

FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

# **Matching emotional states through auditory information: can music induce visual attention?**

**Rita Isabel Belo Duarte Aires Moreira**



Master in Multimedia  
Specialization in Interactive Music and Sound Design

Supervisor: Dr. Inês Salselas Cabral

July 28, 2020



# **Matching emotional states through auditory information: can music induce visual attention?**

**Rita Isabel Belo Duarte Aires Moreira**

Master in Multimedia  
Specialization in Interactive Music and Sound Design

Approved in oral examination by the committee:

President: Gilberto Bernardes de Almeida, Assistant Professor

External Examiner: Daniela da Costa Coimbra, Adjunct Professor

Supervisor: Inês Salselas Cabral, Postdoctoral Fellow

July 28, 2020



# Abstract

The purpose of this study is to investigate whether music, a fundamental element of sound in the movies, can capture the viewer's visual attention.

Sound and image, in the cinematic experience, are two components that become inseparable, providing information and transmitting emotions. Indeed, image has a very strong impact on the viewer. However, sound that is frequently unnoticed and neglected, has too an undeniable potential to convey emotional content and, consequently, be a primal element that captures visual attention. In parallel, there have been major technological changes in cinema that aim to improve the cinematic experience. The concept of immersion has grown and, as these changes affect the way visual content is presented, the role of sound has also been questioned aiming a higher level of involvement of the viewer in the narrative. This creates new challenges to capture the viewer's attention and convey a certain narrative. With this in mind, an experiment was designed and conducted to test if music, in an audiovisual context, could bias visual attention. Accordingly, in an online setup, participants' preferences we have tested over two simultaneous video scenes combined with a selected composed song for this purpose. Results shown that participants' preferences differed with different musical contexts. Therefore, we conclude that music constituted a factor that induced visual attention. This is most likely to happen when the condition of consistency between conveyed emotion by music and video is met.

**Keywords:** Music, emotion, attention



# Resumo

O objetivo deste estudo é investigar se a música, um elemento fundamental do som nos filmes, pode capturar a atenção visual do espectador.

O som e a imagem, na experiência cinematográfica, são dois componentes que se tornam inseparáveis, fornecendo informações e transmitindo emoções. De fato, a imagem tem um impacto muito forte no visualizador. No entanto, o som que é frequentemente despercebido e negligenciado, tem um potencial inegável para transmitir conteúdo emocional e, conseqüentemente, é um elemento primordial que capta a atenção visual. Paralelamente, houve grandes mudanças tecnológicas no cinema que visam melhorar a experiência cinematográfica. O conceito de imersão cresceu e, como essas mudanças afetam a forma como os conteúdos visuais são apresentados, o papel do som também foi questionado, visando um maior nível de envolvimento do espectador na narrativa. Isto cria novos desafios para capturar a atenção do espectador e transmitir uma certa narrativa. Tendo isto em mente, projetamos e conduzimos um experimento para testar se a música, em um contexto audiovisual, poderia influenciar a atenção visual. Conseqüentemente, numa configuração online, testamos as preferências dos participantes em duas cenas de vídeo simultâneas combinadas com uma música composta selecionada para esse fim. Os resultados mostraram que as preferências dos participantes diferiam em diferentes contextos musicais. Portanto, concluímos que a música constituiu um fator que induziu a atenção visual. É mais provável que isto aconteça quando a condição de consistência entre a emoção transmitida pela música e pelo vídeo coincidem.

**Palavras-chave:** Música, emoção, atenção





# Acknowledgements

I want to thank all those who were present and helped me along this path, not only during the development of this dissertation but throughout my master's. Thank you to my supervisor Dr. Inês Salselas that helped me and guided me along this journey.

I would like to express my gratitude to :

- My family, specially to my mom and brothers, João and José, for the unconditional support and for always being available to help me and motivate me. Thank you once again to them for being fantastic actors and an amazing cameraman;
- To my multimedia friends for this fantastic two years;
- To my friend Miguel for always being there for me, we walked this path together;
- To my friends, for always giving me good advices.

Rita Moreira



*“There’s a sadness to the human condition that I think music is good for. It gives a counterpoint to the visual beauty, and adds depth to pictures that they wouldn’t have if the music wasn’t there”*

Mike Figgis



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Structure of the Dissertation . . . . .	1
<b>2</b>	<b>State of Art</b>	<b>3</b>
2.1	Nonverbal communication . . . . .	4
2.2	Sound on Screen . . . . .	6
2.2.1	From the silent era - the arrival of talkies . . . . .	6
2.2.2	Narrative . . . . .	7
2.2.3	How sound and image are perceived . . . . .	9
2.3	Music and emotion . . . . .	10
2.3.1	Music as a form of communication . . . . .	10
2.3.2	Emotion in music . . . . .	11
2.3.3	Musical elements and emotion . . . . .	11
2.3.4	Music Emotion Classification . . . . .	19
2.3.5	Empathetic vs anempathetic music . . . . .	23
2.4	Attention . . . . .	24
2.4.1	Orienting attention . . . . .	24
2.4.2	Executive attention . . . . .	25
2.4.3	Alerting attention . . . . .	25
2.4.4	Emotional attention . . . . .	25
2.5	Summary . . . . .	26
2.6	Problem statement . . . . .	26
2.7	Aims of the study . . . . .	27
<b>3</b>	<b>Methods</b>	<b>29</b>
3.1	Participants . . . . .	29
3.2	Stimuli . . . . .	30
3.3	Design . . . . .	32
3.4	Procedure . . . . .	33
<b>4</b>	<b>Results and analysis</b>	<b>37</b>
4.1	Videos without sound . . . . .	37
4.2	Videos with sound . . . . .	38
4.3	All musical contexts . . . . .	39
4.4	Discussion . . . . .	43
4.4.1	Main findings . . . . .	43
4.4.2	Comparison with existing literature . . . . .	43

<b>5 Conclusion and Future work</b>	<b>45</b>
5.1 Future work . . . . .	45
<b>A Questionnaire</b>	<b>47</b>
<b>References</b>	<b>51</b>

# List of Figures

2.1	The 35 successfully facial expressions used to transmit emotive information across cultures (Srinivasan and Martinez, 2018) . . . . .	5
2.2	An overview of discriminative body movements and postures from Wallbott (1998)	6
2.3	Dynamics terms and symbols from : <a href="https://www.pinterest.pt/pin/268949408973103240/">https://www.pinterest.pt/pin/268949408973103240/</a> . . . . .	12
2.4	Musical articulations from: <a href="https://www.dummies.com/art-center/music/piano/how-to-articulate-your-piano-playing/">https://www.dummies.com/art-center/music/piano/how-to-articulate-your-piano-playing/</a> . . . . .	13
2.5	Gradual dynamic markings from : <a href="https://www.aboutmusictheory.com/music-dynamics.html">https://www.aboutmusictheory.com/music-dynamics.html</a> . . . . .	13
2.6	Strong and weak beats in 4/4 from: <a href="https://blog.landr.com/syncopation/">https://blog.landr.com/syncopation/</a>	14
2.7	Crusis, anacrusis and metacrusis (Cleland and Dobrea-Grindahl (2013)) . . . . .	14
2.8	Beats in a bar with a time signature of C or 4/4 from: <a href="https://www.pianotheoryexercises.com/reading-music/measures-bars/">https://www.pianotheoryexercises.com/reading-music/measures-bars/</a> . . . . .	14
2.9	Types of bars from: <a href="https://www.mightyexpert.com/what-is-a-bar-music/">https://www.mightyexpert.com/what-is-a-bar-music/</a>	15
2.10	Syncopation from: <a href="https://www.teoria.com/es/referencia/s/sincopa.php">https://www.teoria.com/es/referencia/s/sincopa.php</a> . . . . .	15
2.11	Tempo Markings (speed) and changes of tempo from: <a href="https://www.theflutecoach.com/tempo-markings/">https://www.theflutecoach.com/tempo-markings/</a> . . . . .	16
2.12	Annotated circumplex emotion model (Fernández-Sotos et al., 2016) . . . . .	17
2.13	Harmony and melody from: <a href="https://musicskills101.com/what-is-the-difference-between-harmony-and-melody/">https://musicskills101.com/what-is-the-difference-between-harmony-and-melody/</a>	
2.14	Chord inversions from: <a href="https://www.secretsofsongwriting.com/2015/07/15/how-to-use-chord-inversions-to-greatest-effect/">https://www.secretsofsongwriting.com/2015/07/15/how-to-use-chord-inversions-to-greatest-effect/</a> . . . . .	18
2.15	Six of the basic emotional faces (Ekman and Friesen, 2003). From top left: anger, fear, disgust, surprise, happiness, sadness. . . . .	20
2.16	Adjective circle by Kate Hevner (Hevner, 1936) . . . . .	20
2.17	James Russell’s Circumflex Model (Seo and Huh, 2019) . . . . .	21
2.18	CYANITE analysis of several songs . . . . .	22
2.19	Classification of songs by energy and valence based on the human emotional Pleasure-Arousal-Model by Russell (1980) from Helmholtz et al. (2017) . . . . .	22
3.1	M.T. and M.R. CYANITE classification from <a href="https://cyanite.ai">https://cyanite.ai</a> . . . . .	31
3.2	M.T. characteristics by CYANITE from <a href="https://cyanite.ai">https://cyanite.ai</a> . . . . .	31
3.3	M.R. characteristics by CYANITE from <a href="https://cyanite.ai">https://cyanite.ai</a> . . . . .	32
3.4	Study setup . . . . .	32
4.1	Preference in videos without sound . . . . .	38
4.2	Preference in videos with tense music . . . . .	39
4.3	Preference in videos with relaxing music . . . . .	39

4.4	Preference of stressful vs calm videos without music, with tense music and with calm music . . . . .	40
4.5	Preference of stressful scenes with M.T. or M.R. . . . .	40
4.6	Preference of calm scenes with M.T. or M.R. . . . .	41
4.7	Without Sound vs M.R. vs M.T. for all combinations . . . . .	42
A.1	Questionnaire: introduction . . . . .	47
A.2	Questionnaire: sound test . . . . .	48
A.3	Questionnaire: demographic data 1 . . . . .	48
A.4	Questionnaire: demographic data 2 . . . . .	49
A.5	Questionnaire: Part I - example of video without sound . . . . .	49
A.6	Questionnaire: Part II - example of video with sound . . . . .	50
A.7	Questionnaire: Part III - final questions . . . . .	50



# List of Tables

3.1	Video combinations without sound . . . . .	34
3.2	Video combinations with sound . . . . .	35



# Abbreviations and Symbols

AI	Artificial Intelligence
AR	Augmented Reality
bpm	beats per minute
DAW	Digital Audio Workstation
FBAUP	<i>Faculdade de Belas Artes da Universidade do Porto</i>
FCUP	<i>Faculdade de Ciências da Universidade do Porto</i>
FEUP	<i>Faculdade de Engenharia da Universidade do Porto</i>
FLUP	<i>Faculdade de Letras da Universidade do Porto</i>
IBM	International Business Machines
MIDI	Musical Instrument Digital Interface
PAD	Pleasure Arousal Dominance
MR	Mixed Reality
VR	Virtual Reality



# Chapter 1

## Introduction

Music has always been present in our lives through which ideas and emotions are transmitted, being a form of communication used worldwide (Hargreaves et al., 2005). Several studies demonstrate its importance, benefits and effects in various domains (Cohen et al. (2002), North et al. (2000) Petress (2005), Batt-Rawden (2010)). The emergence of new media makes music accessible to a greater number of people, but even so it remains an underestimated media (Iuppa and Borst, 2012), which often goes unnoticed in a society where the visual predominates (Storr, 2015). Despite this, music and sound have always had an important role in multimedia content, even before technological developments, as silent films were accompanied by music which means that silent movies were not truly silent (Miller, 1982).

Cinema is a multisensorial media and music is an important element in the cinematic experience. Music has a huge contribution in the emotional effect that the film has on the audience. Sound for audiovisual content is an area that is constantly growing, from the establishment of talkies with “The Jazz Singer” (Tankel, 1978) until the appearance of *Dolby Atmos* with “Brave” (Wright, 2015). This does not only takes place in the world of cinema, but also in video-games and, more recently, in immersive realities, like augmented reality (AR) and virtual reality (VR).

Nowadays the use of audio in cinema and games is increasing and audiovisual content makers are embracing new technologies to develop immersive experiences that require new skills and new sound design strategies. Thus, difficulties arise in ensuring that the viewer follows a certain narrative or that his attention is focused on what is intended. In this way, this study explores pathways in which music can elevate and focus visual attention through emotion.

### 1.1 Structure of the Dissertation

This dissertation is constituted by five chapters. The first chapter presents an introduction to what will be discussed, researched and analyzed.

In Chapter 2 is presented the state of art of the project, from the silent era to the establishment of sound on screen, where the focus is music and its power of conveying emotions. These three concepts will be related: music, attention and emotion. Although this work is centered on the

sound element, image has an undeniable important role in the cinematographic world. Moreover, before the introduction of synchronized sound, it was largely through non-verbal communication that emotions were transmitted, continuing today to be one of the main means of transmitting information. For these reasons, this chapter begins with a brief connection between emotion and non-verbal communication. At the end of this chapter, the problem and the objective of the thesis are presented.

Methods appear in chapter 3, where the procedure for conducting the study is described in order to answer the research question. Starting with the presentation of the participants, the materials used, such as videos and music, the design of the study and how it was presented to the participants, the information collection instrument and how this data was analyzed.

In chapter 4 the results are presented for the three different musical contexts and the discussion, relating the main findings to the existing literature.

Finally, the last chapter (5) presents the conclusion about the research done and suggestions for future work.

## Chapter 2

### State of Art

This chapter will focus on the current state of art and previous studies related to the aims of this study. The central point is to explore ways in which music, attention and emotion are attached and how they influence each other. Not everything we hear in the movies is actually produced by the source we see. The truth is that sound can affect reality, creating illusions and affecting how we feel and our perception. A research by Konečni concludes that music generates emotions and through them sound engage the viewer ([Konečni, 2008](#)). But why the role of emotion is so important? What we decide to pay attention determines our experiences, “my experience is what I agree to attend to” ([James et al., 1890](#)). In the case of audiovisual content, we think that we can control our attention, but this doesn’t always happen. The audience is influenced to focus attention upon what audiovisual content creators desire. However, what strategies do they have available to capture the audience’s attention?

Music is an important factor in films, as a way of involving the audience with the narrative through emotions. Previous studies conclude that emotional significant stimuli modulate our attention and are much more relevant to us than neutral stimuli, so next will be presenting attention and its relationship with emotion ([Nicks and Carriou, 2016](#)).

## 2.1 Nonverbal communication

Before the introduction of synchronized sound, the activity of acting consisted of communicating emotions through body language. This is the process of telling a story without resorting to the use of words, only communicating and transmitting emotions through physical behavior. Studies indicate that most information is transmitted through facial expression (55%) and only a small part through words (7%). The other percentage is related to the way things are said (38%). This means that 93% of the information corresponds to nonverbal communication (Mehrabian, 1972). Mehrabian et al. (1971) affirms that we can't separate our action from feelings, and we are caught up in the lie when what we say doesn't match our body language, so the body tells the truth. "The fact is that body language is more honest than words" (Pease and Pease, 2008, P.94).

The nonverbal communication can be divided into three distinct parts (Wang, 2016):

- *Kinesics* - body language, facial expression, gestures
- *Proxemics* - body distance between people in communication
- *Haptic* - interaction involving touch

The use of this type of language is essential for any actor to interpret the character, convey emotions and emphasize with the audience. This also can change the way audience interpret the scene. For example exaggerated body language will give a more comic character to the scene so is often seen in comedies. Also other types of body language are made especially for certain non-real characters, as is the case with animated films. This means that the narrative will determine how the actor will interpret the character, or will act. But the audience can usually decode the emotions that the actors are feeling and often perceive them as if they were real (Zacks, 2015).

Facial expressions are really important when we talk about nonverbal communication. Darwin talk about on his book, that some facial expressions are universal referring to humans and animals (Darwin and Prodger, 1998). Since facial expressions are related to the movement of facial muscles, they can often be confusing, so they are associated with universal emotions such as happiness or sadness, thus being more easily decoded by the audience.

A study conducted in 2018 found that "only 35 [facial expressions] are successfully used to transmit emotive information across cultures, and only 8 within a smaller number of cultures" (Srinivasan and Martinez, 2018, n.p.). These 35 expressions are represented in figure 2.1. Of these 35, they found that 17 were associated with emotions of happiness, which demonstrates the complexity of this emotion, showing that humans express happiness in various ways, and only 1 linked to disgust.

