task	Filler task
Stimuli presented in a randomized order	Music 1	Music 1	Music 1
	Music 2	Music 2	Music 2
	Music 3	Music 3	Music 3

Table representing the stimuli presented in each condition.

The order of the presentation of each stimulus was, due to randomization, not necessarily as in the table shown above. The excerpts were randomized to reduce the impact of one stimulus on the following. A loud piece of music high in arousal might for instance have sounded even louder if it was presented after a quiet piece low in arousal, due to the contrast in dynamics and energy. This phenomenon is known as a question order bias and is prevalent in a musical context (Juslin, 2019: 95-97). Randomization did only include the order in which the three music excerpts were played, and the overall procedure was the following: video without music (one of two elements randomly presented) - the three musical excerpts in random order (two accompanied by visuals) - filler task - the same musical extracts in random order.

3.3 Stimuli

3.3.1 Music

The chosen music stimuli were extracts from a slowed down version (800%) of Section 1 from *Music for 18 Musicians* by Steve Reich, the second movement of *Company* by Philip Glass, and *Dream* by Ian Post. Participants were presented with approximately one minute and 20 seconds from each excerpt. The two main criteria for the selection of music excerpts were absence of lyrics and lack of familiarity. Given that an influence of extramusical information from the moving image on music was the object of investigation, it was important

to have extramusical information controlled for. Vocal music with lyrics was therefore excluded due to the possible impact the content of the lyrics could have on emotional reactions. This would otherwise have been one extra variable to take into consideration.

Well known music was avoided due to the possibility of it evoking emotions through episodic memories, or other idiosyncratic associations. This could have had an unwanted influence on the results. Familiarity could also, due to the mere exposure effect, have an impact on ratings of enjoyment. Familiar music has been found to be enjoyed more and to a larger extent be able to induce physiological reactions like chills than unfamiliar music (Panksepp, 1995; North & Hargreaves, 1995; Greve et al., 2007; Greve et al., 2009). Although it is impossible to fully control participants preferences and familiarity with the music excerpts, the impact of these variables can nevertheless more easily be minimized by consciously omitting well known music.

The chosen music excerpts are dissimilar in both style and genre and were predicted to be perceived differently in relation to the emotional character. In the selection of music, it was important to have stimuli that were dissimilar from each other to achieve a variety in the repertoire. There are mainly two reasons for having this variety of music. In the first place, it would probably help to increase the overall enjoyment and reduce the chances of participants getting in a negative mood due to dislike of a specific type of music. Secondly, a variety might also give better insight into the influence of the moving image on music in general.

3.3.1.1 Music Excerpt 1

The first music excerpt was taken from a slowed down version (800%.) of the first section from *Music for 18 Musicians* by Steve Reich. This excerpt was not slowed down for the experiment, but the version was found in this manner and chosen based on its character. It can, partially due to a low tempo and lack of clear rhythm, be characterized by an ambient and calm feeling. Despite *Music for 18 Musicians* being quite well known, this slowed down version was expected to be unfamiliar to the majority of the participants. Due to an extremely low tempo, it is practically impossible to find any clear similarities with the original material. The musical preference of participants is a confounding variable that might have an impact on the experience of this excerpt. Some may for instance dislike ambient music but, due to it being rather unfamiliar, it is not presumed to be of serious matter. In relation to the tonality, the music can be described as rather ambiguous (neither major nor minor), and it consists of chords that are more complex and contain more notes than the traditional minor and major

chords. As indicated by some of the existing research, an influence from the moving image may be more likely due to this ambiguity, (Boltz et al., 2009). This predicted ambiguity was therefore one of the main motivations for choosing this excerpt. When it comes to the perceived arousal, the music was, due to a significantly slow tempo, expected to be given a low rating by participants when played alone.

3.3.1.2 Music Excerpt 2

The second music stimulus was an extract taken from the third movement of Philip Glass's Second String Quartet *Company.* Philip Glass is a composer associated with the minimalist movement, which is part of the western classical tradition. This extract is probably the most well-known and was therefore the one most likely to be recognized by the participants. It may possibly be the excerpt that has been influenced the most by the preference of the participants. This is both due to it partly being the most well-known, as well as a possible dislike for classical music by a selection of the participants. Given that classical music is known for not being the preferred music of parts of the population, it was important to take this into consideration. Despite being the most well-known of the music excerpts, it is relatively unknown compared to works from the classical canon, and the impact of musical preference was therefore not expected to be of serious matter. The piece can be described as quite dramatic and, compared to the former stimulus, this piece was expected to be given a rating of valence located on the negative side of the scale. The opening part of the excerpt is quite low in dynamic, but it has a serious character. With the entry of a new motif after approximately 30 second the dynamic suddenly increases and the mood of the piece becomes more dramatic. Whereas the valence was expected to be assessed as low by participants, the rating of the energy was on the other hand not as predictable. While the tempo is relatively slow the music has a rich texture and both the intensity, and the dynamic level is quite high in the last part of the piece.

3.3.1.3 Music Excerpt 3

The last music excerpt *Dream* by Ian Post can be ascribed to the category of library music and was written with the intention of evoking specific feelings. The stimulus has been given the labels "uplifting" and "happy", and the music was expected to be rated on the upper side of the valence scale when presented alone. The excerpt contains vocals (with no lyrics), synth, drums, piano and guitar, and this instrumentation contributes to giving the music a "pop"

feeling. This piece was, in addition to receiving a positive rating of valence, also expected to be perceived as rather high in relation to arousal. A moderately high tempo and a rich texture were the musical components that this assumption was based on but as with the former example, there is an amount of uncertainty due to neither the dynamic level nor the tempo being remarkably high. Due to it not being written for the purpose of exposure on neither radio, concerts nor streaming platforms, it was the excerpt least expected to be familiar to participants and therefore also most unlikely to be affected by music taste. This however did not fully exclude the possibility of it being disliked by some participants due to genre preferences.

3.3.2 Videos

The two visual stimuli that were chosen for the study were a video depicting a surfer named Mike Parson surfing on a large wave and a video of the reunion between a lion called Christian and his former owners. Each video lasted, as the music excerpts, for approximately 1 minute and 20 seconds. The selection of the videos was grounded in their expected ability to induce strong emotions and physiological response in participants, as well as differ from each other in relation to felt and perceived emotions. The structural matching between the videos and the music was also a crucial point for the selection of the videos. A narrative aspect is evident in the videos with both containing a turning point and a climax that is expected to trigger a surprise reaction as well as emotional intensity. In the first video, the narrative content is related to the gradual increase of a wave and a climax was expected to be experienced when the wave suddenly breaks. Uncertainty related to whether or not the surfer will be able to remain on the board as the wave breaks, is also likely to result in an experience of suspense. In the “Christian the lion video”, narrative elements are connected to how the lion will react when seeing the former owners, and suspense, related to the outcome of the meeting, was also predicted to be experienced for this video. A turning point/climax was expected to be experienced when the lion embraces the owners.

3.3.2.1 Video 1

The “surfing video” starts by depicting Mike Parson surfing on what gradually turns in to a large wave. Due to the camera angle, the size of the wave remains unclear, but as the camera distance itself from Parson and the wave increases in volume, you gradually get an impression of the full size of the wave. As Parson is surfing, the wave starts to break, and we see it

folding over the surfer, expecting it to hit him. After being hidden under the wave for a short amount of time, Parson appears standing on his board and continues surfing while the remainder of the wave gradually disappears.

This video was expected to induce awe in the participants. As mentioned, vastness and accommodation are believed to be the two main attributes to this emotion (Keltner and Haidt, 2003: 303), and they were both considered to be present in the "surfing video" depicting Mike Parson. As vastness refers to anything of a great size, the volume of the wave is expected to result in a perception of vastness. Vastness of a social size might also be experienced in the form of admiration for the courage and the expertise the surfer expresses through his actions. For accommodation to occur, an event or stimuli must be of an unexpected or challenging nature that creates a need for adjusting existing mental schemas. The breaking of the vast wave, and the surfer's ability to remain on the board can be characterized as an extraordinary event predicted to lead to surprise and a need for accommodation. Several of the themes expected to flavor the experience of awe were also expected to be perceived in the video. Elements of threat are present due to the danger involved in surfing on the wave and the surfer also shows great skills by being able to remain on the board in the encounter with the wave. Beauty can also be ascribed to the large waved due to its shape and grand size. A tendency to experience awe in the nature that was found in a study from 2007 might also support the hypothesis that the "surfing video" might evoke awe in participants (Shiota et al., 2007). This video was not expected to be familiar to most of the participants, but it is possible that some could have seen it before.

3.3.2.2 Video 2

The "Christian the lion video" begins with the lions' former owners observing him as he comes walking down a hill. As the lion approaches the men, they start to smile, and we see shifting images between the two men and Christian. After walking slowly, the lion suddenly increases his speed and embraces them with a big hug. The rest of the video depicts the owners and the lion hugging each other. The backstory to this encounter is that the two former owners named John Rendall and Anthony Burke had bought the lion from a zoo when he was very young. After taking care of him for a year, the owners decided to release Christian into the wild nature in Kenya. This video shows the reunion between the owners and Christian a year later and there was a lot of uncertainty as to whether or not the lion would recognize his

former owner, but as the video shows he clearly did. (A lion called Christian The whole Documentary, 2020).

Whereas the video of Mike Parson was predicted to evoke awe, this video was expected to move participants. As mentioned earlier, being moved is a mixed emotion of a social nature and can amongst other things be triggered by a reunion (Menninghaus et al., 2015: 5). The theme of the video is undoubtably the reunion between Christian and his former owners, and although the backstory was not told to participants, the hugging clearly indicates the there was a strong social bond between the owners and the lion. Emotions were also expected to be induced by the mechanism of emotional contagion. The video shows many close-up shots of the owners smiling and laughing and these facial expressions were hypothesized to elicit positive emotions in participants through mimicry. Because this video has millions of views on YouTube and was a part of a documentary, it was expected to be the stimulus recognized the most by participants in the current study.

3.3.2.3 Limitations in the Selection of the Videos

In the choice of videos, the presence of strong emotions was (amongst other things) a governing factor. Videos conveying mixed and complex emotions were therefore chosen, due to their expected ability to evoke strong emotions (Juslin, 2019: 491; Gabrielsson 2010) as well as physiological responses such as chills (Konecni, 2005; Wassiliwizky et al., 2015; Schurtz et al., 2012; Schubert et al., 2018). Basic emotions were avoided in the video selection due to an expected weaker ability to induce strong emotions and psychological responses compared to more complex emotions. This choice was not unproblematic, and although the chosen videos might be more suitable in terms of felt emotions, they were probably not the most adequate choice regarding perceived emotions. The two-dimensional model, that an organization of perceived emotion can be based on, is more applicable in the measurement of basic emotions, due to a clearer preference both in relation to valence as well as arousal. As mentioned, both awe and being moved are emotions that are harder to map onto this scale than basic emotions due to the conflicting nature of these emotions.

Due to the presence of both positivity and negativity in complex emotions, the valence dimension may be victim to somewhat ambiguous ratings. Although this might be the case with the selected videos, they were still expected to be rated differently on this dimension. Being moved is an emotion that contains both sadness and happiness, but the ratio between these two emotions depends on the triggering event. A reunion is in that sense an event likely

to contain more positivity than negativity. Despite the fact that many reunions contain elements of absence that are often expressed through tears, the reunion between Christian and the former owner's reunion is more of a positive kind, expressed through smiles and laughter in the former owners. This video was therefore likely to be rated on the positive side of the valence dimension.

Because of the dramatic content of the "surfing video" and the possibility of death or serious accident in the encounter with the wave, the experience was expected to be flavored more by fear than by beauty and the video was therefore hypothesized to be given a significantly lower valence rating than the Christian the lion video. The videos might therefore have had a significant impact on the perceived valence of the music when the two were paired.

In relation to the arousal dimension, it is unclear whether a significant difference between the two videos might be found. The "surfing video" was hypothesized to be rated on the higher end of this dimension, due to the intensity present in awe, but to what extent this would be the case was unclear. The "Christian the lion video" was also expected to be given a rating on the higher end of the scale particularly due to the activity in the video, but there was also some uncertainty connected to the ratings on this dimension. In both cases the videos might be given a rating on the higher end of the arousal spectrum, but it was unclear whether or not this would be significantly different from the ratings of the music.

3.3.3 Combining Music with Videos

All the videos were edited and combined with music in iMovie. Sounds coming from the natural environment (diegetic sounds) of the videos were erased and they were shown with no sound when presented alone. This was done to fully isolate the moving image variable from the music variable and prevent any unwanted interactions from occurring. Although necessary, this choice may have had some negative impacts on the experience of the videos when presented alone, as watching events occur without the accompaniment of sounds is a quite unusual experience in our day and age. The lack of sound from the video may also have reduced the perception of congruency, as the two stimuli became less integrated.

The selection of videos and music was to a large extent governed by structural similarities between the two. Given that congruency between visuals and music is important for sustaining the unity assumption, which is further believed to be a prerequisite for cross-modal interactions (Chen and Spence, 2017), it was important to have videos and music that were

matched with each other. Particularly formal congruency related to the spatial order, was prioritized when combining music with videos, but semantic congruency related to emotions was also expected to be perceived in most of the combinations. In addition to helping sustain the unity assumption, a matching may possibly also have made narrativization more likely to occur due to a correspondence in the structure of the two. External narratives might in addition to cross-modal interactions be a source of emotional influence from the moving image on the music.

In the editing process, the expected peaks in the music excerpts were synchronized with the predicted climaxes in the two videos. In the first musical excerpt (ambient track), a climax was assumed to be experienced when a high frequent tone enters the musical texture and helps to increase the dynamic level. In the second excerpt (string quartet) the climax was expected to be perceived when a new heavier and louder motif appears. In the last excerpt (pop tune) the climax was also expected to be the result of a new more intense musical part. The climaxes of the music excerpts were edited to co-occur with the breaking of the wave in the first video and the lion hugging of the men in the second. Due to a pairing of all the music extracts with the videos, a perfect synchronization was not possible to achieve, and in some cases the climax in the video appears slightly before the music reaches its peak.

In the combination of the extract from *Company* by Philip Glass and the "Christian the lion video", congruency between visuals and music was not expected to be experienced by participants. This was due to a lack of both semantic and formal congruency in this combination. Whereas the video with the lion was hypothesized to evoke a feeling of being moved, the "string quartet" was not expected to induce these emotions due to its dramatic character. The dramatic content of the music may also have manipulated participants into expecting the encounter to be more violent in the beginning. A confusion may then have occurred when this prediction was debunked. Also, the two stimuli were not structurally matched with each other, and the embrace in the video happens before the new and heavier section in the "string quartet" appears. This pairing was done consciously to see if incongruence between the two stimuli could have a negative effect on the experience of the music during both the first and the second presentation.

3.4 The Internet Survey

The experiment was conducted through an online survey from the Qualtrics platform. The test was estimated to take approximately 15 minutes to complete, and participants had the opportunity of completing the survey in either Norwegian or in English. Although the majority of the participants were expected to be from Norway, the survey was translated to English so that people outside of Norway had the opportunity to participate in the study. People who completed the survey were given the opportunity to take part in a drawing of two movie tickets.

The reason why an internet survey was chosen over a lab experiment was that it made recruitment of participants less demanding and more flexible, in addition to being less time consuming, with regard to the planning as well as the execution of the experiment. An experimenter was for instance not needed and issues related to *experimenter bias* were eliminated (Cozby and Bates, 2015: 192).

Despite being the most suitable option, this choice posted some disadvantages. The environment where the survey was conducted would not be something that could be controlled. Although participants were able to complete the survey in a comfortable environment of their own choosing, there was also the possibility that the survey would be conducted in a noisy environment with distractions or disturbances. This could further have had a negative influence on the mood of the participants. To reduce this unwanted influence from surroundings, participants were asked to use headphones throughout the survey (headphones were also requested for better sound). Participants were also asked to do the test on a computer. This request was primarily due to the survey not being optimized for mobile phones, but it also reduced the probability of participants completing the survey in a noisy environment, as for instance on a train or a bus. Although these requests may have minimized the influence from unwanted variables, these instructions were not followed by all participants. One participant commented that bad sound from the speakers of the computer probably had a negative influence on the emotional involvement with the music, while another participant reported some issues which were related to the use of a mobile phone instead of a computer. It is also plausible that other participants did not follow all the instructions.

3.4.1 Consent Form

All instructions, in addition to information about the study and the procedure of it, were given in a consent form that was shown to participants in the opening of the survey. Participants were not given any information about the aim of the study, but they were informed about the general theme of the study, being an examination of the relationship between film and music. It was assumed that very few if any participants would figure out the aim of the study, due to it being a rather unconventional investigation of the relationship between music and the moving image. Problems related to demand characteristics were therefore not regarded as something of serious matter, and the use of deception was not believed to be necessary (Cozby and Bates, 2015: 190-91). The fact that an inverse examination of this relationship (the influence from music on the moving image) is the most conventional approach, makes it likely that some participants believed this to be the case also with this experiment. After completing the survey, two of the participants commented that they found it to be interesting how music could have an impact on the moving image, supporting this assumption.

3.4.2 Questions

After being presented with a stimulus, either video alone, music alone, or music and video together, participants were set to answer questions related to the specific stimulus. For the majority of the questions, a slider was applied. Whereas for instance multiple choice questions can only offer categorical data, a slider provides numeric data that is easier to statistically analyze, and it provides more nuanced answers.

Free descriptions were on the whole omitted due to the high amount of information that comes from such reports. These types of answers would also produce a lot of information that would be quite difficult to structure and analyze. For each stimulus, one question nevertheless gave an opening for freer description. This question was a yes/no multiple chose question related to the experience of physical responses. Participants who answered yes were requested to a write down their response in a blank text box below. Despite not being written explicitly in the instructions, this question gave room for freer description, but these replies were not expected to be very long.

Questions asked after each stimulus:

Did you enjoy the music?

Not at all — Very much

Did the music fit together with the film?

Not at all — Very well

Did you experience chills/goosebumps?

Not at all — Very much

Did you experience any other form of bodily reaction (warmth in chest, increase in heart rate, lump in throat, tears etc.)?

- ◯ Yes
- ◯ No

If yes: Please describe your reaction.

Did you experience a feeling of awe (mixture of admiration, surprise and fear)?

Not at all — Very much

To what extent were you moved/touched?

Not at all — To a large extent

How would you describe the mood of the music?

Negative — Neutral — Positive

How would you describe the energy of the music?

Very low (calm) — Neutral — Very high (energetic)

Almost all of the questions above were presented after each stimulus. The question related to how well the music and film fitted together was of course only present when music was presented in combination with visuals and the enjoyment question was not asked after the video alone stimulus. This was mainly because enjoyment for the dependent variable (music) was of interest, but the lack of sound was also believed to have a stronger negative impact on ratings of enjoyment than other variables, such as induced and perceived emotions. The questions were identical in each condition.

3.4.2.1 Enjoyment

The question related to enjoyment was asked since enjoyment seems to be associated with emotional engagement, in particular states like happiness, tenderness and sadness (Thompson, 2006; Vuoskoski et al., 2011). A positive influence on induced emotions from the moving image might therefore result in an increase in enjoyment. This question was also asked due to enjoyment being a quite intuitive term not expected to be very challenging to answer precisely. By incorporating this question, the investigation was in addition to induced emotions extended to involving preference, which enjoyment is associated with. It is also possible that the moving image might help to increase the level of attention and involvement given to the music, which in turn might result in the music being enjoyed more. In the studies conducted by Geringer and colleagues, video accompaniment to music did result in higher ratings for both involvement and liking, as well as increased attention although with quite small effects, (Geringer et al., 1996; Geringer et al., 1997).

