UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Università degli Studi di Padova

Dipartimento di Studi Linguistici e Letterari

Corso di Laurea Magistrale in
Strategie di Comunicazione
Classe LM-92

Tesi di Laurea

Emotions in everyday sounds. An experimental study on the influence of colour in emotional knocking sounds.

Relatore
Prof. Antonio Rodà
Dipartimento di Ingegneria
dell'Informazione

Laureanda
Asia Turato
n° matr.2010524 / LMSGC

Anno Accademico 2021 / 2022

We gestate in Sound, and are born into Sight

Cinema gestated in Sight, and was born into Sound.

Walter Murch in the preface of

Audio-Vision. Sound on Screen

by Michel Chion, 1994.

CONTENTS

INTRODUCTION

This thesis is an experimental project on sound design, aimed at exploring the world of sound and its artistic and communicative applications concerned with emotion elicitation. When designing any type of communication, whether it is an informative content, a marketing campaign or a multimedia product, it is important to know how to express specific concepts, what their effect on the audience might be and how they can be implemented and reinforced by means of communication tools. Therefore, this study aspires to give some directions on the design of multimedia products aimed at generating emotional reactions in the audience, that can be applied to more artistic fields such as the cinema and videogames industries, as well as marketing campaigns.

Sound and emotions are apparently two very different concepts.

Sound is the element that completely revolutionised the cinema industry in the 1920s. Through sound it became possible not only to give life to the actors and objects on screen, but also to express numerous other information. Indeed, by paying particular attention to the making of a sound, this can become a character itself, an instrument to unify different scenes and time spans, as well as a tool to generate the overall tone of a movie.

However, sounds in movies, such as sound effects and musical soundtrack, are not the only ones able to achieve these goals. Everyday sounds, for instance walking and knocking on a door, are more powerful means of information that one may think. It is this last type of sounds, more precisely knocking sounds, the object of the present study. In particular, it will be discussed the potential sound has, especially when correlated with the elicitation of emotions.

The field of emotions is one of the most complicated yet studied ones in science. Not only is it difficult to determine what emotions are, but also to identify and classify them is a quite complicated task. Many different researchers tried to give a definition of this term, as well as provide a list of basic/primary/fundamental emotions. Many recent studies investigating emotion elicitation refer to Ekman's system of basic emotions (Ekman, 1999), but a common agreement on this topic still has to be found among the scientific community.

Emotions can be elicited by many different stimuli, whether they are haptic, olfactory, tasty, visual or audible. In particular, the interaction between these two latter senses is the object of the present study, which aims at investigating the influence of colours in knocking sounds. In previous researches, both colours and everyday sounds proved to be elements capable of eliciting a set of different emotions based on their parameters: hue, saturation and brightness in the first case, and pitch, intensity, frequency, timbre, duration and more in the second one.

The perception of emotions through these two channels, visual and aural, will be studied by means of an empirical experiment aiming at investigating the influence of certain colours, proved to be associated with specific emotions, on emotionally performed knocking sounds.

This study is based on the hypothesis that the association between knocking sounds and emotions, already validated by Pauletto and Iop (2021), changes if the knocking sound is performed on a door of a different colour. This study draws inspiration from the mentioned one since audio stimuli from the same dataset, also used in Barahona-Ríos and Pauletto (2020), have been used and, as a consequence, the same set of emotions has been selected among Ekman's basic list. However, the present study differs from the previous one for mainly three aspects, concerning the colour-emotion associations, the visual stimuli and the participant's method of judgement. In the first case, the colour-emotion combinations we expect to find are: yellow and happiness, red and anger, grey and sadness, and purple and fear. These associations were made on the basis of an analysis of several studies on the

emotional perception of colours (§ 2.3). Secondly, this study uses videos of a hand knocking on a door of different colours, while theirs used still images of differently coloured doors presented together with a knocking sound. Finally, participants of the present study judged the emotion portrayed on the videos using four sliders built on a continuous scale from 0 (total disagreement) to 10 (total agreement), each representing one of the chosen emotions. Whereas in Pauletto and Iop (2021), participants had to judge the emotion through a single-choice question.

Moreover, we included some follow-up questions to understand if some of the participants are synesthetes.

Therefore, we expect to have different results from the study conducted by Pauletto and Iop (2021), which did not report significant influence of colours, showing how the aural channel dominated on the visual one. To collect data, we used the free online software PsyToolkit[1] (Stoet, 2010, 2017), while for the analysis, we analysed the mean values and their standard deviation.

Before presenting the experiment, described in Chapter 3, an in-depth analysis of sound, auditory perception and emotions will be carried out in the first two chapters.

More specifically, the first chapter will examine the definition and characteristics of sound, and how these can be used to achieve many different purposes: from telling information and constructing a story, to eliciting emotions. Different examples of how sounds are used in the movie industry, as well as how they are perceived in everyday life will be discussed, along with the importance of the work of people working in the sound industry. To this end, particular attention will be given to the figure of the Foley artist and its fundamental role in the history of cinema and in the development of a film's sound effects. While discussing these topics, specific focus will be put in the relationship between sound and emotions.

[1] https://www.psytoolkit.org/

Finally, some examples of illusions created through the combination of the visual and the aural channels will be examined.

Chapter 2 will, on the other hand, present auditory and emotional perception from a more psychological point of view. The theories from Charles Darwin, Robert Plutchik and Paul Ekman will be discussed since they present the theoretical foundations and starting point of the research reported in Chapter 3. Furthermore, the topic of emotions will be put in relation to that of colours and specific colour-emotion associations will be presented. Numerous studies that investigated this process will be brought to the reader's attention, as they were fundamental in the colour choice for the present experiment.

Finally, Chapter 3 presents the experiment conducted. The definition of the questionnaire's structure has been done in collaboration with Prof. Emery Schubert and Diana Zhang of the University of New South Wales Sydney, and Dr. Anthony Chmiel of the Western Sydney University.

1. THE SOUND

"Anything in our world that moves vibrates air. If it moves in such a way that it oscillates at more than about 16 times a second this movement is heard as sound. The world, then, is full of sounds. Listen." This is how R. Murray Schafer introduced his chapter on the sonic environment in his book *The New Soundscape. A Handbook for the Modern Music Teacher* (1969, p.5). Indeed, everyday, in every moment, we are continuously, consciously or unconsciously, surrounded by sounds, whether they arrive from a person, for instance speech, from nature, for example wind, rain, leaves, from objects, like a car, a device, a creaking chair, from media products, e.g. a movie, a song, a video game, or from an action we make, e.g. walking, knocking, typing on a computer and many more.

If seen from a technical point of view, sound is actually a vibration, a wave of energy, that our auditory system, composed of our ears and brain, translates into neural impulses and allows us to perceive and hear a certain and identifiable sound. Depending on the way in which this vibration propagates in space and reaches the receiver, it can provide useful information about its source, its medium but also the space in which it was emitted and the distance of the listener from the source.

Therefore, sound is much more than just a vibration: it is also a very powerful source of information about our surroundings, emotional states, movements and other indicators that sometimes are taken for granted but can also be fundamental for some people. Indeed, sound plays an essential role for blind people in helping them to perceive the surrounding world with all its elements, spaces, obstacles and actors. For the blind, the soundscape is one of the most important means of obtaining information about the world. In fact, by not being dominated by the more powerful sense of vision, visually impaired people have developed the ability to perceive and hear sonic elements and features that normal-vision people can not (Southworth, 1967).

As the following sections will examine more in depth, sound is much more than noise (term used, in this case, as a general reference to everyday sounds and not with its technical connotation of "unpitched sound without partials") that only accompanies what we see.

Because of its perceptive qualities and functions, which will be discussed in the present Chapter, as well as in Chapter 2, sound can initiate networks of associations, both bodily and cognitive (Fahlenbrach, 2008), making it the perfect instrument for the transmission of information, storytelling, elicitation and intensification of emotions, besides the creation of perceptual illusions when combined with the visual channel. It is with the union of this last element that sound creates an additional value, to paraphrase Michel Chion (1994), in creating an unique experience that is greater than the sum of the two parts and that would, therefore, not be perceived by the audience in the same way as if it was expressed by the two channels alone.

All of these several practices of sound are what make it the real game changer and the more versatile instrument in the media production industry with the introduction and development of specific working figures dedicated to the analysis, the creation and the production of sound.

1.1. SOUND AS A SOURCE OF INFORMATION

As Schafer (1969) suggested, the world we live in is full of sounds, and, for this reason, the soundscape we are immersed in our everyday life is composed of a myriad of different sounds such as music, speech, animal and other nature noises and a multitude of others generated by the environment, which are the so-called *everyday sounds*. These latter types of sounds can include, for example, our fingers leafing through a book or a newspaper, car keys, a car door, a bathroom or kitchen tap, fingers typing on a keyboard, our feet walking on concrete and our hands knocking on a door. All of these sounds are generated by the action of humans (Susini, Hoiux & Misdariis, 2014) and in the present study *knocking sounds* have been the focus of investigation (Chapter 3). This type of sounds, as well as the walking ones, are the so called *impact sounds*, which are impulsive signals generated by a short time interaction between two objects (Visell et al., 2009): a hand and door in the first case and feet and an horizontal surface in the second one.