Gestures are also a form of nonverbal communication where the hands, face, arms and other parts are used to transmit messages and "can be a valuable key to an emotion a person may be feeling at the time" (Pease and Pease, 2008, p.11). Gestures are used in sign language to replace speech. This means that we can transmit or recognize a message without the use of speech. Some gestures, like move the head left to right that means no, or shrug that means that the person doesn't





Figure 2.1: The 35 successfully facial expressions used to transmit emotive information across cultures (Srinivasan and Martinez, 2018)

understand are some examples of universal gestures. But other factors like the distance between actors or their posture will convey certain emotions.

A study conducted by Wallbott (1998) analysed body movements and gestures from 224 videos where actors express several emotions. They conclude that certain gestures are related to specific emotions but the posture and movements are also linked to the intensity of emotions (figure 2.2).

Although sound has been established in cinema, as will be presented later, and dialogue is part of the performance, previous studies show that non-verbal communication continues to be extremely important. As the relationship between image and sound will be presented, verbal and non-verbal communication goes hand in hand but the non-verbal is more impacting than the verbal.

Emotion	Upper body	Shoulders	Head	Body Parts Arms	Hands	Movement quality
	Erect	Differences from 'standard' movement and posture pattern observed, the 'standard' pattern being:				No extreme judgments
		Hanging	Toward camera	Hanging	No specific activities	
Cold anger Hot anger		Up		Stretched out frontal Stretched out frontal	Lateralized movements Lateralized movements, pointing, opening/closing, back of hands sideways, many illustrators	High movement activity, expansive movements, high movement dynamics
Elated joy		Up	Backward	Stretched out frontal or upward	Opening/closing, many illustrators	High movement activity, expansive movement, high movement dynamics Low movement dynamics Inexpansive movements
Happiness Disgust	Collapsed	Backward or forward	Downward	Crossed in front of chest		Low movement activity, low movement dynamics Expansive movements
Contempt Sadness Despair Fear	Collapsed	Forward			Opening/closing Opening/closing, many self-manipulators	Low movement activity, low movement dynamics Expansive movements
Terror				Stretched out sideways		High movement activity
Shame Interest Pride	Collapsed		Downward Backward	Stretched out frontal Crossed in front of chest	Many self-manipulators Lateralized movements	
Boredom	Collapsed		Backward			Low movement activity, inexpansive movements, low movement dynamics

Figure 2.2: An overview of discriminative body movements and postures from [Wallbott \(1998\)](#)

## 2.2 Sound on Screen

The advances in technology have revolutionized not only the way films are made, but also the way we experience them, making the movies more realistic and engaging and trying to establish a deeper connection with the audience. According to George Lucas, this connection is centered “on conveying emotions through cinematic experience, not necessarily through the narrative” ([Jones, 2016](#), n.p.). As well, Walter Murch considers that emotion, the first of his “rule of six” for film editing and what he considers most important “is the thing that you should try to preserve at all costs”. Actually, the emotion is directly related to the nature of the human being along with the instinct and logic, people form emotional connections, so the filmmakers need to motivate the user to form these associations, creating emotional experiences ([WebofStories, 2017](#)).

Sound is a powerful tool that can enhance the emotions of any scene and puts the audience into the story ([Holman, 2012](#)). This shows how powerful sound can be in terms of emotional charge and, according to Gary Rydstrom, “sound is still the best way to experience emotion” ([Dogwoof, 2019](#)).

### 2.2.1 From the silent era - the arrival of talkies

Since the inception of film-making, movies were silent, what means no synchronized sound, but usually accompanied by an orchestra, a single player or a phonograph recording. This was used to reinforce the atmosphere of a film and to give more emotion and emphasize important moments since it was not possible to communicate emotion through dialogue or sound design. One of the most famous personalities of the silent movies era was Charlie Chaplin. Although there is no synchronized sound, the music was essential to improve the experience of the viewer [Kobel](#)

(2007), which reinforced the fact that “the silent films were never really silent” (The Telegraph, 2014) and the music had an essential role before the arrival of talkies.

*“There is a point where we just let the music take over everything”*

— Christopher Nolan, *Inception*

The idea of combining synchronized sound with visual image started from the beginning of films, but due to technical issues this was not possible. The synchronized musical score and sound effects came in 1926 in “Don Juan”, a film directed by Alan Crosland, who also directed the first film with synchronized speech, “The Jazz Singer” in 1927. This event marks the transition into the sound era. The age of talkies began and sound became one of the two main stimuli that people experience in cinema. Thus, the filmmakers have not only to tell a good story but have to do it in a way to make it more engaging, this time, using sound as a tool (Kobel, 2007).

The impact of “The Jazz Singer” was enormous, but the first all-talking feature film was produced in 1928, called “The Lights of New York”. In the same year the first cartoon with synchronized sound was released, “Steamboat Willie with Mickey Mouse”, created by Walt Disney, where the use of the technique “Mickey Mousing” starts, a strict correspondence between music and movements.

### 2.2.2 Narrative

We are continually stimulated with information that comes to us through various sensory pathways, but we only pay attention to a very small amount of information. Even with a lot of information coming to us, we are able to center our attention in what is important because our brain acts like a filter (Talsma et al., 2010).

In this way, how do we manage to focus attention to what is really important? We process a lot of information in the form of narrative. Making an audience forget everything and be transported into a story shows the power of a narrative in human world. This experience of being transported by a narrative is called narrative transportation, coined by Gerrig (Herman, 1997).

According to Green and Brock (2000), transportation carries a feel of absorption, involves catching attention and incorporates emotional engagement and loss of information related to the real world. People lose themselves in a narrative because the story can change how people see the world.

Humans are sociable and like to have contact with others because this help them in many different ways, like being understood. This is possible because of communication between humans, machines or between a human and a machine (Wiener, 1950). For Adam Skelter (Courage, 2018), people use stories to comprehend the world to transmit information, making them an universal feature of communication (Costabile et al., 2018).

The way we communicate with others is changing extremely fast, because of the number of platforms where narratives can be told. This growth is due of the advances in technology, as “every new medium has given rise to a new form of narrative” (Rose, 2012, p. 12). First it was

done orally, the people get together and pass oral messages through singing, music, and poetry (Reporter, 2015). Then new forms emerge with the invention of the printing press that although the invention date is not known, The Diamond Sutra is the oldest printed book to date, 868 A.D. (Editors, 2019). Finally in the year 1800 appeared the audiovisual, photography, motion picture, radio, telephones, television and social media (Rose, 2012).

### 2.2.2.1 Narrative forms

Narrative is one of the most important elements in media. The story can be present as a linear narrative, the most common form of narrative, if the story follows a singular line or non-linear if the order of the events is alternated (Miller, 2004). *Un Chien Andalou*, a surrealist silent short film created by Luis Buñuel and Salvador Dalí was the first film that dismantles the linear narrative and took a non-linear approach (Dancyger and Rush, 2007).

In other ways the narrative besides being linear or non-linear can also be interactive. This means that the story changes based on the player's decisions, the user stops being passive to have some power of choice or the illusion of it, in the case of linear narrative. To allow the user to have real choice is necessary to use other types of technologies, like Artificial Intelligence, developed by Alan Turing (Miller, 2004). This has been implemented in games like *Avery*, that uses IBM cloud and Watson technology (MixedBag, 2018).

The first interactive experiences took place long before the rise of internet, but with the recurrent improvement of technologies that made possible the creation of more deeper interactive experiences. Beyond sound, making the audience a part of the narrative makes them feel a part of something, like an accomplishment where they are emotionally involved in the story. For example, in Japan, the use of live narrators, known as *Benshi* made the silent movies interactive (Mannoni, 2019). Also in theatre, the "fourth wall" convention was destroyed. The play "Six Characters in Search of an Author" by Luigi Pirandello uses a non-linear approach to tell the story, where the "actors in an uncompleted play talk and refer to themselves as if they were real people" (Miller, 2004).

### 2.2.2.2 Narrative and sound

Sound has the potential to make the story impactful and remarkable. As we saw earlier the narrative has an important part in films to tell the story, as a "diegetic" way the narrative world refers to the screen world, and in a "non-diegetic" way indicates what is underscore, does not belong to the world (Buhler et al., 2010). These terms are used also to classify sound, if the sound source is evident on the screen the sound is "diegetic", if not the sound is "non-diegetic". In other words the barrier that separates this classification is very tenuous, making it possible for one to become the other, it's a "trans-diegetic" sound. "Play it, Sam" scene from *Casablanca* movie is an example where the music progresses from "diegetic" to "non-diegetic" (Casablanca, 1942). Thus, sound has narrative roles, such as (a) direct narrative role where the sound is used to make the audience understand the story, (b) subliminal narrative role, "the key to an important storytelling power of

sound”, where the sound is perceived unconsciously, an example of this is the film soundtrack, that bring different emotions to the audience, and (c) grammatical role, where the sound has the power of cohesion (Holman, 1999).

Sound for linear narratives has a start and an end, don't change depending on user choices. Karen Collins distinguish the terms interactive, adaptive and dynamic in non-linear video games audio. When the sound is triggered to a player action and it stays the same whenever he repeats it, is interactive. An example of this form is when every time the player makes the character jump, a specific sound is generated. In other hand, adaptive sounds are those who occur due to adjusts in the game and not because of player actions. Dynamic is the combination of both. The player is no longer passive and is linked to the sound of the game.

She also discusses the use of diegetic and non-diegetic sounds and explores how the dynamic activity in games audio can change during the game. On the one hand, diegetic sounds are real sounds that take place in the scene like dialogue and sound effects, and, on the other hand, non-diegetic sounds are normally the music in the background. Just like in movies, in games sometimes non-diegetic sounds can become diegetic and vice-versa. In the words of Collins relative to the implications of dynamic sound, “the direct participation between a player and the audio takes on a new role in kinetic gestural games” and “this changing role from passive to active listening is an important element of sound in games”.

Collins keeps referring that music is used to focus the player and guide is attention through the game, giving cues, predict something and helping the player follow the narrative. The music has an important paper in games. In certain games, the player also has the choice of what music he wants to listen, like in Euro Truck Simulator.

The use of sound is a key element to immerse the player into the narrative and some important elements are use to enhance this immersion, like dialogue, “in a game is the suspension of disbelief, adding realism or creating illusion” (Collins et al., 2007, p. 9) and the sound is used to transport the user from the local where they are into the experience environment, Collins refers this as a way to camouflage the distractions and noises and focus the player in the action.

Finally, the way the player gets emotionally involved is like in the movies, “sound works to control or manipulate the players emotions, guiding responses to the game” (Collins et al., 2007, p. 10) , but in games is a little bit different, since the player is responsible for the character actions and needs to deal in the consequences of his choices (Collins et al., 2007)

### 2.2.3 How sound and image are perceived

Dakic (2009) concludes that “the basic tools of sound design are the imperfections of human cognition and perception of sound” . Several research works have been developed exploring how music affects our perception of visual information. As mentioned earlier, emotions have a huge impact on the interaction between sound and visual perception, as described by Cohen (1999). The combination of musical cues and visual information enhances the emotional response, according to Davidson (1993). In his study, he investigated the body movements of performers while they played the same piece but with differences in artistic behaviors. He concluded that “not only

is vision a useful source of information about manner but that is actually more clearly specifies manner than the other modes” after analyzing the response of the observers to “three modes: vision alone, sound alone and sound and vision together”.

Furthermore, [Chion \(2019\)](#), demonstrates that both, audio and image, influence and transform each other. This "symbiosis" is termed as the “audio-visual contract”. This expression appoints an agreement, where sound and image became a single “entity”.

## 2.3 Music and emotion

### 2.3.1 Music as a form of communication

Music is used as a form of communication, people can share emotions and interact with others through this universal language. Humans resort to music to express different kinds of emotions. For example, “Grândola, Vila Morena” has had a huge impact in *Revolução dos Cravos*, which was used as a password to initiate the military operations to end the dictatorship in Portugal ([R7, 2011](#)). Many famous musicians pass messages through music to their fans, helping people connect with each other, even being from different cultures. Currently with the development of technology and streaming services the facility to communicate with the other side of the world is improved. This happens in movies, where the sound is a really powerful tool for storytelling, and to emotionally connect us ([Cross, 2009](#)).

#### 2.3.1.1 Functions of music in multimedia

[Cohen \(1999\)](#) describes eight functions in multimedia. The first one is *Masking* that hide undesirable sound, “an essential component of the silent film”. “*Provision of continuity* is the second function, that explains that music provides cohesion between sounds referring that “the role of continuity or breaking the action is a standard technique of film music”. Focusing the attention of the viewer is one of the roles of music, thus *direction of attention* is the third function that the author presents. The fourth is *mood induction*. This shows that music can modify how the listeners feel. The emotion part is associated with the fifth function, *communication of meaning*, the music has the power to transform the meaning of a moving image, or “can take on meanings through association with events”. The sixth function such as attention, “*music as a cur for memory*”. This is especially used in advertisements or associate with a specific film. *Arousal and focal attention* is the seventh, where the author refers that music enhances the “attention focus and filter out distractions”, and “improve the intended impact of the multimedia experience, making the experience seem real”. The last one is *Music aesthetics*, the use of music indicates a much better experience since “music is an art form” .

### 2.3.2 Emotion in music

Music plays an important role in our lives, often used to interact with the auditor. Previous researches affirm that music can guide our attention, memory and perception of visual cues. For example, persuade us, enhancing the impact of advertisements (Tan et al., 2013). Music and video are linked together, and both of them influence each other. Music evokes emotion that will influence the viewer's attention and "is "the most efficient code" for emotional expression in film" according to Kalinak (Cohen, 2011, p. 267). Cohen demonstrates that sound is a key element to create emotion in films using the film *Witness* as an example. In this case, the realistic sound was replaced by an unrealistic music, "departures from reality via music makes an episode "more real", more vivid, more emotionally relevant". The author "concludes that music, owing in large part to the explicit knowledge and skills of the composer, provides one of the strongest sources of emotion in film" (Cohen, 2011, p. 268). This is the main intention of the music, transmit emotions Cooke (1959). This process consist of three parts (Juslin and Laukka, 2004) :

- Expression : the emotion that is expressed, for example by the musician that is playing;
- Perception : the emotion that is perceived by the audition, that can be different from the one that is expressed;
- Induction : the emotion that is induced in the listener.

### 2.3.3 Musical elements and emotion

According to Grekow (2018), "timbre, dynamics, rhythm, tempo, and harmony"(p.1) are the musical elements that affect emotions. So to understand the relationship between these elements and emotions we will further develop these music dimensions.

#### 2.3.3.1 Timbre

Timbre is the attribute that allows us to distinguish sounds with the same pitch, loudness and duration. So, timbre, allows to differentiate instruments, being a multidimensional complex property that depends of several parameters of sound such as attack time, decay time, spectral centroid (i.e., brightness), amplitude, spectral flux, among others (Wu et al., 2014; Santoboni and Belardinelli, 2005; Hailstone et al., 2009). Previous studies connected instruments to basic emotions and discovered that some are related with sadness such as plucked string, the piano is neutral and some are tend to be positive, like harpsichord and mallet percussion (Chau et al., 2014). Others relate the violin, trumpet and clarinet to happy emotions and heroic character, whereas emotions like sad and depressed are present by flute and horn. Also found out a correlation between the average spectral centroid with emotional judgment (Wu et al., 2013).

### 2.3.3.2 Dynamics

Dynamics is the variation of loudness between notes, a subjective measure related to the intensity of the sound (Carterette and Friedman, 1978). This element is not measured as the amplitude through decibels since the dynamics are relative, so dynamic symbols (see table 2.3) are used. Dynamics will depend on several aspects such as the dynamic range of the instrument itself, the previous dynamics, the genre of music, the acoustics of the place and the skills of the performer. Therefore, it is understood that in a given instrument a *forte* will not be equal to a *forte* in a different instrument.

The "fortepiano" instrument is the predecessor of the instrument that now called piano, since unlike the other keyboard instruments of the time, he could play dynamics (Paul, 2012). The *forte* means loud and the *piano* quiet. Additionally, the composer can use markings accents when the intention is to change the dynamics of a specific note (2.4). Since dynamics depends on factors such as the type of instrument this will also have implications on the note timing.