It is important to address the fact that although similar, some distinction can be made between preference/liking and enjoyment. While the former term is concerned with «making a decision about the stimulus as a whole» (Brattico and Pearce, 2013: 54), enjoyment is on the other hand more associated with pleasure. The two are nevertheless strongly connected (Vuoskoski and Eerola, 2017) and the intensity of enjoyment often lays the foundation for preference (Brattico and Pearce, 2013: 54). With the phrasing «Did you enjoy the music», the question was, due to this strong connection, as well as the focus on one specific listening experience, believed to be interpreted in a similar manner as an optional phrasing where the term enjoyment was substituted with liking.

3.4.2.2 Congruency

One of the questions was also devoted to the experience of congruence/incongruence between music and videos. Although this question was only shown in the video and music combinations, it was important in relation to theories on cross-modal perception. If a positive influence from the moving image on music were to be found, then it would be interesting to see if this correlated with congruence. To make the question understandable for every participant, the term congruency was avoided, and participants were asked if they thought the video and music fitted with each other. One possible weakness in this question is that it might be too open in terms of what is meant specifically. It does not give any insight into whether or not the level of congruency is of a formal or a semantic nature. It is therefore impossible to know if an experienced incongruence would be due to a lack of temporal order, or rather due to a mismatch in terms of associations connected to the paired stimuli. While for instance the combination of *Company* was edited to be synchronous with the "surfing video", it may still have been experienced as incongruent by participants, due to classical music traditionally being associated with high culture and surfing with popular culture as well as musical styles like pop and rock.

3.4.2.3 Felt Emotions and Physiological Responses

The occurrence of felt emotions after exposure to each stimulus was measured through questions related to the experience of awe and being moved, as well as physiological and other forms of intuitive responses. Since the two videos were hypothesized to evoke either awe (Mike Parson surfing on large wave) or being moved (Christian the lion), one question for each emotion was asked. A more nuanced mapping of induced emotion could have been possible with the inclusion of questions related to other emotions, as for instance sadness, joy, and anger, but the test would then have been too time-consuming, and this could have reduced the number of responses obtained.

Because the existing literature suggests that being moved and awe are predictors of the occurrence of chills (Konecni, 2005; Schurtz et al., 2012; Wassiliwizky et al., 2015; Schubert et al., 2018), a question related to this phenomenon was also asked. The presence of chills is a good indication of autonomic states of emotional involvement, but it is at the same time worth noting that the occurrence of chills is largely dependent on certain personality traits. Some people also never seem to experience these sensations regardless of emotional involvement (Goldstein, 1980; Sloboda, 1991; McCrae, 2007; Nusbaum and Silvia, 2011).

The question related to other physiological responses/bodily reactions was asked because both awe and being moved are known for being accompanied by other response than just chills (Braud, 2001; Huron, 2006; Menninghaus et al., 2015; Schubert et al., 2018). The "surfing video" may in the same manner result in tears and gasps in participants, since these responses have been suggested to accompany awe. Due to the overwhelming nature of this emotion, as well as elements of surprise and threat clearly present in the video, an increase in heart rate was also predicted to occur. This response is a direct result of activation of the sympathetic nervous system and associated with both fear responses and arousal (Ekman et al., 1983; Juslin, 2019: 25).

The "Christian the lion video" was hypothesized to elicit a warm sensations and tears, as these responses are associated with being moved. Smiles from the former owners of Christian might also be imitated by participants through the mechanism of emotional contagion. These assumptions are of course based on the predicted ability of the video to evoke feelings of being moved. In retrospect this question could have been phrased differently to better include a wider range of responses than those coming from the autonomic nervous system, for instance responses such as smiling. Participants nevertheless tended to include them in their responses.

One important reason for including both chills and other responses in the questionnaire is that they were expected to be reported more accurately. Whereas the occurrence of emotional states such as awe and being moved might be harder to report accurately, physiological responses such as chills are more based on specific sensations that suddenly appear and may in some cases also be visible (goosebumps and tears). Although more accurate, the reliability of the reports to this question can also be discussed. One participant for instance reported generally feeling quite cold and was therefore not sure if she experienced chills or not.

3.4.2.4 Perceived Emotions

Perceived emotions were measured through two question based on a two-dimensional model of emotions. One question was related to the valence dimension while another was connected to the arousal dimension. The term valence was substituted with the more commonly known term mood, and although this term is not equivalent to valence, it was expected to be understood correctly with the inclusion of negativity and positivity at each end of the scale. Arousal was substituted with energy, since it to a lesser degree was expected to be associated with felt emotions than arousal. Given that people have a tendency to mix perceived emotions

with felt (Juslin, 2019: 212-123), it was extra important to avoid confusion and make it clear what the purpose of these questions were. Participants were in that regard asked how they would describe the mood (valence) and the energy of the music, making sure the focus was on the music and not the listener.

An alternative approach to the two-dimensional model in which emotions are treated as distinct categories was not used for two main reasons. First, a two-dimensional model would offer more nuanced ratings of emotional reactions that are easier to statistically analyze. Second, questions based on a categorical approach would demand many more questions than the two used to assess variations in perceived emotions. One disadvantage with choosing a dimensional approach is that it offers little insight into the specific emotions recognized. It is also unclear whether or not it is possible to distinguish emotions with similar ratings in both dimensions (Brattico and Pearce, 2013). Fear and anger are for instance two emotions quite different in character that are both rated on the low end of the valence dimension and the high end of the arousal dimension, albeit with different location on each dimension. Participants were not likely to be familiar with an organization of emotions in terms of these two dimensions, and it is therefore quite challenging, if possible, to draw conclusions on what specific emotions were perceived, based on ratings of valence and arousal.

3.4.3 Demographics and Comments

After the presentation of all the stimuli, participants were set to answer some questions related to their age, gender, nationality, musical background, as well as familiarity with the stimuli. The last question might be relevant in terms of expected reactions to the stimuli. Familiarity with the videos may have had a big influence on the results, since elements of surprise, which is clearly present in both, may be weaker when the stimuli is familiar. Participants were also asked to report if they experienced any problem with playing any of the excerpts. One participant reported that one music excerpt stopped several times throughout the playback, which may have had a negative influence on the experience of the excerpt. Two other participants reported that music files took a while to upload, but this is not expected to have had much of an impact on the result. Participants were finally given the opportunity to give general feedback on the study.

3.5 Participants

Participants were mainly recruited through social media (Facebook and Twitter) and a link to the survey was shared on these platforms. A total number of 68 participants completed the internet survey, but 2 participants were removed due to blank replies to several of the questions. Of the number of participants that were included in the analysis (N = 66), 33 were females, 32 were males, while one participant chose the category "other". The mean age of the population was 30,80 (SD = 11,56) and participants of different ages were represented in the survey, with the youngest participant being 20 years old and the oldest being 67 years. The study had participants from numerous countries, but the majority of the population came from Norway (72,7% N = 48). The other nationalities that were represented in this study were Denmark (10,6% N = 7), Sweden (4,5% N = 3), United Kingdom (4,5% N = 3), France (1,5% N = 1), America (1,5% N = 1), Mexico (1,5% N = 1), New Zealand (1,5% N = 1), and Japan (1,5% N = 1).

Regarding musical background, 21 participants described themselves as musicians (31,8%), 16 participants reported playing an instrument at a competent level (24,2%), 14 participants reported playing an instrument at an amateur level (21,2%), 14 participants chose the category «I like music a lot» (21,2%), while one participant chose the category «I think music is okay» (1,5%).

22 of the participants had not seen or heard any of the videos or music excepts, while 32 participants were familiar with the Christian the lion clip, and 5 participants with the video of Mike Parson surfing on a large wave. 9 participants had heard the 3. Movement of the String quartet *Company* by Philip Glass, and 5 participants reported being familiar with the slowed down version of *Music for 18. Musicians* by Steve Reich. No participants were familiar with *Dream* by Ian Post. In addition, 5 participants reported that they might be familiar with some of the stimuli.

21 participants were assigned to the first condition, 20 to the second condition, and 25 to the third condition. 31 one of the participants were shown the first video (surfing video) alone while 35 participants were exposed to the second video (Christian the lion).

3.6 Data Analysis

Data from all the participants that completed the survey was exported from Qualtrics and imported to Microsoft Excel. Two participants did not complete the survey in the proper manner, but all questions were answered, and data was therefore also imported to excel. Raw data was systematically organized and further imported into different statistical programs. For blank columns in the numeric data, a rating of 50 was applied, rather than an alternative approach where mean ratings for each question was applied. The default position of the slider was 50, and for an answer to be registered, participants needed to click on the slider. It is therefore highly likely that participants who did not touch the slider still intended it to be a rating of 50. Whether or not an empty column meant a rating of 50, or no rating at all, was however not something that it could be certain of. A threshold of answers to at least 80% of the questions was therefore created, which further resulted in two participants being excluded from the data analysis (the same participants as mentioned earlier in the *Participants* section).

Slider questions were imported to RStudio for analysis. A one-way ANOVA model was applied to the analysis of the music excerpts, due to it being suitable for experiments that contains more than two conditions and have an independent group design. T-test comparison were conducted on data from for questions related to congruency (how well music and video fitted together) and the videos presented alone. A two-way ANOVA analysis was, in addition to the one-way, conducted on the numeric data to assess if a possible interaction effect from the video alone on the music conditions was present.

Mean ratings for slider questions related to video alone, music alone, and music and video stimulus were imported to SPSS for a Linear Regression analysis. This analysis was applied to assess how much the music and the video variables each contributed to the ratings of the music and video together.

No statistical analysis was conducted on the open question related to physiological response, since the number of responses from each condition was much lower than for the slider ratings. Responses were nevertheless systematically mapped and responses from different stimulus categories compared with each other.

4 Results

4.1 Slider questions

To test for the possible effects from the videos on music, a one-way ANOVA was run for each musical excerpt, comparing differences between stimulus categories (music alone, music and video 1, and music and video 2). For the significant findings, a post hoc analysis, more specifically the Tukey HSD (Honesty Significant difference) was applied due to it being good at controlling the Type I error rate, and an adequate choice for populations of an equal size (Field et al., 2012: 431-32). P values for the Tukey HSD were adjusted in the package used in RStudio. Given that questions related to how well music excerpts and videos fitted together (congruency variable) were only asked after music in combination with visuals, the one-way ANOVA model was substituted with t- tests comparisons. For the videos presented alone, t-tests comparisons were conducted on the variables: chills, awe, being moved, valence and arousal, to test for significant difference between the two videos. Mean ratings for every stimulus were also calculated to see if potential effects were positive or negative. Effect sizes were calculated for all the significant results, with the Eta squared being applied to the significant ANOVAs, and the Cohen's d to post hoc and t- tests.

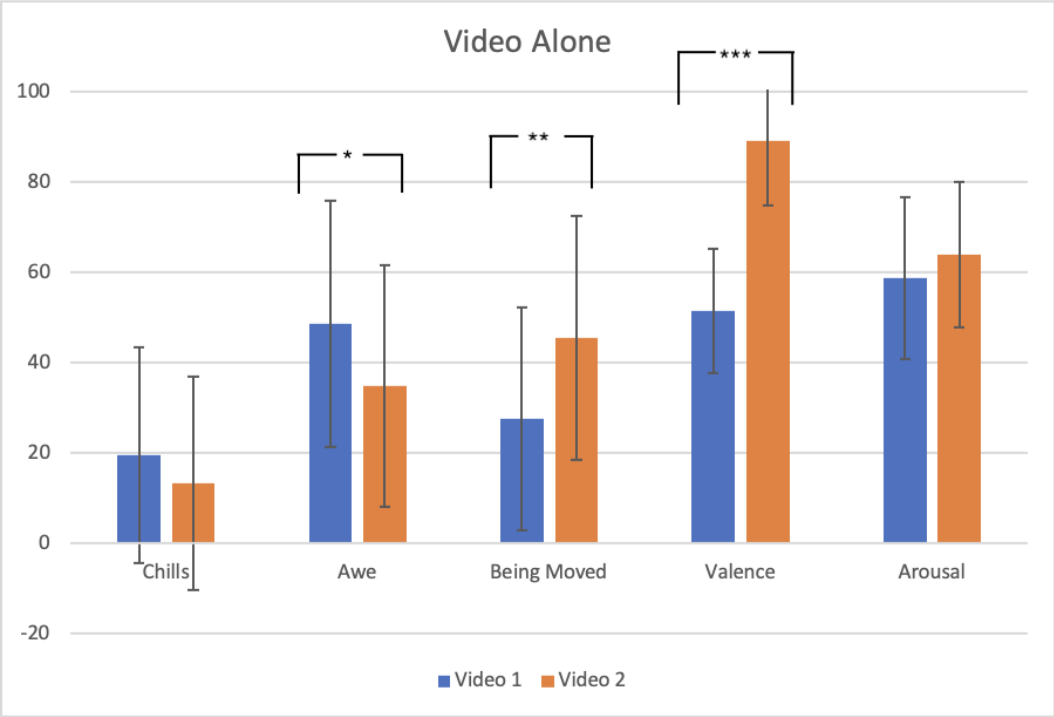
4.1.1 Videos Alone

As predicted, the rating of awe was higher for video 1 (surfing video), ($M = 48.48$, $SD = 27.24$) compared to video 2 (Christian the lion), ($M = 34.86$, $SD = 26.76$), and t-testing revealed a significant difference $t(62.75) = 2.044$, $p = 0.04$, $d = 0.49$ (small effect). The second video had a significantly higher rating for the being moved variable compared to the first one $t(63.95) = 2.821$, $p = 0.006$, $d = 0.65$ (medium effect) (video 1: $M = 27.52$, ($SD = 24.58$), (video 2: $M = 45.46$, $SD = 27.08$). The largest effect was found for the valence variable $t(63.66) = 10.857$, $p < 001$, $d = 1.6$ (large effect) with the second video being given a substantially higher rating than the first one (video 1: $M = 51.48$, $SD = 13.72$), (video 2: $M = 89.14$, $SD = 14.44$). No significant difference was found between the videos for either the chills variable $t(62.89) = 1.059$, $p = 0.29$, or the arousal variable $t(60.17) = 1.266$, $p = 0.21$.

As hypothesized, the "surfing video" had the biggest impact on the experience of awe while feelings of being moved were to a larger extent evoked by the "Christian the lion video" than the "surfing video". A wide agreement between participants was nevertheless not present, since neither of the videos had a mean rating over 50 in relation to the induced emotion, and quite high standard deviations. A general consensus was found in the experience of valence for the Christian the lion video, having the highest mean rating of all the variables and a quite low standard deviation.

	T	P value
Chills	1.059	0.29
Awe	2.044	0.04
Being Moved	2.821	0.006
Valence	10.857	< .001
Arousal	1.266	0.21

T-test for each variable for video alone.



Mean ratings for video 1 and 2 presented alone. The error bars are representing standard deviation. p<.05 = *, p<.01 = **, p<.001 = ***

4.1.2 First Presentation of Music Excerpts

In the analysis of the music excerpts, dependent variables were ratings of enjoyment, chills, awe, being moved, valence, and arousal, while the video accompaniment (no video, video 1, or video 2) was the independent variable. ANOVAs were conducted on all the dependent variables (enjoyment, chills, awe, etc.) for all the music excerpts, assessing the effects of video accompaniment on ratings of them.

4.1.2.1 Enjoyment

For the enjoyment variable, a significant effect was found for the second music excerpt (string quartet), $F(2,63) = 11.57$, $p < 001$, $\eta 2 = 0.27$ and post hoc analysis (Tukey HSD) revealed a negative effect of video 2 (Christian the lion) on the enjoyment of the music. The music in combination with the second video was significantly lower than the music presented alone, $p < .001$, (music and video 2: $M = 39.9$, $SD = 26.51$), (music alone: $M = 73.15$ $SD = 16.9$), and calculations of effect size using the Cohen's d formula revealed this effect to be very large, $d = 1.19$. The second video also had a significantly lower rating than the first video (surfing video) when paired with the second music excerpt $p = 0.02$, $d = 0.71$ (medium effect), (music and video 1: $M = 58.16$, $SD = 21.91$).

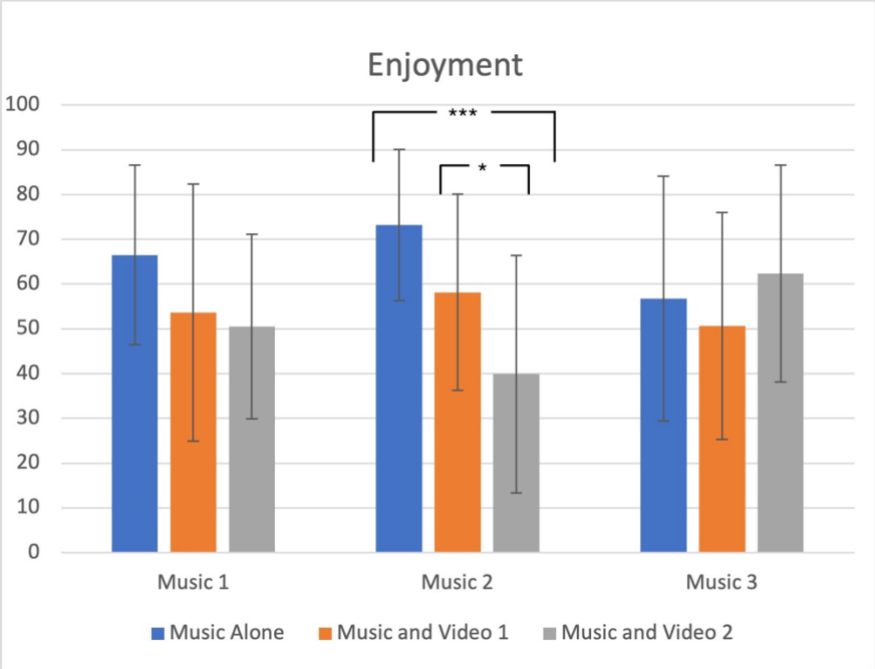
A similar negative effect was also found for music 1 (ambient track) with the music alone stimulus being higher than the experimental stimuli, but this was not of significance, $F(2,63) = 3.05$, $p = 0.05$, (Post- hoc testing: music alone - music and video 2, $p = 0.06$). A general negative trend for the effect of the moving image on music was found on the enjoyment variable of several of the excerpts (se bar chart for a more detailed illustration).

	F value	P value
Music 1	3.052	0.05
Music 2	11.57	<.001
Music 3	1.161	0.32

Table 1: Results from one-way ANOVA.

Music 2	P Value
Video 1 - Video 0	0.07
Video 2 - Video 0	<.001
Video 1 - Video 2	0.02

Table 2: Post hoc analysis (Tukey HSD).