When hearing a sound, multiple information can be acquired thanks to its perceptual and physical parameters such as loudness, pitch, timbre, reverb and frequency and humans have the ability to identify objects or the event that produced that sound just by hearing it. As it will be further discussed in § 2.1, this human and, more in general, animal ability to discriminate between the origin of the sound source is due to evolution necessities such as danger awareness.

The type of information that these characteristics reveal can be, for instance, about the space from which the sound comes, the time, the intention, the emotional state of a person and the relationship this one can maintain with someone else.

Regarding space, depending on the loudness, the distance of a sound event from the receiver can be deduced: the louder the sound is, the closer the source or the event will be to the

listener. Additionally, reverberation and echo are indicators of the size of the space and the amount of elements present in it. For example, a sound emanated in a big room with little to no furniture will produce a longer reverb than the one emitted in a small one, and a scream on the top of a mountain will have a longer echo than one emitted on a street full of skyscrapers. This happens because everytime a sound wave hits an obstacle, e.g. a building or a wall, part of its energy is absorbed or transmitted, therefore its amplitude decreases. In addition, the material of the obstacle, and in general of all the elements present in the environment, affects these two parameters as well: different materials have different absorption coefficients, which lead to a longer or shorter reverb. For example, recording studios are made of isolating and very absorbing materials in order to get the perfect voice quality and not create sound reflections that will act as interference. Furthermore, the size of the source usually determines its fundamental frequency, thus, from the analysis of the fundamental frequency, one can figure out the size of the sound source as well as its characteristics (Li et al., 1991). The importance of the space and material information that sound conveys is essential in humans and it is clearly visible in blind people, as Southworth (1967) highlights. As a matter of fact, smaller and hard surfaced places "resonate with the sounds of one's voice and footsteps, and 'sound shadows' can inform one of the size of a space, its contents, materials, and also about one's location in the space" (*ibidem*, p. 10) and the height of the frequencies rebounding from an object gives clues whether the object hit was made of glass (higher frequencies) or was a brick (lower frequencies).

Concerning time, the regular beat of a sound, for instance the seconds ticking on a clock, or the amount of times a sound is repeated, e.g. a church bell ringing three times means it is 3 o'clock, can keep track of time and, as it will be further described in section 1.2, the rhythm of music as well as of ticking sound effects contribute to the speed of a scene: making it seem frenetic or calm.

Furthermore, some specific sets of physical and perceptual sound parameters can indicate the intention of an object. For instance, a fast rate, a high pitch, and irregular harmonics are indicators of urgency (Susini, Hoiux & Misdariis, 2014). When hearing an alarm, to give an example, its volume, pitch and irregularity tell us that we need to act quickly and the identification of such source characteristics of sound can be decisive in critical situations or to be able to do specific tasks. Particular attention must therefore be put in the design of such devices (Suied et al., 2008). In the case of the everyday sound of knocking on a door, knocks performed with a high frequency and with a high volume might indicate that the person knocking is scared, while few low intensity knocks would probably suggest that the person is sad (Bedenko et al., 2020). The same reasoning can be done for any action made by humans, such as walking. Indeed, everyday sounds are very explicit tellers of the relationship between people and their emotions, as well as the physical characteristics of the person doing the action. This aspect is once again clearly visible in the case of walking and knocking, the two most studied human-produced sounds.

In the first case, the shoe size, the type of shoe, the ground material, the pace and the presence of high frequency components can be important cues to determine the gender of the walker, its weight, its posture as well as its emotional state (Li et al., 1991; Visell et al, 2009; De Gotzen et al., 2013; Turchet & Rodà, 2017). It is interesting to notice how the same object, a shoe, can actually emit such different sounds by being of a different brand, model, owner and, of course, by using it to walk on different ground materials. As Ament (2009) herself noticed, she was fascinated by the fact that all of her shoes sounded so different and never really paid too much attention to it before.

In the case of knocking on a door, we can not only understand the state of the person knocking and the material of which the door is made, but the former indicator can also lead to understand the relationship between the knocker and the person on the other side of the door

(Houel, Arun et al., 2020; Pauletto & Iop, 2021; Bedenko at al., 2020;). The information that the person listening acquires is that there is someone outside the room that would like to enter. Then, the way in which the knocks are performed, like the duration, the number of knocks, the rate and the regularity can indicate whether the person knocking shares familiarity with the one on the other side, or whether he or she is happy, angry, sad and so on. For instance, a more neutral knocking may indicate that there is no intimacy between the two or more subjects, whereas a more emotionally performed sound, such as happy or angry, suggests that the people involved in the situation share some previous experiences and closeness.

A further way in which the listener can perceive the emotion of the other person is through the voice: once again, based on the variation of prosodic parameters like pitch and duration, but also loudness and voice quality, people can judge if their interlocutor is happy, angry, sad or surprised. However, the parameter that contributes the most in the perception and transmission of emotions in the voice is pitch (Murray & Arnott, 1993; Frick, 1985; Lieberman & Michaels, 1962 all as cited in de Gelder & Vroomen, 2000): a high pitched voice is likely to be an index of happiness, while a low pitched one may indicate anger or even sadness.

Another aspect that is interesting to note is that, depending on the perceptual parameters of sound, humans are also able to envision specific visual entities or associate the way a sound sounds to an object. This is the case of the *Maluma/Takete* experiment proposed by Köhler (1920s, as cited in Görne, 2017) who demonstrated that people associate the word “Maluma” to a round figure and “Takete” to an edgy one, even though these two words do not exist and therefore mean absolutely nothing in real life. This phenomenon, related to phonosymbolism, which is the emotional meaning of a phoneme (Whissell, 1999), is what Görne calls *crossmodal metaphors*: a term to indicate the connection between audio stimuli and the

consequent perception of visual metaphors generated by the sound. Thanks to this association, it is possible to communicate with sound design. For instance, edgy objects such as weapons, e.g. knives or swords, are characterised by a sharp and high pitched noise, which is also connected to light images, whereas low pitched sounds are associated with rounder objects and may give a feeling of power. Despite some scholars acknowledging the natural emotional connotation of certain phonemes (Tsur, Fónagy, Jakobson as cited in Whissell 1999), others believe that the sounds are arbitrarily associated to symbols whose meanings derive from conventions (Chomsky, Skinner, Saussure, as cited in *ibidem*). However, this distinction is not so strict, since sound has actually the possibility to be both expressive and arbitrary, even if, in the first case, subjectivity plays an important role so the meaning of a phoneme is not to be intended as absolute.

Phonosymbolism can also create associations with colours (Esposito, Sini & Castagneto, 2018) and emotions, but the relationship between these last two aspects will be further examined in § 2.3. Sounds and the metaphors they create in human minds are also present in other contexts apart from the movie or multimedia industry. Indeed, in marketing, the way a brand name or a product sounds may influence the potential customers in the consideration and purchase phases of the customer's journey if the product is new and therefore not well established in the consumer's mind yet.

These different kinds of information, for instance the material and the emotional state, are related to the intention of sounds. As a matter of fact, it is fundamental to distinguish the form of a sound from its function (Susini, Hoiux & Misdariis, 2014). On one side, intention of form is related to the quality of the object and of the sound, and to express this concept in the best and most correct way possible, coherence is a crucial aspect: indeed, when creating sounds for any kind of product, they must be coherent with it regardless of their function. In the case of knocking on a door, the form is related to the material of the door, thus a wooden

door cannot sound like a metal or a glass one. On the other side, intention of function considers sound as a "communicator of necessary and useful information" (*ibidem*, p. 108) in order to use or interpret the object and its messages. Once again, in the case of knocking on a door, the knock tells us someone wants to enter a room. Some examples of functions can be notification, to warn or to guide, feedback, to confirm, and interaction, to facilitate the user's interaction with the object and these purposes of sound can be easily recognisable in digital devices, such as computers, tablets and smartphones.

The entirety of the above-mentioned sound properties, characteristics and information must be taken into account when creating the sound for a movie, a video game, a virtual reality environment as well as industrial products. In the cases of media products, sound and its features can become very important storyteller devices. A great example of how sound alone can provide a myriad of information is the work *The design of an audio film for the visually impaired* by Lopez and Pauletto (2009), in which they demonstrated how creating a movie for blind people, therefore using only sound events with no need for a narrator, is doable. To achieve this, they based their story on Roald Dahl's *Lamb to the Slaughter* (1954) and through sound effects, reverberation and music they were able to convey information about the actions, specific spaces and layouts, feelings and tension to their listeners, respectively.