Term	Symbol:	Effect:
<b>pianissimo</b>	<i>pp</i>	very soft
<b>piano</b>	<i>p</i>	soft
<b>mezzo piano</b>	<i>mp</i>	moderately soft
<b>mezzo forte</b>	<i>mf</i>	slightly loud
<b>forte</b>	<i>f</i>	loud
<b>fortissimo</b>	<i>ff</i>	very loud
<b>fortepiano</b>	<i>fp</i>	loud then soft
<b>sforzando</b>	<i>sfz</i>	sudden accent
<b>crescendo</b>		gradually louder
<b>diminuendo</b>		gradually softer

Figure 2.3: Dynamics terms and symbols from : <https://www.pinterest.pt/pin/268949408973103240/>

Changes in dynamics can also be made gradually, and for that they are used *crescendos* or *diminuendos* (figure 2.5). consequently, the transition from one dynamic to another is less abrupt.

Several researches studied the influence of dynamics in music emotions. Not all studies come to similar conclusions, Ilie and Thompson (2006) concludes that the participants consider the loudest music to be more pleasant, while other studies don't consent. Despite this, both agree that the loudest music is more tense and energetic. Gundlach (1935) assign the terms "exalted, worthy, melancholy, sad and gloomy" to soft dynamics and "whimsical, glad, and flipp" to louder



Musical Articulations		
Symbol	Name	How to Play the Note
•	Staccato	Short
–	Tenuto	Long
>	Accent	Hard
Λ	Accent (Housetop)	Harder
>•	Accent with staccato	Hard and short
≧	Accent with tenuto	Hard and long

Figure 2.4: Musical articulations from: <https://www.dummies.com/art-center/music/piano/how-to-articulate-your-piano-playing/>

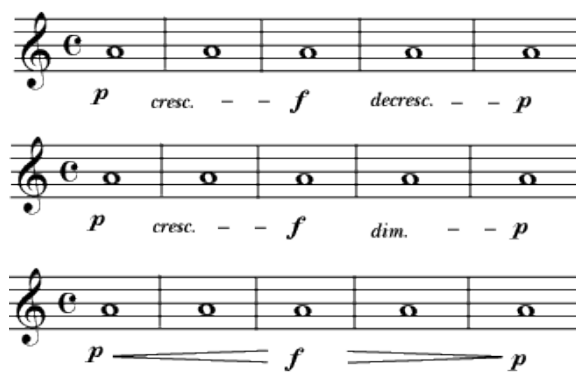


Figure 2.5: Gradual dynamic markings from : <https://www.aboutmusictheory.com/music-dynamics.html>

dynamics (Gundlach, 1935, p.641). With this, researches associate soft dynamics to more calm and sad music, and louder dynamics to more energetic, happy and tense music. Regarding variations in intensity, wider range is related with fear and smaller range with happiness. On the other hand, rapid changes in dynamics are connected with fear and enjoyment, but not many changes with peace and sadness (Gabrielsson and Lindström, 2010).

### 2.3.3.3 Rhythm

Rhythm, although a basic element of music, is a complex concept. According to Deutsch (2013), rhythm is "the order in the movement" (p.150). Other definitions by London (2012) refer that "Rhythm involves patterns of duration that are phenomenally present in the music" (p.4), and when these patterns are associated with a measure, they are called meter, "the measurement of the number of pulses between more or less regularly recurring accents" (Cooper et al., 1963, p.4). Concluding, rhythm is a structured pattern that is repeated all over the music.

Several terms are associated with Rhythm :

- Beat: regular pulse, this term is used when we take into account the metric (Cooper et al. (1963)). The beat is constituted by a "strong", accented, and a "weak" beat, not accented

(figure 2.6). The "downbeat", is the first beat of the measure and the "upbeat" is the last one of the previous measure. Music usually starts with a "downbeat", but it also happens to start with an "anacrusis", note that precedes the first strong tempo, or "crusis" (Cleland and Dobrea-Grindahl (2013)).



Figure 2.6: Strong and weak beats in 4/4 from: <https://blog.landr.com/syncopation/>

Other terms related to beat are "on-beat" and "off-beat". For example in a time signature of 4/4, that is with 4 beats, the first tempo, the "downbeat", is usually the strongest one, and the second strongest is the third beat, these are called the "on-beats". The second and fourth, the weak ones, are called off-beats. And the subdivisions that are between beats are even more weak.

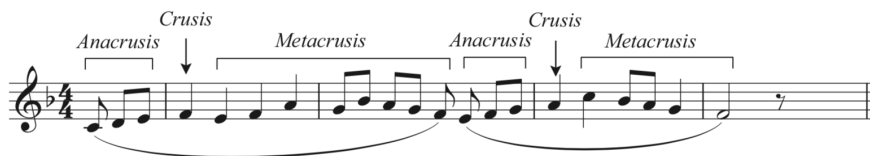


Figure 2.7: Crusis, anacrusis and metacrusis (Cleland and Dobrea-Grindahl (2013))

- Measure or bar: This beats constitute a measure or bar and are equally spaced. The length of the measure is defined by the time signature that appears at the beginning of the piece. If the time signature is 4/4, the measure has a length of 4 beats (figure 2.8). The time signature can change in the middle of a music piece, so the length of the measure will change according to the time signature.



Figure 2.8: Beats in a bar with a time signature of C or 4/4 from: <https://www.pianotheoryexercises.com/reading-music/measures-bars/>

- Syncopation: When the accent appears not in a strong beat it's called syncopation (figure 2.10) (Fitch and Rosenfeld (2007)).

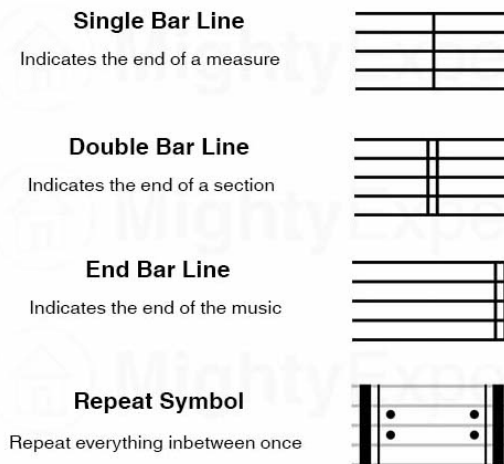


Figure 2.9: Types of bars from: <https://www.mightyexpert.com/what-is-a-bar-music/>



Figure 2.10: Syncopation from: <https://www.teoria.com/es/referencia/s/sincopa.php>

Regarding the connection between rhythm and emotion, Hevner (1937) says that the effects of this musical element, although not strong, are easy to understand, referring that "a flowing rhythm is light, happy, graceful, and sentimental" (Hevner, 1937, 626) and in contrast a firm rhythm conveys "sadness, dignity and vigour" (Gabrielsson and Lindström, 2010, p.391). In turn, a regular rhythm is linked to positive emotions like happiness and an irregular rhythm to negative emotions like anger (Gabrielsson and Lindström, 2010).

### 2.3.3.4 Tempo

Tempo is related with the speed of the music. Can be measure in beats per minute (bpm) or in a subjective way using terms. This marks appear at the beginning of the piece. Just as there are markings for a gradual change in dynamics, there are also for tempo (figure 2.11).

Word	Common Abbreviations	English Definition and Description	Recommended beats per minute (bpm)
Accelerando	Accel.	Gradually getting faster	
Adagio		At ease.Slowly	66 - 76
Allargando		Broaden. A slower version of rallentando	
Allegretto		Moderately fast	100
Allegro		Lively and fast	120 - 140
Andante		At an easy walking pace	56 - 88
Andantino		Between adagio and andante	
A tempo		Return to the former speed	
Largo		Broadly; Slowly	40 - 60
Larghetto		Less slowly than largo	60 - 66
Larghissimo		Very, very slow	<20
Lento	Lent	Slowly	40 - 60
Moderato		Moderately	100 - 120
Mosso		Movement. Slightly more lively	
Prestissimo		As fast as possible	>200
Presto		Very fast	150 - 200
Rallentando	Rall.	Gradually slower	
Ritardando	Rit/ Ritard	Gradually slower (but not as slow as rallentando)	
Ritenuto	Riten	Holding back	
Stretto		Quickening	
Stringendo		Tightening. Gradually faster	
Vivace		Lively. Faster than allegro	140
Vivo		Lively	

Figure 2.11: Tempo Markings (speed) and changes of tempo from: <https://www.theflutecoach.com/tempo-markings/>

According to [Gabrielsson and Lindström \(2010\)](#), tempo demonstrates to have more value than mode in terms of influencing emotions such as happiness and sadness, referring that fast tempo is associated with happiness and slow tempo with sadness. The author also compares tempo with arousal ([Gabrielsson and Lindström, 2010, p.381](#)), pitch ([Gabrielsson and Lindström, 2010, p.388](#)) among other musical elements.

A research developed by [Fernández-Sotos et al. \(2016\)](#) conclude that the values for happiness and surprise increase when we went from 90 to 150 bpm and the sadness value decrease. Also concludes that the emotion of sadness is the most affected when we go from a slower tempo to a faster one. The figure 2.12 shows the Circumflex model with the three emotions studied in this research and three combinations of tempo (90, 120 150 bpm).

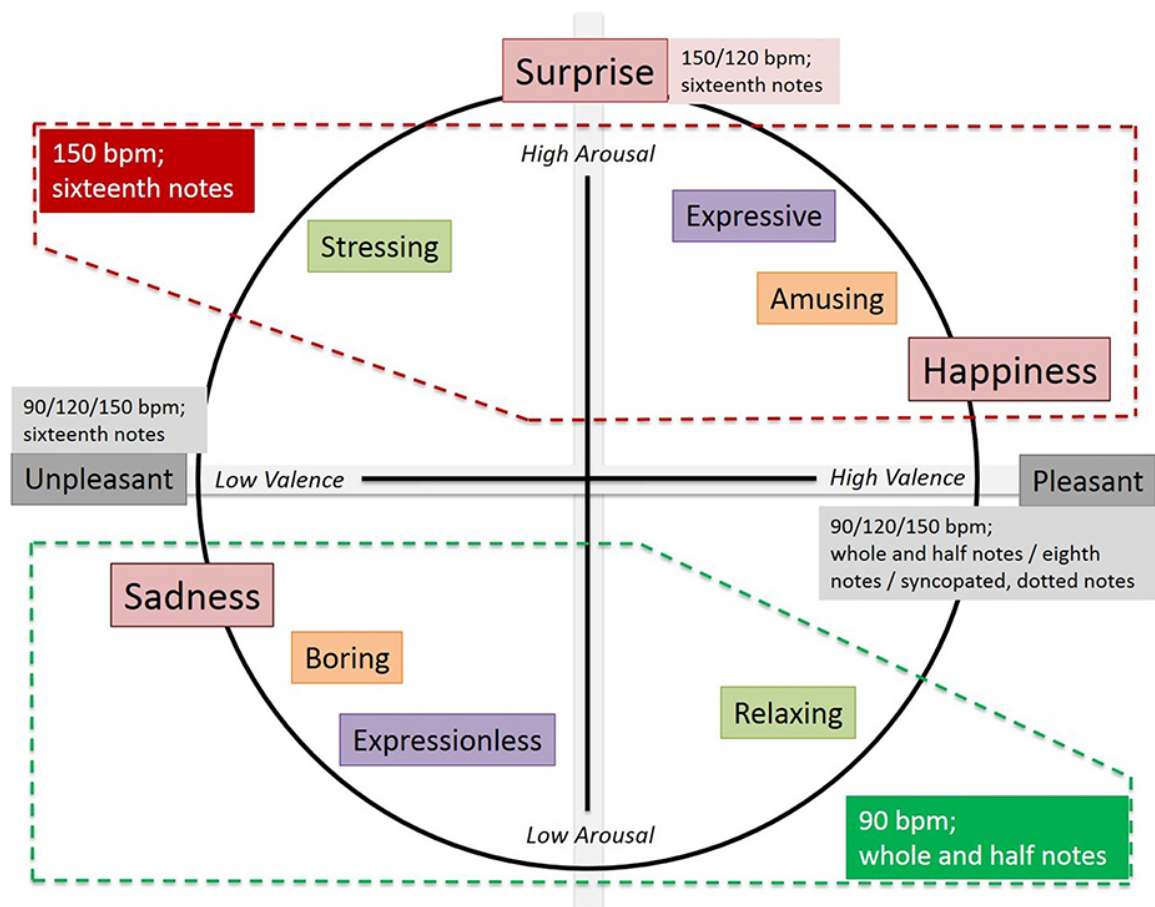


Figure 2.12: Annotated circumplex emotion model (Fernández-Sotos et al., 2016)

### 2.3.3.5 Harmony

Harmony is a group of several notes (pitches) that are analyzed in a vertical way, instead of an horizontal way used in melody (figure 2.13). The set of three or more notes is a chord, that can be played together or separately (arpeggio). To the group of chords that forms the harmony we call a chord progression.



Figure 2.13: Harmony and melody from: <https://musicskills101.com/what-is-the-difference-between-melody-and-harmony/>

To analyze the harmony it is necessary to break down the chords, from low (lowest note) to top (highest note). On a three-note chord the first one is the root note, then we have the third (makes a third interval with the root note) and fifth (makes a fifth interval with the root note). Of course, the chords are not always in the root position, and the note that gives the name to the chord may not be that one that has the lower pitch in the chord, in this case, we are facing so-called inversions (figure 2.14).

This does not mean that we need have chords to have harmony, from the moment we have two notes played simultaneously it is already called harmony (Benward (2014)).

	"Stable & Strong"	Slightly less stable	Least stable
	Root Position	1st Inversion	2nd Inversion
	Chord root on the bottom	3rd of the chord on the bottom	5th of the chord on the bottom

Figure 2.14: Chord inversions from: <https://www.secretsofsongwriting.com/2015/07/15/how-to-use-chord-inversions-to-greatest-effect/>

According to Hevner (1937) it is easier to understand the effect of simpler harmonies than of more complex ones. The author associates simple harmonies with "happy, graceful, serene, and

sentimental" emotions while more complex ones are "sad and vigorous" but also exciting (Hevner, 1937, p.626). For (Gabrielsson and Lindström, 2010, p.390), "simple, consonant harmony may be associated with expressions like happy/gay, relaxed, graceful, serene, dreamy, dignified, serious, and majestic; complex and dissonant harmony with excitement, tension, vigour, anger, sadness, and unpleasantness". Finally Cook and Fujisawa (2006) affirms that major chords are linked to positive emotions while minor ones are linked to negative emotions.

### 2.3.4 Music Emotion Classification

These musical elements (timbre, dynamics, rhythm, tempo and harmony) change along with the music so the musical composition can have more than one emotion associated. Therefore, we need to track the dominant emotion in a musical composition (Grekow, 2012), although it is difficult to distinguish one emotion from another (Weigand, 2004). Different models were created to classify emotions, grouping in two main classifications : the Categorical (discrete) model and the Dimensional model.

#### 2.3.4.1 The Categorical model

The Categorical model or discrete model utilizes words to distinguish emotions that all humans recognize. These basic emotions are described as discrete because "they are produced and recognized pan-cultural" (Russell, 1994, p.102). Several studies have been conducted in order to find out which are the basic emotions.

One of the most common categorical approach's was created by Ekman (1992) where they define six basic emotions: happiness, anger, sadness, surprise, disgust, and fear (figure 2.15). These emotions have specific characteristics associated and can be distinguished for example from facial expression (Ekman, 1992).

Hevner (1936) greatly contributed to the understanding of the relationship between music and emotion. She created a model "for the characterization of mood/emotional responses to music" (Hargreaves et al., 1986, p.110). The proposed "adjective circle" (figure 2.16), is composed by eight groups of adjectives, categorized by similarity and circularly arranged, with a total of 66 different adjectives. However, there are some inherent problems with this model, according to Juslin (2019). Hence, it is difficult to describe music through words. However, this difficulty is related to the way we perceive music and not to emotions. Other authors have improved this model like Farnsworth (1954). Later, Schubert (2003) also made an improvement to the model proposed by Hevner (1936), since according to Juslin (2019) the adjectives chosen by Hevner do not refer to "emotional categories" (p.106).