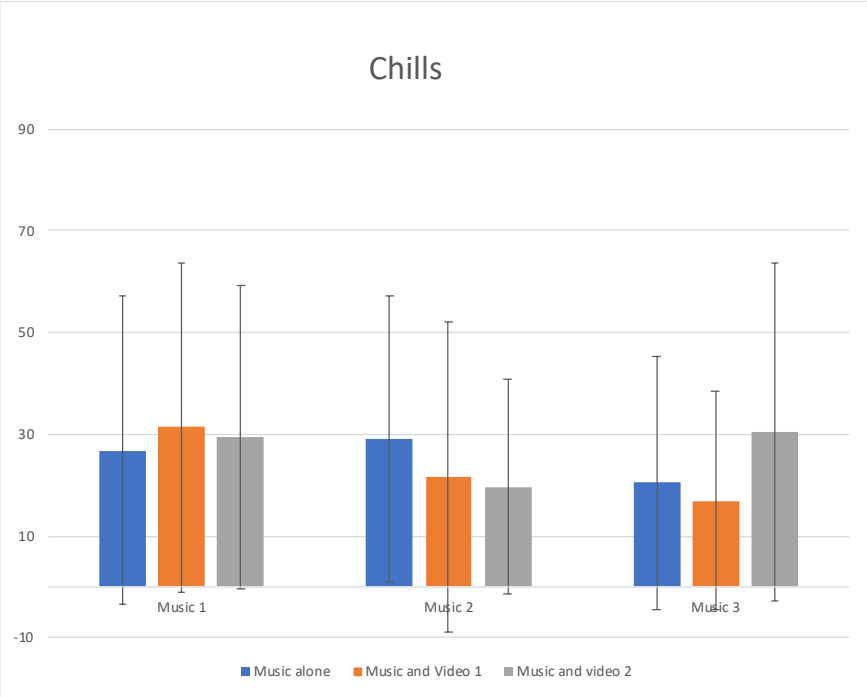
Mean ratings for each stimulus category of all the music excerpts. The error bars are representing standard deviation. p<.05 = *, p<.01 = **, p<.001 = ***

4.1.2.2 Chills

No significant difference was found between stimulus categories in relation to the occurrence of chills. This was the case for all the three music excerpts, and ratings were more or less similar between control and experimental stimuli. The standard deviation was in general very high indicating major variations between participants.

	F value	P value
Music 1	0.124	0.88
Music 2	0.696	0.50
Music 3	1.531	0.22

Table 3: Results from one-way ANOVA.



Mean ratings for each stimulus category of all the music excerpts. The error bars are representing standard deviation. p<.05 = *, p<.01 = **, p<.001 = ***

4.1.2.3 Awe

A significant effect was found for both music 1 (ambient track), $F(2,63) = 4.995$, $p = 0.009$, $\eta 2 = 0.14$, and music 3 (pop tune), $F(2,63) = 4.877$, $p = 0.01$, $\eta 2 = 0.13$, for the experience of awe. A post hoc analysis revealed that for the first music excerpt, the music in combination with the first video (surfing video) had a significantly higher rating compared to the music alone, $p = 0.007$, $d = 0.89$ (large effect), (music and video 1: $M = 53.52$, $SD = 32.25$), (music alone: $M = 26.68$, $SD = 22.17$). For the third music excerpt, a significant positive effect was found for both the first video compared to music alone, $p = 0.01$, $d = 0.87$ (large effect), and the second video compared to the music alone, $p = 0.05$, $d = 0.7$ (medium effect), (music and video 1: ($M = 45.8$, $SD = 30.46$), (music and video 2: $M = 39.96$, $SD = 30.26$), (music alone: $M = 19.76$, $SD = 23.37$). As hypothesized, the "surfing video" had an ability to increase the level of awe when paired with music, but more surprisingly this effect, although smaller, was also present in the "Christian the lion video" when paired with the "pop tune".

	F value	P value
Music 1	4.995	0.009
Music 2	2.194	0.12
Music 3	4.877	0.01

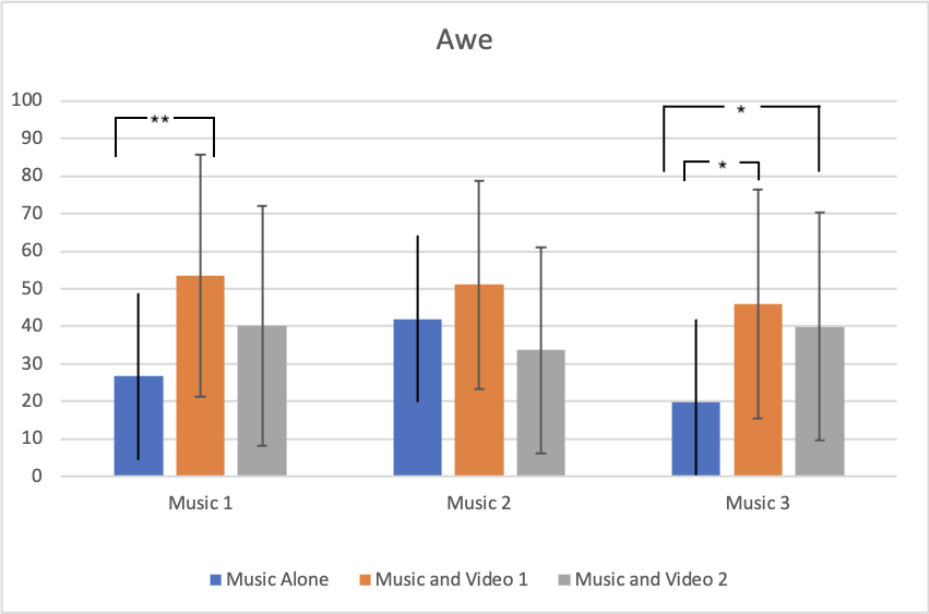
Table 4: Results from one-way ANOVA.

Music 1	P value
Video 1 - Video 0	0.007
Video 2 - Video 0	0.27
Video 1 - Video 2	0.30

Table 5: Post hoc analysis (Tukey HSD).

Music 3	P value
Video 1 - Video 0	0.01
Video 2 - Video 0	0.05
Video 1 - Video 2	0.77

Table 6: Post hoc analysis (Tukey HSD).



Mean ratings for each stimulus category of all the music excerpts. The error bars are representing standard deviation. p<.05 = *, p<.01 = **, p<.001 = ***

4.1.2.4 Being Moved

For the being moved variable, a significant effect was only found for the third music excerpt (pop tune), but this was also the largest effect in the whole study regarding the influence of the moving image on music, $F(2,63) = 15.14$, $p < 001$, $\eta 2 = 0.32$. Post hoc comparisons showed a significant positive effect for the music in combination with the second video (Christian the lion) when this category was compared to both music alone, $p < .001$, $d = 1.19$

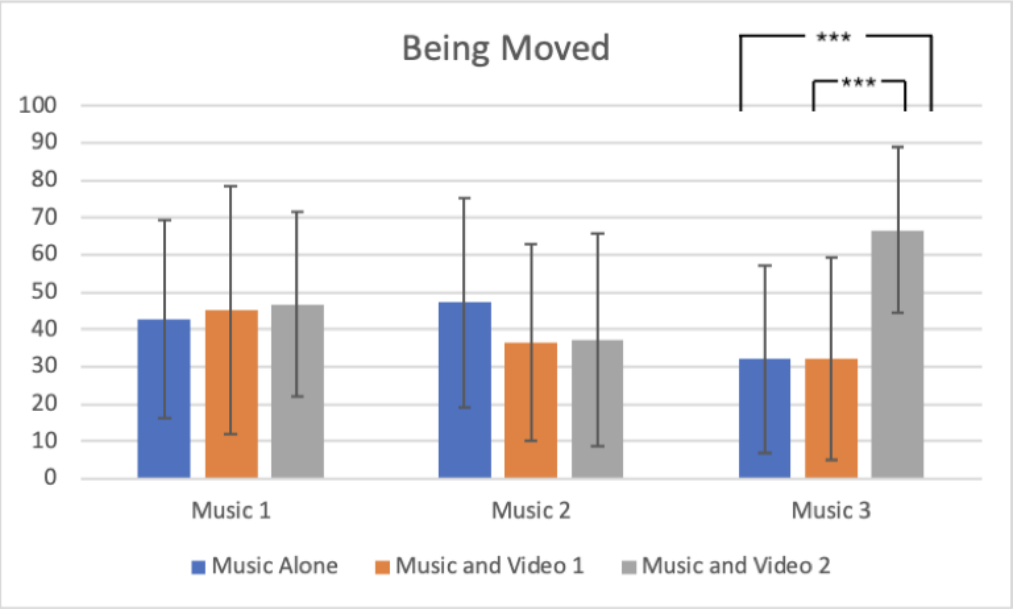
(large effect), as well as music accompanied by video 1, $p < .001$, $d = 1.15$ (large effect). Mean ratings for music alone ($M = 32.04$, $SD = 25.00$) and video 1 ($M = 32.2$, $SD = 27.24$) were more or less the same, while the mean rating for video 2 was over twice as high ($M = 66.64$, $SD = 22.26$), further demonstrating the significant positive effect for the second video on the experience of being moved.

	F value	P value
Music 1	0.115	0.89
Music 2	0.999	0.37
Music 3	15.14	<.001

Table 7: Results from one-way ANOVA.

Music 3	P value
Video 1 - Video 0	0.99
Video 2 - Video 0	<.001
Video 1 - Video 2	<.001

Table 8: Post hoc analysis (Tukey HSD).



Mean ratings for each stimulus category of all the music excerpts. The error bars are representing standard deviation. p<.05 = *, p<.01 = **, p<.001 = ***

4.1.2.5 Valence

For the second music excerpt (string quartet), a significant effect was found for the valence variable, $F(2,63) = 7.159$, $p = 0.001$, $\eta 2 = 0.18$, and further post hoc testing revealed a significant negative effect from the second video (Christian the lion) compared to the music alone, $p = 0.001$, $d = 0.99$ (large effect), (music alone: $M = 51.45$, $SD = 17.45$), (music and video 2: $M = 28.78$, $SD = 22.39$). This negative effect was similar to the one found for the enjoyment variable between music 2 paired with video 2 and music 2 alone.

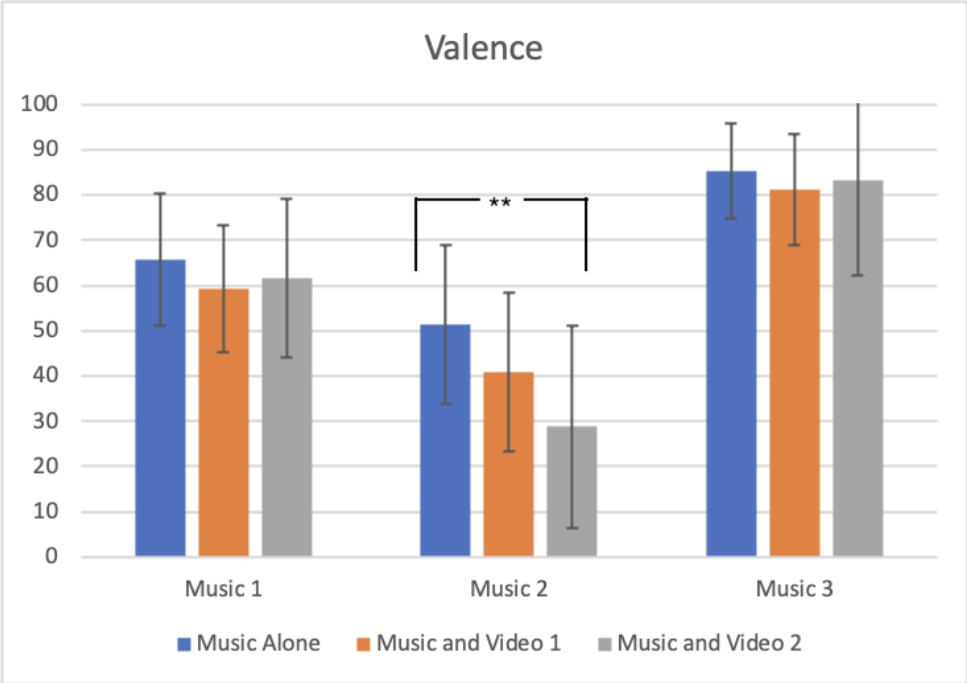
A clear distinction between valence ratings was, regardless of stimuli categories, found between all the music excerpts. The stimulus categories for the third music excerpt (pop tune) had the highest ratings, while the categories for the second except had the lowest ratings (see bar chart). A one-way ANOVA comparing valence rating between music excerpts further revealed this distinction to be of significance, $F(2,195) = 99.71$, $p < .001$, $\eta 2 = 0.91$ (large effect), indicating that the music excerpts were perceived significantly different in terms valence.

	F value	**P value**
Music 1	1.044	0.36
Music 2	7.159	0.001
Music 3	0.349	0.70

Table 9: Results from one-way ANOVA.

Music 2	**P value**
Video 1 - Video 0	0.17
Video 2 - Video 0	0.001
Video 1 - Video 2	0.09

Table 10: Post hoc analysis (Tukey HSD).



Mean ratings for each stimulus category of all the music excerpts. The error bars are representing standard deviation. p<.05 = *, p<.01 = **, p<.001 = ***

4.1.2.6 Arousal

For the arousal variable, a significant effect was found between the categories in the third music excerpt (pop tune), $F(2,63) = 4.836$), $p = 0.01$, $\eta 2 = 0.13$. A difference was not found between the music alone and any of the video pairings, but rather between the first and second video when accompanying the third music excerpt $p = 0.008$, $d = 0.9$ (large effect). Music and video 2 ($M = 65.56$, $SD = 15.07$) had a significantly higher rating than music and video 1 ($M = 47.95$, $SD = 20.34$).

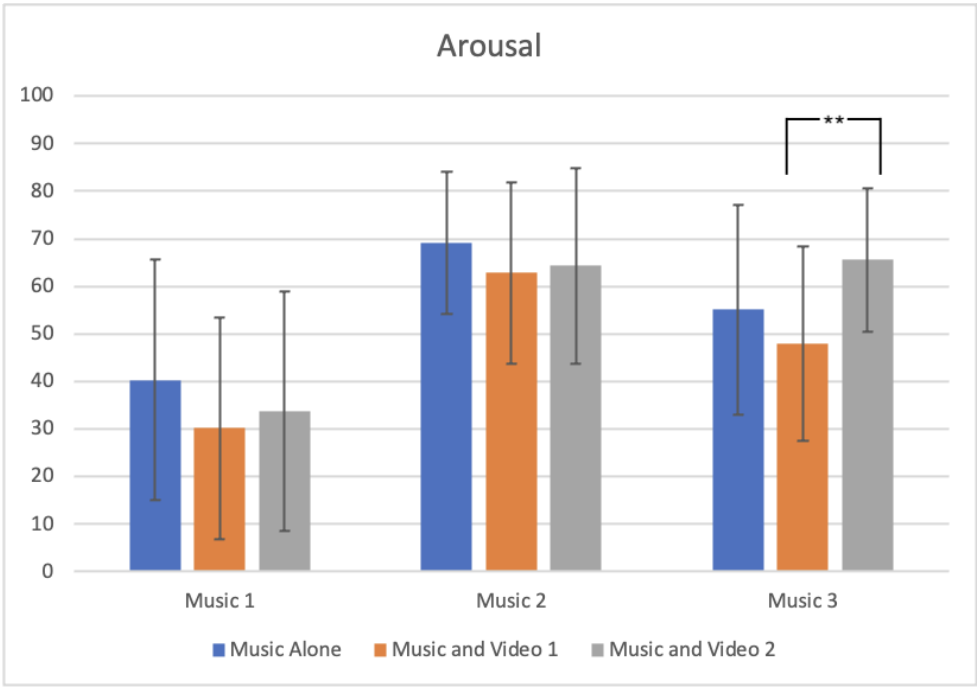
Similarly to the valence variable, all the music excerpt were perceived significantly different in terms of arousal, $F(2,195) = 35.33$, $p < .001$, $\eta 2 = 0.53$. The first music excerpt had a lower rating than the two remaining excerpts regardless of stimulus categories, and the highest rating for arousal was found in the second music excerpt as well as the music and video 2 pairing for the third excerpt.

	F value	P value
Music 1	1.015	0.37
Music 2	0.668	0.52
Music 3	4.836	0.01

Table 10: Results from one-way ANOVA.

Music 3	P value
Video 1 - Video 0	0.45
Video 2 - Video 0	0.16
Video 1 - Video 2	0.008

Table 12: Post hoc analysis (Tukey HSD).



Mean ratings for each stimulus category of all the music excerpts. The error bars are representing standard deviation. p<.05 = *, p<.01 = **, p<.001 = ***

4.1.2.7 Impact of Videos Alone on the Music Excerpts

To assess the possible impact the video alone (presented in the beginning of the experiment) may have had on the experience of the three music excerpts during the first presentation (particularly music excerpts combined with the video previously presented alone), a two-way ANOVA was applied. The independent variables were video alone (video 1, or video 2), and video accompaniment to the music (no video, video 1, video 2). After running tests for all the dependent variables (enjoyment, chills, awe, ect.) on all music excerpt, a significant interaction effect was found for the awe variable in the second music excerpt (string quartet), $F(2,60) = 3.64$, $p = 0.03$, $\eta 2 = 0.1$. Further ANOVA testing revealed that a significant effect between conditions was present in the group of participants that had been presented with the second video (Christian the lion) alone in the beginning of the test, $F(2,32) = 6.093$, $p =$

0.005, η2 = 0.27. Post hoc comparisons showed that the accompaniment of the first video (surfing video) had a significant positive effect on awe compared to the second video, p = 0.004, d = 1.43 (large effect), (music and video 1: M = 62.5, SD = 19.38), (music and video 2: M = 19.38, SD = 27.22). Further t- testing also revealed that the awe ratings for music 2 in combination with video 1 was significantly different between the group of participants that was presented with video 1, and the group that was presented with video 2 in the beginning of the test, t(20.5) = 2.1726, p = 0.04, d = 0.78 (medium effect). The rating of awe was significantly higher for the group that was presented with video 2 alone in the beginning (M = 62.5), compared to the group that saw video 1 alone (M = 40.46). It can therefore be suggested that the familiarity with the first video may have resulted in a decreased experience of awe when the video was repeated in combination with the second music excerpt.

An interaction effect was also found for the energy variable on music 2, F(2,60) = 3.692, p = 0.03, but neither ANOVAs conducted within each group (video 1 alone /video 2 alone) or t-test conducted between each group could find any significant effects. All in all, the presentation of videos in the beginning of the test did not seem to have a big impact on the second experience of the videos in combination with music.

Awe	F value	P value
Music 2 (Video 1 alone)	0.385	0.70
Music 2 (Video 2 alone)	6.093	0.005

Table 13: ANOVA tests conducted separately on participants presented with video 1 and video 2 alone in the beginning of the study.

Music 2 (awe)	P Value
Video 1 - Video 0	0.39
Video 2 - Video 0	0.09
Video 2 - Video 1	0.004

Table 14: Post hoc analysis (Tukey HSD) of participants presented with video 2 alone.

Music 2/ Video 1	T	P value
Music 2 alone - Video 1 alone	2.172	0.04

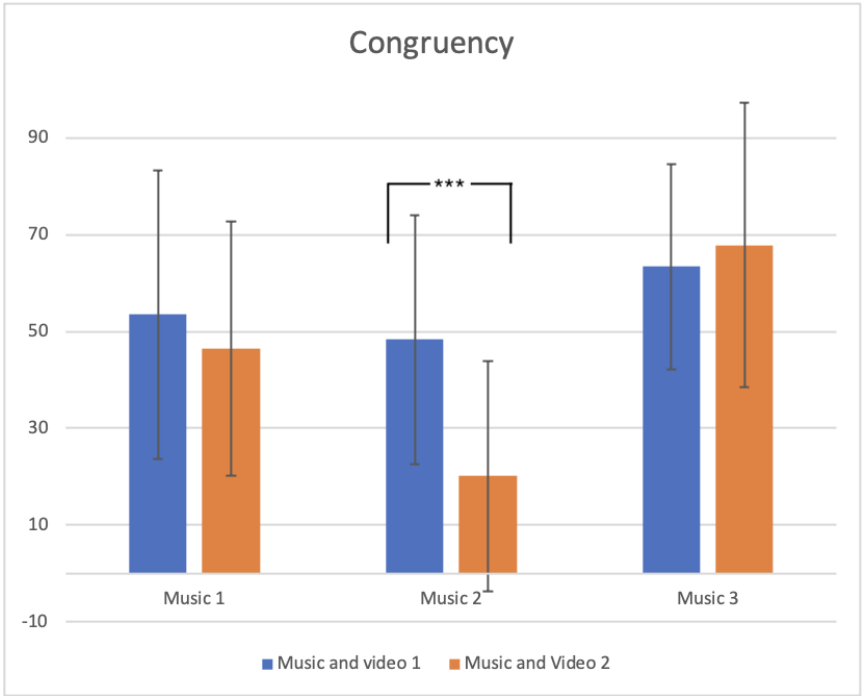
Table 15: T test comparison from the music 2/ video 1 condition, between participants presented with video 1 and video 2 alone.