Other two great and well-known examples of how sound can become a storyteller device, if not its driver, are the movies *Dunkirk* (2017) and *C'era una volta il West* (1968). These two movies will be further examined in the following sections § 1.2 and § 1.3 as instances of the informational and innovative ways of using sound in movie productions. The first movie is remarkable for its masterful use of sound, sound effects and musical score causing it to win the Oscar for Best Sound Editing and Best Sound Mixing in 2018, while the second one represents a milestone in the cinema industry, also for its careful and intelligent use of sounds and soundtrack.

1.2. THE AREAS OF SOUND

In their work *The Craftsman: The use of sound design to elicit emotions* (2014), Hillman and Pauletto tried to develop sound producing techniques in order to trigger specific audience responses. The researchers underlined the importance of the context in which the sound is placed, since no sound is ever heard in isolation and in their research they proposed the concept of the four "areas of sound". These areas are the Logical, the Abstract, the Temporal and the Spatial and the balance of these four areas in the soundscape of a movie is what contributes to the elicitation of emotions in the viewers/listeners.

Firstly, the Logical area contains dialogue, commentary, diegetic music and symbolic and signalling sounds, therefore it is the area that usually carries the most meaning.

Secondly, the Abstract area presents atmospheres, room tones and synchronous and non-synchronous sound effects.

Thirdly, as the name suggests, the Temporal area deals with the rhythm, pace and punctuation of the soundtrack and it is characterised by a contrast between high and low temporal features. Within this category, the distinction between *empathetic* and *anempathetic* music made by Chion (1994) can be placed. Indeed, for the author, there are two ways in which music in film can create a specific emotion. The first one, the empathetic one, is by directly expressing the feeling of the scene, for instance by taking on the scene's rhythm, tone, and phrasing. This type of music is what Cristopher Nolan and Hans Zimmer used in *Dunkirk,* which is a movie that makes sound its glue and protagonist and if sound had not been used in such an innovative and intelligent way, the output would have not been the same. In one of the scenes, the steps of a soldier running on the beach are matched to the rhythm of a ticking sound, which is also a recurring theme in the whole movie (Rapan, 2018).

The second one, the anempathetic one, creates conspicuous indifference to the situation depicted on the screen, by progressing in a steady, undaunted and ineluctable way, in contrast to the visual scene. This juxtaposition of action and indifference has the effect of intensifying the emotion. In other scenes of *Dunkirk*, the opposite process happens. The leitmotif of the movie is "waiting": the British soldiers are stuck on the beach and they have nowhere to go, making some scenes of the movie quite static from the visual point of view. However, the musical score has the ability to make those static scenes full of tension, for example when announcing in a *non-diegetic* way (see below) the arrival of the enemies (Knight-Hill, 2019).

Lastly, the Spatial area is about the position of sounds inside the soundscape.

Some examples to better understand this concept are:

- Logical: footsteps on a certain material, a creaking door, a dialogue, a camera flash, the wind and a gunshot;
- Abstract: atmosphere sounds such as murmur, the traffic and bar sounds;
- Temporal: breathing, ticking, the musical soundtrack and passing traffic;
- Spatial: room reverb, murmur and nature sounds

(Hillman and Pauletto, 2014).

Knocking, the object of this research, can be a mixture of Logical (predominant), Temporal (depending on the pace of the action) and Spatial (if the material of the door and the size of the room produce reverb).

As a result, when creating the sounds in the post-production of a movie, multiple factors must be taken into account, consequently, professional figures are crucial in the making of a quality product. This aspect will be further examined in §1.3.1.

When using sounds in media products other "areas" of sound placement must be taken into account. Indeed, two positions in which sound can be placed are inside the story and outside of it. More specifically, an audio that is inside the story is called *diegetic,* for instance a

dialogue, an object falling or a piano played by a character, while what comes from the outside is called *non-diegetic*, for example a narrating voice or the soundtrack. However, this does not mean that only audio registered during the shooting of a movie is diegetic: sounds made in the post-production, for instance *Foley sounds*, which will be discussed at a greater length in §1.3.1, are diegetic even though they were edited, mixed and put in the scene after it was filmed. To make the difference between these two types of sounds clearer, one must remember that if a character in the movie, video or video game is likely to hear the sound, this one is diegetic, whereas if only the spectator will hear it, then it is a non-diegetic sound.

Based on whether the sound is diegetic or non-diegetic, the information rate that the sound brings changes. A sound design uniquely composed of diegetic sounds is called *documentaristic* or *naturalistic*, and its information rate is low given the redundancy between image and sound: what we hear is what we see, while a sound design that presents only non-diegetic sounds is called *surreal* or *mystical* since it enriches the meaning of the picture by adding information that otherwise would not be given (Görne, 2017). Apart from these two levels, the author suggests two more categories of informational sounds. The first one being the *quasi-documentaristic* sound design, made by mainly diegetic sounds and its purpose is to guide the audience's attention to particular parts of the scene, and the second one being the *supernatural* one. This latter type of sound design is made by mainly non-diegetic sounds and it carries a lot of meaningful information being made of evident sound effects.

To summarise, for the author, the more diegetic a sound is, the lower information rate it carries, whereas the less diegetic it is, the higher the information rate it brings to the audience. However, this distinction is not always true, since diegetic sound can also provide some meaningful information that would not be perceived by the audience if it wasn’t for the sound. This is the case of emotional sound effects, such as knocking and walking, which can provide information about the emotional state of the person knocking or walking even if the

audience does not see his or her face. Many researches have been made on this topic (Iop & Pauletto, 2021; Bedenko et al., 2020; Houel, Arun et al., 2020; Turchet & Rodà, 2017; Vitale & Bresin, 2008) and the present work aims at investigating it even further.

Nevertheless, the distinction between diegetic and non-diegetic sounds is not that strictly defined: a diegetic sound can become a non-diegetic one and vice versa. This is the case of *transdiegetic* sounds, which can be used, for instance, to transition from a scene to the other but that are still part of the same sequence, thus they are somehow related. An example of a diegetic sound becoming non-diegetic is when the audience hears the song playing on the earphones of a character slowly turning into a soundtrack, as shown in a scene from *The Amazing Spiderman 2 (2014)* directed by Marc Webb, where at some point the main character removes his earphones while the audience keeps hearing the song *Gone, Gone, Gone* by Philip Philips that was initially playing only inside the story. Another great example, and possibly the best-known and famous one, is from *C'era una volta il West* (1968) by Sergio Leone. In the opening sequence, there are three hired assassins waiting for someone to get off a train. Once the person in question (the character called "Harmonica") arrives, he starts playing the harmonica and Ennio Morricone used this *diegetic* sound to introduce his *non-diegetic* music composition (Valle, 2011).

When considering everyday sounds such as knocking, we are talking about *diegetic* sounds and they are hardly ever used as non-diegetic, since they come from the world surrounding us or are directly made by us.

Furthermore, depending on the kind of messages conveyed by the sound and by the image and by their being diegetic on non diegetic, Görne proposes a scheme that puts in relation the information coming from the two channels. In the case in which the sound and the image carry the exact same meaning, we have redundant information and therefore this category is called *paraphrase*. Even if the sound communication does not give additional information, it

can be used as a tool for the emotional climax of the narration. Then, if the image conveys a message and the sound conveys another one but that it is still related to the visual one, we have *parallelization* or *complement*. This is used when further connected information wants to be given in a way that avoids redundancy and condensates the story. Moreover, there are situations in which the image does not carry any emotional information and everything relies on sound and it is the case of *polarization*. There is also the possibility of *incongruence* in which what is expressed through the image is negated by the sound, as for instance in the case of the above-mentioned use of anempathetic music or when, in horror movies, positive images are matched with a dark and scary sound. Finally, the last combination is *irritation* and it happens when the image transmits a message and the audio transmits one that is completely independent. This latter union is usually used when a sound anticipates the following scene by overlapping.

Apart from the four areas mentioned above, the division between diegetic and non-diegetic, and the Görne scheme, sounds in movies can also be classified as Voice, Music and Sound Effects. However, someone argues that these last two types of divisions are limiting the true expressive potential of a sound (Knight-Hill, 2019) and that more aspects should be taken into account. For the author, this classification assigns "limited functional roles to specific types of sounds" (*ibidem*, p. 1), whereas every kind of sound, whether it is music, voice or sound effects, can embody the expressive potential of sound.

1.3. SOUND DESIGN

The principles of sounds, as well as its characteristics and informational power in storytelling, are at the basis of sound design. With this term it is meant the craft of creating

"new sounds with the intention that they will be heard in a given context of use" (Susini, Houix & Misdariis, 2014, p. 108) and particular attention is given to the newness, therefore originality, of the sound. A sound can be considered new when it cannot be found in any existing database, library, and it cannot be recorded or it cannot be used without being edited and modified first.