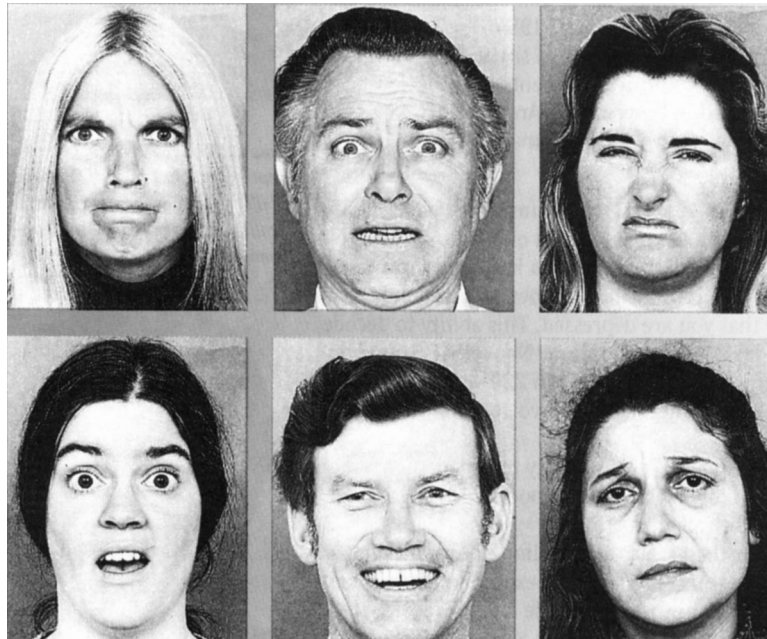


Figure 2.15: Six of the basic emotional faces (Ekman and Friesen, 2003). From top left: anger, fear, disgust, surprise, happiness, sadness.

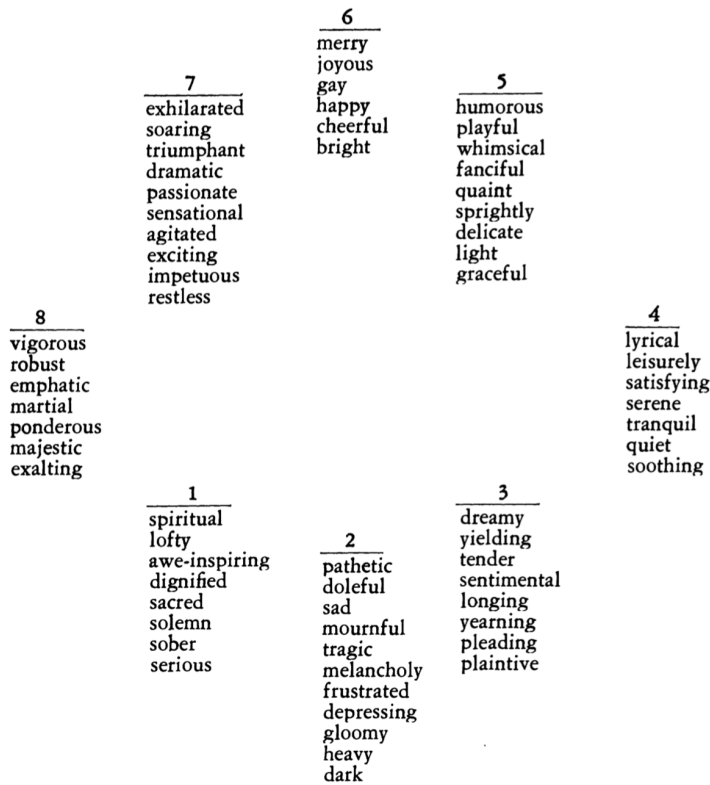


Figure 2.16: Adjective circle by Kate Hevner (Hevner, 1936)



### 2.3.4.2 The Dimensional model

The Dimensional model has two or three dimensions to distinguish emotions. Most of them contain valence (pleasure) and arousal (alertness) dimensions. This model differs from the theory of basic emotions since it proposes "that all affective states arise from two fundamental neurophysiological systems" (Posner et al., 2005, n.p.). One of the most famous is *The circumflex model* which was created by James Russell (figure 2.17) (Russell, 1980). In his model, emotions are spread out in a circular two-dimensions space, where the x-axis corresponds to valence and y-axis to arousal. Unlike the categorical model, where the research was essentially done on animals in order to arrive at discrete emotions, Russell refers to the difficulty in describing emotions. Therefore, it is not possible that "individuals do not experience, or recognize, emotions as isolated, discrete entities" (Posner et al., 2005, p.5). Emotions, in this model, are represented far from the center (figure 2.17), since the values of arousal and valence cannot be neutral. If we divide the *Circumflex Model* into quadrants, in the first one we can find emotions related with happiness, the second one with anxiety, the third one with sadness, and the last one with satisfaction (Posner et al., 2005).

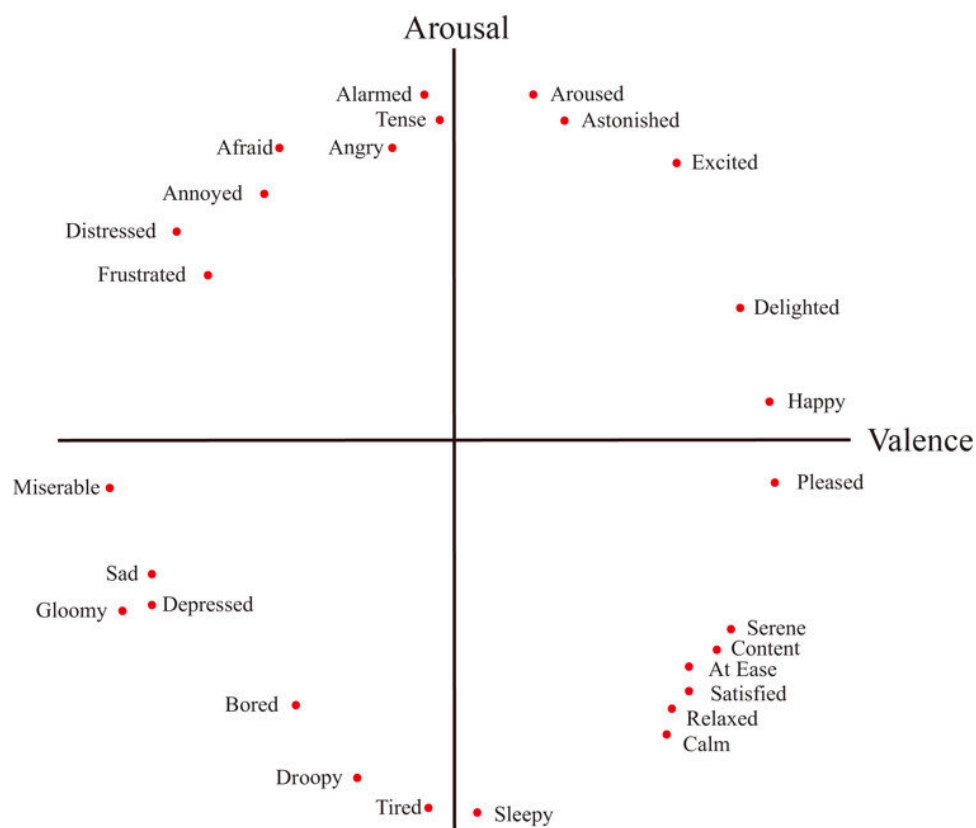


Figure 2.17: James Russell's Circumflex Model (Seo and Huh, 2019)

Several programs analyze music and can recognize the emotions in them. One example is CYANITE<sup>1</sup> that uses AI (Artificial Intelligence) to find the mood of the song. This tool also allows you to find songs with a similar emotional effect (figure 2.18). Figure 2.19 shows the classification

<sup>1</sup>CYANITE : <https://cyanite.ai>

of music by energy and arousal according to James Russell’s *Circumflex model* Helmholz et al. (2017).

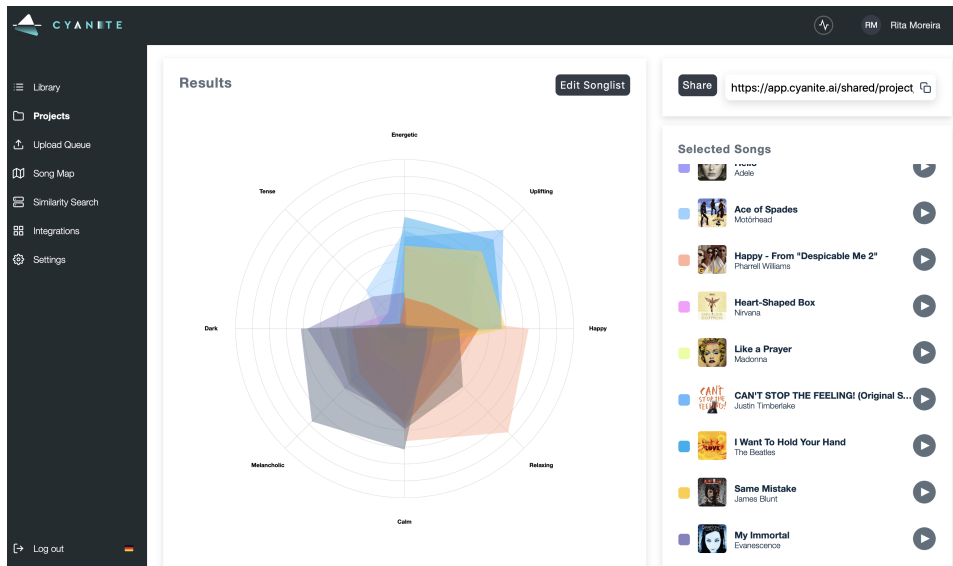


Figure 2.18: CYANITE analysis of several songs

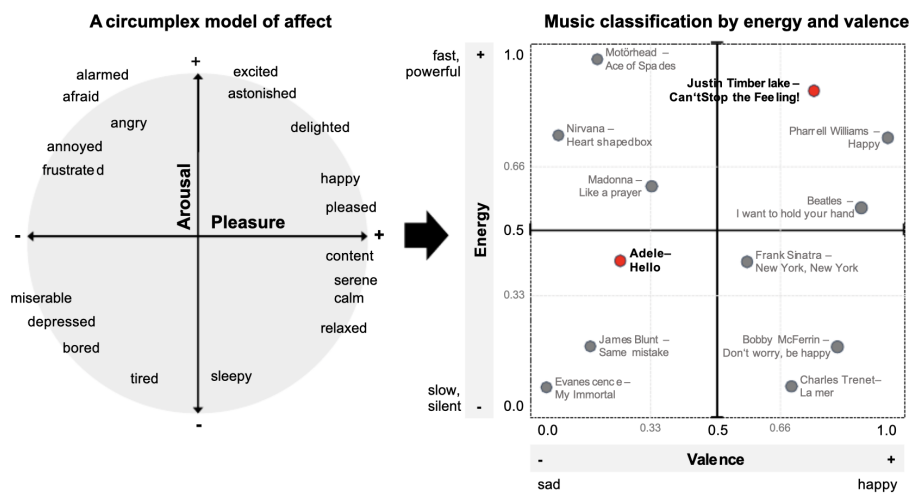


Figure 2.19: Classification of songs by energy and valence based on the human emotional Pleasure-Arousal-Model by Russell (1980) from Helmholz et al. (2017)

One example of a model that utilizes three-dimensional dimensions (pleasure, arousal, dominance) to describe emotions is the PAD emotional state model developed by Mehrabian and Russell (1974). The pleasure dimension measure if an emotion is positive or negative, the arousal is related to the energy/intensity of emotion and dominance is associated with a sense of control (Bakker et al., 2014).

### 2.3.5 Empathetic vs anempathetic music

As we saw earlier in this chapter, from the era of silent film that music has an important role to drive the emotions. Nowadays, music has become increasingly important, becoming an essential element that constitutes a film.

However, how big is the effect of music in a movie? The importance of music in movies is related to emotion and it can affect our mood and the perception of the scene, changing completely its meaning and making us see the scene in different ways (Herget, 2019)

For Chion (2019), in order to create a desired emotion, music can be empathetic or anempathetic. Empathetic means that the music is in agreement with the image, both express the same feelings, if we have a sad or depressing scene, the music will match this feeling. But also music can be in counterpoint with image conveying a different emotion that passed by the image. This happens for example when we have a happy song playing during a tragic scene when someone dies.

Many decisions needed to be done when composing music for a film whereas “rhythm, melody, harmony, tempo, volume, and instrumentation of the music can strongly affect a viewer’s emotional reactions” (Chion, 2019, p.277). The type of music that accompanies the image will urge different emotions and make us see the image in a different way, and create “different impressions of the footage”(Chion, 2019, p.277). Additionally, sometimes a specific melody is associated with an actor or someplace, or can predict dangerous or that something is going to happen. It is a musical phrase that we recognize, occurs constantly and has some identity.

In every genre of movies, we have different sounds and music which creates a specific atmosphere to enhance the experience emotionally. Sometimes we can connect the soundtrack to a movie, like in *Gladiator*, or *Braveheart* or even in *Pirates of the Caribbean*.

The first film with an original non-diegetic score enhancing the narrative using techniques that changed film scoring forever was *King Kong* (1933), composed by Max Steiner. These techniques continued to be used in horror films, like mickey mousing or leitmotif, a short musical phrase associated with a character where we can “observe a clear and consistent relationship between a musical idea and its onscreen counterpart” Hickman (2017). This last technique was used by John Williams in *Jaws*, one of the most memorable film scores of all time, the quicker it gets, the closer we know the shark is. The feeling that the music transmits is related to the appearance of the shark and remains in the audience’s minds as an alert signal Petrikis (2015).

## 2.4 Attention

Every day, we are bombarded with more information than we can take in, eventually failing to notice in the presence of information considered important. Our brain can't process all of the information, so our conscious awareness depends on what we choose to pay attention, and what we select. This is called selective processing of incoming sensory information (Driver, 2001).

We miss many information because we are paying attention to something else, and sometimes it is hard to focus on what is important. One example of selective attention is "The cocktail party effect". This phenomenon explains the ability of humans to focus on a specific stimulus even when they are receiving various pieces of information. The process of selecting what is important and disregard the rest is called attention, the ability to focus on specific information.

*Working memory, competitive selection, top-down sensitivity control* and *saliency filters* are essential components of attention. *Working memory* is a system that store data for a period of time, is short-term memory. Knudsen (2007) uses the example of an animal when facing an object, information about that object go into his working memory, "thus, working memory represents the objects attention" (Knudsen, 2007, p. 60).

Before reaching working memory, information needs to pass through a process called *competitive selection*, where more relevant stimuli are elected from the rest in order to divert unimportant information. To balance the power of each information that is trying to enter the *working memory top-down sensitivity control* is used, an approach to enhance the quality of the information. This is related to *saliency filters*, that select stimuli that are important depending of past memories, previous experiences and related to biological components (Knudsen, 2007). We can identify two types of attention, bottom-up and top-down attention. In bottom-up or stimulus-driven the stimulus influences what we perceive and "alerts us to salient items in our environment" (Connor et al., 2004, n.p.).

In contrast, top-down attention, uses our knowledge to influence perception, it is cognition driven, and this cognition modifies our perception. This type of attention is voluntary, the viewer decides where to pay attention. Paying attention to a small region in space is called top-down spatial attention and when choose to look to all objects is called top-down feature attention (Pinto et al., 2013).

On the other hand, orientation, executive monitoring and alerting responses are attention capacities, that correspond respectively to selective attention, divided attention and sustained attention (Geva et al., 2013). As mentioned before, selective attention is the process of focusing on relevant information, as Revlin (2012) explains "We must be selective in our attention by focusing on some events to the detriment of others. This is because attention is a resource that needs to be distributed to those events that are important" (Revlin, 2012, n.p.).

### 2.4.1 Orienting attention

The direction in which attention is conducted is called orienting. For Posner (1980), orienting means that the source where the information comes is lined up with attention so to orient humans

and select a location in the space to focus attention. The orienting systems that is correlated with areas of the parietal and frontal lobes directs the visual attention to a stimuli. This cue manipulates orienting so that the attention is directed to the location of the stimulus (Fan et al., 2002). Although the eyes act in accordance with the direction of attention, orienting doesn't mean that there must be eye movements (Kubovy, 2017).

Exogenous and endogenous are the two main types of orientating. We call it exogenous orienting when the relocate of attention is due to a striking change, involve a passive process. When the attention is directed to a location by a conscious decision is endogenous orienting, an active process (Salselas and Penha, 2019).