4.1.2.8 Congruence between Music excerpts and Videos

The main aim with the question related to congruency was to see if ratings of this variable correlated with other potential significant results, but t-tests comparisons were also conducted on each music excerpt to test for potential differences of significance between the two video categories. After running t-tests on all the excerpts, a significant finding was only present for the second music excerpt (string quartet), $t(43.55) = 3.84$, $p < 001$, $d = 0.99$ (large effect). Ratings for the music in combination with the second video (Christian the lion), ($M = 20.19$, $SD = 23.8$) were significantly lower than the combination of the music and the first video (surfing video), ($M = 48.32$, $SD = 25.7$). Music 2 paired with video 2 also had the lowest mean rating of all the excerpts. The negative effect video 2 had on music 2 regarding the enjoyment and valence variables may therefore be explained by a perceived incongruence between music and visuals. Ratings were highest for the third music excerpt (pop tune) when combined with both videos (video 1: $M = 63.4$, $SD = 21.28$), (video 2: $M = 67.88$. $SD = 29.45$).

	T	P value
Music 1	0.793	0.43
Music 2	3.84	<.001
Music 3	- 0.494	0.62

Table 16: T-test scores.



Mean ratings for each experimental stimulus category (video accompaniment) of all the music excerpts. The error bars are representing standard deviation. p<.05 = *, p<.01 = **, p<.001 = ***

4.1.3 Regression Analysis

Since an ANOVA model does not give any information about the contribution of individual variables to the main results (Field et al., 2012: 259), a multiple regression model was applied to assess the contribution of both the video alone and the music alone (predictor variables) on music paired with videos (outcome variable). Mean ratings for the video alone and for the music alone were included as independent variables in the model, while the mean ratings for music and video in combination were included as dependent variables. The independent variables were entered simultaneously (forced entry) as it is considered a more suitable method than for instance stepwise methods (Field et al., 2012: 264). A regression analysis was only conducted on variables in which a significant effect was found in the ANOVA tests and the chills variable was therefore not included. Given that questions related to enjoyment were not asked after the video alone stimuli, this variable was also excluded. Variables left for analysis were: awe, being moved, valence and energy.

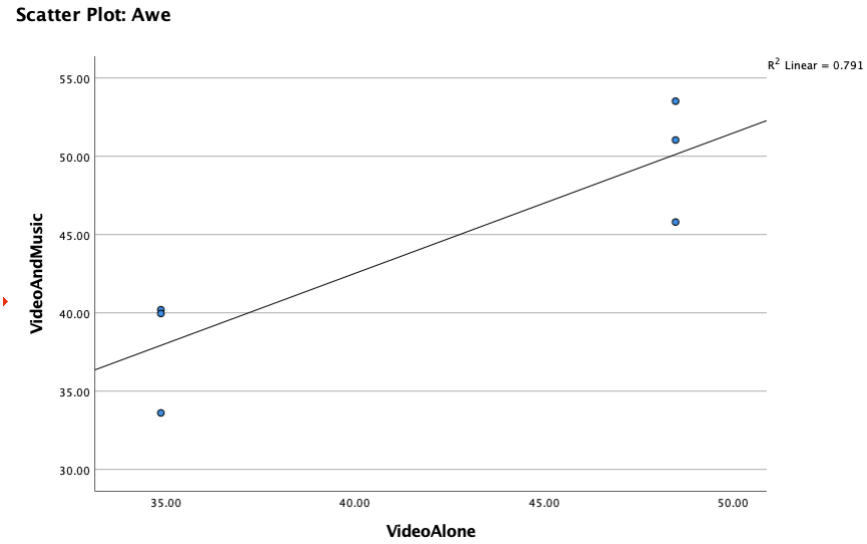
Video alone	Music alone	Video and Music
Video 1	Music 1	Video 1/Music 1
Video 2	Music 1	Video 2/Music 1
Video 1	Music 2	Video 1/Music 2
Video 2	Music 2	Video 2/Music 2
Video 1	Music 3	Video 1/Music 3
Video 2	Music 3	Video 2/Music 3

Predictors and outcome variables: Organization of data for the multiple regression analysis in SPSS. The data was organized in this manner for each variable (awe, being moved, valence and arousal). Each cell consists of the mean rating of a specific variable for a stimulus.

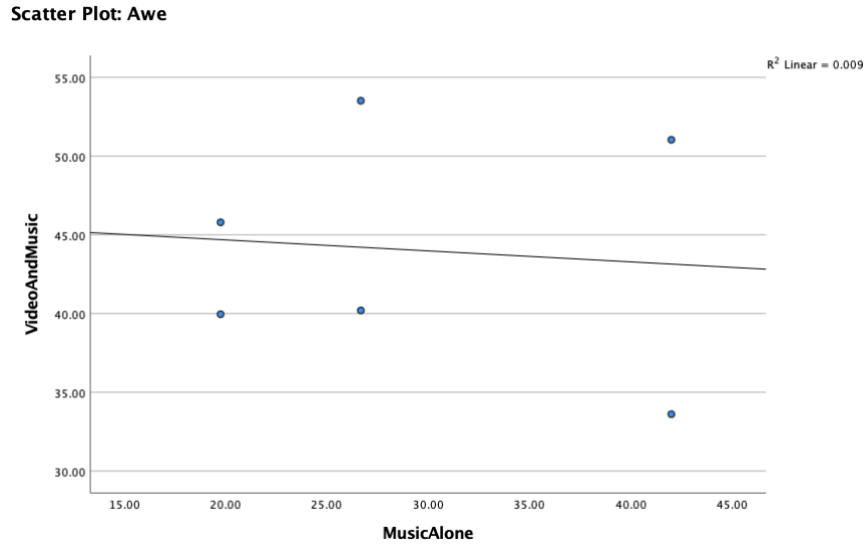
4.1.3.1 Awe

For the analysis of the awe variable, the beta value of the video alone variable was 0.889, $p = 0.04$, while the music alone variable was - 0.095, $p = 0.74$. The significant results for the awe variable were therefore mostly explained by the video variable. The independent variables further explained 80% of the variance for the dependent variable (R square = 0.8), suggesting

that the positive effect of the music and video stimuli to a large extent was explained by the video variable.



Scatter plot for video and music (dependent variable) and video alone (independent variable).

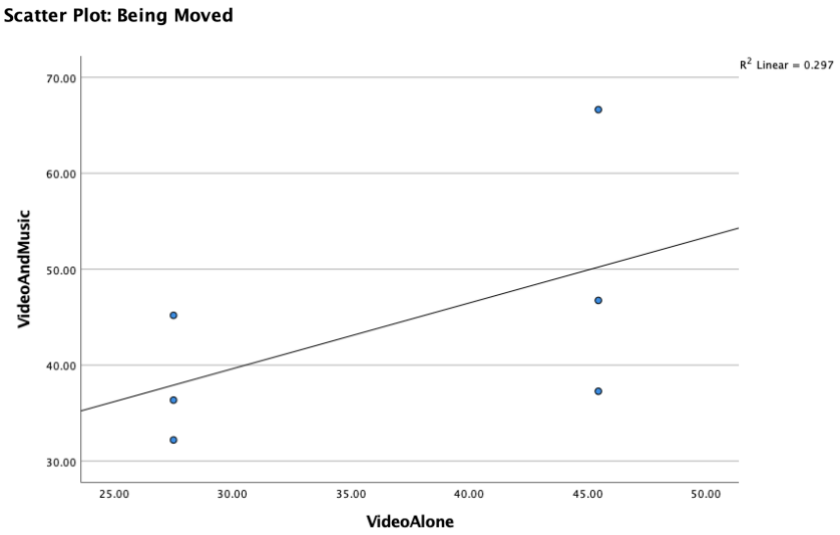


Scatter plot for video and music (dependent variable) and music alone (independent variable).

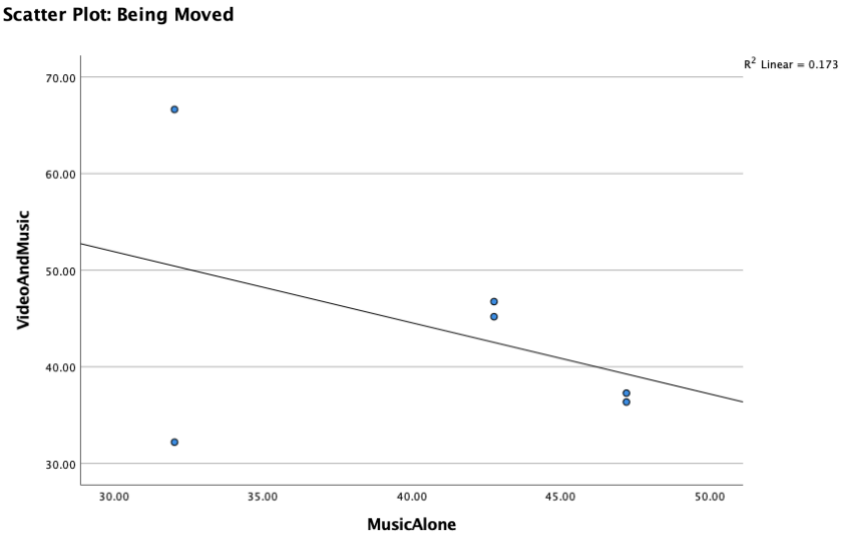
4.1.3.2 Being Moved

The beta value of the video alone variable was 0.545, p = 0.29, while the value of the music alone variable was - 0.416, p = 0.4, for the being moved variable. Although the video variable explained more of the variance than the music alone variable, the result was not significant

and over half of the variance did not seem to be explained by any of the independent variables (R square = 0.47). The positive effect the music and video stimulus had compared the music alone stimulus may therefore be explained by unknown variables. Individual mean ratings reveal that this is particularly evident for the third music excerpt (pop tun) and the second video (Christian the lion). Individually the video had a mean rating of 45.46 and the music a rating of 32.04, but the mean rating for the combination was 66.64 which is significantly higher than the two individual variables.



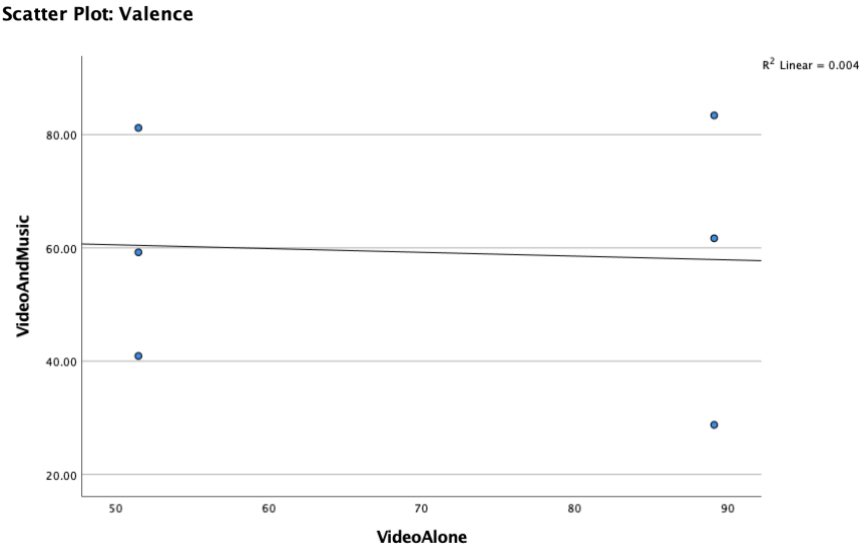
Scatter plot for video and music (dependent variable) and video alone (independent variable).



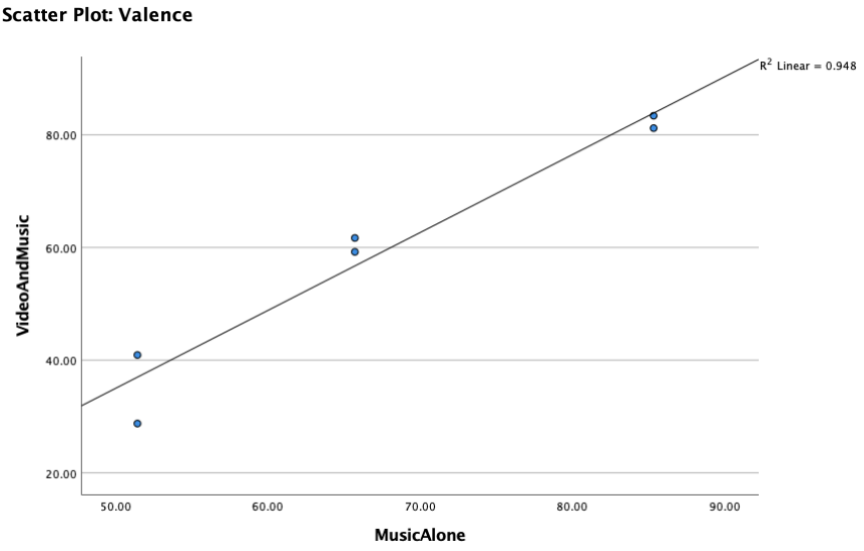
Scatter plot for video and music (dependent variable) and music alone (independent variable).

4.1.3.3 Valence

For the valence variable, the music alone was shown to contribute the most to the music and video ratings with a beta score of 0.974, p = 0.005 for music alone, while the beta value of video alone was -0.063, p = 0.65. The valence variable also had the highest R square = 0.952 which indicates that almost all of the variance in the dependent variable (music and video) can be explained by the music alone. Contrary to the hypothesis, the moving image in general had little impact on the dependent variable.



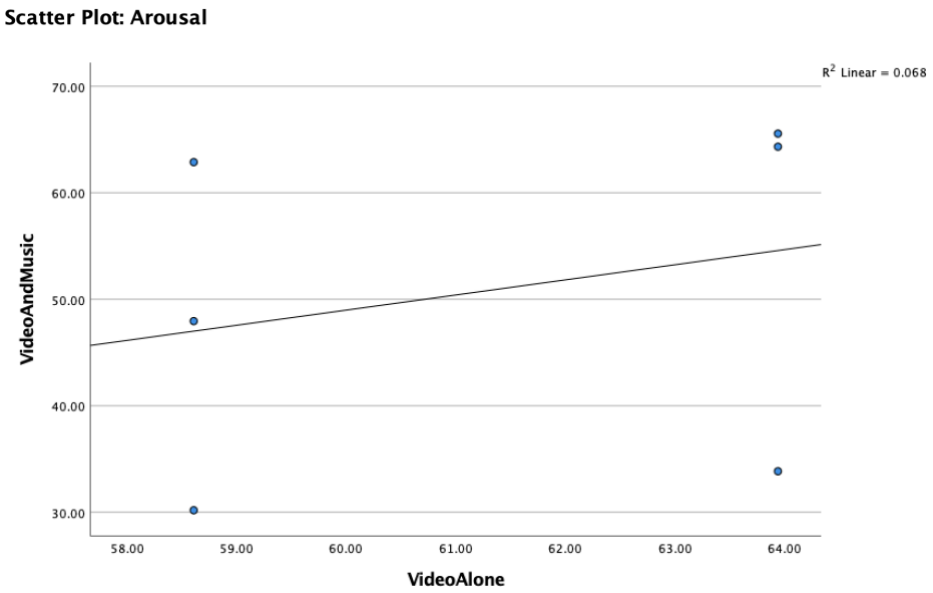
Scatter plot for video and music (dependent variable) and video alone (independent variable).



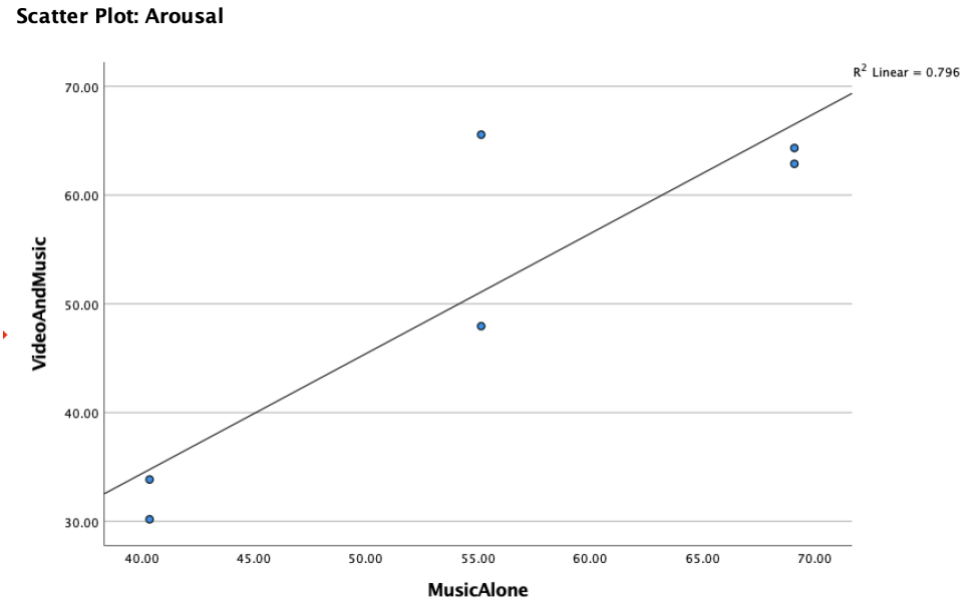
Scatter plot for video and music (dependent variable) and music alone (independent variable).

4.1.3.4 Arousal

As with the valence variable, music contributed more to the rating of music combined with video for the arousal variable. The beta value for music alone was 0.892, p = 0.03, while the beta level for video alone was 0.261, p = 0.3. The model was quite good at explaining the variance in the dependent variable with an R square of 0.864 and the variance in the dependent variable was mostly explained by the music alone variable.



Scatter plot for video and music (dependent variable) and video alone (independent variable).



Scatter plot for video and music (dependent variable) and music alone (independent variable).

4.1.4 Second Presentation of Music Excerpts

Although music excerpts were presented alone in all the conditions under the second presentation, the analysis of the data was conducted in the same manner as under the first presentation. The independent variable was the video accompaniment under the first presentation of the music, and the stimulus categories were: music (previously) presented alone, music (previously) paired with video 1, and music (previously) paired with video 2.