There are a lot of areas in which sound design can be applied such as artefacts production, for instance an alarm, car technologies, navigation systems and advertising, but there is one art that is capable of incorporating all the possible uses of sound design: cinema.

However, before examining in depth the different uses of sound design in cinema, it is necessary to understand when and how sound design was born.

Indeed, the concept of sound design was born with theatre and radio drama in the 1920s. In his book *Radio Drama. Theory and Practice* (2001), Tim Crook states that radio plays are a form of art that have not received enough recognition in the past century, despite their fundamental contribution to the birth and development of sound design. As a matter of fact, it was under this form that storytelling through sound, or "sound art" as Crook defines it, was brought to the audience first, instead of with film.

Radio dramas are forms of narration that use sound to convey information. The visual channel is obviously the missing element in radio, hence all the information, from space surroundings, objects, characters movements and emotions must be conveyed through sound and sound only. William Ash (1985, p. 1, as cited in Lopez & Pauletto, 2009), identifies radio dramas as "a story told in dramatic form by means of sound alone" and Lopez and Pauletto extract from this definition the reliance of storytelling on audio only and the absence of the visual part is not considered to be a necessary element for the listeners in order to understand the story. The two scholars applied this principle in their previously discussed work *The design of an audio film for the visually impaired* (2009).

Nevertheless, the history of radio drama is poorly documented, however Crook outlined what can be considered the first radio plays in America, United Kingdom and Australia. The common trait that these countries share is that, even if geographically distant, their first documented radio plays were made in the 1920s. Some references established that 1922 was the year in which the first sound effects were used in the play *The Wolf*, broadcasted by the station WGY in Schenectady, New York. Whereas concerning Australia, the first radio play broadcasted was *The Myth of Sweeney Todd, The Barbarous Barber,* by the Melbourne station 3LO in 1925. On the other side, in the United Kingdom, BBC Radio shall be considered as one of the most important radio broadcasters when it comes to the development of radio dramas.

Crook reports that the first play written for radio is generally considered to be *A Comedy of Danger Radio drama: theory and practice* by Richard Hughes, in 1924, whereas Lord Asa Briggs, in his history series of the BBC, stated that the transmission of extracts of Shakespearian plays such as *Julius Caesar*, *Henry VIII*, *Much Ado About Nothing* and *Twelfth Night*, transmitted two years before, in 1923, marks the beginning of radio dramas in the United Kingdom.

Still in the 1920s, with the creation by the italian futurist Luigi Russolo of machines that simulate nature sounds, for example the sound of the wind, of rain and of thunder, as well as train sounds and bomb explosions, avant-gard artists such as Jean Epstein and Eugène Deslaw could insert sounds in their productions for the renowned Parisian cinema Studio 28 (Susini, Houix & Misdariis, 2014).

However, it was not before the 1970s that the term *sound designer* made its first appearance and was recognised in the North-American film industry (*ibidem*). It was Walter Murch, with his work on *Apocalypse Now* (1979) by Francis Ford Coppola, who coined this term. Using his words:

> [...] if an interior designer can go into an architectural space and decorate it interestingly, that's sort of what I'm doing in the theatre. I'm taking the three-dimensional space of the theatre and decorating it with sound. [...] *Apocalypse Now* [...] that was where *sound designer*, the word, came from. (Jarrett, 2000 as cited in Susini, Houix & Misdariis, 2014).

In particular, nowadays the term "sound designer" is used to refer to the sound director of a movie, who is the person that supervises and manages that the sound outputs of a movie production are coherent. Different categories of *sound workers* exist, such as the editor, the mixer and the Foley artist, but they will be further examined in § 1.3.1.

In reality, more generally, a sound designer is the person who supervises the sound for any product, whether it is a movie, an artefact, a song, an advertisement, a video game or an immersive environment. In order to do this job, an extended amount of training, experience and knowledge is required (Selfridge & Pauletto, 2021).

When referring to the cinema industry, sound is usually the element in movies that the audience tends to not consciously notice, unless there are some technical problems that make it noticeable. Nevertheless, in media production sound is arguably the best and more powerful storytelling device: it can be used to merge different scenes, to transition from one scene to another, to set the overall tone of a product and consequently elicit emotions in the audience, as a means of communication and also as a way to introduce and characterise specific characters (Pauletto & Iop, 2021; Bedenko et al., 2020; Houel, Arun et al., 2020, Knight-Hill, 2019, Valle, 2011). As an example, knocking actions can be used as a transition element since they are a way in which the emotions of the person who knocks can be represented, and besides, they can create expectations in the audience about the possible

emotional and physical reaction of the person hearing the knock from inside the room (Barahona-Ríos & Pauletto, 2020).

In the preface of the book *Audio-Vision. Sound on Screen* (Chion, 1994), Walter Murch talks about the "cinematic inversion of the natural order" comparing humans to cinema when underlining how these two things are created inversely. With the sentence "We gestate in Sound, and are born into Sight, Cinema gestated in Sight, and was born into Sound" Murch expresses how the human hearing develops before we are even born, being this the dominant sense, for then to be replaced by sight at birth. On the other hand, cinema was born with sight only, suffice it to know that movies were silent for the first 30 years, and this has been the dominant sense since 1927, the year in which the first movie with synchronised sound, *The Jazz Singer* by Alan Crosland, was made. The progress sound has made in the last 95 years is noteworthy, for two reasons above all: one being from a technical point of view and the other one from a more human point of view. The first one is the development of new technologies such as magnetic recorders and the other one is the creativity of the people working in the industry. However, this is an improvement that was not meant to increase the attention of the audience towards the sound in the movie, but to enhance the visual elements: "the better the sound, the better the image" (*ibidem*, p. VIII). This enrichment of the visual qualities given by the expressiveness and informativeness of sound is what Chion calls *added value*. With this term the author refers to the impression, obviously wrong, that sound is unnecessary, that it simply duplicates what is shown and this illusion is especially present in the case of *synchresis*. This word was coined by Chion to make reference to the creation of a relationship between picture and sound that can be generated independently of any rational logic or existing link between the two. This concept will be better explained in § 1.3.1 referring to the Foley effect.

To conclude the general discourse on sound design and the important role it holds in storytelling, two movies from different decades must be considered. The first one is *C'era una volta il West* (1968) by Sergio Leone. In this movie, sound plays a very important role and from the first scene it reveals its predominant presence also as a way for telling apart different characters. The beginning of the movie is indeed for the most part silent, meaning that only few words are spoken and ambient sounds are the real protagonists. In the first scene, the train station one, there are three hired assassins and, as Fawell highlighted (Valle, 2011), to each one of them corresponds a specific sound: water/drops, cracking (of knuckles) and the buzzing of a fly. These three sounds help create the identity of the three characters, who are therefore described through sound, thanks to an intensified repetition of certain gestures and movements rather than being presented to the audience through the explicit words of a narrator or of the characters themselves.

The other movie example is once again *Dunkirk* (2017), by Cristopher Nolan and, even in this case, sound is the real protagonist of the movie, to the point that it is the element that creates tension, unifies different time spans and becomes a character itself. The movie shows three different time frames: one week before the ending of the Dunkirk evacuation, one day before, and one hour before. Although these moments are set in different moments in time, they are all linked together by means of the score and other sound elements present in the movie to the point that they even might be mistaken as actions happening at the same time. In general, the soundtrack of this movie can be divided into two parts: repetitive sounds, such as clocks, heartbeats and other "ticking" elements and the sound related to the Shepard tone (Rapan, 2018), which is an acoustic illusion that will be better described in § 2.1. The ensemble of all of these elements follows the progression of visual shots and they contribute immensely to the creation of tension, as mentioned in § 1.2, which is a fundamental aspect of the movie that would have not been the same without or with different sound details.

1.3.1 THE FOLEY EFFECT

When talking about sound design, it is impossible to not mention the Foley effect. Not many people know or realise that when they are watching a movie, the majority of the sound that they hear has been made in post production and it is hardly ever the sound recorded on set during the shooting of the movie. The reason for this is because the resulting sound may not be of the highest quality due to interfering noises or because not perceptible. This is when the work of Foley artists comes into action. To give a brief definition of Foley effect, one can say that it is the creation and recording of sound effects in synchronisation with the scene during the post production editing of the movie.

Nevertheless, Foley is actually more than this: it is a true form of art. Suffice it to know that the people performing Foley are called *artists*. Foley artists create and perform *ad hoc* sounds for movies, from steps at different paces and materials, to door slams, glasses tinks, sea waves, clothes movements and many more, while watching the scene being rolled on a screen in front of them and, in order to do this job properly, few abilities are requested.