### 2.4.2 Executive attention

Executive attention is a system that trigger pertinent stimuli and constrain distractions including processes like conscious detection, inhibition, and conflict processing (Rueda et al., 2015). Thus, this system has a role on managing information and conducting our behaviour. According to Kellogg (2015), “executive attention is always needed when (a) planning or making decisions, (b) correcting errors, (c) the required response is a novel or not well-learned, (d) conditions are cognitively demanding or dangerous, and (e) an automatic response must be inhibited and overcome” (Kellogg, 2015, p. 83). When we talk about “emotional, social and motivated” although executive functions be related with cognitive functions, this types of behaviors are complex to control because they “require a different kind of mechanism”. However, recent research associates executive attention with the increment of emotional understanding and regulation, and with progress of conscience (Rueda and Paz-Alonzo, 2013).

### 2.4.3 Alerting attention

Alerting attention is linked with sustained vigilance depending most of the time on spontaneous processes. This state can vary endogenously if it is related to a motivation shift, or exogenously, if we have a sudden change in stimulus (Rueda et al., 2015). Mezzacappa says that “alerting attention in the most basic sense refers to the capacity to maintain a state of alert arousal” (Mezzacappa, 2004) Accordingly, “alerting inhibits executive control and enhances orienting” (Boncompagni and Casagrande, 2019).

### 2.4.4 Emotional attention

What can engage the viewer by directing him to what is important? How does the brain select what is important? Recent research explored the role of emotion on driving attention, “people more readily pay attention to emotional than neutral stimuli” (Vuilleumier, 2005; Desimone and Duncan, 1995). Emotion affects our attention, and when the emotional stimulus, positive or negative, is relevant the events are detected faster, becoming the spotlight of our attention (Brosch et al., 2008). According to Tim Dalgleish and Mick Power (2000) “emotion and attention are intimately

linked” (Dalglish and Power, 2000, p. 171). The same happens in visual search “when the target has some emotional value” the viewer discovers him more faster (Vuilleumier, 2005, p. 1).

In other hand, according to Easterbrook, negative emotions like stress and anxiety reduce the extension of attention, (Huntsinger, 2013) or in some cases like identifying a color of a illustration or word when having an emotional connotation, the task will be done more slowly(Vuilleumier, 2005).

## 2.5 Summary

This analysis demonstrates that music, attention and emotion are concepts that are strongly related. Music has the power to modulate our attention through emotion giving a sense of immersion in audiovisual contexts and playing a crucial role in the audience’s experience. Additionally, sound design and music are tools used to create illusions and change the perception of the spectator. Although the visual part of the film receives most of the attention, is the sound that will affect what the audience sees, driving the viewers to what the producers want them to pay attention to.

Through this literature review it is noticeable that due to the appearance of new technologies and platforms where narratives can be told, the process for creating sounds are changing over the years in order to guide the attention of the viewer.

## 2.6 Problem statement

*“The primary reason for music listening is the power that music has in stirring our emotions.”*

— Marko Ahtisaari and Ketki Kranam

Throughout the previous sections of this chapter, we have seen that the way stories are told is constantly changing. Specifically, the rise of new technologies such as virtual environments is revolutionizing the way we experience audiovisual content, creating new forms of engaging the audience. In an attempt to approximate to reality, these environments bombard users with stimuli that can be overwhelming (Chueng and Marsden, 2002), which means that new challenges arise to focus the attention of the viewer and make him follow a certain narrative. Therefore how can a narrative be followed?

Additionally, we have observed that sound is a powerful tool to foster narratives since it can influence our attention through emotion and transport the viewer into action, reinforcing the immersive experience. Consequently, can music, through its emotional charge, induce visual attention and contribute to the following of a given narrative? Can music be a tool for orienting through emotional engaging? The consistency between the correspondence of the visual and auditory emotional state, according to Salselas et al. (2020), may be an effective factor in driving visual attention?

## **2.7 Aims of the study**

One of the main characteristics and riches of music is its emotional power (Juslin and Laukka, 2004). The main goal of this research is to explore how emotion, driven by music, can be a powerful tool for inducing visual attention in the audiovisual context. In this sense, the following procedures were established:

- Design a pilot study and test it in order to understand if any video is more interesting due to some visual factor, to receive suggestions regarding the study and to find out about any mistake that could compromise the study itself;
- Design an experimental study based on the conclusions drawn from the pilot test, which aims to test if the music through its emotional charge can induce attention;
- Analyse the data obtained in the study. The preference of the participants in the three musical contexts (without sound, with tense music and with relaxing music) will allow to find the answer to the aim of this study.





## Chapter 3

# Methods

This chapter will provide a detailed description of the steps underlying the methodology of this investigation. In this sense, an experimental study will be presented which aimed to test the subject's visual attention when confronted with two simultaneous video scenes combined with a selected composed song for this purpose.

Initially, the plan was to test subjects in an on-site experimental design, involving the use of eye-tracker equipment. However, given the circumstances of being in a pandemic, it was not possible to conduct this study in person with the participants. Therefore, due to the COVID-19 restrictions, the experimental design was completely adapted to be online. A preference questionnaire (appendix A) was designed where the participants had to select the video that captured their attention through multiple choice. The capture of the stimuli had also to be adapted to these circumstances, as will be mentioning bellow.

### 3.1 Participants

This study involved a total of 287 participants, already excluding those who were not eligible for this study. In this population, 191 were female, 95 male and 1 did not identify themselves as being of one gender or another. The subjects were older than 15 years old, with the most varied areas of interest, such as health, sports, multimedia, policy, among others, but 25% of the sample show that have interest in music.

The recruitment procedure was made online, using social media and email. This email was sent to students from *FEUP*, *FBAUP*, *FLUP* and *FCUP*. Although everyone can answer the questionnaire, for inclusion in the study it is necessary to meet the eligibility criteria. This consists of not having hearing problems or difficulties in viewing the videos presented, either in terms of vision problems or technical problems that compromised the quality of the videos and consequently the final results.

## 3.2 Stimuli

The scenes used in this study were recorded using an iPhone camera. Due to the situation we were in, quarantined, we had to use the material we had available at home. The videos consisted of a static camera and are accompanied by three musical contexts, no sound, relaxing music, and tense music. They were thought to be as neutral as possible in terms of color, with the fewest number of objects so that the viewer's attention does not fall on these aspects. The same for the actors' clothing. The scene is constituted by a table with two chairs, where each actor is sitting, a window behind the table and a curtain. The actors also interact with a computer, a phone and a book depending on each scene. In terms of framing was used the "american plan" or "Cowboy shot", with origin in western movies, so that we could see both actors' heads and the gun (Thompson and Bowen, 2009). This allows seeing all the actor's actions in the scenes, from the shaking leg to the opening of the curtain. The stimuli are very distant from what could be considered an experimental laboratory one. Conversely, and having in mind the COVID-19 adaptations that had to be done, we consider that the stimuli has become less artificial and, thus, more representative of the diversity and richness of the sensory information encountered in the real world.

The videos consist of four different scenes :

S1 : stressful scene 1<sup>1</sup>

S2 : stressful scene 2<sup>2</sup>

C1 : calm scene 1<sup>3</sup>

C2 : calm scene 2<sup>4</sup>

These videos differ in the actions of the actors and in the emotions they express. In S1 and C1, the male actor is playing on a computer and the female actress is reading a book. In this case, the female actress keeps the same expression whether we are in a calm or stressful scene. The male actor is calm in C1 and is stressed in S1.

In S2 and C2, the male actor is touching the phone and shows the same emotion regardless of the scene. The female actress interacts with the curtain and with the window, and is calm in the case of C2 and stressed in the case of S2.

The videos had no dialogue and all sound from the scene was removed. Two different pieces of music were composed according to the emotion that was intended to induce through the video, which is empathetic music or to convey the opposite emotion, anempathetic music.

M.T. : tense music<sup>5</sup>

<sup>1</sup>S1 : [https://www.youtube.com/watch?v=Td\\_5kyX3TLc](https://www.youtube.com/watch?v=Td_5kyX3TLc)

<sup>2</sup>S2 : <https://www.youtube.com/watch?v=21MdTc-vROY>

<sup>3</sup>C1 : <https://www.youtube.com/watch?v=kR4m1M5pQqo>

<sup>4</sup>C2 : <https://www.youtube.com/watch?v=06ZpHKtaEXo>

<sup>5</sup>M.T. : <https://youtu.be/25GeOD4BcLo>

M.R. : relaxing music<sup>6</sup>

Both pieces of music were composed using *Ableton Live*. All the instruments were played in a MIDI (musical instrument digital interface) controller that allows to communicate between devices, in this case controlling *Ableton Live*<sup>7</sup> vst (virtual studio technology). For the sounds of the instruments we have used Kontakt<sup>8</sup>, a sampler that has several available libraries with a wide selection of sampled instruments. Relaxing music consists of a simple piano base in C major, which is then completed by the entry of pizzicato violins and the xylophone that gives rhythm to the music and an almost fantasy feel. The melody is also present on the piano in conjunction with the glockenspiel. Finally, strings were added to fill and Celeste also give a sense of time.

For the tense music was utilized strings sounds in a crescendo of tension and percussion to give rhythm and a feeling that something is going to happen, almost an anticipation that makes us nervous. With the aim of having an external evaluation and test the emotions conveyed by the composed music, as being relaxed or tense, an analysis was performed using CYANITE<sup>9</sup> which classified M.T. as tense and ambient music (figure 3.1) and M.R. as a calm and ambient (figure 3.1).



Figure 3.1: M.T. and M.R. CYANITE classification from <https://cyanite.ai>

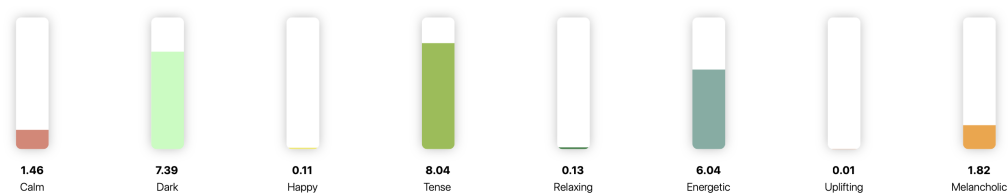


Figure 3.2: M.T. characteristics by CYANITE from <https://cyanite.ai>

To avoid technical issues regarding audio, a sound test was presented before starting the questionnaire itself. Then the participants were asked to listen to the presented audio and if it was working correctly go to the next section.

Other variables that could compromise the audio quality or the participant's attention were more difficult to control, as each responded using their own resources/materials. However, at the beginning of the questionnaire, the use of headphones was advised, as well as being asked to watch

<sup>6</sup>M.R. : [https://youtu.be/QA\\_L9L9JPW0](https://youtu.be/QA_L9L9JPW0)

<sup>7</sup>Ableton Live : <https://www.ableton.com>

<sup>8</sup>Kontakt : <https://www.native-instruments.com/en/products/komplete/samplers/kontakt-6/>

<sup>9</sup>CYANITE : <https://cyanite.ai>

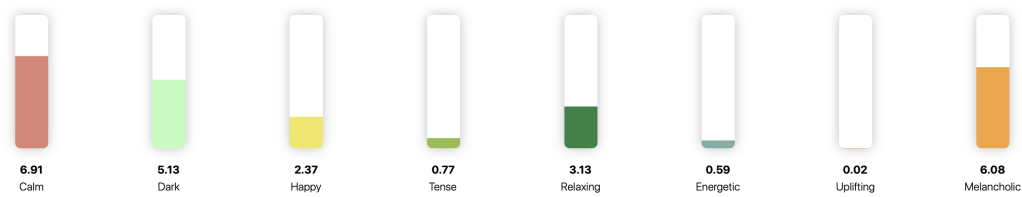


Figure 3.3: M.R. characteristics by CYANITE from <https://cyanite.ai>

the videos until the end, as it was necessary to know the one who got the most attention and not the one they looked at first. Participants were also told to avoid distractions and that the duration of the questionnaire is 25 minutes.

### 3.3 Design

The subjects had to follow an online questionnaire (using Google Forms) that consisted of the presentation of two videos simultaneously, side by side, with a duration of 30 seconds each one. In a two-alternative forced choice paradigm, participants expressed their preference by choosing the video that most called their attention. Figure 3.4 shows how the videos were presented to the participants. The participants were informed with the minimum about the purpose of the study, neither the title of the thesis was presented in order to ensure that they were not influenced. They were informed that the questionnaires consist of the visualization of two videos side by side, and they have to choose their preference. It was also advised to avoid distractions and to use headphones (figure A.1).

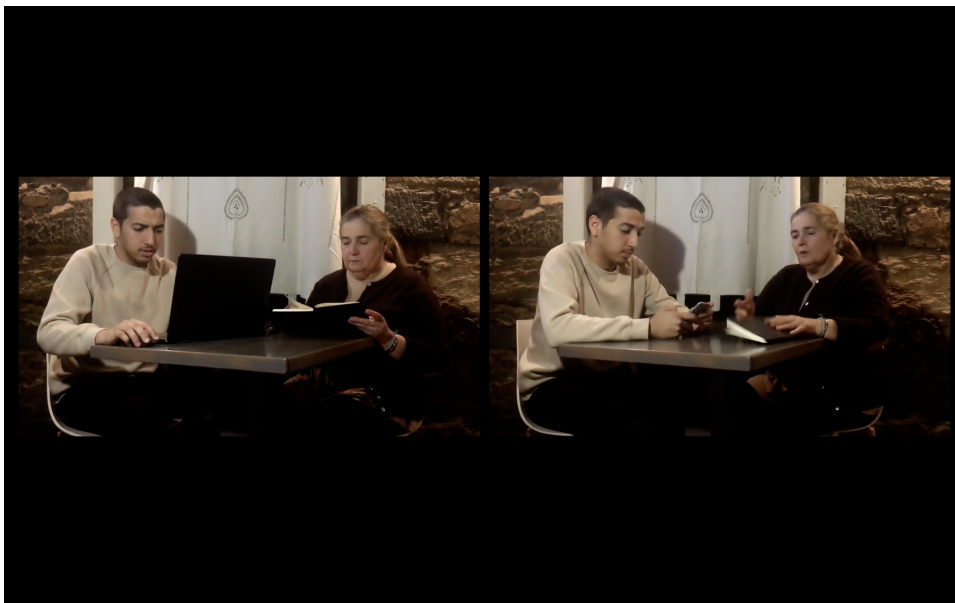


Figure 3.4: Study setup

After the visualization of all the combinations, the participants had to answer two open-ended questions, the first relating to the reasons that led to the preference of one video over the other in the case of no music, and the second in the case of music. Both questions were not mandatory.

### 3.4 Procedure

Before the study itself, a pilot study was carried out with 11 participants in order to ascertain any error or bias that might have occurred in the experimental design and stimuli, whether the videos and music would be adequate and also to collect suggestions.

In order to test any bias introduced by the video, in the first part of the study, the videos were presented without any sound. In the second part, videos were accompanied by relaxing or tense music.

In order to ensure that no bias was introduced due to the order of the stimuli presentation it was necessary to randomize the order of the questions. After extensive research on what could be the best tool to perform this randomization, we have faced a few problems. In spite of many having this feature, the inability to link a question to a video made the randomness of the questions not possible automatically.

Consequently, five different questionnaires were created using Google Forms tool, distinguished by the order of the questions presented in part I and II. Although this was not the preferred solution, it was the possible one, in order to avoid this bias. With all the data collected, it was necessary to find out if this difference in order had an influence on the participant's response and if not then add the data together as if it were a single questionnaire.

Division of the questionnaire :

- Introduction figure (A.1)
- Sound test (figure A.2)
- Demographic data (figure A.3;figure A.4)
  - Gender
  - Age
  - Academic qualifications
  - Interest areas
  - Vision and hearing problems
- Part I (no sound) (figure A.5)
  - Which video caught your attention the most
- Part II (with sound) (figure A.6)
  - Which video caught your attention the most

- Part III (figure A.7)
  - Reasons that led him to prefer one video to another

Part I of the questionnaire consist of viewing all combinations of the four videos ( $n=4$ ) combined two by two ( $p=2$ ) considering the sides. Which means, according to the formula 3.1, simple arrangement, we have a total of 12 combinations (table 3.1).

$$A_{n,p} = \frac{n!}{(n-p)!} \quad (3.1)$$

In this part, the videos do not have any sound serving not only to know which of the videos the participants prefer only having the images but also to mislead preferences per side.