4.1.4.1 Enjoyment

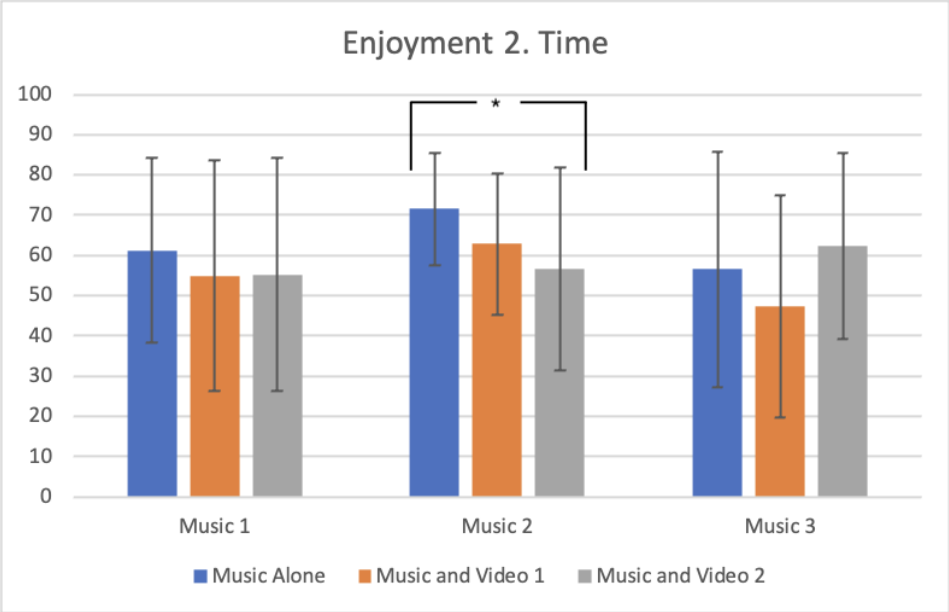
In the repetition of the music excerpts, the lowest p value for the enjoyment variable was found for the second music excerpt (string quartet), $F(2,63) = 3.071$, $p = 0.05$, $\eta 2 = 0.09$. Although the p value was not under 0.05, further post hoc testing revealed that, as in the first presentation, ratings for the music and second video (Christian the lion) were significantly lower than rating for the music alone, $p = 0.04$, $d = 0.69$ (medium effect), (music alone: $M = 71.6$, $SD = 13.94$), (music and video 2: $M = 56.61$, $SD = 25.13$). Although a smaller effect, it therefore seems like a significant carry-over effect from the second video on the second music excerpt was present. As for the first presentation, the first music excerpt (ambient track) was also rated higher when presented alone compared to the two remaining categories, but the difference was not of significance and even lower than in the first presentation. Mean ratings of the music excerpts were, all in all, quite similar to the ratings for the first presentation, but with smaller differences between categories.

	F value	P value
Music 1	0.418	0.66
Music 2	3.071	0.05
Music 3	1.815	0.17

Table 18: Results from one-way ANOVA.

Music 2: Enjoyment	P Value
Video 1 - Video 0	0.30
Video 2 - Video 0	0.04
Video 1 - Video 2	0.52

Table 19: Post hoc analysis (Tukey HSD).



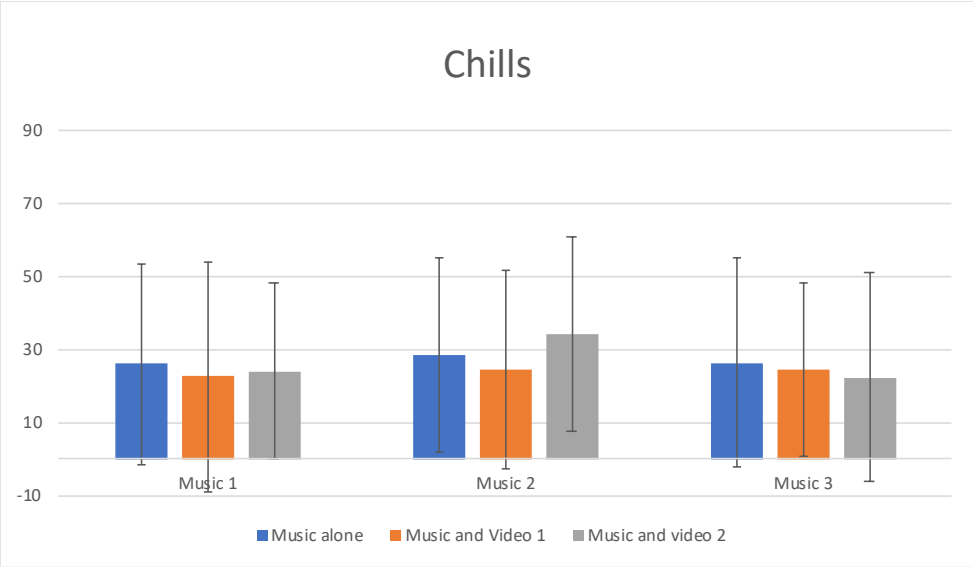
Mean ratings for each stimulus category of all the music excerpts. The error bars are representing standard deviation. p<.05 = *, p<.01 = **, p<.001 = ***

4.1.4.2 Chills

As in the first presentation of the music, no significant findings were found for the chills variable. Ratings were quite similar to those for the first presentation, but ratings for all categories of the first music except, as well as the combination of music 3 and video 2, were lower for the second presentation of the music.

	F value	P value
Music 1	0.086	0.92
Music 2	0.718	0.49
Music 3	0.124	0.88

Table 20: Results from one-way ANOVA.



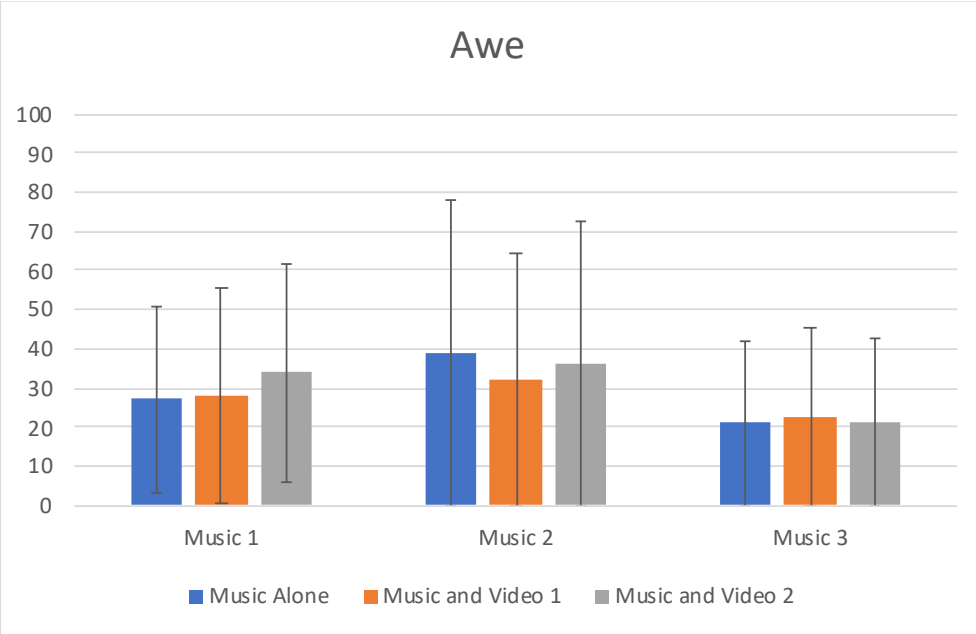
Mean ratings for each stimulus category of all the music excerpts. The error bars are representing standard deviation. p<.05 = *, p<.01 = **, p<.001 = ***

4.1.4.3 Awe

For the awe variable, none of the significant findings from the first presentation were replicated in the second presentation and ratings were more or less the same for each stimulus category of the music excerpts. While the rating of the music alone was more or less the same as for the first presentation, all the positive effects from the video were absent, and rating were more or less similar between categories. It can therefore be concluded that no carry-over effects were found for the awe variable on the music.

	F value	P value
Music 1	0.443	0.65
Music 2	0.346	0.70
Music 3	0.033	0.97

Table 21: Results from one-way ANOVA.



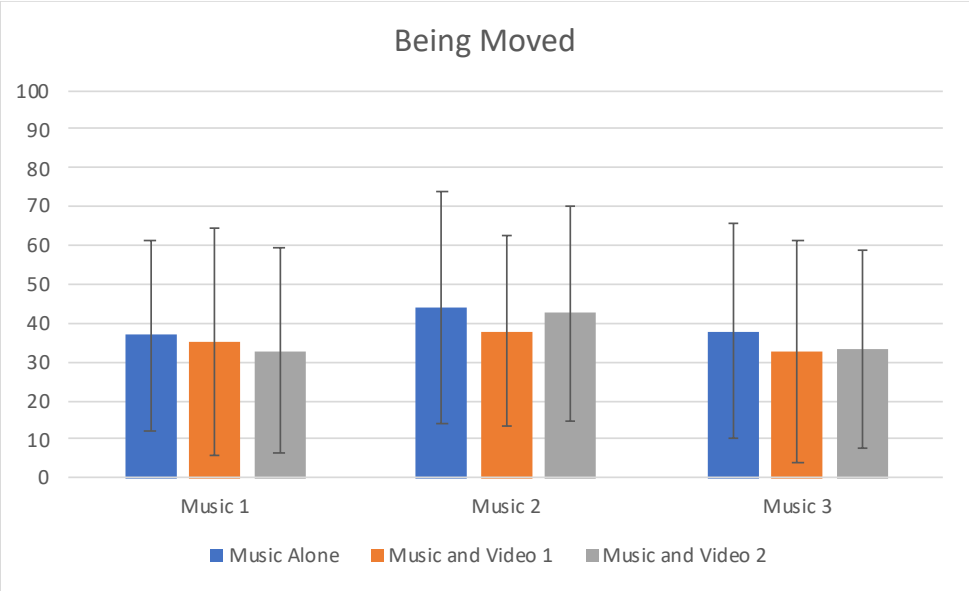
Mean ratings for each stimulus category of all the music excerpts. The error bars are representing standard deviation. p<.05 = *, p<.01 = **, p<.001 = ***

4.1.4.4 Being Moved

No significant findings were found between the conditions in any of the excerpt for ratings of being moved. The significant findings from the first presentation between the music and video 2 category and the two remaining ones for the third music excerpts (pop tune), were not present. The music alone category was even rated higher than the second video category, but this difference was small. A carry-over effect was therefore not present for the being moved variable.

	F value	P value
Music 1	0.124	0.88
Music 2	0.302	0.74
Music 3	0.252	0.78

Table 22: Results from one-way ANOVA.



Mean ratings for each stimulus category of all the music excerpts. The error bars are representing standard deviation. p<.05 = *, p<.01 = **, p<.001 = ***

4.1.4.5 Valence

For the valence variable, the results from the first presentation were to a large extent replicated with a significant effect for the second music excerpt (string quartet), $F(2,63) = 6.32$, $p = 0.003$, $\eta 2 = 0.17$. As in the first presentation, the second video (Christian the Lion) had a negative effect on the music and the effect was almost as significant as in the first presentation, $p = 0.002$, $d = 0.99$ (large effect), (music alone: $M = 51.3$, $SD = 17.72$), (music and video 2: $M = 31$, $SD = 18.23$). A carry-over effect from the video 2 on the second music excerpt was therefore clearly present for the valence variable.

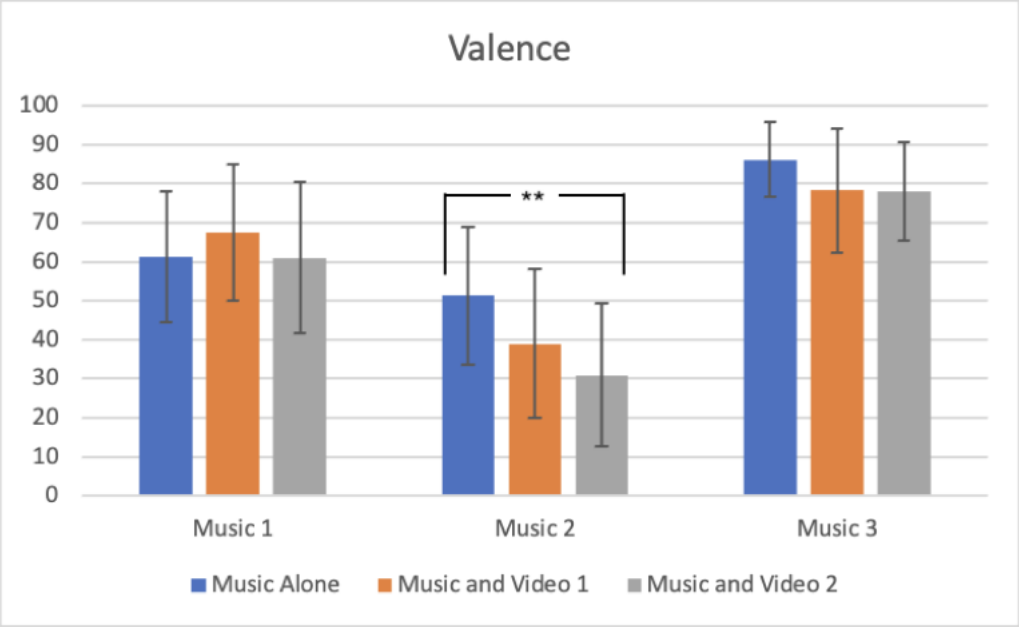
The significant difference between music excerpt, regardless of stimulus categories, was also present for the repetition, but the effect size was considerably smaller $F(2,195) = 92.45$, $p < .001$, $\eta 2 = 0.49$.

	F value	P value
Music 1	0.93	0.40
Music 2	6.325	0.003
Music 3	2.792	0.07

Table 23: Results from one-way ANOVA.

Music 2: Valence	P Value
Video 1 - Video 0	0.07
Video 2 - Video 0	0.002
Video 1 - Video 2	0.30

Table 24: Post hoc analysis (Tukey HSD).



Mean ratings for each stimulus category of all the music excerpts. The error bars are representing standard deviation. p<.05 = *, p<.01 = **, p<.001 = ***

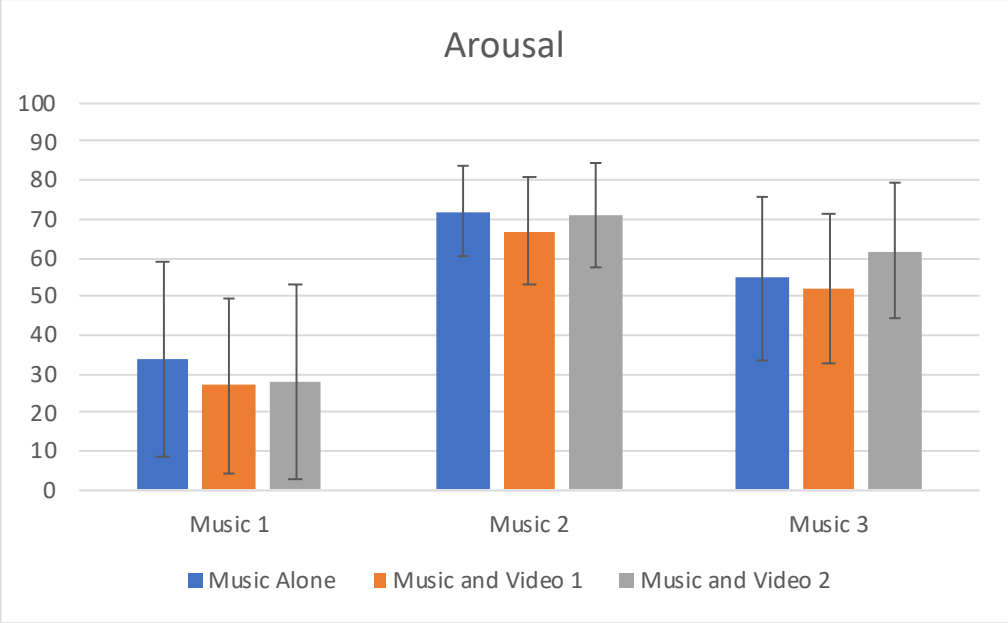
4.1.4.6 Arousal

No significant findings were found for the arousal variable. The ratings of the music excerpts were quite similar to the ratings for the first presentation, but the significant difference between video accompaniment categories for the third music excerpt was not found and a carry-over effect was not present for the arousal variable.

The significant difference between music excepts was, as for the valence variable, also present in the second presentation of the music, $F(2.195) = 71.85$, $p < .001$, $\eta 2 = 0.42$.

	F value	P value
Music 1	0.543	0.58
Music 2	0.991	0.38
Music 3	1.515	0.23

Table 24: Results from one-way ANOVA.



Mean ratings for each stimulus category of all the music excerpts. The error bars are representing standard deviation. p<.05 = *, p<.01 = **, p<.001 = ***

4.2 Physiological and Intuitive Responses

For the presentation of the first video alone, 15 of 31 participants reported experiencing other physiological/bodily responses, while 18 of 35 participants presented with video 2 reported answering yes to this question. Of the 21 participants assigned to the first condition, between 5 and 8 participants reported other physiological responses to the different music stimuli they were exposed to. For the second condition, between 3 and 8 of the 20 participants gave reports and for the third condition consisting of 25 participants, the number was between 7 and 19. Due to a quite low response rate to several of the questions related to the music stimulus, the significance of many of the reports can be discussed and the results may not be

very reliable. They may nevertheless give some insight to the variety of physiological responses. In addition to bodily reaction, many participants wrote down emotions they experienced.

4.2.1 Videos Alone

For the first video (surfing video), eight participants reported experiencing an increase in heart rate. A lump in throat and in the stomach occurred in two participants, and one reported experiencing some nausea. The shared experience of an increase in heart rate suggests that the video, through the activation of the sympathetic nerves system, had an ability to induce emotional states high in arousal. This was further supported by answers from four participants who experienced feelings of excitement. The reports indicate that emotions experienced tended to be of a positive nature with only three participants reporting experiencing states that alone could indicate the experience of displeasure (fear, and nausea). Neither gasps, nor tears, which in the existing literature are associated with awe, were experienced by any of the participants when presented with the "surfing video".

For the second video (Christian the lion) a general trend towards the occurrence of responses connected to positive emotions was found. As predicted, both smiles and a warm sensation were associated with the second video. Eight participants reported smiling, and six participants reported experiencing warmth in the chest. Two participant reported laughing (one of these also reported crying), and one additional participant experienced a lump in the throat. Many of the participants also reported experiencing joy and happiness, suggesting an increase in induced emotions connected to high valence.

4.2.2 Music Excerpt 1

For all the categories of the first music excerpt (ambient track), participants tended to report an experience of calmness and peace, suggesting that the excerpt had a tendency to induce emotions low in arousal and on the high side of the valence dimension. A warm sensation was also experienced across several conditions.

For the "ambient track" paired with the "surfing video" two participants experienced a warm feeling and responses were, compared to other conditions, characterized by an experience of admiration. Some negative emotions were experienced by two of the participants, with one reporting feeling of nervousness and another describing the experience as a mixture of calmness and distress. While eight participants reported an increase in heart

rate for the presentation of the video alone, only one participant reported experiencing this for the presentation of music 1 and video 1. One participant who experienced an increase in heart rate during the first presentation of the video 1 alone, reported a lower pulse when experiencing the same video in combination with music, further suggesting that the music may have reduced the activation of the sympathetic nervous system. For the second presentation of the first music excerpt in this condition, the characteristics of the first presentation were not present. Reports were mainly characterized by the experience of calmness further supporting the absence of a carry-over effect found in the ANOVA analysis.

When exposed to the "ambient track" in combination with the "Christian the lion video", five participants reported warmth in the chest, two reported watery eyes, and one participant reported an increase in heart rate. The prevalence of warmth in chest and tear responses may indicate that several of the participants were moved by the music and video combination. Compared to the reports to video 2 alone, no participants reported smiling when the music was accompanying the visuals. When the music was repeated, only four participants from this condition gave an answer to the question. None of them were related to a warm sensation, suggesting little influence from the video on the second presentation. One of the participants who reported experiencing watery eyes experience the same reaction for the second presentation of the music, which indicates that a carry-over effect may have occurred in one participant.