First of all, these artists need to have a sense of rhythm, since they have to perfectly synchronise their actions with the ones they see on the monitor, so people with a background in dancing or acting will seem more effortless (Ament, 2009) since sometimes what Foley artists have to replicate is the choreography to a dance and, in general, they are considered to be “sound actors”. Indeed, as on-screen-actors, these apparently invisible figures have the responsibility to render perfectly not only the actions of the characters, but their personality and essence as well in order to make their job unnoticeable. As Karen Baker Landers, the supervising sound editor for *The Bourne Ultimatum* (2007) said about the work of Foley artists:

“when it's good you don't even notice. When it's bad, though, it's distracting and maybe you don't feel the presence of the character, or his size, or his speed, or his dexterity." (Jackson, 2007).

Second of all, the other key element when performing Foley is creativity. As a matter of fact, these artists can use any object to create any kind of sound, regardless of the relation between the object they are using and the real object depicted on screen. This process is what Chion (1994) called *synchresis*, a word he coined joining *synchronism* and *synthesis,* to refer to “the spontaneous and irresistible weld produced between a particular auditory phenomenon and visual phenomenon when they occur at the same time”. This is a perceptual phenomenon that we hold to be true because we are convinced that what we hear is what we see.

However, synchresis does not always mean perfect match: when we watch a dubbed movie, so not in its original language, we still do not perceive the difference between the lip movements and the words that the dubbing actor pronounces, unless we pay particular attention. In some cases, the discrepancy between audio and visual stimuli can give life to perceptual phenomena such as the McGurk effect, which will be discussed in the next section (§ 1.3.2).

Synchresis allows Foley artists to make the most subtle yet astonishing audiovisual configurations by unleashing their creativity. Some of the most famous and used examples of Foley and synchresis are:

- the sound of the *lightsaber* in the Star Wars movies. With the lightsaber being an object that does not exist in real life, it was necessary to create a new and unique sound. Ben Burtt, the sound designer for the series, explained[2] how the change in pitch of a humming sound coming from an old projector motor together with the buzz

[2] For video reference of the interview: https://www.youtube.com/watch?v=tZj1mYLC7h0 (last visited May on 1st 2022)

created by carrying a microphone near an old television set was the perfect sound for the new non-existing object and this discovery was actually made by accident.

- The sound of snow. When characters are walking on snow or the hiker in a documentary is doing this action, it is most likely that the sound of his feet crunching snow are made in post production where the Foley artist is simply walking on a sand pit[3] or on a leather pouch filled with cornstarch (Doyle, 2013).
- The flapping of birds' wings. Even for this type of sound, the rendering depends on the creativity of the artists. For instance, some flat a pair of gloves (*ibidem*) while others use a duster made of feathers[4].
- The sound of walking. While walking can be simply reproduced by walking on the same ground material that the actor is walking on, especially in animated movies there can be other techniques such as bending a leather wallet or hitting your hand with a baseball glove[5].
- The sound of breaking a bone. When a character breaks a bone, one of the most popular Foley props used to better represent that sound is snapping a celery stick in half (Wright, 2014).

This principle is used because what is most important is not the perfect matching of the new sound with the original one by using the same tool, but rather the evocation of a feeling connected to the real sound (Donaldson, 2014). Indeed, Foley artists, apart from recreating the various prop sounds and actors' movements, must also take into account the feel of the object or of the character that is portrayed in the scene. For instance, the genre of the movie

[3] Great Big Story x The Weather Channel, *The Magic of Making Sound*, 2017, https://www.youtube.com/watch?v=UO3N_PRIgX0 (last visited on May 1st 2022)

[4] See note 3; Wired, *How the Sounds From Your Favorite Movies Are Made*, 2016, https://youtu.be/0GPGfDCZ1EE?t=81 (last visited on May 1st 2022) .

[5] Wired, *How the Sounds From Your Favorite Movies Are Made*, 2016, https://youtu.be/0GPGfDCZ1EE?t=81 (last visited on May 1st 2022)

influences how Foley is rendered: exaggeration in comedy in order to make people laugh or subtle in horror movies to create tension in the audience.

The importance of sound in a movie is especially evident in the horror movies *A Quiet Place* (2018) and the sequel *A Quiet Place II* (2020) both directed by John Krasinski. These movies are set in a post-apocalyptic world inhabited by blind aliens with an acute sense of hearing that will kill whoever makes some noise, thus the characters are forced to avoid making any sound, to the point that they communicate using sign language. However, the sound design process for this movie was a central part in the making and it is what contributed to make this short series a box-office success. In an interview for *Insider*[6], the Foley artists of the movies described how the way they designed and performed the sounds was fully dictated by the logic of the scene: if they could hear the sounds from a distance, that would have meant that the monsters could hear it as well. Hence, if the shot showed the movement from a close up, the audience could hear it, otherwise it was not audible.

This aspect highlights another crucial peculiarity of Foley: attention to detail. Paying attention to even the smallest details of a scene is fundamental in making the audience believe in what they are seeing and therefore hearing (Doyle, 2013). Hence, representing details does not mean that every sound present in the scene has to be reproduced by Foley artists, but it implies understanding what kind of sounds are the ones that need to be enriched and those that can be eliminated in order to make the others stand out more.

Some of the Foley sounds created for these movies with the principle of synchresis are crab legs for the monster's walk, celery and lettuce for the opening of their ears, and an edited sound of teaser on grapes for the movements of the monsters[7].

[6]Insider, How the sound effects in "A Quiet Place" were made, 2019, https://www.youtube.com/watch?v=WnozP8OWeik (last visited May on 1st 2022)

[7] *ibidem*.

Even if the absence of sound is fundamental for the characters, its subtle presence performed by the Foley artists and then edited by the sound editors and mixers is crucial to create the right tension in the audience, since a completely silent movie would have not been believable. Therefore, even if subtle, Foley sounds contribute to create the environment, atmosphere and mood of movies (Weiss, 2012 as cited in Doyle, 2013).

As all of the examples mentioned above can prove, Foley is a procedure that completely changed the history of cinema, while at the same time its biggest aim is to not be noticeable, and *A Quiet Place* is a great example. As Foley artist David Yewdall said:

> If the audience is aware of what we do, we have pulled its attention out of the film reality we are creating, and therefore we have failed" (1999 as cited in Doyle, 2013, p. 10).

The words of the artist are in line with the ones expressed above by Chion (1994) as he referred to audio as an enhancement of the picture. Sound does not have to distract the audience and prevail on the picture, but it has to reinforce it: "the better the sound, the better the picture".

However, even if Foley can create an infinite number of sounds using multiple props, its origins lay in walking sounds. Footsteps are in fact the classic Foley noise, such that at the beginning of Foley, the people practising it were called Foley *walkers*. The transition from *walker* to *artist* is due to the recognition of the artistic part of the job, which requires a broad creativity and made them transition from "mechanical labourers to creative decision-makers" (Wright, 2014, p. 205). Even though its role in movie-making is fundamental, this job category was not recognised as one, but its workers were classified under the tag of sound editors, until the summer of 2006, when *Foley artist* was included as an official category of

the Motion Picture Editors Guild (Ament, 2009). The simplest explanation of the difference between a Foley artist and a sound editor is that the first one is more concerned with what the character is doing on the scene, while the second one usually edits effects that deal with the action or with the environment. On the other hand, sound mixers have the task to see if the sound created by the artists fits the scene well both in synchronisation and in realism.

To come to the origins of this technique, we must go back to North-America in the mid-1920s, when the first movie with synchronised sound, *The Jazz Singer* (1927) by Warner Bros Studios, came out. This movie represented a milestone in the history of cinema and all the other film production companies had to keep up with it, or their silent movies would have been perceived as obsolete (Doyle, 2013; Ament, 2009). This is when Universal Studios relied on Jack Foley and his team to come up with a way to put sound in their movie *Show Boat* (1929) and they started to synchronise live sounds to the projected picture: that represented the birth of Foley sounds. After World War II, thanks to the construction of soundstages with specific acoustic qualities to maximise the sound, the use of different directional microphones and the invention of magnetic recording techniques made possible by the advancement of technology, the way in which sound and images were produced truly changed (Doyle, 2013). Nowadays, the Foley Stage is set up of a monitor or projector in which the artists watch the scene while they play the sounds, one or more pits, for instance a sandpit, a woodpit, and a water tub, microphones and a prop room with all kinds of objects that might be necessary to produce sounds (Trento & de Götzen, 2011).

The work of Foley artists is undoubtedly essential, but it could not be possible without other two figures that contribute in the making of the final sound output of the movie: the Foley editor and the Foley mixer.