Combination 1	S1 vs C1
Combination 2	S1 vs S2
Combination 3	S1 vs C2
Combination 4	C1 vs S1
Combination 5	C1 vs S2
Combination 6	C1 vs C2
Combination 7	S2 vs S1
Combination 8	S2 vs C1
Combination 9	S2 vs C2
Combination 10	C2 vs S1
Combination 11	C2 vs C1
Combination 12	C2 vs S2

Table 3.1: Video combinations without sound

In part II the preference for side is already excluded, thus presenting 6 different combinations and each of them accompanied either with relaxing music or with tense music (table 3.2)

In order to analyze the results of this study, Microsoft Excel was used, version 16,30 for macOS. After organizing the data from the 5 questionnaires that were distinguished by the order of appearance of the videos it was investigated if this order had an influence on the choice in videos without sound and with sound. For that, the 2D column chart was created for each combination. Also, this method was used to see if there would be a preference per side. Since it was proven that there would be no difference, the data from the 5 questionnaires were joined in one.

The next step was to create a second column chart as well to the videos without sound and to the videos with sound to compare side by side each case.

The final step was to compare the three musical contexts, no sound, relaxing music, tense music for each combination and ascertain the percentage of participants who preferred the calm scene or the stressful scene without and with sound. Also find out what percentage preferred the

Combination 1	S1 vs S2	M.T.
Combination 2	C1 vs C2	M.T.
Combination 3	S1 vs S2	M.R.
Combination 4	C1 vs C2	M.R.
Combination 5	S1 vs C1	M.T.
Combination 6	S1 vs C2	M.T.
Combination 7	S1 vs C1	M.R.
Combination 8	S1 vs C2	M.R.
Combination 9	S2 vs C1	M.T.
Combination 10	S2 vs C2	M.T.
Combination 11	S2 vs C1	M.R.
Combination 12	S2 vs C2	M.R.

Table 3.2: Video combinations with sound

calm scene with tense music, a calm scene with relaxing music, a stressful scene with relaxing music and a stressful scene with tense music.





## Chapter 4

# Results and analysis

The main focus of this study was to explore if music could capture the viewer's visual attention. The preference of the participants was investigated in three different musical contexts, without sound, with relaxing music and with tense music. After verifying that there was no order or side bias this variable has been ignored. Results relative to this analysis are presented using 2D column charts, as follows.

### 4.1 Videos without sound

When both videos were presented side by side in the condition of conveying the same emotion (S1 vs S2, C1 vs C2), in the case of S1 vs S2, the participant's preference fell more on S2 (91%) while only 9% preferred S1. In C1 vs C2, 86% preferred C2 and 14% preferred C1. In both combinations, scene 2 was the most chosen.

When videos expressed different emotions and we had the same scene (S1 vs C1, S2 vs C2), the preference was for stressing scenes. 84% of the subjects selected S1, while only 16% preferred C1 and 80% choose S2 over C2 (20%).

The same did not happen with different scenes and different emotions (S1 vs C2, S2 vs C1). In the case of S1 vs C2, the calm scene was preferred (63%) with 37% choosing S1. But with S2 vs C1, 93% of the participants choose S2 while 7% preferred C1. The referred results can be observed in Figure 4.1.

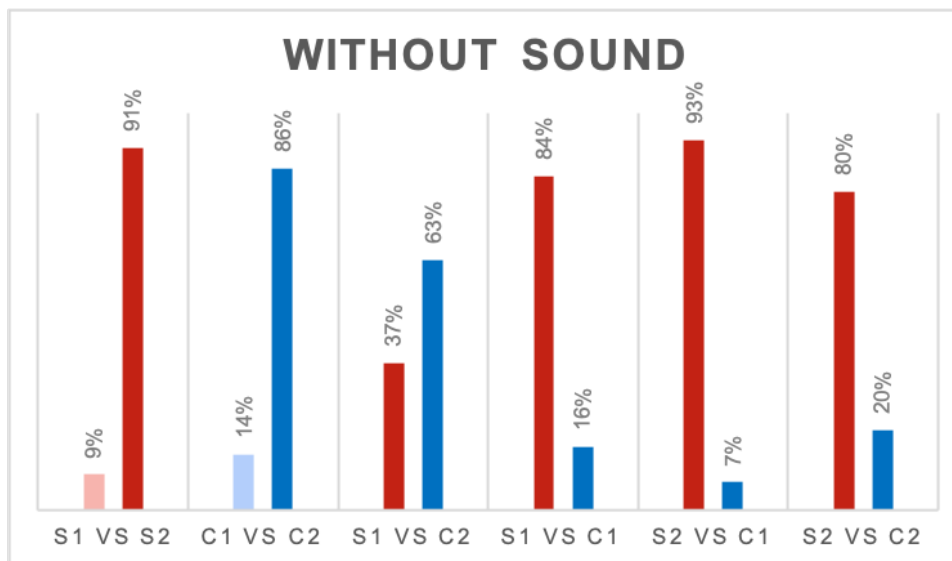


Figure 4.1: Preference in videos without sound

## 4.2 Videos with sound

When adding sound, i.e., music to the videos, results show different patterns. The results can be observed in Figure 4.2 and Figure 4.3 that we summarize as follows:

- **S1 vs S2** : Preference is maintained on S2 (89%) with 11% choosing S1 with **tense music**. Although with relaxing music the preference remains in S2, this increases to 60%, and the preference for S1 rises to 40%.
- **C1 vs C2** : Preference in both musical conditions falls on C2, with 88% with tense music and 67% with relaxing music. C1 has a percentage of 12% with tense music and 33% with relaxing music.
- **S1 vs C1** : With tense music S1 had the preference of the participants (88%) and C1 had 12%. These preferences were reversed with relaxing music with S1 having 47% and C1 with 53%.
- **S1 vs C2** : Despite the difference being small, with tense music the participants preferred S1 (55%) with 45% choosing C2. With relaxing music, there was a significant increase in preference for C2 with 77%, with 23% choosing S1.
- **S2 vs C1** : Almost all participants preferred S2(92%) over C1(8%) with tense music. On the other hand, with relaxing music, the preference was similar, with 52% preferring S1 and 48% preferring C1.
- **S2 vs C2** : With tense music the preference was greater in S2(80%) than in C2(20%). The same did not happen with relaxing music, with S2 going down to 47% and C2 going up to 53%.

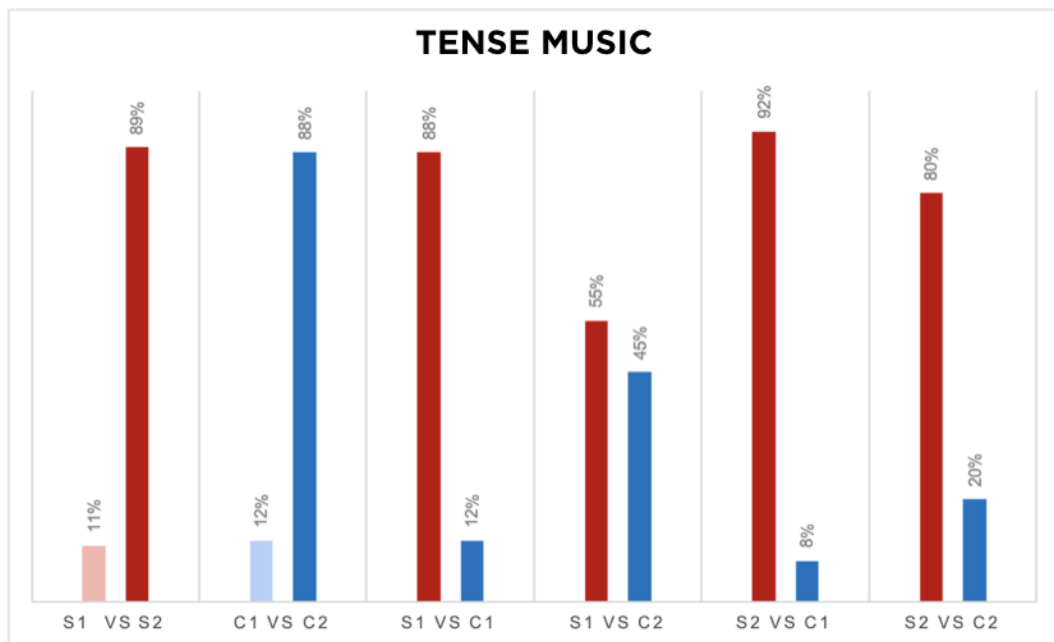


Figure 4.2: Preference in videos with tense music

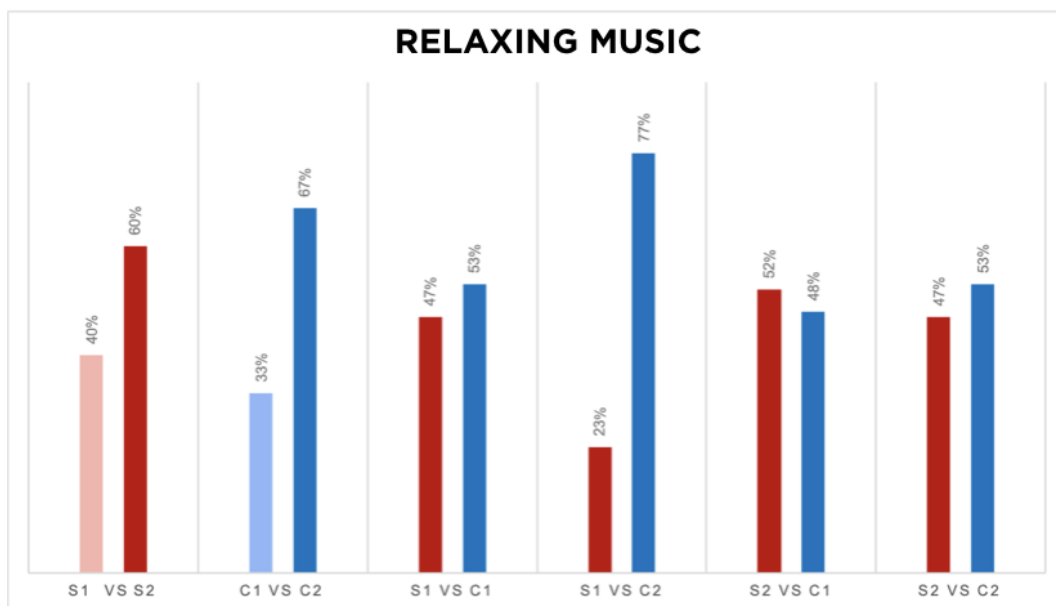


Figure 4.3: Preference in videos with relaxing music

### 4.3 All musical contexts

By excluding the combinations between stressful videos and calm videos and adding the preference for calm or stressful videos from the other combinations, in both contexts of without and with sound (M.T. and M.R.), we obtain the graph in the figure 4.4.

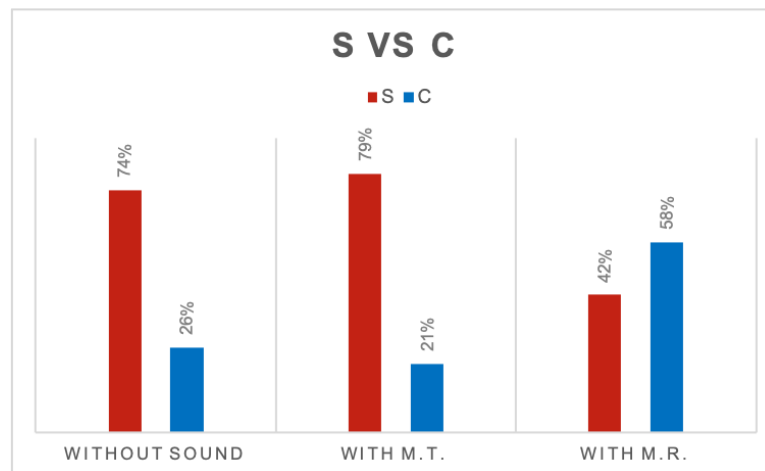


Figure 4.4: Preference of stressful vs calm videos without music, with tense music and with calm music

This graph shows that the trend was to choose stressful scenes in both situations, and in the case of no sound, the percentage was 74% for stressful scenes and 26% for calm. In the case of sound, the percentage for stressful scenes dropped to 60% and the preference for quiet scenes increased to 40%.

Now separating the preference for stressful scenes or for calm scenes depending on the type of music we obtain the two graphics below (4.5, 4.6) that shows that the preference for stressful scenes was 65% with M.T. and 35% with M.R., and 73% choose calm scenes with M.R. and only 27% with M.T.



Figure 4.5: Preference of stressful scenes with M.T. or M.R.

Finally comparing side by side the three musical contexts for the six different combinations (figure 4.7) :

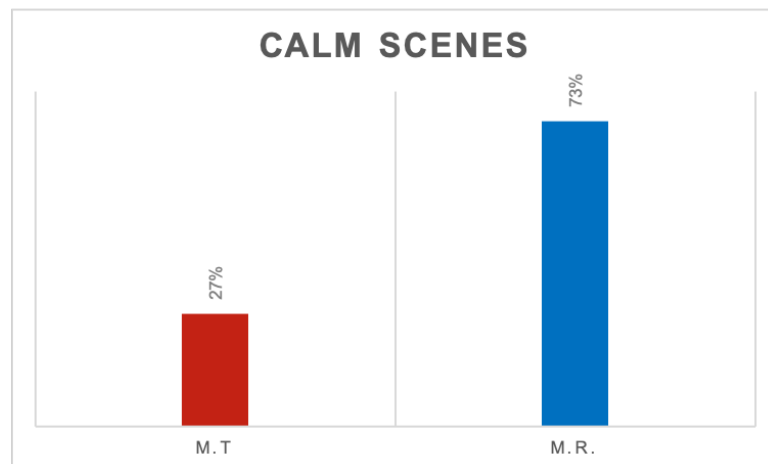


Figure 4.6: Preference of calm scenes with M.T. or M.R.

- **S1 vs S2** : In the context of no music and with M.T. the preference fell for S2, the percentages being similar. With M.R. in spite of the greater preference being for S2 (60%), there was an increase for S1 (40%).
- **C1 vs C2** : In the three contexts, preference was greater in C2, this dropped from 86% (without sound) to 67% with M.R..
- **S1 vs C2** : In this case, the S1 preference went from 37% (without sound) to 23% (M.R.) to 55% (M.T.). This shows us that the participants preferred the calm scene (C2) in cases of no sound and with M.R., but with M.R. it increased, however this reversed with M.T., who despite the difference being small, preferred the stressful scene (S1).
- **S1 vs C1** : The preference was greater in S1 in the case of no sound (84%) and with M.T. (86%). On the other hand, with M.R. they preferred the calm scene (53%).
- **S2 vs C2** : This case is similar to the previous one, with participants preferring S2 in the context of no music (80%) and with M.T. (80%). With M.R. they prefer C2 (53%).
- **S2 vs C1** : The preference in the three contexts is the stressful scene (S2), which goes from 93% without music to 53% in the case of M.R. With M.T. the stressful scene (S2) has a 92% preference.



Figure 4.7: Without Sound vs M.R. vs M.T. for all combinations

## 4.4 Discussion

### 4.4.1 Main findings

One of the main aspects that can be deduced from the analyzed results is that the preference in the case of videos without sound falls on the one in which the actors show more agitation, where there is more movement. This preference is mostly similar to the case of videos presented with tense music, either with videos that convey the same emotion (ex: S1 vs S2) or with videos with different emotions (ex: S2 vs C1). It is also noticeable the preference for the video in which the emotion transmitted by it is equal to the emotion transmitted by the music, demonstrating that the participants prefer empathetic music.

### 4.4.2 Comparison with existing literature

The main focus of film making remains to tell the story visually utilizing different techniques that will influence how the audience perceives the story (Brown, 2016). In fact, the moving image is the component of the film that most focuses the viewer's attention, since, according to Green (2010), day by day the stimuli reach us mostly through images, dialogue, sound. Also mentioning that the form of sound that the audience focuses most of their attention on is diegetic, like dialogues. This kind of information besides being a form of communication, is more natural than nondiegetic, and music is mostly presented this way. On the other hand, according to Green, music plays an essential role in a film, being "a tool that can expose the inner feelings and thoughts of characters and can shape the way that viewers feel about what's happening on screen" (Green, 2010, p.81). This reinforces the idea that music can be an important component with regard to communication and as a source of information. Just as there are several techniques for the construction of the cinematographic image, music can also be used to "control the way that the audience interprets a film" (Green, 2010, p.82), taking advantage of several techniques such as leitmotif or congruence.