Participants who were presented with the "ambient track" alone had significantly fewer physiological responses than participants in the other conditions, with the only response being an increase in heart rate which was experienced by one participant. Many of the participants tended to experience calmness and relaxation and one experienced the music as uplifting. During the second presentation of the music, many reports tended to be more focused on the music being uplifting. One participant reported experiencing a warm sensation and another experienced restlessness.

4.2.3 Music Excerpt 2

For the second music excerpts (string quartet), the most common response was an increase in heart rate, which occurred in all the conditions. Some participants also reported experiencing negative emotions like stress, fear, and anxiety, suggesting that the music had a tendency to induce negative emotions low in valence and high in energy.

When the music was paired with video 1 (surfing video), the majority of participant who wrote down a response experienced an increase in heart rate (four out of eight responses). One participant experienced warmth in the chest while another experienced fear and tension from the music. For the second presentation, responses were less negative than the first presentation, but three of the eight participants who gave a response during the second presentation also experienced an increase in heart rate (two had the same experience during the first presentation of the music) and one of them also reported an increase in respiratory rate. This may suggest a carry-over effect in some participants, but it is also possible that the music alone accounted for these responses. One participant also experienced a lump in throat, and the two last responses were connected to an experience of curiosity in one participant and being moved in the last.

Participants who heard the music in combination with video 2 (Christian the lion) tended to report a dissatisfaction with the combination of music and video. One participant commented on the music being too dramatic for the content of the video, and three other participants also reported dissatisfaction and confusion with the pairing of the music and the video. Three participants nevertheless reported experiencing joy, presumably over the content in the video, suggesting that the mismatch between video and music did not erase the occurrence positive emotions in all the participants. For the second presentation of the music excerpt, reports were more focused on physiological response, with an increase in heart rate being experienced by two participants and warmth in chest by one. Two participants also reported dissatisfaction during the second presentation and one participant reported discomfort, which may imply the presence of a carry-over effect.

For the music alone, three participants experienced a warmth in chest, goosebumps, and an increase in heart rate (one person for each response). Stress and tension were also experienced by two participants. For the second presentation, only three participants gave a repose to the question. Two were concerned with physiological responses, (shivers down the spine and feeling of excitement in chest), while the last participant reported experiencing a feeling of happiness.

4.2.4 Music Excerpt 3

A general trend was found for the experience of happiness for all the stimulus categories in the last music excerpt (pop tune), indicating emotional states high in valence. The most

common physiological response was a warm feeling, being experienced by participants in all conditions.

For the music in combination with video 1 (surfing video), the most common response was an increase in heart rate, which occurred in three out of the six participants who responded to the question. One participant experienced warmth and excitement, another reported laugher, while the last participants responding was annoyed by the music. During the second presentation no participants reported experiencing an increase in heart rate and responses were connected to an experience of joy and a warmness in one participant. The participant who was annoyed by the music during the first presentation was even more so during the repetition demonstrating a negative effect for the repetition.

Most answers were given to the combination of the "pop tune" and the "Christian the lion video" with a total number of 19 responses. A warm sensation was experienced by six participants and another six participants reported smiling when presented with the stimulus. One participant experienced a lump in throat and another an increase in heart rate. None of the participants seemed to have had any negative experiences with the stimulus, and most of the participants reported happiness and comfort. During the repetition, four participants experienced a warm sensation (three of them had the same experience during the first presentation) suggesting that a carry-over effect may have occurred for some participants. One participant also reported feeling happy because the music made her think of the lion. Reports were in general more focused on the character of the music with three participants experiencing a driving rhythmic feel from the music. None of the participants reported smiling during the second presentation of the music, which suggests that the contagion effect was not carried over to the music. One participant also expressed boredom with the music.

For the music alone, two out of the six participants who responded experienced a warm sensation, one participant reported happiness and laughter, and another participant expressed a wish to move to the music. Emotions of a positive nature were experienced by all the participants who responded yes to the question. During the repetition, two participants experienced a warm sensation (one had the same experience during the first presentation of the music) and all descriptions expressed were of a positive nature like in the first presentation.

5 Discussion

5.1 First Presentation of Music Excerpts

The present study investigated the influence the moving image could have on musical emotions, covering both perceived and induced emotions, as well as enjoyment. The moving image had, in accordance with the hypothesis, a positive influence on the experience of awe (ambient track and pop tune) and being moved (pop tune). Ratings of awe were, as predicted, mainly influence by the "surfing video" while an impact on ratings of being moved was only found for the "Christian the lion video". Despite this positive influence on induced emotions, ratings of chills were unaffected by the presence of the moving image. Positive effects were neither found for the enjoyment variable and induced emotions did, contrary to predictions, not seem to enhance the enjoyment of the music. The moving image neither had a positive (mood congruent) influence on perceived emotions, but a significant difference was found, in relation to the arousal variable, between the "surfing video" and the "Christian the lion video" when combined with the "pop tune". A negative (mood incongruent) influence on enjoyment and valence was, on the other hand, found for the "string quartet" when paired with the Christians the lion video.

5.1.1 Enjoyment

The hypothesis that a positive influence from the moving image on felt emotion (awe and being moved) would result in the music being enjoyed more was not verified, and no positive effects were found for any of the music excerpts. These results do therefore not support findings from studies conducted by Geringer and colleagues in which the moving image had a positive influence on both liking and induced emotions. It is nevertheless important to have in mind that the empirical basis for this assumption is quite weak, as results from these studies were not very robust (Geringer et al., 1996; Geringer et al., 1997). The negative effect found for the pairing of the "Christian the lion video" and the "string quartet" was, on the other hand, expected. Although no predictions were made in relation to which variables would be negatively influenced, a negative effect was in general expected, due to a predicted perception of the pairing as incongruent. This prediction was verified, and the pairing of the "string quartet" with the "Christian the lion video" was both experienced as the least congruent and

given significantly lower congruency ratings than the pairing of the “surfing video” and the “string quartet”.

The negative effect from the incongruent pairing may be the result of a lack of emotional engagement with the music due to a perceived mismatch between visuals and music. The enjoyment of music is quite often predicted by the presence of positive emotions (Juslin, 2019: 468), and negative emotions may therefore have an opposite influence. In a study conducted by Vuoskoski and colleagues, scary music experts elicited the most intense emotions, but were significantly less liked than happy, tender, and sad music excerpts, indicating a clear negative influence from negative emotions on enjoyment (Vuoskoski et al, 2011). Negative emotions such as irritation and confusion, that might have been experienced for the incongruent pairing, are therefore likely to have had an opposite influence, explaining the decrease in enjoyment. Reports from some participants related to the experience of bodily responses indicated that the incongruent pairing was experienced as dissatisfying.

Despite being in conflict with the hypothesis, the lack of a positive influence on enjoyment may, have a logical explanation. Even though there is a link between enjoyment and induced emotions, some independence between the two is nevertheless expected to exist. The enjoyment of music may in addition to induced emotions also be grounded in *aesthetic judgments*. To what extent enjoyment is based on felt emotion or the aesthetic quality of music may vary between people. Emotions may for instance be induced through mechanisms such as contagion or evaluative conditioning without music being judged to have a high aesthetic value or being liked (Juslin, 2019: 465-66). In the current study, participants may in that regard have relied more on genre preference and the aesthetic quality of the music, than felt emotions when they evaluated the enjoyment potential of the music. Although possible, this does not seem to explain ratings of the incongruent pairing which seems to be based more on contextual information from the moving image and induced negative emotions. One possible explanation for this is that negative emotions may have a stronger impact on the enjoyment of music than positive emotions, as well as complex emotions such as awe and being moved.

5.1.2 Chills and other Physiological and Intuitive Responses

For the chills variable, no significant results were found in any of the music excerpts. Even though the videos had an influence on ratings of awe and being moved, which is expected to trigger the occurrence of chills, a connection between these emotions and chills was not found

in the current study. In spite of this, major variation within each condition was found, which is likely to be explained by the role individual differences plays in the experience of chills (McCrae, 2007; Nusbaum and Silvia, 2011).

The lack of influence from the videos may partially be explained by specific elements in the music. A sudden change in dynamics, as well as the entry of the human voice or high frequent sounds resembling it, have been found to elicit chills in several studies (Sloboda, 1991; Panksepp, 1995; Greve et al., 2007; Bannister, 2020). The predicted climaxes in the music excerpts do contain these features. In the first music excerpt a high pitch tone enters the soundscape and some changes in dynamics also accompany this tone. A sudden increase in dynamics and intensity occurs in the second music excerpt, and in the third music excerpt the human voice is present throughout the whole piece. The lack of any significant influence from the moving image on the experience of chills may not necessarily mean that the videos were not able to trigger chills. The lack of influence may be the result of the music already containing features connected to the occurrence of chills.

Since other physiological and intuitive responses than chills were not rated on a slider, it was not possible to statistically analyze differences between stimulus categories. Some differences in responses were nevertheless found. Particularly instances of smiling only occurred during the presence of the "Christian the lion video", both alone and in combination with the "pop tune". This finding further supports the presence of mechanisms related to emotional contagion, and the smiling response was likely an intuitive imitation of the men smiling in the video. A similar pattern was also found for the experience of a warm sensation, which was evident in many of the reports for the "Christian the lion video" either alone or paired with music (with the exception of the string quartet). This response was nevertheless also present in other music excerpts, although not as evident as in the ones in which the "Christian the lion video" was present. Increased heart rate was mostly connected to the viewing of the "surfing video", verifying the hypothesized outcome, but as with the experience of a warm sensation some participants also experienced this response in the absence of this film. Some responses such as gasps, were not reported, but the connection made between awe and gasps is not empirically based but rather founded in evolutionary theories (Huron 2006).

5.1.3 Awe

The awe variable provided the highest number of significant results, with both the "ambient track" and the "pop tune" being influenced by the first video in relation to the experience of awe. The positive influence the "surfing video" had on awe was further supported by significantly higher awe ratings for this video compared to the "Christian the lion video", when presented alone. Results from the regression analysis suggests that the videos explained the majority of the variance in the combination of music and videos. The positive influence from the "surfing video" on the experience of awe was in accordance with the hypothesis, as it was believed to possess certain qualities associated with the experience of awe such as vastness and accommodation.

The interaction effect that was found for the pairing of the "string quartet" and the "surfing video" may be explained by a lack of surprise due to a familiarity with the video. For participants presented with the "surfing video" alone at the beginning of the test, the sudden break of the wave and the surfers' ability to remain on the board came as no surprise, and this may have had an impact on the experience of awe. This may be due to the fact that an absence of surprise can result in a decrease in the experience of accommodation, which awe is believed to rely heavily on. When the events in the video are familiar, they are not likely to be perceived as unexpected and challenging, and accommodation may not be as urgent (Keltner and Haidt, 2003: 303-04). Although a significant interaction effect was found for the second music excerpt, it is not clear why it was not present for any of the other excerpts. It is possible that the dramatic character of the music to a larger extent may have enhanced the surprise element, triggering more awe than the other music excerpts, but this is just a speculation.

The "Christian the lion video" also rather unexpectedly had a positive effect on awe when paired with the "pop tune", although not as significant as the result for the "surfing video" in combination with this music excerpt. This may partially be explained by the uncertainty connected to what would happen after the lion started approaching the men. For participants unfamiliar with the video, it is possible to assume that participants would fear a more violent outcome of the encounter. There are nevertheless several arguments that can be raised against this assumption. Many of the participants were familiar with the video, and participants who were presented with the "Christian the lion video" alone at the beginning of the test did not give significantly lower ratings of awe compared to those unfamiliar with the video. In addition to this, the "pop tune" did not in any way support prediction of a violent outcome, as it was given a very high valence rating in all conditions. There is therefore some uncertainty

as to why the “pop tune” paired with the “Christian the lion video” was given a significantly higher rating than the music alone.

5.1.4 Being Moved

For the variable being moved, the “Christian the lion video” had a positive influence on the “pop tune”. This was the only significant finding, but the effect turned out to be larger than all the positive effects from the awe variable. A large effect size was found for the “Christian the lion video” and music when compared to both the music alone and the music paired with the “surfing video”, demonstrating the robustness of this finding. It was no surprise that fewer music excepts were positively influenced by the “Christian the lion video” for the being moved variable, compared to ratings of awe for the “surfing video”. While all the music excerpts were expected to be positively influenced by the first video in relation to awe, a positive effect from the second video on ratings of being moved were only predicted to occur in two of the three music excerpts. This was due to a mismatch between the “string quartet” and the “Christian the lion video”.

The “Christian the lion video” did not, contrary to the hypothesis, have a positive effect on the “ambient track” for the being moved variable, as ratings were more or less similar for all the different stimulus categories. Although a difference was not found for this pairing, self-reports revealed that several of the participants experienced physiological responses that are associated with feelings of being moved, like a warm sensation in the chest as well as tears/watery eyes. These are in addition to chills expected to be the most common responses accompanying this emotion. It is therefore likely that several participants were moved by the combination, even though a statistical difference was not found.

Compared to the “surfing video”, the “Christian the lion video” was given a significantly higher rating for the being moved variable when presented alone, supporting the hypothesis that this video the largest extent would be able to elicit emotions of a moving nature. Interestingly, the regression analysis revealed that a large amount of the variance for the dependent variable (music paired with video) could not be explained by the independent variables (music alone and video alone). In particular, the “pop tune” and the “Christian the lion video” were given much higher ratings in combination, compared to when the two stimuli were presented alone. This finding is somewhat similar to the result from the brain imaging study conducted by Elder and colleagues. In this study the pairing of music and films resulted in far more activity in emotional eras of the brain, as compared to the presentation of

each stimulus separately (Eldar et al., 2007). It is therefore possible that the combination of the "pop tune" and "Christian the lion video" similarly resulted in a form of synergy effect, in which the addition of concrete information to the music, as well as emotions from the music to the visual, enhanced the overall experience. For the incongruous paring of the "string quartet" and the Christian the lion video, the pairing was, not surprisingly, given a lower rating compared to the two stimuli presented alone. This may also partially explain why a large amount of the variance from the dependent variable was not explained by the independent variables.

5.1.5 Valence

As with the result for the enjoyment variable, no significant positive (mood congruent) influence from the videos was found for the valence variable, but a negative (mood incongruent) influence of significance was found for the pairing of the "string quartet" and the Christian the Lion video. Further regression analysis on the valence variable revealed that the music variable predicted most of the results from the combination of the music and video, supporting the general lack of a mood congruent influence from the moving image on music. All the music excerpts were perceived significantly different regardless of stimulus categories, demonstrating that the selection of music excerpts was as dissimilar from one another as intended. A significant difference was also found between the two videos, with the "Christian the lion video" being perceived as more positive than the "surfing video".

Although differences were found between music excerpts and the two videos, the results were not completely in accordance with the prediction. The "ambient track" was expected to be perceived as the most ambiguous excerpt, but the mean rating of the music alone was above 60, giving it a rating on the higher side of the valence scale. The "string quartet", which was expected to be perceived as negative, was rather given a mean rating around 50. The "pop tune" was the only one in which ratings were corresponding to the prediction, with a relatively high valence rating above 80. The "surfing video", which was expected to be given ratings on the negative side of the valence scale, received a mean rating around 50. The "Christian the lion video" was on the other hand given a high valence score as predicted.

The fact that the videos were not given contrasting valence ratings when presented alone, may be part of the reason why the results from the study by Marilyn Boltz and colleagues were not replicated (Boltz et al., 2009). With the first video being perceived as rather ambiguous it may not have had the ability to have an influence on the music in the same

manner as those selected in the study from 2009. In addition, the lack of ambiguity in the "ambient track", which was expected to be particularly prone to an influence from the moving image, also likely explains why no impact was found. Interestingly, the combination of the "Christian the lion video" and the "ambient track" was even given a slightly lower rating than the music alone. Since the video on its own had a very high valence rating above 80, it was expected to have a positive influence on the music excerpt. For the "pop tune", no difference of significance was found between categories, which further suggest that an influence on valence depends on the ambiguity of a music stimulus. The absence of a positive influence from the videos may also be explained by other factors than the amount of ambiguity in the stimuli. Perceived emotions are for instance expected to be less dependent on situational factors than induced emotions, which may partially explain why a positive influence was found for the experience of awe and being moved, and not for the perception of valence (Gabrielsson, 2001:137).

The significant negative effect from the pairing of the "string quartet" and the "Christian the lion video" was somewhat interesting and unexpected. While the negative impact on enjoyment was some extent predicted, the negative effect on valence was not expected to occur, based on what existing theories suggests. Since audio-visual interactions in the perception of music are expected to rely on the presence of congruency (Cohen, 2013; Boltz, 2013), it is rather surprising that the perceived valence of the music was influenced, although negatively, by the least congruent paring of music and visuals. A perceived incongruence was rather expected to prevent interactions between video and visuals from occurring. This was found to be the case in a study conducted by Annabel Cohen, although with a focus on how music could influence the moving image (Cohen, 2013: 27). In the current study, the mismatch between the positively valenced video and the music made the music sound even more negative in terms of valence. It therefore seems like an interaction of a negative nature between music and visuals can occur in the absence of congruency.

Another finding open to discussion, is the rating of the "string quartet" when presented alone, which may be interpreted to indicate a perception of the music as ambiguous in terms of valence. This assumption in many ways conflicts with the incongruence in the pairing with the Christian the lion video. Although this mismatch experienced by participants may be explained by a formal incongruence, it also seems like a mismatch in terms of valence was perceived. This is supported by the open question related to bodily responses in which the incongruent pairing was described as both unfitting and too dramatic for the content of the film. Based on theories by Cohen and Boltz, the valence of the music excerpt would have

been positively influenced by the video due to the ambiguity of the music in terms of valence. Although given a mean rating around 50, the strong perceived incongruence and the negative influence suggest that a characterization of the music as ambiguous, may not be very suitable.

5.1.6 Arousal

Little influence was found from the videos on the perception of arousal in the music. The only significant finding was between the two video pairings for the "pop tune", with a significantly higher rating for the Christian the lion pairing. In the regression analysis, as with the valence variable, music explained most of the variance in the dependent variable, and the three music excerpts were perceived significantly different regardless of stimulus categories. This indicates that the influence of the moving image was limited for this variable.

No significant difference was found between the videos when presented alone, and the "Christian the lion video" was only given a slightly higher mean rating than the "surfing video". These mean ratings were located around 60 on the scale and the videos did therefore not have a very stable character in terms of arousal. Particularly the "surfing video" was expected to be given a much higher rating, due to the predicted intensity associated with the surfer's encounter with the large wave. The lack of influence from the moving image on music may therefore partially be explained by the video's lacking a clear preference for any side of the arousal scale. The majority of the music excerpts were on the whole given a slightly lower rating when combined with videos, as compared to when they were presented alone.

The significant difference found in two of the categories for the "pop tune" may be due to the music being given a rather ambiguous rating in terms of arousal when presented alone. It may therefore have been more susceptible to an influence from the videos. This significant result nevertheless conflicts with the ratings of the videos alone. While the videos in the absence of any music were given quite similar ratings, they did influence the music in different ways. The "Christian the lion video" and music was given significantly higher arousal ratings than the combination of the music and the "surfing video". Compared to the music alone category, the "Christian the lion video" increased arousal ratings while the "surfing video" decreased ratings of arousal. Why this was the case when the two videos were not rated significantly different from one another is rather surprising, and further investigation would be needed to provide an answer to this.

5.1.6 Congruence

As predicted, the combination of the “string quartet” and the “Christian the lion video” was perceived as the least congruent pairing, being given a significantly lower rating than the pairing of the “string quartet” and the “surfing video”. Apart from this combination, mean ratings were in general lower than predicted. This was particularly evident for the “ambient track” in combination with both the videos, as well as the pairing of the “string quartet” and “surfing video”. Relatively high standard deviations also revealed a lack of consensus in terms of level of congruency.