With the advancement of technologies, this job has been suffering from the creation of digital libraries as well as the creation of synthetic sounds, thanks to which thousands of sound

effects are already made and available to sound designers. However, ready-made sounds are not the same as the crafted ones, especially in the rendering of emotions. As seen before, a crucial aspect of Foley is following the actions of the actor and putting all of their expressiveness in the sound to better create the mood and atmosphere of that specific scene. With digital sound libraries, what misses is the human touch of the artist and the exact correspondence with the feeling and movement of the scene. These factors contribute to creating unique sounds that make the final product stand out more from the rest. Consequently Foley is not only a functional activity but it is above all an artistic one (Wright, 2014) and especially important in the transmission of emotions to the audience. This concept is valid for all the sounds present in the picture: from original ones to everyday sounds: the way a character walks or knocks on a door, for example, says a lot about his or her mood and behaviour. This latter example will be the object of the experimental study described in Chapter 3.

In conclusion, even if all the examples cited above refer to cinema, Foley sounds can actually be applied to other kinds of products that require sound such as video games, advertisements and many more.

1.3.2 CONFLICTING STIMULI: THE MCGURK EFFECT

As previously discussed, audiovisual integration is a phenomenon that, in some cases, is able to trick human perception. The interaction between sound and sight can lead to the reinforcement of the information or an interference between the two senses. Concerning the Foley Effect, hearing reinforces sight, it enhances it, but while doing so it leads our mind to think that what we see corresponds to what we hear, but as seen above with the principle of synchresis, that is proved to not always be true.

However, there is also another way in which combining audio and visual stimuli can lead to perceptual phenomena and this is the case of the McGurk Effect: a phenomenon in which sound interferes with sight and vice versa. This effect was discovered in 1976 by Harry McGurk and John MacDonald. The two researchers investigated the processes that underlie speech perception and they emphasised the importance and the influence of vision: an aspect that had not been considered before. In brief, the McGurk effect is an illusion that happens when we are presented with incongruent audiovisual stimuli, which, in its most straightforward form, implies the visual signal of a phoneme, e.g. the syllable "ga", dubbed onto the acoustic signal of another phoneme, e.g. the syllable "ba" (Alsius, Paré & Munhall, 2018). Another example is the auditory "baba" combined with the visual "gaga" was perceived by most participants as "dada" (de Gelder & Vroomen, 2000). Thus, the simplest definition of this effect is that when incongruent audiovisual stimuli are presented to humans, these last ones tend to merge the information from the two modalities into a new element that is not present in either one of them. However, this "conservative" definition does not include a broad number of possible other responses to incongruent audiovisual speech stimuli (Alsius, Paré & Munhall, 2018). The experiment, thanks to which McGurk and MacDonald supported their discovery, consisted in showing a film of a woman's head lipsyncing the syllable "ga" while being dubbed with the syllable "ba" to a group of participants. As a result, the individuals reported hearing a third syllable "da" and, in the case of the reverse process, hearing "bagba" or "gaba" (McGurk & MacDonald, 1976). Before their discovery, speech perception was believed to be unimodal, a purely acoustic process, until the two researchers highlighted the fundamental role of the visual modality as well. Indeed, as sound is an instrument that conveys additional information about the environment, the situation, as well as the emotional state of a person, visual elements, for instance the face, can add information about the sound we are hearing. Hence, if in movies sound adds information to the visual

part, that is considered to be the most important one, in speech perception the opposite mechanism happens: in a process considered to be mainly, if not only, auditory, the visual channel carries additional information as well as power.

However, the McGurk effect can be experienced also outside of laboratory investigations, and it is more common than one actually may think. When watching a dubbed movie, the discrepancy between what we hear and how we see the lips of the actors moving can be more or less evident, giving the audience the impression that the information coming from the two channels comes from different sources (Munhall, Gribble, Sacco & Ward, 1996). The four authors demonstrated also that strict temporal synchrony between auditory and visual speech stimuli is not required in order for the McGurk effect to show, but when the two stimuli are presented with a delay of more than 180 ms, the visual stimuli influenced the response the most. However, in the case of watching a dubbed movie, the spectators are not under test and therefore their eyes are not uniquely concentrating on the lips of the actor, but they are paying attention to the whole scene and the stimuli that come with it, thus the McGurk effect may become less prominent.

This effect has been mostly studied using vowels and consonants (Nicholls et al., 2004; Massaro, 1998; Munhall et al., 1996; McGurk & MacDonald, 1978, 1976), with a stronger illusion detected in the cases in which weak auditory consonants, such as “b”, were present, since they tend to be more confusable than others (Alsius, Paré & Munhall, 2018). Nevertheless, some scholars have tried to test the McGurk effect with simple words or more complex sentences as well as in different languages (Li et al., 2013; Fagel, 2006; Abelin, 2004, 2007; Hietanen et al., 2004; de Gelder & Vroomen, 2000; Massaro & Egan, 1996). However, as these studies highlighted, language is not the only element that is perceived by the two channels under consideration, but the perception of emotions appears to happen in the similar bimodal way (de Gelder & Vroomen, 2000). What these latter studies added to the

original McGurk and MacDonald experiment is, indeed, the use of emotionally expressive phrases in order to verify if the said illusion can be further affected by the expression of emotions through the face and the voice. By presenting conflicting audiovisual stimuli with the two channels expressing different emotions, for example a happy face and an angry voice, or a surprised voice and a disgusted face, participants tended to merge the two emotions together and give a third emotion as a result, exactly like in the syllable experiments.

What these studies have demonstrated is that emotions are conveyed both through the visual channel, for instance the face, and through the auditory one, for example the voice. This bimodal process is particularly evident when taking in consideration visually and auditory impaired people. In fact, in cases in which the auditory channel is damaged or completely not functioning, the face and the overall body movements, therefore the visual channel, are the only means that manage to convey the emotional information. On the other hand, in cases in which the visual system is fully or partly compromised, the quality of the voice together with its sound characteristics, as illustrated in the sections above such as changes in pitch, are the only indicators thanks to which blind people can perceive the emotion of their speaker.

In the study conducted by de Gelder and Vroomen (2000), the visually and auditory-abled participants of the experiments were presented with static photographs of a happy and angry face together with an emotionally neutral sentence (the Dutch translation for “his girlfriend came by plane”) and they were asked to ignore the facial expressions in the first case and the auditory signal in the second one. Results showed that, even if asked to exclude one of the two modalities, the participants were still influenced by it. A similar experiment was carried out two years earlier by Massaro (1998), in which it was established that the way in which instructions were given to the participants, thus bimodal or auditory instructions such as “use the information from *both* the face and the voice” in the first case and “make the judgement on the basis of what you *heard the voice* expressing” in the second one, influenced their

emotion perception. These kinds of experiments highlight the bidirectionality of emotion perception.

In general, the entirety of the visual modality, thus not only the face but also gestures and movements, is able to express emotional information. The relevance of the movements of the face and the overall body is also evident in infants or when someone tries to learn a new language: when learning to speak, humans do not only try to reproduce the same sounds, but the same lip, tongue, throat and other movements of the speaker as well (Paget as cited in Vernon, 1934).

The basic idea behind the principle of the emotional McGurk effect, thus emotional contrasting audiovisual integration, is at the basis of the experimental research of the present study, which aims to verify if certain colours, that have been proved to be associated with specific emotional states, have the ability to increase, decrease or even change the perception of an emotion conveyed through an emotional sound. Nonetheless, in this case, the emotional sound element will not be speech but rather five emotionally different knocking actions and the visual element will consist in a video of a hand knocking on a door of different colours. This experiment will be extensively discussed in Chapter 3.

2. PERCEPTION AND EMOTIONS

Our bodies are the means through which we perceive the world surrounding us. From hearing a walking sound, for example, we can understand if the person is walking on a soft, solid or aggregate material, as well as recognise the specific one. As seen in Chapter 1, the same process happens for other everyday actions such as knocking on a door. Apart from the characteristics of objects, humans can also identify the gender of the person doing the action, as well as his or her emotional state (Tajadura-Jiménez et al., 2015).

Depending on the characteristics of a certain sound, as mentioned above, these can activate different physical, as well as cognitive, responses in humans: from the perception of danger to that of rest. As Sonnenschein stated, hearing a sound may have many physical effects on our body temperature, blood pressure and circulation, pulse rate, heating and sweating (as cited in Donaldson, 2014) and these changes can consequently elicit a set of different emotions. As it will be deeply examined in § 2.2, the first person to theorise the different physiological changes happening to our bodies after being exposed to a stimulus was Charles Darwin in his book *The Expression of the Emotions in Man and Animals* (1872). His theories have been of inspiration for other researchers, who tried to identify a set of universal basic

emotions. However, in emotion literature there is little agreement on the topic of basic emotions, for different reasons that will be further examined in § 2.2.