According to Zacks (2015) we tend to interpret what we see in a film as if it were real, getting immersed in the action. Zacks demonstrates this through "mirror rule" and "success rule". The first rule describes that when we see some behavior we tend to imitate it, "do what you see" (Zacks, 2015, p.5). For example, when we see the actress from the recorded videos for this study feeling stress and making a lot of movements, we tend to imitate this behavior, act like her, however it can also go beyond that, making us feel stressed. The "success rule" on the other hand will seek the experiences of the past, for example, if in the past I acted in this way and managed to get rid of the danger then now in the same situation we will "do what has worked" (Zacks, 2015, p.9)

In this study, the actors use nonverbal cues such as body language and facial expressions to transmit what they feel. In the case of no sound, the trend of preference, as the results show, was for stressful videos with more movement, and previous studies show that we pay more attention to faces with emotional cues essentially when they are negative, such as fear. The same is true for other types of non-facial stimuli (Vuilleumier, 2002). The preference in videos without sound shows that emotional stimuli capture more attention than neutral, where the actor appears

calmer. Other studies suggest that memory is an important factor regarding attention, and we tend to remind negative stimuli over neutral or positive so we pay more attention to them (Friedenberg and Silverman, 2011). Also, not only emotional stimuli make fixation time longer, but also uncomfortable pictures catch visual attention (Ono and Taniguchi, 2017).

Most participants said that the characters' agitation made them prefer one video to another, International (2011) refers that movement capture attention from the viewers and that "the eye is inevitably attracted to a moving object" (International, 2011, p.11). According to Clayton (2012) some scenes can automatically attract our attention, like stressful or natural scenes. In the case of stressful scenes and mention again the "success rule", when we watch some events that we consider sad or gives us the sensation of danger/fear, the amygdala makes us focus our attention in the scene since the amygdala "is at the center of evaluating sensory information of social and emotional significance"(Marzluff and Angell, 2013, p.139).

Regarding the congruence between music and image, other studies using the same approach as this one, presenting scenes that convey different emotions accompanied by three different musical contexts. They arrived at the same result, the condition without music was considered the most appropriate in all the clips, but for some, one of the other alternatives was also considered appropriate or almost as appropriate as having no music. This alternative refers to music that conveys the same emotion as the video (Wallengren and Strukelj, 2015). Also, the situation with music may make the viewer more alert, since it was concluded that an unexpected event is comprehended more notably with music (Auer et al., 2012).



## Chapter 5

# Conclusion and Future work

During this study, the cinematographic universe was explored, from silent films to the appearance of talkies in 1926 and music has always been an essential element. Just as previously, there was no synchronized audio, emotions were transmitted through body language, whether through gestures, facial expression, posture, among others. This type of communication manages to change the audience's perception of the scene and even nowadays its dominance on the part of the actors is fundamental since it manages to capture our attention.

Music is also used as a form of communication, which takes the visual to another level, managing to affect the way we feel and our perception. We have the illusion that we have control over our attention, but this may not be the case, we are always being influenced by external stimuli, either by the visual or by the sound. The fact is that emotionally charged stimuli have greater strength in guiding our attention than neutral stimuli.

With this and with the developments that have been done since the silent era new challenges arise to guide the attention of the viewer throughout the narrative so a study was developed to find out if music, through its emotional charge, induce visual attention and contribute to the following of a given narrative.

We conclude that music can capture our attention essentially when emotion conveyed by the music is consistent with the emotion that the video is transmitting, that is, when the music is empathetic. In the situations without sound the body language of the actors proves to be the factor that captures attention of viewers, where agitation, more sudden movements, feeling of concern led the audience to prefer one video to the other.

### 5.1 Future work

With regard to future work, several aspects could enrich studies in this area:

- The use of an eye-tracker that was originally intended to be used, but under the circumstances was impossible, would be an added value in this type of study as it is a tool for studying visual attention. It is possible to know not only the video that captured the most

attention, but whether certain factors caused the attention to focus on something, or even what led to the attention.

- The number of emotions can be expanded, in this case, two quite different ones were used. In addition, more than one emotion could be transmitted by video, extending them so that the emotion changes after some time since in audiovisual terms we see this passage of emotions constantly.
- Change the presentation of the videos to an immersive environment such as *VR* because this technology along with *AR* and *MR* are the future.
- This problem is also present in videogames, and since this study was held by cinema it would be an added value to adapt it to the videogames universe and consequently explore adaptive music.

# Appendix A

# Questionnaire

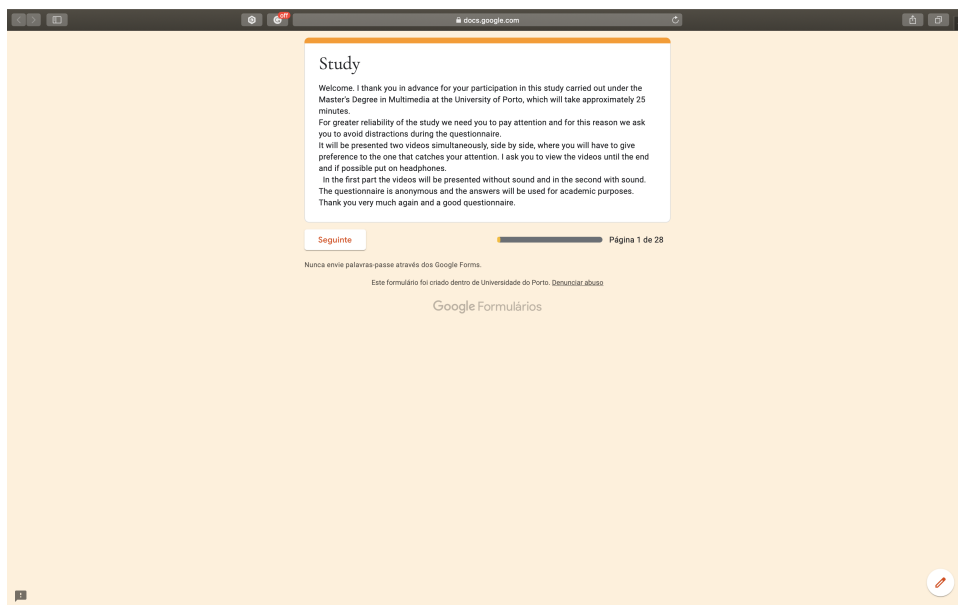


Figure A.1: Questionnaire: introduction

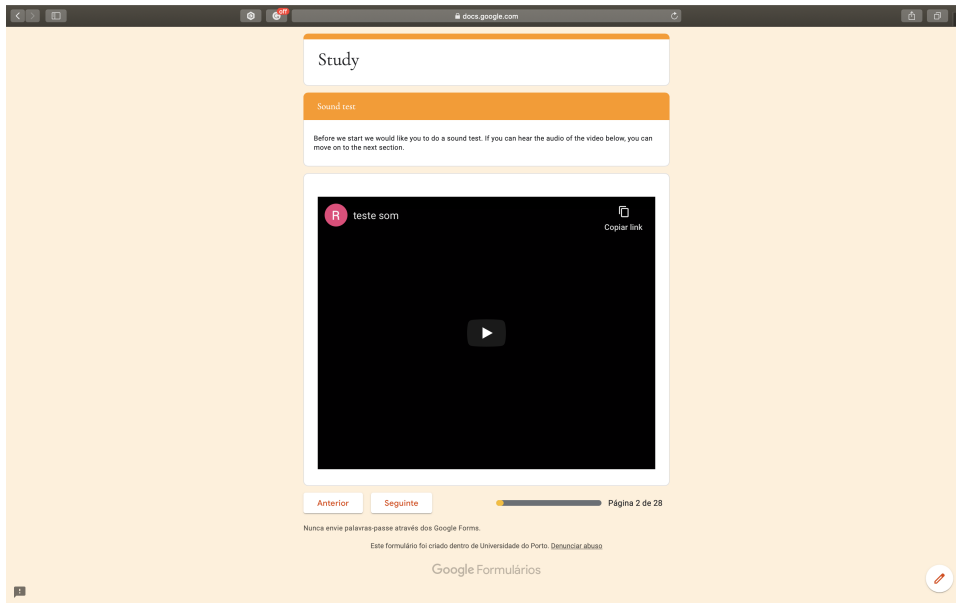


Figure A.2: Questionnaire: sound test

Study

\*Obrigatório

Gênero: \*

Female

Male

Outra: \_\_\_\_\_

Idade: \*

<15

15 - 30

30 - 45

> 45

Academic qualifications: \*

Bachelor

Master

Doctorate

Outra: \_\_\_\_\_

What are your areas of interest? (ex: engineering, design, education, music, games, health ...)

.....

Figure A.3: Questionnaire: demographic data 1

The screenshot shows a Google Forms interface with the following content:

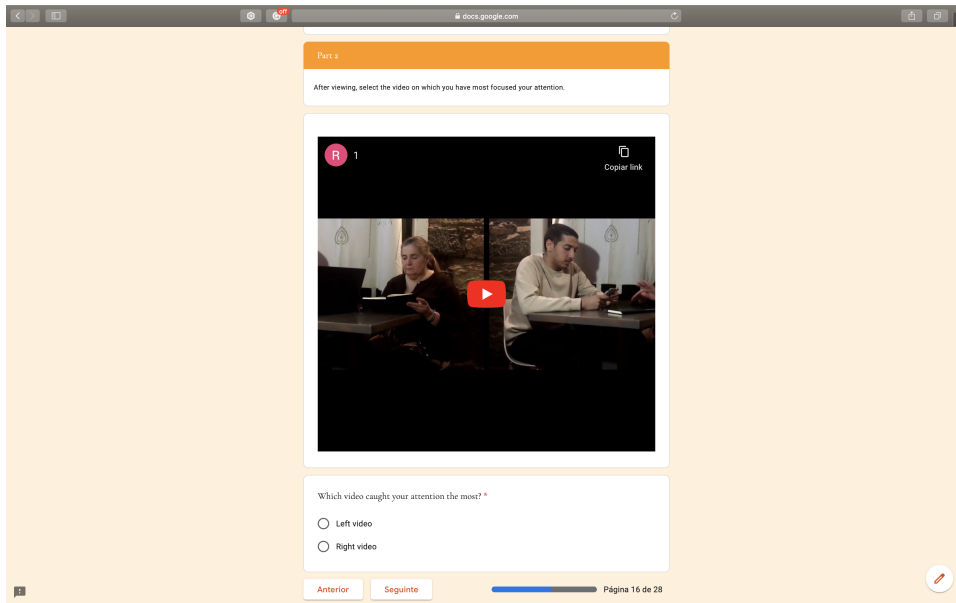
- Radio button options:  Doctorate,  Outra: \_\_\_\_\_
- Text question: "What are your areas of interest? (ex: engineering, design, education, music, games, health.....) \*"  
Response field: "A sua resposta" \_\_\_\_\_
- Text question: "Do you have a vision problem that makes it impossible to see the videos in this questionnaire? \*"  
Radio button options:  Yes,  No
- Text question: "Do you have a hearing problem? \*"  
Radio button options:  Yes,  No
- Text question: "If so, which one? "  
Response field: "A sua resposta" \_\_\_\_\_
- Navigation buttons: "Anterior" and "Seguinte"
- Page indicator: "Página 3 de 28"
- Footer: "Nunca envie palavras-passe através dos Google Forms. Este formulário foi criado dentro de Universidade do Porto. [Denunciar abuso](#) Google Formulários"

Figure A.4: Questionnaire: demographic data 2

The screenshot shows a Google Forms interface with the following content:

- Section header: "Part I"
- Text instruction: "After viewing, select the video on which you have most focused your attention."
- Video player: A video titled "Combinação\_1" showing two people at a desk. The left side shows a woman looking at a laptop, and the right side shows a man looking at a laptop. A play button is in the center.
- Text question: "Which video caught your attention the most? \*"  
Radio button options:  Left video,  Right video
- Navigation buttons: "Anterior" and "Seguinte"
- Page indicator: "Página 4 de 28"

Figure A.5: Questionnaire: Part I - example of video without sound



Part 3

After viewing, select the video on which you have most focused your attention.

1

Copy link

Which video caught your attention the most? \*

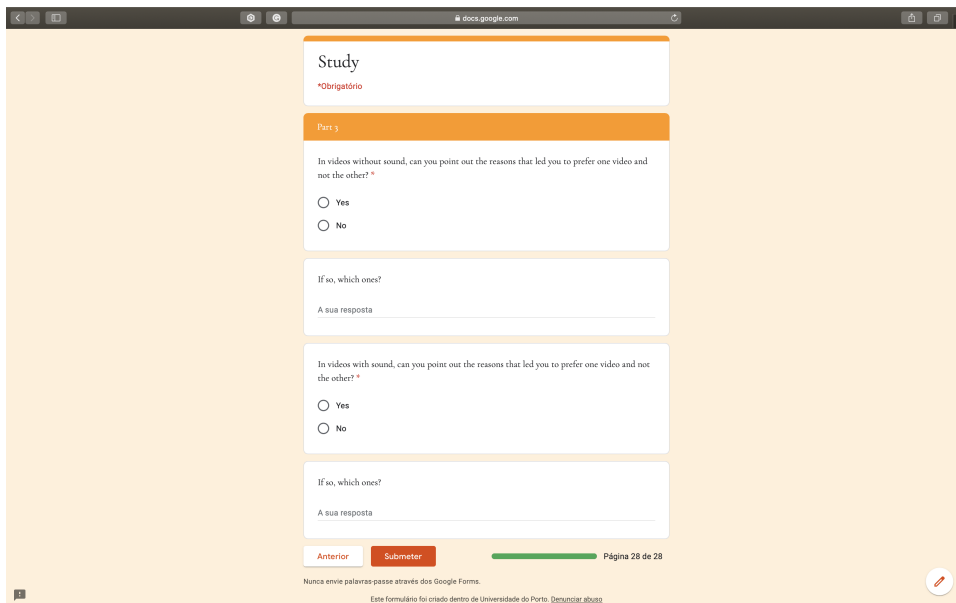
Left video

Right video

Anterior Seguinte

Página 16 de 28

Figure A.6: Questionnaire: Part II - example of video with sound



Study

\*Obrigatório

Part 3

In videos without sound, can you point out the reasons that led you to prefer one video and not the other? \*

Yes

No

If so, which ones?

A sua resposta

In videos with sound, can you point out the reasons that led you to prefer one video and not the other? \*

Yes

No

If so, which ones?

A sua resposta

Anterior Submeter

Página 28 de 28

Nunca envie palavras-passe através dos Google Forms.

Este formulário foi criado dentro de Universidade do Porto. Denunciar abuso

Figure A.7: Questionnaire: Part III - final questions

# References

- Karin Auer, Oliver Vitouch, Sabrina Koreimann, Gerald Pesjak, Gerhard Leitner, and Martin Hitz. When music drives vision: Influences of film music on viewers' eye movements. In *Conference Proceeding of the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music*, pages 23–28, 2012.
- Iris Bakker, Theo Van der Voordt, Peter Vink, and Jan de Boon. Pleasure, arousal, dominance: Mehrabian and russell revisited. *Current Psychology*, 33(3):405–421, 2014.
- Kari Bjerke Batt-Rawden. The benefits of self-selected music on health and well-being. *The Arts in Psychotherapy*, 37(4):301–310, 2010.
- Bruce Benward. *Music in Theory and Practice Volume 1*, volume 1. McGraw-Hill Higher Education, 2014.
- Ilaria Boncompagni and Maria Casagrande. Executive control of emotional conflict. *Frontiers in psychology*, 10:359, 2019.
- Tobias Brosch, David Sander, Gilles Pourtois, and Klaus R Scherer. Beyond fear: Rapid spatial orienting toward positive emotional stimuli. *Psychological science*, 19(4):362–370, 2008.
- Blain Brown. *Cinematography: theory and practice: image making for cinematographers and directors*. Taylor & Francis, 2016.
- James Buhler, David Neumeyer, and Rob Deemer. *Hearing the movies: music and sound in film history*. Oxford University Press New York, 2010.
- Edward C Carterette and Morton P Friedman. Handbook of perception: Hearing (vol. iv), 1978.
- Casablanca. Play it, sam. [https://www.youtube.com/watch?time\\_continue=3&v=7vThuwa5RZU](https://www.youtube.com/watch?time_continue=3&v=7vThuwa5RZU), 1942. In comment section, accessed on 2020/01/28.
- Chuck-jee Chau, Bin Wu, and Andrew Horner. Timbre features and music emotion in plucked string, mallet percussion, and keyboard tones. In *ICMC*, 2014.
- Michel Chion. *Audio-vision: sound on screen*. Columbia University Press, 2019.
- Priscilla Chueng and Phil Marsden. Designing auditory spaces to support sense of place: the role of expectation. In *CSCW Workshop: The Role of Place in Shaping Virtual Community*. Citeseer, 2002.
- Susan D Clayton. *The Oxford handbook of environmental and conservation psychology*. Oxford University Press, 2012.