Congruency between visuals and music seems to be a determining factor for the overall experience and perception of the music. In particular the negative effect from visual on music, both present for the enjoyment and the valence variable, was strongly correlated with a low level of congruency. Why this negative effect was only significant for the valence and enjoyment variable and not for the experience of for instance being moved, awe, chills or perceived arousal, is worthy of discussion. Given that induced emotions are involuntary responses that may be triggered by mechanisms like for instance contagion or a brain- stem reflex, the “Christian the lion video” or the “string quartet” may, regardless of the incongruent combination, have resulted in induced emotions. Enjoyment is on the other hand more associated with the overall experience and the evaluation of a stimulus. The music may therefore be more susceptible to being negatively influenced by the incongruent accompaniment of the video for this variable. The reason why a negative influence was found for the valence, but not the arousal variable, might be that the incongruence seems to be mostly related to the valence of the two stimuli. While the video and music were relatively similar in terms of arousal, when presented alone, the difference was quite big for the valence rating.

The level of congruency may to some extent also explain several of the other findings. The pairing of the “ambient track” and the “Christian the lion video”, which was given the second lowest mean rating in relation to congruency, contrary to the hypothesis had no significant influence on the experience of being moved. This combination was also rated lower, compared to the other categories for the enjoyment variable, and slightly lower than the music alone category for the valence variable. For the pairing of the “Christian the lion video” and the “pop tune”, which was given the highest rating for congruency, a significant positive effect on the experience of being moved and awe was found. This combination was also given a slightly higher enjoyment rating than the other categories. This pattern seems to further

indicate that a positive influence from the moving image on music may rely on congruency between the two stimuli. This finding is in accordance with research by Marilyn Boltz and Annabel Cohen, which focuses on how unification between music and visuals is fundamental for the character of one stimulus to be transferred to another (Cohen 2013, Boltz 2013), as well as the study from 2015 conducted by Vuoskoski and Eerola, in which the description intensifying induced emotions was emotionally congruent with the music (Vuoskoski and Eerola 2015).

5.2 Second Presentation of Music Excerpts

For the repetition of the music, no positive carry- over effects were found for any of the excerpts. For the awe and being moved variables in which significant findings were present under the first presentation, mean ratings were more or less similar across the different stimulus categories. This suggests that induced emotions evoked by the moving image, were not sustained and carried over to the music when presented later. The significant difference between the experimental categories (music paired with videos) for the arousal variable was not found for the repetition either. The significant negative effects related to enjoyment and valence were on the other hand present for the second presentation of the music, although with a lower effect size for the enjoyment variable. The last finding demonstrates that music may have a persistent negative influence on the experience of music.

Ratings of the music excerpts in the repetition were quite similar to those found in the first presentation, indicating that the ratings given were of a certain reliability. Particularly ratings of excerpts from the music alone category were in general the same when repeated for all the variables, suggesting that neither a mere exposure effect nor a boredom effect had much of an impact on the second experience of the music. The music excerpts were in relation to valence and arousal also rated significantly different from one other in the repetition.

The lack of any significant positive effects in the repetition was not in accordance with the hypothesis. Since emotional events of high intensity were expected to be remembered more easily than non-emotional events, positive effects found for the experience of awe and being moved in the first presentation were expected to be sustained. Although one participant explicitly wrote down being moved as a direct result of thinking of the lion when rehearing a music excerpt, previously accompanied by the Christian the lion video, it seems like a

tendency to recollect the videos was not present for the majority of the participants. The lack of a positive (mood congruent) influence on perceived emotions conflicted with the original hypothesis, but since no positive (mood congruent) effect was found for the first presentation of the music, the lack of influence in the repetition was expected.

The persistent negative influence on the enjoyment and valence variables indicates that negative effects were more prone to being sustained as compared to positive effects found in the music. It is possible to imagine that the mismatch between visuals and music made a bigger impact than other pairings, as it may have conflicted with participants' anticipation of how the two stimuli would fit together. One participant reported being confused as a result of this, and it is plausible that others had a similar experience. The recollection of stimuli associated with negative emotions has also been found to be less dependent on attention than stimuli connected with positive emotions (Talmi et al., 2007), which may suggest that this type of stimuli is remembered more easily. This may further explain the present findings, as the ratings of enjoyment, as well as some reports related to physiological responses, indicated that the experience of the "string quartet" paired with the "Christian the lion video" was negative.

The persistence of the negative effect is in many ways also in accordance with theories related to evaluative conditioning. The pace of conditioning, as the recollection of past experiences, is to a large extent dependent on the intensity of a stimulus and negative experiences may result in a faster conditioning than positive ones. This is particularly evident in fear conditioning, in which conditioning may occur after only one pairing when an unconditioned stimulus is very unpleasant, like a shock (LeDoux, 1995). It is important to have in mind that the displeasure experienced for the incongruent pairing is unlikely to be of the same intensity as unconditioned stimuli used in fear conditioning experiments. Conditioning after only one pairing is therefore not very probable to have occurred in the current experiment. It should nevertheless not be fully excluded as a possible explanation.

Ratings of enjoyment are in particular likely to be explained by the impact negative experiences can have on memory. Since enjoyment tends to be associated with pleasure and is predicted by positive emotions (Thompson, 2006; Juslin, 2019: 236), a low rating of this variable may indicate the presence of negative emotions. The possible recollection of the video may have elicited many of the same negative feelings that participants experienced during the first presentation, further impairing the enjoyment of the music. It is also possible to imagine that mechanisms related to evaluative conditioning may have been responsible for the sustained effect on ratings of enjoyment, although this is less likely.

The negative influence on valence was in many ways more surprising than the negative effect on enjoyment. This was also surprising for the first presentation (perceived emotions are less expected to be influenced by contextual information than induced emotions: Gabrielsson, 2001:137), but even more so for the repetition. This is because episodic memories are mostly associated with induced emotions. This finding is not in accordance with conditioning either, as changes in perceived emotions are expected to demand more repetition than induced emotions to occur. The influence from the video did neither consist of the valence from the video being carried over to the music, which is how traditional forms of conditioning operates, but an opposite pattern was found. This finding is of great interest, as it was more robust than the significant difference for enjoyment in the repetition, and at the same time not in accordance with established theories.

5.3 Summary of Main Findings

The present study suggest that the moving image may intensify emotions induced while listening to music. This study is most likely the first to investigate the influence the moving image may have on complex emotions like awe and being moved, and the results do indicate that visuals information in the form of videos may have an impact on strong emotional experiences during music listening. Narrative elements in the videos, particularly related to suspense, may also have helped to increase the intensity of emotions induced in a similar manner as mechanisms related to musical expectancy.

A significant influence was on the other hand not found for the experience of chills. This was a bit surprising given that an influence for both awe and being moved was found, but this may partially be explained by elements in the music excerpts that are associated with chills. It is also important to stress that the occurrence of chills is to a large extent predicted by personality traits such as openness to experience, which makes it more challenging to empirically investigate this phenomenon (McCrae, 2007; Nusbaum and Silvia, 2011). Some difference, although not statistically analyzed, seems to have been present between conditions in relation to other physiological and intuitive response such as smiling, increased heart rate, and sensations of warmth. In particular, smiling found in the viewing of the “Christian the lion video” indicates the presence of emotional contagion.

The lack of ambiguity in the music excerpts, as well as the ambiguity present in the "surfing video", may explain why a positive influence from the moving image on perceived emotions was more or less absent. This further indicates that unambiguous music may be resistant to an impact from the moving image. It would therefore seem like an influence in terms of perceived emotions may be of a somewhat limited ability.

The negative effects found for the enjoyment and perceived valence of the pairing of the "string quartet" and the "Christian the lion video" is of particular interest. No existing study has, to the best of our knowledge, examined the impact incongruent pairings may have on the experience of music, and this topic has earlier been suggested as an area for future research (Boltz et al., 2009). It seems like an influence on perceived emotions can occur even in the absence of congruence, contrary to previous assumptions (Cohen, 2013). An influence from the moving image may not only be of a positive nature but is rather a versatile process and may in some cases even diminish the experience of music.

The significant findings from the second presentation of the music indicates that the negative influence on enjoyment and valence related to the incongruent pairing was more robust and persistent compared to the positive effect found for awe and being moved. The moving image may, despite a lack of a positive (mood congruent) carry-over effect, therefore have a lasting impact on the experience of music.

The level of congruency between visuals in general seems to have to have played a crucial role for the influence of the moving image on music. This is particularly reflected through the negative effect found for the incongruent pairing, but it also seems like a positive influence from the videos to a large extent depended on a perceived congruency between the two stimuli. A positive influence may not only rely on the emotional strength of the moving image, but also on how well it is integrated with the music.

5.4 Limitations

The current research does contain some limitations. The presentation of the videos on their own was important to better assess for responses that they alone could account for, but this choice was not completely free from problems. The lack of sound in the videos, although necessary in order to fully isolated visuals and sounds, may have reduced felt emotions. Since visuals and sounds are intertwined in our perception of the world, the experience of the

moving image without sound may be a somewhat strange experience. Videos and films are almost always accompanied by either diegetic, or non-diegetic sounds (often in the form of music), and this further strengthens an expectation of sound in encounters with the moving image. The violation of this expectancy may have had a negative impact on the experience of the videos. This may possibly explain why mean ratings of both chills, awe, and being moved were not very high and this can, to some extent, have resulted in uneven comparisons between the videos alone and music alone.

Although the present study found the moving image to have a positive effect on the experience of awe and being moved, it is not possible to know if emotions induced by the moving image were carried over to the music when the two stimuli were combined. While enjoyment and perceived emotion are more a result of reflections and judgments of a stimulus, induced emotions are on the other hand the result of an involuntary response. It is therefore harder to track down the source of an induced emotion and know if the music was experienced as more emotional. Since we tend to seek unification between visuals and music, the music was expected to be connected to emotions trigged by the videos, but this is not something that the current study gives insight into. One solution to this might have been to ask a question in which the focus was on emotions induced by the music, but it was not expected to be easy, if possible at all, to separate emotions experienced from the visuals and those from the music. This question might also have been confusing for participants as it more easily could have led to felt emotions being confused with perceived emotions, since the focus would have been more on the music than the listener.

Translation issues may, to some extent, have had an unwanted influence on the results. In particular the translation of awe into Norwegian is not completely unproblematic, as the Norwegian term is not entierly equivalent to the English in meaning. Although both terms are usually applied to sublime and overwhelming experiences, the Norwegian translation of awe which is æ*refrykt* (English = honor and fear) has connotations that are of a less positive valence. *Ærefrykt* tends to be more associated with deep respect and fear and is less concerned with wonder and admiration than awe. In addition to this, the Norwegian term is less commonly used in daily language than awe and therefore not as intuitive. An explanation of the term was for this reason given in both Norwegian and English, but it is still quite possible that the term was interpreted differently by Norwegian participants compared to those from other countries. This may further have had a negative influence on the validity of the result from the awe variable.

Another limitation with the current study is the lack of distinction between formal and semantic congruency in the question related to how well the music fits together with the film. The pairing of the "string quartet" and the "Christian the lion video" was for instance predicted to be perceived as incongruent both in terms of semantic and formal congruency, but it is impossible to fully know the level of contribution from the two forms of congruency on participants' rating. The current study therefore cannot fully explain whether the negative impact on enjoyment and valence was mostly due to semantic incongruence due to a mismatch in affect, or formal incongruence related to the lack of synchrony in the structure of the two stimuli. It is however not unlikely that both forms played a part in participants' rating. The phrasing of the question related to congruency «Did the music fit with the visuals» could also have been phrased differently, as it does not offer an equal focus on to the two stimuli. «Did the music and film fit together» would in retrospect have been a better phrasing.

The fact that none of the excerpts were pilot tested is also likely to have put some limitations on the present research. With a large sample of music excerpt and videos in an initial phase of the study (pilot test), it would have been possible to filter out excerpts that did not meet the criteria for the stimuli selection. Chill-triggering features, the experience of awe and being moved, as well as perceived emotions (valence and arousal), would then to a larger extent have been controlled for in both the videos and the music excerpts.

The interaction effect found for participants presented with the "surfing video" in combination with the classical excerpt is of limited reliability. The number of participants that were presented with the "Christian the lion video" alone in the beginning of the study was only around 10.

5.5 Future Directions

Future investigations of the influence of the moving image on music may consider choosing videos that are more contrasting both in terms of valence and arousal, for instance having one video expected to be given high ratings of both valence and arousal, and another video expected to be given low ratings of these dimensions. For the music excerpts, more effort should be put into controlling the valence and arousal and making sure that the experiment contains music that is neutral in both dimensions. Having videos with a clearer affective character and more neutral musical excerpts, in line with the study conducted by Boltz et al.

(2009), would be a more suitable choice and increase the possibility of obtaining positive effects, particularly in relation to perceived emotions. A pilot study could in that regard also be applied, to better assess the valence of potential music and video excerpts. This type of stimuli selection would also be more in line with conditioning research in which a conditioned stimulus tends to be neutral and an unconditioned stimulus more potent. The chances of finding a persistent positive carry-over effect are therefore likely to increase with more ambiguous music and unambiguous videos.

Regarding induced emotions, future studies might look into other emotions than just awe and being moved. In particular induced emotions that correspond to those expected to be perceived should also be examined more systematically, since the two are likely to co-occur. A video that for instance is perceived as high in valence and arousal is also likely to induce emotions such as happiness and joy in listeners. This seems to have been the case for the "Christian the lion video" in the current study, reflected through written responses to the open questions. Moreover, a positive effect on perceived and corresponding induced emotions may give a study more convincing results. Although basic emotions such as happiness, joy, fear, and sadness are less likely to induce strong emotions and chills, they should in general be given more attention in future research. This is because they are able to be both perceived and induced in a musical context (Gabrielsson, 2001). Negative emotions such as displeasure or annoyance might also be investigated in future studies, since they are likely to accompany incongruent pairings.

Although no significant findings were found for the occurrence of chills, this is an area that might be examine further. Music excerpts should in that case be chosen more carefully in relation to physiological responses. Excerpts that contain element that are associated with the occurrence of chills, such as a sudden increase in dynamic, modulation, a high pitch voice, or an instrument that resembles it, should preferably be avoided.

The negative influence from the incongruent pairing is an area which in particular should be given more attention. The negative influence provided the most robust results, being present also for the second presentation, and this is also the first study to date to examine this. The impact of congruency should be examined more systematically with a clearer division between semantic and formal congruency. It would for instance be interesting to see if pairings that are semantically congruent but temporally incongruent, have the same negative influence as music excerpts that are semantically congruent but lacking in terms of temporal synchrony. This would further give more insight into the influence of the separate forms of congruency.

5.6 Conclusion

This study has investigated the role visual information in the form of the moving image may play in relation to enjoyment and emotions induced and perceived, while listening to music. The findings from this study give insight into ways in which extramusical information and a visual context may influence our encounters with music.

Although an opposite investigation (the impact music can have on the moving image) at first may seem to be more useful due to the importance of music in for instance movies, a reverse investigation may also have a utility value. Music-videos in which the moving image serves the music is one area that quickly comes to mind. A conscious pairing of videos with a strong emotional content and music, in which congruency is taken into account, may possibly result in an enhanced musical experience. The love of film music, reflected through releases of music soundtracks separately from films, is also worth highlighting. It is possible that the popularity of pop songs like *My Heart Will Go On* from *Titanic*, or *Don't You (Forget About Me)* from *The Breakfast Club* is partially due to the emotional content of the film they accompany.

This investigation may also, to some extent, have a relevance in music-related settings where the moving image is absent but where visual information surrounding the listener is potent. A panoramic view of an impressive landscapes may for instance help to evoke a feeling of awe resulting in a listening experience with a deeper emotional impact.

The present research has offered a detailed investigation of the possibilities of an influence from the moving image on music. The role the moving image plays in our encounters with music seems to be of great relevance, and this is an area which will hopefully be given more attention in future research.

References

Athanasopoulos, G., Eerola, T., Lahdelma, I., & Kaliakatsos-Papakostas, M. (2021). Harmonic organisation conveys both universal and culture-specific cues for emotional expression in music. *PloS one*, *16*(1), e0244964.

Baeyens, F., Field, A. P., & Houwer, J. D. (2005). Associative learning of likes and dislikes: Some current controversies and possible ways forward. *Cognition & Emotion*, *19*(2), 161-174.

Bannister, S. (2020). A survey into the experience of musically induced chills: Emotions, situations and music. *Psychology of Music*, *48*(2), 297-314.

Baumgartner, T., Lutz, K., Schmidt, C. F., & Jäncke, L. (2006). The emotional power of music: how music enhances the feeling of affective pictures. *Brain research*, *1075*(1), 151-164.

Bernardini, L., & Niccolini, A. (2015). Does music calm the dog?. *DOG BEHAVIOR*, *1*(2), 13-17.

Blair, M. E., & Shimp, T. A. (1992). Consequences of an unpleasant experience with music: A second-order negative conditioning perspective. *Journal of Advertising*, *21*(1), 35-43.

Boltz, M. G. (2013). Music videos and visual influences on music perception and appreciation: Should you want your MTV. *The psychology of music in multimedia*, 217-234.

Boltz, M. G., Ebendorf, B., & Field, B. (2009). Audiovisual interactions: The impact of visual information on music perception and memory. *Music Perception*, *27*(1), 43-59.

Boltz, M., Schulkind, M., & Kantra, S. (1991). Effects of background music on the remembering of filmed events. *Memory & cognition*, *19*(6), 593-606.

Brattico, E., & Pearce, M. (2013). The neuroaesthetics of music. *Psychology of Aesthetics, Creativity, and the Arts*, *7*(1), 48.

Braud, W. (2001). Experiencing tears of wonder-joy: Seeing with the heart's eye. *Journal of Transpersonal Psychology*, *33*(2), 99-112.

Carroll, N., & Carroll, N. E. (1996). *Theorizing the moving image*. Cambridge University Press.

Chaffin, R., Logan, T. R., Begosh, K. T., Cross, I., & Thaut, M. (2009). Performing from memory. *The Oxford handbook of music psychology*, 352-363.

Chan, L. P., Livingstone, S. R., & Russo, F. A. (2012). Facial mimicry in response to song. *Music Perception: An Interdisciplinary Journal*, *30*(4), 361-367.

Chen, Y. C., & Spence, C. (2017). Assessing the role of the 'unity assumption'on multisensory integration: A review. *Frontiers in psychology*, *8*, 445.

Cohen, A. J. (2001). Music as a source of emotion in film. *Music and emotion: Theory and research*. Oxford University Press.

Cohen A. J. (2008). Music in performance art: Film, Theatre, and dance. *The Oxford Handbook Of Music Psychology.* Oxford University Press.

Cohen, A. J. (2013). Congruence-association model of music and multimedia: Origin and evolution. *The psychology of music in multimedia*, 17-47.

Cozby Paul C., Bates, Scott C. (2015). *Methods in Behavioral Research*. McGraw-Hill Education.

Dickson, G. T., & Schubert, E. (2020). Music on Prescription to Aid Sleep Quality: A Literature Review. *Frontiers in psychology*, *11*.

Dimberg, U., Thunberg, M., & Elmehed, K. (2000). Unconscious facial reactions to emotional facial expressions. *Psychological science*, *11*(1), 86-89.

Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental brain research*, *91*(1), 176-180.

Doyen, S., Klein, O., Pichon, C. L., & Cleeremans, A. (2012). Behavioral priming: it's all in the mind, but whose mind?. *PloS one*, *7*(1), e29081.

Eerola, T., & Vuoskoski, J. K. (2011). A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music*, *39*(1), 18-49.