Moreover, in everyday life, humans do not perceive sound alone, but they integrate it with the other senses in a multisensory process that may lead to phenomena such as synesthesia and the already discussed synchresis and McGurk Effect. In particular, in recent years, specific attention has been brought to the influence of colour in the elicitation of emotions and the multiple research findings have been applied to a variety of different fields such as marketing, interior and web design, colour therapy, artificial intelligence and media production like movies, e.g. Disney Pixar *Inside Out* (2015), television and video games. Nevertheless, as it will be further discussed in the following sections (§ 2.2 and § 2.3), the association between colours and emotions is an old concept that became especially relevant starting from the second half of the twentieth century thanks to the work of the American psychologist Robert Plutchik. This topic is, however, still investigated, since there is not a large degree of agreement on the definition and identification of emotions, as well as in the consequent relation they have with colours.

Considering that sound is another element that has the power to elicit emotions and feeling states in the listener, the experiment presented in Chapter 3 aims at examining even further these associations, combining multiple emotion eliciting stimuli to see if there is a channel, aural or visual, that prevails on the other, if there is integration between the two, or if they create a fusion that generates a third emotion. Results may constitute the basis for future works on media products concerned with emotion elicitation.

2.1 AUDITORY PERCEPTION

As P. E. Vernon underlined in his study (1934), auditory perception had not been studied as in depth as visual perception, which can find exhaustive explanation in the Gestalt theory. Hearing, being a very complex process, had been confined to the perception of single tones and elementary vocal sounds.

To give a brief explanation of how the auditory system works, one must keep in mind that there are three aspects of sound: the source, the medium and the receiver. Source characteristics have already been described in the previous Chapter, but it is important to mention that the smallest entity of auditory perception is called “auditory object”, which consists in a simplified interpretation of the complicated data that is collected by the ear. During this process, the auditory system translates the sensation of sound into an hypothesis of an object in space as the source of the sound (Görne, 2017).

Concerning the medium, for example the air in which sound propagates, its characteristics change the way that sound reaches the receiver, thus the waveform generated by an object, a voice or a musical instrument, is not the same one that reaches the tympanum, due to the multiple reflections, interferences and also the position of the listener.

The listener’s, or more technically the receiver’s, perception of sound stimuli has been at the centre of Vernon’s study. The auditory system is composed of the ear, which consists of the external, the middle and the inner ear, and the brain, and its aim is transforming the acoustic pressure into nervous impulses that are then transformed into a perception called sound. The human hearing domain ranges from 20 to 20.000 Hz and outside this threshold it is not possible to hear. Sound intensities between 0 and 140 dB can be heard with no problems, whereas sounds above this domain can cause permanent damage to the auditory system.

Apart from these technical facts on how the auditory system works, it is interesting to examine how it has developed since the beginning of life on Earth.

Following Vernon's analysis, the organs of hearing, of position sense and of angular rotation sense are all placed in the labyrinth and the development of sense organs was primarily driven by reactions to environmental changes, thus they do not provide the organism with knowledge of the external world, as one may think. Hence, any organism's auditory system was born with the intention of adapting to environmental changes, namely, to maintain equilibrium, orientation, posture, to detect warnings or for reproduction purposes. Therefore tones or noises initially had no significance for them.

The importance of orientation is actually the best example to reinforce these findings. Concentrating on humans, since the development of the auditory system was due to evolution purposes, the two ears are placed as far as possible in order to give men the sense of localisation. If the system had developed for musical purposes, one ear placed in the centre of the face, like the nose, would have been enough according to Vernon (1934).

Speech and music, which are the latest evolved auditory functions, are present only at the human level, being them cortical functions that require rapid and complex learning, and not mechanisms of the inner ear. Indeed, injuries to some parts of the cortex may have an effect on music and speech perception, but they do not alter the more primitive and general responses to the "brightness" of a noise. This brightness is what makes many think that other animals, such as dogs, can react and therefore understand musical stimuli. What really happens is that when non-humans hear a musical piece, depending on its properties, such as pitch, amplitude, rapidity of change and number of high partials, this one resembles the vibration patterns of certain noises that require actions such as flight, activity, rest, approach and so on. These kinds of reactions caused by sound properties are the same desired actions

that sound designers want to generate in the listener both when creating an artefact, for instance an alarm, but also when generating the atmosphere for a movie scene (Chapter 1).

In the case of the object of the present study, knocking on a door with different emotional states will produce different sounds that will help the person on the other side understand whether he or she has, for instance, to worry or not about the person knocking or the message he or she brings.

The phenomenon of recognising specific acoustic patterns that lead to action or rest can be represented in the same way in which the figure-ground Gestalt theory for vision works. Those sounds that stand out from the rest of the acoustic environment are the "figures", while the others that lay on the background and do not elicit particular attention to the listener are the "ground". The discrimination between these two types of sounds is given by the higher pitch, greater intensity, different timbre and different dynamics of the figure from those of the ground (*ibidem*). This differentiation also depends on the intentions of the listener, a concept that will be further described in § 2.1.1.

With this discourse on how the auditory system works and has evolved, it is now possible to examine how auditory perception integrates with the other senses in everyday life.

Firstly, it is interesting to notice how people use words from other senses whenever they have to describe a sound. Words such as "warm", "cold", "soft", "smooth", "rough" are borrowed from touch, "bright", "dark", "high", "low", "deep", "edgy", "round", "flat", "colourful" from sight, and "sweet" from taste, thus our auditory perception operates through crossmodal metaphors coming from visual, haptic and tasty property of auditory objects (Görne, 2017). These metaphors are what contributes to give sound a meaning and they are a way in which a sound designer can communicate through sound.

When it comes to multisensory integration, there is a perceptual phenomenon called *synesthesia*, which is the Greek word for *union* ("syn") and *sensation* ("aisthesis") and it

symbolises the process for which the stimulation of one sense produces an involuntary perception in another sense (Cytowic, 1989). This innate and involuntary condition is quite rare, it is predominant in women (*ibidem*) and it can be elicited by any kind of stimulus and any kind of sense: letters and words, days of the week, sounds and voice, numbers as well as taste, smell and pain (Whitelaw, 2008; Hupka et al., 1997).

The most common sense integration is sight with sound, also called "coloured hearing". Thus, for some people, sounds can represent a visual experience, in addition to the most obvious auditory one (Fernay, Reby & Ward, 2012). Therefore, in this case, sound automatically elicits visual images as well as auditory ones (Goller, Otten & Ward, 2008). Some examples of "coloured hearing" may be the visualisation of a dark-velvet colour in a tree trunk-like texture when hearing a cello playing or pastel colours when hearing the sound of a flute (*ibidem*), or, as some interviews to synesthetes reported by Cytowic underline: "what first strikes me is the color of someone's voice [...] like a flame with protruding fibers", "spearmint tastes like cool, glass columns. Lemon is a pointed shape [...]", "the name Paul is such an ugly color [...]" (Cytowic, 1989, p. 849).

It has been shown that some people react only to speech stimuli, while others are potentially triggered by any kind of sound (Fernay, Reby & Ward, 2012; Goller, Otten & Ward, 2008). Different sounds can then evoke a multitude of visual percepts such as colours, shapes, movements and size located in different positions in space. However, colour and geometry are the most prominent associations coming from multisensory integration.

The first medical case of synesthesia was registered in 1710, when Thomas Woolhouse, an English eye doctor, reported the case of a blind man who was able to perceive sound-induced coloured vision. Still in the 18th Century, Sir Isaac Newton and Erasmus Darwin tried to find a correlation between the energy of sound and colour and they managed to create instruments

that play sound and light simultaneously. High-pitched sounds tend to be perceived as lighter, higher and smaller images than low-pitched sounds (Goller, Otten, Ward, 2008).

One of the effects of synesthesia is the elicitation of strong emotions, thus, once again, the combination of sound and vision underlines its important emotional power.

To conclude, perception of acoustic stimuli can be used in the movie industry to create the right tension and the overall tone of a scene or movie. In particular, musical scores are usually believed to be distant from the world of the movie (Knight-Hill, 2019) and are seen as commentary of the visual actions described on the scene. Music is therefore considered to be separated from the action, an abstract element that accompanies the visuals, when in reality the score itself can be the one item that gives meaning to the visual scenes, as seen in the above-mentioned examples of C*'era una volta il West* and *Dunkirk*. It is this last movie that proves to be one of the best examples of how the soundtrack can become an active element for the creation of the overall tone and mood of the movie and consequently for the creation of emotions in the audience thanks to a perceptual phenomenon called *Shepard tone*. The Shepard tone is a special type of sonic articulation that contributes to the production of meaning and it has been used in recent cinematic works. Its discovery belongs to Roger Shepard, who, in 1964, published an article on relative pitch in which he introduced a new visual representation of pitch, designed as a spiral in which all the octaves of the same pitch were connected on a vertical axis. Due to the circular shape, by moving up or down in a *glissando* along the spiral in different parallel octaves, a sort of "auditory Penrose Stairs" is produced, creating an auditory illusion of never ending and giving the listener a sensation of always rising/ascending or always falling/descending (Rapan, 2018). Even though Shepard was the one who discovered this perceptual phenomenon, its fame was given by Jean-Claude Risset, who was one of the first artists to apply this concept in his composition *Computer Suite From Little Boy* in 1968. In *Dunkirk*, this principle was used to connect both the three

different timelines and the sound effects, such as the sound of aeroplanes, in a way in which the spectator could feel a continuous and increasing intensity, mirroring the feelings of the soldiers in the movie (Nolan, as cited in Rapan, 2018).