- Kent D Cleland and Mary Dobrea-Grindahl. *Developing musicianship through aural skills: A holistic approach to sight singing and ear training*. Routledge, 2013.
- Annabel Cohen, Betty Bailey, and Thomy Nilsson. The importance of music to seniors. *Psychomusicology: A Journal of Research in Music Cognition*, 18(1-2):89, 2002.
- Annabel J Cohen. Functions of music in multimedia. *Music, mind, and science*, pages 40–68, 1999.
- Annabel J Cohen. Music as a source of emotion in film. 2011.
- Karen Collins et al. An introduction to the participatory and non-linear aspects of video games audio. *Essays on sound and vision*, pages 263–298, 2007.
- Charles E Connor, Howard E Egeth, and Steven Yantis. Visual attention: bottom-up versus top-down. *Current biology*, 14(19):R850–R852, 2004.
- Norman D Cook and Takashi X Fujisawa. The psychophysics of harmony perception: Harmony is a three-tone phenomenon. 2006.
- Deryck Cooke. *The language of music*. 1959.
- Grosvenor W Cooper, Grosvenor Cooper, and Leonard B Meyer. *The rhythmic structure of music*. University of Chicago Press, 1963.
- Kristi A Costabile, Randi Shedlosky-Shoemaker, and Adrienne B Austin. Universal stories: How narratives satisfy core motives. *Self and Identity*, 17(4):418–431, 2018.
- Film Courage. People use story to make sense of the world by adam skelter. <https://www.youtube.com/watch?v=noMIhvpV5VY>, 2018. In comment section, accessed on 2020/01/28.
- Ian Cross. Music as a communicative medium. 04 2009. doi: 10.1093/acprof:oso/9780199545872.003.0005.
- Vesna Dakic. *Sound Design for Film and Television*. GRIN Verlag GmbH, München, 2009. ISBN 3640454790.
- Tim Dalglish and Mick Power. *Handbook of cognition and emotion*. John Wiley & Sons, 2000.
- Ken Dancyger and Jeff Rush. *Alternative scriptwriting: successfully breaking the rules*. Taylor & Francis, 2007.
- Charles Darwin and Phillip Prodger. *The expression of the emotions in man and animals*. Oxford University Press, USA, 1998.
- Jane W Davidson. Visual perception of performance manner in the movements of solo musicians. *Psychology of music*, 21(2):103–113, 1993.
- Robert Desimone and John Duncan. Neural mechanisms of selective visual attention. *Annual review of neuroscience*, 18(1):193–222, 1995.
- Diana Deutsch. *Psychology of music*. Elsevier, 2013.
- Dogwoof. Making waves, 2019. URL <https://www.makingwavesfilm.co.uk>.



- Jon Driver. A selective review of selective attention research from the past century. *British Journal of Psychology*, 92(1):53–78, 2001.
- History.com Editors. Printing press. <https://www.history.com/topics/inventions/printing-press>, 2019. In comment section, accessed on 2020/01/28.
- Paul Ekman. Are there basic emotions? 1992.
- Paul Ekman and Wallace V Friesen. *Unmasking the face: A guide to recognizing emotions from facial clues*. Ishk, 2003.
- Jin Fan, Bruce D McCandliss, Tobias Sommer, Amir Raz, and Michael I Posner. Testing the efficiency and independence of attentional networks. *Journal of cognitive neuroscience*, 14(3): 340–347, 2002.
- Paul R Farnsworth. A study of the hevner adjective list. *The Journal of Aesthetics and Art Criticism*, 13(1):97–103, 1954.
- Alicia Fernández-Sotos, Antonio Fernández-Caballero, and José M Latorre. Influence of tempo and rhythmic unit in musical emotion regulation. *Frontiers in computational neuroscience*, 10: 80, 2016.
- W Tecumseh Fitch and Andrew J Rosenfeld. Perception and production of syncopated rhythms. *Music Perception*, 25(1):43–58, 2007.
- Jay Friedenberg and Gordon Silverman. *Cognitive science: An introduction to the study of mind*. Sage, 2011.
- Alf Gabrielsson and Erik Lindström. The role of structure in the musical expression of emotions. *Handbook of music and emotion: Theory, research, applications*, 367400, 2010.
- Ronny Geva, Michal Zivan, Aviv Warsha, and Dov Olchik. Alerting, orienting or executive attention networks: differential patters of pupil dilations. *Frontiers in behavioral neuroscience*, 7: 145, 2013.
- Jessica Green. Understanding the score: Film music communicating to and influencing the audience. *The Journal of Aesthetic Education*, 44(4):81–94, 2010.
- Melanie C Green and Timothy C Brock. The role of transportation in the persuasiveness of public narratives. *Journal of personality and social psychology*, 79(5):701, 2000.
- Jacek Grekow. Mood tracking of musical compositions. In *International Symposium on Methodologies for Intelligent Systems*, pages 228–233. Springer, 2012.
- Jacek Grekow. Audio features dedicated to the detection and tracking of arousal and valence in musical compositions. *Journal of Information and Telecommunication*, 2(3):322–333, 2018.
- Ralph H Gundlach. Factors determining the characterization of musical phrases. *The American Journal of Psychology*, 47(4):624–643, 1935.
- Julia C Hailstone, Rohani Omar, Susie MD Henley, Chris Frost, Michael G Kenward, and Jason D Warren. It’s not what you play, it’s how you play it: Timbre affects perception of emotion in music. *The quarterly Journal of Experimental psychology*, 62(11):2141–2155, 2009.

- David J Hargreaves, Raymond MacDonald, and Dorothy Miell. How do people communicate using music. *Musical communication*, 1:1–26, 2005.
- David J Hargreaves et al. *The developmental psychology of music*. Cambridge University Press, 1986.
- Patrick Helmholz, Dominik Siemon, and Susanne Robra-Bissantz. Summer hot, winter not!—seasonal influences on context-based music recommendations. 2017.
- Ann-Kristin Herget. On music’s potential to convey meaning in film: A systematic review of empirical evidence. *Psychology of Music*, page 0305735619835019, 2019.
- David Herman. *Experiencing narrative worlds: on the psychological activities of reading*, 1997.
- Kate Hevner. Experimental studies of the elements of expression in music. *The American Journal of Psychology*, 48(2):246–268, 1936.
- Kate Hevner. The affective value of pitch and tempo in music. *The American Journal of Psychology*, 49(4):621–630, 1937. ISSN 00029556. URL <http://www.jstor.org/stable/1416385>.
- Roger Hickman. *Reel music : exploring 100 years of film music*. W.W. Norton & Company, New York, 2017. ISBN 0393937666.
- Tomlinson Holman. Roles of sound. *Retrieved on April, 6:2005*, 1999.
- Tomlinson Holman. *Sound for film and television*. Routledge, 2012.
- Jeffrey R Huntsinger. Does emotion directly tune the scope of attention? *Current Directions in Psychological Science*, 22(4):265–270, 2013.
- Gabriella Ilie and William Forde Thompson. A comparison of acoustic cues in music and speech for three dimensions of affect. *Music Perception*, 23(4):319–330, 2006.
- Toastmasters International. *GESTURES:YOUR BODY SPEAKS*. TOASTMASTERS INTERNATIONAL, 2011.
- Nick Iuppa and Terry Borst. *Story and simulations for serious games: Tales from the trenches*. CRC Press, 2012.
- William James, Frederick Burkhardt, Fredson Bowers, and Ignas K Skrupskelis. *The principles of psychology*, volume 1. Macmillan London, 1890.
- Brian Jay Jones. *George Lucas*. Hachette UK, 2016.
- Patrik N Juslin. *Musical emotions explained: Unlocking the secrets of musical affect*. Oxford University Press, USA, 2019.
- Patrik N Juslin and Petri Laukka. Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *Journal of new music research*, 33(3): 217–238, 2004.
- Ronald T Kellogg. *Fundamentals of cognitive psychology*. Sage Publications, 2015.
- Eric I Knudsen. Fundamental components of attention. *Annu. Rev. Neurosci.*, 30:57–78, 2007.

- Peter Kobel. the library of congress. *Silent Movies: The Birth of Film and the Triumph of Movie Culture*, 2007.
- Vladimir J Konečni. Does music induce emotion? a theoretical and methodological analysis. *Psychology of Aesthetics, Creativity, and the Arts*, 2(2):115, 2008.
- Michael Kubovy. *Human Perception*. Routledge, nov 2017. ISBN 081538954X. URL <https://www.xarg.org/ref/a/081538954X/>.
- Justin London. *Hearing in time: Psychological aspects of musical meter*. Oxford University Press, 2012.
- Laurent Mannoni. Silent cinema: A guide to study, research and curatorship. *Journal of Film Preservation*, (101):155–157, 2019.
- John Marzluff and Tony Angell. *Gifts of the crow: how perception, emotion, and thought allow smart birds to behave like humans*. Simon and Schuster, 2013.
- Albert Mehrabian. *Nonverbal communication*. Transaction Publishers, 1972.
- Albert Mehrabian and James A Russell. *An approach to environmental psychology*. the MIT Press, 1974.
- Albert Mehrabian et al. *Silent messages*, volume 8. Wadsworth Belmont, CA, 1971.
- Enrico Mezzacappa. Alerting, orienting, and executive attention: Developmental properties and sociodemographic correlates in an epidemiological sample of young, urban children. *Child development*, 75(5):1373–1386, 2004.
- Carolyn Miller. *Digital storytelling : a creator's guide to interactive entertainment*. Elsevier/Focal Press, Amsterdam Boston, 2004. ISBN 0415836948.
- Patrick Miller. Music and the silent film. *Perspectives of New Music*, pages 582–584, 1982.
- MixedBag. Avery, May 2018. Disponível em <http://mixedbag.it/portfolio/avry/>.
- Gailynn Nicks and Yannick Carriou. Emotion , attention and memory in advertising. *Warc*, (November), 2016.
- Adrian C North, David J Hargreaves, and Susan A O'Neill. The importance of music to adolescents. *British journal of educational psychology*, 70(2):255–272, 2000.
- Yohei Ono and Yasutomi Taniguchi. Attentional capture by emotional stimuli: Manipulation of emotional valence by the sample pre-rating method. *Japanese Psychological Research*, 59(1): 26–34, 2017.
- Pamela Mia Paul. *The impact of the development of the fortepiano on the repertoire composed for it from 1760–1860*. PhD thesis, Citeseer, 2012.
- Barbara Pease and Allan Pease. *The definitive book of body language: The hidden meaning behind people's gestures and expressions*. Bantam, 2008.
- Ken Petress. The importance of music education. *Education*, 126(1), 2005.
- Titus Petrikis. Creating a new sound world for dracula (browning 1931). *The New Soundtrack*, 5: 35–54, 03 2015. doi: 10.3366/sound.2015.0067.

- Yair Pinto, Andries R van der Leij, Ilja G Sligte, Victor AF Lamme, and H Steven Scholte. Bottom-up and top-down attention are independent. *Journal of vision*, 13(3):16–16, 2013.
- Jonathan Posner, James A Russell, and Bradley S Peterson. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and psychopathology*, 17(3):715, 2005.
- Michael I Posner. Orienting of attention. *Quarterly journal of experimental psychology*, 32(1): 3–25, 1980.
- R7. Grândola, Vila Morena": o símbolo da música que libertou os portugueses, March 2011. Disponível em <https://www.telegraph.co.uk/culture/film/classic-movies/8406944/Silent-films-were-never-really-silent.html>.
- Reporter. The evolution of storytelling. <https://reporter.rit.edu/tech/evolution-storytelling>, 2015. In comment section, accessed on 2020/01/28.
- Russell Revlin. *Cognition: Theory and practice*. Macmillan, 2012.
- Frank Rose. *The art of immersion: How the digital generation is remaking Hollywood, Madison Avenue, and the way we tell stories*. WW Norton & Company, 2012.
- M ROSARIO Rueda and PM Paz-Alonzo. Executive function and emotional development. *contexts*, 1:2, 2013.
- M Rosario Rueda, Joan P Pozuelos, and Lina M Cómbita. Cognitive neuroscience of attention from brain mechanisms to individual differences in efficiency. 2015.
- James A Russell. A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161, 1980.
- James A Russell. Is there universal recognition of emotion from facial expression? a review of the cross-cultural studies. *Psychological bulletin*, 115(1):102, 1994.
- Inês Salselas and Rui Penha. The role of sound in inducing storytelling in immersive environments. In *Proceedings of the 14th International Audio Mostly Conference: A Journey in Sound on ZZZ*, pages 191–198, 2019.
- Inês Salselas, Rui Penha, and Gilberto Bernardes. Sound design inducing attention in the context of audiovisual immersive environments. *Personal and Ubiquitous Computing*, pages 1–12, 2020.
- Riccardo Santoboni and Marta Olivetti Belardinelli. Influence of timbre on emotions and recognition memory for music. 2005.
- Emery Schubert. Update of the hevner adjective checklist. *Perceptual and motor skills*, 96(3\_suppl):1117–1122, 2003.
- Yeong-Seok Seo and Jun-Ho Huh. Automatic emotion-based music classification for supporting intelligent iot applications. *Electronics*, 8:164, 02 2019. doi: 10.3390/electronics8020164.
- Ramprakash Srinivasan and Aleix M Martinez. Cross-cultural and cultural-specific production and perception of facial expressions of emotion in the wild. *IEEE Transactions on Affective Computing*, 2018.

Anthony Storr. *Music and the Mind*. Simon and Schuster, 2015.

Durk Talsma, Daniel Senkowski, Salvador Soto-Faraco, and Marty G Woldorff. The multifaceted interplay between attention and multisensory integration. *Trends in cognitive sciences*, 14(9): 400–410, 2010.

Siu-Lan Tan, Annabel J Cohen, Scott D Lipscomb, and Roger A Kendall. *The psychology of music in multimedia*. Oxford University Press, 2013.

Jonathan D Tankel. The impact of the jazz singer on the conversion to sound. *Journal of the university film association*, 30(1):21–25, 1978.

The Telegraph. Silent films were never really silent, April 2014. Disponível em <https://noticias.r7.com/internacional/grandola-vila-morena-o-simbolo-da-musica-que-libertou-os-portugueses-23042014>

Roy Thompson and Christopher Bowen. *Grammar of the Shot*. Taylor & Francis, 2009.

Patrik Vuilleumier. Facial expression and selective attention. *Current Opinion in Psychiatry*, 15 (3):291–300, 2002.

Patrik Vuilleumier. How brains beware: neural mechanisms of emotional attention. *Trends in cognitive sciences*, 9(12):585–594, 2005.

Harald G Wallbott. Bodily expression of emotion. *European journal of social psychology*, 28(6): 879–896, 1998.

Ann-Kristin Wallengren and Alexander Strukelj. Film music and visual attention: A pilot experiment using eye-tracking. *Music and the Moving Image*, 8(2):69–80, 2015.

Victor CX Wang. *Encyclopedia of strategic leadership and management*. IGI Global, 2016.

Web of Stories. Walter Murch - Film-maker, March 2017. Disponível em <https://www.webofstories.com/play/walter.murch/1>.

Edda Weigand. *Emotion in dialogic interaction: advances in the complex*, volume 248. John Benjamins Publishing, 2004.

Norbert Wiener. *The Human Use of Human Beings, Cybernetics and Society*. Norbert Wiener,... Houghton Mifflin Company, 1950.

Benjamin Wright. Atmos now: Dolby laboratories, mixing ideology and hollywood sound production1. *Living Stereo: Histories and Cultures of Multichannel Sound*, page 227, 2015.

Bin Wu, Simon Wun, Chung Lee, and Andrew Horner. Spectral correlates in emotion labeling of sustained musical instrument tones. In *ISMIR*, pages 415–420, 2013.

Bin Wu, Andrew Horner, and Chung Lee. Musical timbre and emotion: The identification of salient timbral features in sustained musical instrument tones equalized in attack time and spectral centroid. In *ICMC*, 2014.

Jeffrey M Zacks. *Flicker: Your brain on movies*. Oxford University Press, USA, 2015.