Eifert, G. H., Craill, L., Carey, E., & O'Connor, C. (1988). Affect modification through evaluative conditioning with music. *Behaviour research and Therapy*, *26*(4), 321-330.

Ekman, P. (1992). Are there basic emotions? *Psychological Review*, 99(3), 550–553.

Ekman, P. (1999). Basic emotions. *Handbook of cognition and emotion*, *98*(45-60), 16.

Ekman, P., Levenson, R. W., & Friesen, W. V. (1983). Autonomic nervous system activity distinguishes among emotions. *science*, *221*(4616), 1208-1210.

Eldar, E., Ganor, O., Admon, R., Bleich, A., & Hendler, T. (2007). Feeling the real world: limbic response to music depends on related content. *Cerebral Cortex*, *17*(12), 2828-2840.

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Sage publications.

Flanagan, O. J. (1992). *Consciousness reconsidered* (Vol. 250). Cambridge, MA: mit Press.

Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., ... & Koelsch, S. (2009). Universal recognition of three basic emotions in music. *Current biology*, *19*(7), 573-576.

Gabrielsson, A. (2001). Emotion perceived and emotion felt: Same or different?. *Musicae scientiae*, *5*(1_suppl), 123-147.

Gabrielsson, A. (2010). Strong experiences with music. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, applications* (pp. 547–574). New York, NY: Oxford University Press.

Gabrielsson, A. (2011). *Strong experiences with music: Music is much more than just music.* Oxford University Press.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, *119*(2), 593-609.

Ganis, G., Thompson, W. L., & Kosslyn, S. M. (2004). Brain areas underlying visual mental imagery and visual perception: an fMRI study. *Cognitive Brain Research*, *20*(2), 226-241.

Garcia, E. E. (2005, January). Scriabin's Mysterium and the birth of genius. In *Mid-winter meeting of the American psychoanalytic association*.

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.

Geringer, J. M., Cassidy, J. W., & Byo, J. L. (1997). Nonmusic majors' cognitive and affective responses to performance and programmatic music videos. *Journal of Research in Music Education*, *45*(2), 221-233.

Goerlich, K. S., Witteman, J., Schiller, N. O., Van Heuven, V. J., Aleman, A., & Martens, S. (2012). The nature of affective priming in music and speech. *Journal of cognitive neuroscience*, *24*(8), 1725-1741.

Goldberg, M. E., Chattopadhyay, A., Gorn, G. J., & Rosenblatt, J. (1993). Music, music videos, and wear out. *Psychology & Marketing*, *10*(1), 1-13.

Goldstein, A. (1980). Thrills in response to music and other stimuli. *Physiological Psychology*, *8*(1), 126-129.

Gorn, G. J. (1982). The effects of music in advertising on choice behavior: A classical conditioning approach. *Journal of marketing*, *46*(1), 94-101.

Grewe, O., Nagel, F., Kopiez, R., & Altenmüller, E. (2007). Listening to music as a re-creative process: Physiological, psychological, and psychoacoustical correlates of chills and strong emotions. *Music Perception*, *24*(3), 297-314.

Grewe, O., Kopiez, R., & Altenmüüller, E. (2009). The chill parameter: Goose bumps and shivers as promising measures in emotion research. *Music Perception*, *27*(1), 61-74.

Guldbrandsen, Erling E. (2012). *Vestens Musikkhistorie: fra 1600-vår tid.* Cappelen Damm AS.

Huron, D. (2006). *Sweet Anticipation: Music and the Psychology of Expectation*. Massachusetts Institute of Technology.

Iwamiya, S. I. (2013). Perceived congruence between auditory and visual elements in multimedia. *The psychology of music in multimedia*, 141-164.

Juslin, P. N. (2019). *Musical emotions explained: Unlocking the secrets of musical affect*. Oxford University Press, USA.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code?. *Psychological bulletin*, *129*(5), 770.

Juslin, P. N., Liljeström, S., Västfjäll, D., Barradas, G., & Silva, A. (2008). An experience sampling study of emotional reactions to music: listener, music, and situation. *Emotion*, *8*(5), 668.

Juslin, P. N., & Vastfjall, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and brain sciences*, *31*(5), 559.

Keltner, D., & Haidt, J. (2003). Approaching awe, a moral, spiritual, and aesthetic emotion. *Cognition and emotion*, *17*(2), 297-314.

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A. D. (2004). Music, language and meaning: brain signatures of semantic processing. *Nature neuroscience*, *7*(3), 302-307.

Konecni, V. J. (2005). The aesthetic trinity: Awe, being moved, thrills. *Bulletin of Psychology and the Arts*, *5*(2), 27-44.

Krahé, C., Hahn, U., & Whitney, K. (2015). Is seeing (musical) believing? The eye versus the ear in emotional responses to music. *Psychology of Music*, *43*(1), 140-148.

Krugman, H. E. (1943). Affective response to music as a function of familiarity. *The Journal of Abnormal and Social Psychology*, *38*(3), 388.

Küssner, M. B., & Eerola, T. (2019). The content and functions of vivid and soothing visual imagery during music listening: Findings from a survey study. *Psychomusicology: Music, Mind, and Brain*, *29*(2-3), 90.

Lalanne, C., & Lorenceau, J. (2004). Crossmodal integration for perception and action. *Journal of Physiology-Paris*, *98*(1-3), 265-279.

Lavy, M. M. (2001). Emotion and the experience of listening to music: A framework for empirical research (Doctoral dissertation, University of Cambridge).

LeDoux, J. E. (1995). Emotion: Clues from the brain. *Annual review of psychology*, *46*(1), 209-235.

Livingstone, S. R., Thompson, W. F., Wanderley, M. M., & Palmer, C. (2015). Common cues to emotion in the dynamic facial expressions of speech and song. *The Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 68, 952–970.

Logeswaran, N., & Bhattacharya, J. (2009). Crossmodal transfer of emotion by music. *Neuroscience letters*, *455*(2), 129-133.

Maess, B., Koelsch, S., Gunter, T. C., & Friederici, A. D. (2001). Musical syntax is processed in Broca's area: an MEG study. *Nature neuroscience*, *4*(5), 540-545.

Margulis, E. H. (2017). An exploratory study of narrative experiences of music. *Music Perception: An Interdisciplinary Journal*, *35*(2), 235-248.

Margulis, E. H., Wong, P. C., Simchy-Gross, R., & McAuley, J. D. (2019). What the music said: Narrative listening across cultures. *Palgrave Communications*, *5*(1), 1-8.

McCrae, R. R. (2007). Aesthetic chills as a universal marker of openness to experience. *Motivation and Emotion*, *31*(1), 5-11.

Menninghaus, W., Wagner, V., Hanich, J., Wassiliwizky, E., Kuehnast, M., & Jacobsen, T. (2015). Towards a psychological construct of being moved. *PloS one*, *10*(6), e0128451.

Molenberghs, P., Cunnington, R., & Mattingley, J. B. (2009). Is the mirror neuron system involved in imitation? A short review and meta-analysis. *Neuroscience & biobehavioral reviews*, *33*(7), 975-980.

Molnar-Szakacs, I., Green Assuied, V., & Overy, K. (2011). Shared affective motion experience (SAME) and creative, interactive music therapy. *Musical Imaginations: Multidisciplinary Perspectives on Creativity, Performance and Perception.*, 313-331.

Moutsopoulou, K., Pfeuffer, C., Kiesel, A., Yang, Q., & Waszak, F. (2019). How long is long-term priming? Classification and action priming in the scale of days. *Quarterly Journal of Experimental Psychology*, *72*(5), 1183-1199.

Mull, H. K. (1957). The effect of repetition upon the enjoyment of modern music. *The Journal of Psychology*, *43*(1), 155-162.

Musch, J., & Klauer, K. C. (Eds.). (2003). *The psychology of evaluation: Affective processes in cognition and emotion*. Psychology Press.

North, A. C., & Hargreaves, D. J. (1995). Subjective complexity, familiarity, and liking for popular music. *Psychomusicology: A Journal of Research in Music Cognition*, *14*(1-2), 77.

Nusbaum, E. C., & Silvia, P. J. (2011). Shivers and timbres: Personality and the experience of chills from music. *Social Psychological and Personality Science*, *2*(2), 199-204.

Panksepp, J. (1995). The emotional sources of" chills" induced by music. *Music perception*, *13*(2), 171-207.

Pashler, H., Coburn, N., & Harris, C. R. (2012). Priming of social distance? Failure to replicate effects on social and food judgments. *PloS one*, *7*(8), e42510.

Pavlov, P. I. (2010). Conditioned reflexes: an investigation of the physiological activity of the cerebral cortex. *Annals of neurosciences*, *17*(3), 136.

Peretz, I., Gaudreau, D., & Bonnel, A. M. (1998). Exposure effects on music preference and recognition. *Memory & Cognition*, *26*(5), 884-902.

Piff, P. K., Dietze, P., Feinberg, M., Stancato, D. M., & Keltner, D. (2015). Awe, the small self, and prosocial behavior. *Journal of personality and social psychology*, *108*(6), 883.

Polkinghorne, D. E. (1988). Narrative knowing and the human sciences. Suny Press.

Quittner, A. (1983). on Visual Imagery: A Multiple Measures Approach. *Journal of Mental Imagery*, *7*(1), 105-120.

Razran, G. (1954). The conditioned evocation of attitudes (cognitive conditioning?). *Journal of Experimental Psychology*, *48*(4), 278.

Rizzolatti, Giacomo, et al (1996). "Premotor cortex and the recognition of motor actions." *Cognitive brain research* 3.(2), 131-141.

Rubin, D. C. (2006). The basic-systems model of episodic memory. *Perspectives on psychological science*, *1*(4), 277-311.

Satoh, M., Takeda, K., Nagata, K., Hatazawa, J., & Kuzuhara, S. (2001). Activated brain regions in musicians during an ensemble: a PET study. *Cognitive Brain Research*, *12*(1), 101-108.

Schacter, D. L., & Buckner, R. L. (1998). Priming and the brain. *Neuron*, *20*(2), 185-195.

Schmithorst, V. J. (2005). Separate cortical networks involved in music perception: preliminary functional MRI evidence for modularity of music processing. *Neuroimage*, *25*(2), 444-451

Schmithorst, V. J., Holland, S. K., & Plante, E. (2006). Cognitive modules utilized for narrative comprehension in children: a functional magnetic resonance imaging study. *Neuroimage*, *29*(1), 254-266.

Schmithorst, V. J., Holland, S. K., & Plante, E. (2007). Development of effective connectivity for narrative comprehension in children. *Neuroreport*, *18*(14), 1411.

Schubert, T. W., Zickfeld, J. H., Seibt, B., & Fiske, A. P. (2018). Moment-to-moment changes in feeling moved match changes in closeness, tears, goosebumps, and warmth: Time series analyses. *Cognition and Emotion*, *32*(1), 174-184.

Schurtz, D. R., Blincoe, S., Smith, R. H., Powell, C. A., Combs, D. J., & Kim, S. H. (2012). Exploring the social aspects of goose bumps and their role in awe and envy. *Motivation and Emotion*, *36*(2), 205-217.

Schutz, M., & Lipscomb, S. (2007). Hearing gestures, seeing music: Vision influences perceived tone duration. *Perception*, *36*(6), 888-897.

Shiota, M. N., Keltner, D., & Mossman, A. (2007). The nature of awe: Elicitors, appraisals, and effects on self-concept. *Cognition and emotion*, *21*(5), 944-963.

Sloboda, J. A. (1991). Music structure and emotional response: Some empirical findings. *Psychology of music*, *19*(2), 110-120.

Snyder, B. (2000). *Music and memory: An introduction*. MIT press.

Sollberge, B., Rebe, R., & Eckstein, D. (2003). Musical chords as affective priming context in a word-evaluation task. *Music Perception*, *20*(3), 263-282.

Tan, S. L., & Kelly, M. E. (2004). Graphic representations of short musical compositions. *Psychology of Music*, *32*(2), 191-212.

Talmi, D., Schimmack, U., Paterson, T., & Moscovitch, M. (2007). The role of attention and relatedness in emotionally enhanced memory. *Emotion*, *7*(1), 89.

Tillmann, B., & Bigand, E. (2002). A comparative review of priming effects in language and music. *Advances In Consciousness Research*, *35*, 231-240.

Thompson, S. (2006). Audience responses to a live orchestral concert. *Musicae Scientiae*, *10*(2), 215-244.

Thompson, W. F., Graham, P., & Russo, F. A. (2005). Seeing music performance: Visual influences on perception and experience. *Semiotica, 2005*(156), 203-227.

Thompson, W. F., Russo, F. A., & Quinto, L. (2008). Audio-visual integration of emotional cues in song. *Cognition and Emotion*, *22*(8), 1457-1470.

Tiippana, K. (2014). What is the McGurk effect?. *Frontiers in Psychology*, *5*, 725.

Vansteenwegen, D., Francken, G., Vervliet, B., De Clercq, A., & Eelen, P. (2006). Resistance to extinction in evaluative conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, *32*(1), 71.

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., & Levitin, D. J. (2006). Cross-modal interactions in the perception of musical performance. *Cognition*, *101*(1), 80-113.

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., Dalca, I. M., & Levitin, D. J. (2011). Music to my eyes: Cross-modal interactions in the perception of emotions in musical performance. *Cognition*, *118*(2), 157-170.

Vuoskoski, J. K., & Eerola, T. (2015). Extramusical information contributes to emotions induced by music. *Psychology of Music*, *43*(2), 262-274.

Vuoskoski, J. K., & Eerola, T. (2017). The pleasure evoked by sad music is mediated by feelings of being moved. *Frontiers in psychology*, *8*, 439.

Vuoskoski, J. K., Thompson, W. F., McIlwain, D., & Eerola, T. (2011). Who enjoys listening to sad music and why?. *Music Perception*, *29*(3), 311-317.

Was, C., Woltz, D., & Hirsch, D. (2019). Memory processes underlying long-term semantic priming. *Memory & cognition*, *47*(2), 313-325.

Wassiliwizky, E., Wagner, V., Jacobsen, T., & Menninghaus, W. (2015). Art-elicited chills indicate states of being moved. *Psychology of Aesthetics, Creativity, and the Arts*, *9*(4), 405.

A lion called Christian The whole Documentary Full length. Seen: 21.04.2020. https://www.youtube.com/watch?v=EZ-da0AZcRU

Appendix

Consent form

Dear participant,

Thank you for taking part in this study

The following test is a part of a study about the relationship between film and music. The Study is conducted by Elias Gram-Nilsen, student at the Department of Musicology (IMV) at the University of Oslo. This study is a part of my master's thesis.

Through the course of this survey you'll be presented with one video without sound, some videos accompanied by music, as well as some musical examples without video. After watching/listening to an example you'll be required to answer questions related to your experience of it. Please don't skip to the questions before you're done seeing, or listening to each example. You'll also be asked to solve a small task in the middle of the test, which has nothing to do with music or film. You'll get more information about the aims of the study after you have completed all the questions.

The whole test takes approximately 15 minutes.
The videos are not optimized for mobile, so take the test on a computer or Mac if possible.
Please use headset or earplugs under the whole survey.

The study is anonymous and the answers will, together with data from other participants, be a part of a larger statistical collection. No personal information will be stored, and your identity can't be traced.

You have the right to withdraw from the study at any time, without offering a reason for withdrawal.

If you have any questions, please contact Elias Gram-Nilsen: eliasg@student.imv.uio.no

◉ I have understood what my participation entails and my right to withdraw from the study.

Filler task presented in the middle of the survey

Before you'll be presented with new examples, you need to do this task to refocus your mind.

What is the next number in the line?

- ○ A= 26
- ○ B= 21
- ○ C= 22
- ○ D= 23
- ○ E= 24
- ○ F= 25

Demographics and Musical Background

What is your gender?

- ◯ Mann
- ◯ Women
- ◯ Other

What is your age?

What is your nationality?

What is your musical background?

- ◯ I dont like music
- ◯ I think music is okay
- ◯ I like music a lot
- ◯ I like music a lot and play an instrument at an amateur level
- ◯ I like music a lot and play an instrument at a competent level
- ◯ I like music a lot and am a musician

Familiarity with Stimuli and Comments about the Study

Have you seen any of the videos or heard any of the musical examples before?

- ☐ Christian the Lion (video with a lion embracing to menn)
- ☐ Mike Parson (video with a man surfing a large wave)
- ☐ Steve Reich: Music for 18 musicians slowed down 800% (Ambient music example)
- ☐ Phillip Glass: Company 3. Movement (Classical music example)
- ☐ Ian Post: Dream (Poptune)
- ☐ I may have heard some of the examples abowe
- ☐ No, i have not seen or heard any of the examples before

Did you experience any problems with playing any of the examples?

Do you have any comments about the study

Stimuli

Music Excerpt 1 (music excerpts were uploaded as mp3 files directly on Qualtrics)
https://www.youtube.com/watch?v=6rqfdMW5ys0&feature=youtu.be&ab_channel=Masterproject

Music Excerpt 2 (music excerpts were uploaded as mp3 files directly on Qualtrics)
https://www.youtube.com/watch?v=8Vp_eC0LdvE&ab_channel=Masterproject

Music Excerpt 3 (music excerpts were uploaded as mp3 files directly on Qualtrics)
https://www.youtube.com/watch?v=mXFzl84SV74&feature=youtu.be&ab_channel=Masterproject

Video 1 Alone
https://www.youtube.com/watch?v=kPzWRlbhtt4&ab_channel=Masterproject

Video 2 Alone
https://www.youtube.com/watch?v=8KYWaUgcmZs&ab_channel=Masterproject

Music 1 and Video 1
https://www.youtube.com/watch?v=sXwP6wz46xE&ab_channel=Masterproject

Music 1 and Video 2
https://www.youtube.com/watch?v=9gv-TEeSPI0&ab_channel=Masterproject

Music 2 and Video 1
https://www.youtube.com/watch?v=Ox6M3S90qWk&ab_channel=Masterproject

Music 2 and Video 2
https://www.youtube.com/watch?v=-5D8rlOXv8o&ab_channel=Masterproject

Music 3 and Video 1
https://www.youtube.com/watch?v=pLv47Vr80E0&feature=youtu.be&ab_channel=Masterproject

Music 3 and Video 2
https://www.youtube.com/watch?v=-tVO8MNebpE&ab_channel=Masterproject

Tables from the Regression Analysis

Awe:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.894[a]	.800	.666	4.33889

a. Predictors: (Constant), MusicAlone, VideoAlone

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	8.765	12.337		.710	.529
	VideoAlone	.895	.260	.889	3.443	.041
	MusicAlone	-.070	.191	-.095	-.366	.738

a. Dependent Variable: VideoAndMusic

Being Moved:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.686[a]	.470	.117	11.62054

a. Predictors: (Constant), MusicAlone, VideoAlone

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	49.012	36.250		1.352	.269
	VideoAlone	.686	.529	.545	1.297	.285
	MusicAlone	-.737	.746	-.416	-.989	.396

a. Dependent Variable: VideoAndMusic

Valence:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.976[a]	.952	.920	6.11136

a. Predictors: (Constant), MusicAlone, VideoAlone

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	-29.523	15.493		-1.906	.153
	VideoAlone	-.066	.132	-.063	-.500	.651
	MusicAlone	1.383	.180	.974	7.701	.005

a. Dependent Variable: VideoAndMusic

Arousal:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.929[a]	.864	.773	7.59158

a. Predictors: (Constant), MusicAlone, VideoAlone

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	-96.878	72.784		-1.331	.275
	VideoAlone	1.421	1.163	.261	1.222	.309
	MusicAlone	1.105	.264	.892	4.182	.025

a. Dependent Variable: VideoAndMusic