In an interview for *Gold Derby*[8], Hans Zimmer explained how in *Dunkirk*, he and the director Nolan tried to completely merge the images with the sound and with the music in a way in which it seems like the audience is listening to the movie with their eyes.

Hence, it is once again brought to the attention the incredible power the combination of sound and image has in generating reactions to the audience and manipulating their feelings.

2.1.1 THE WAYS OF LISTENING

Due to the natural way in which our ears are made and work, it is not possible for us to naturally prevent sound perception. If the ears had some closing mechanism like the eyes do, we could decide when to let sound come in and when to stop it, but, given the absence of some "earlids", the omnidirectionality of hearing and the physical nature of sound (Chion, 1994), we are forced to listen every day in every moment. Nevertheless, humans still have some power in deciding how to listen, since we do not always pay the same attention when listening to sounds.

As previously mentioned, sound intention is a very important aspect that helps individuals to understand its meaning and it is especially useful in emergency situations. However, the source is not the only character in sound perception that expresses intentions, but the receiver does it as well.

[8] Gold Derby, Composer Hans Zimmer: 'Dunkirk' was the hardest thing we ever did', 2018, https://www.youtube.com/watch?v=DXMfm0jU2K0 (last visited 23rd May 2022).

In languages, there are different terms used to refer to the act of perceiving a sound such as "to hear" and "to listen" in English, "ascoltare" and "sentire" in Italian, and "entendre" and "écouter" in French. These words tend to be used as synonyms, although they express slightly different concepts related to how one person perceives a sound depending on his or her intentions. Researchers discussing the theme of the *ways of listening* utilise different terms to describe the various possibilities.

Firstly, to remain on the purely linguistic side of the definitions, in his book *Traité des Objets Musicaux* (1966), Pierre Schaeffer made a list of four different ways in which we could listen to the same sound (as cited in Knight-Hill, 2019). The first one is *listening* ("écouter") and it refers to the identification of the sound source, for example recognising an ambulance siren or a cruise horn. Then, the second one is *perceiving* ("ouïr") and it is less specific than the previous one since it consists of simply being aware of the sounds surrounding us, without concentrating on them and trying to find their source. For instance, in the case of the ambulance, we are aware of the sirens and other traffic sounds just as a part of a more general city sound environment. Thirdly, there is *hearing* ("entendre") which is seen as a more "technical way" of listening to a sound. In this case, we pay attention to physical and perceptual parameters such as pitch, timbre, duration, intensity and regularity of a sound. Lastly, *comprehending* ("comprendre") goes beyond the simple recognition of the sound and it involves the identification of its meaning. In this case, sound is treated as a sign, as a meaningful object that is understood by the receiver. For instance, ambulance sirens mean that something serious has happened.

However, these four modes are not absolute, meaning that the receiver can shift between them depending on his or her listening intentions. For example, if we are walking outside, we perceive different sounds coming from the street and from the surrounding environment in general. If we focus a little more, we may listen to a siren, which means that there is an

ambulance, therefore, by concentrating even further, we comprehend that an accident or an emergency has happened. Then, by hearing the intensity of the sirens, we can understand how close it is from us, if it is coming towards our direction or if it is going in the opposite one.

The same reasoning can be applied to any kind of sound, including that of knocking on a door. In this latter case, the door qualities together with the intentions of the listener have the ability to convey different impressions. If seen as "hearing", it is recognised as the sound of a person knocking on a door; if seen as "perceiving" it is simply a background noise, whereas as "comprehending", through the door material, the strength and the regularity in which the knocks are performed, we can understand the characteristics of the door, such as heavy or light, and the emotions of the person knocking as well.

A second category of ways of listening is the one proposed by Chion (1994). Here, the author suggested three listening modes: *casual listening*, *semantic listening* and *reduced listening*.

Casual listening is the most common way of listening and it consists of listening to a sound with the aim of gathering information about its cause or its source, and it could correspond to Schaeffer's *écouter*. Depending on the visibility of the cause, the sound can provide additional information, in case the source is visible, or it could be the only piece of information we have of the event, in case its source is out of the field of view. Despite casual listening being the most common among the three modes proposed by Chion, the author notices how it is also the most easily influenceable and misleading one. Therefore, the information we gather from it depends exclusively on the analysis we make of the sound and, consequently, we are rarely able to recognise a unique source. Indeed, humans possess the ability to recognize either the specific sound source or its general category of belonging. In the case of the human voice, if we are familiar with the speaker, we will be able to specifically recognise his or her voice, whereas if we do not personally know the speaker, we may only categorise the voice into male or female, young or old. The same process is applied

to immaterial and mechanical objects and can also be found in everyday actions, such as walking, since it is possible to recognise the identity of the walker by the steps, if we are very familiar with him or her, and only generally identify the category of "steps" if we are not. Furthermore, casual listening can be especially misleading in cinema with the above-mentioned principle of synchresis. In this case, indeed, the sounds we are hearing do not come from the object depicted on screen, the real source, but they are coming from an off-screen object used to render the sound of the one on-screen, causing this phenomenon to deceive our casual listening.

Semantic listening, on the other hand, is the code used to interpret a message and it is the most complex one to investigate. Codes can be languages of all kinds: from spoken languages, to sign language, Morse code and a lot more. In the case of spoken languages, sentences are composed of a multitude of phonemes, which are not listened to separately, but they are listened to in a more general way, as a part of a system: the sentence. As a matter of fact, even differences in pronunciation, for example when speaking a foreign language, can lead to the right interpretation of the sentence said. Thus, semantic listening is not too sensitive to variations in the pronunciation of phonemes.

Finally, reduced listening is a term that Chion borrowed from Schaeffer and it may be the correspondent for the previously mentioned *entendre*. With this last mode, Schaeffer highlighted the focus on sound traits, independently from its source and meaning. If the previous mode is the most difficult one to study, this one is the hardest to figure out from the receiver's point of view. In fact, in order to identify the specific characteristics of a certain sound, one takes more than one trial and has to listen to it multiple times. For example, given a melody, a musician can reproduce on the first try its general outline, but not the exact tone, with the exception of people possessing "absolute pitch", a rare ability to identify and recreate a given musical tone without having to go through a trial and error process. If

musicians struggle to identify the exact characteristics of a sound, in particular a musical tone, normal people may face more difficulties.

As in Schaeffer's ways of listening, also the ones proposed by Chion are not mutually exclusive, but they overlap and combine, especially in a film soundtrack.

Lastly, another classification of listening intentions is the one proposed by Vernon (1934). In this particular case, the author analysed the way in which listeners perceive music, which can be *indefinite* or *definite*. However, the author specifies that every individual listens in a different way, but all the possibilities can be summarised in one of the two categories.

Indefinite listening is characterised by a passive reception of sound and it happens, for example, when hearing a foreign language we do not know, in infants, who cannot comprehend what they are being told, and when we are hearing a song in public but we are not paying attention to in for different reasons. In all these cases one can at least understand the mood of the speaker thanks to the prosodic parameters of the voice and the rhythm of the song. Some responses that can follow indefinite music listening are therefore emotional moods or interpretations, visualisation of images or daydreaming caused by the perception of emotions, awareness that there is a sound, and wandering of thoughts to topics that are not connected to the meaning of the musical piece. This mode of listening can be associated with Schaeffer's "perceiving".

Definite listening, on the other hand, requires active attention to understand the different components of a song, as well as its meaning and other characteristics such as title and author recognition. The receiver is not only aware that there is a sound in the external world, but he or she also tries to identify it along with its meaning and source. This type of mode is similar to the union of Schaeffer's reduced listening, "hearing" and "comprehending". A definite-listening-musician can, for example, pick up a tune from just hearing it and recognise a familiar melody. In this case, the responses this listening mode can generate in the receiver

are: kinaesthetic body reactions, synesthetic processes, visualisation of the notes on a musical instrument, verbalisation of the note names, emotional responses as well as auditory imagery.

In general, to go back to the Gestalt principle of figure-ground mentioned in § 2.1, definite listening corresponds to the figure, along with Schaeffer's "listening", "hearing" and "comprehending", and Chion's casual, semantic and reduced listening, whereas indefinite listening corresponds to the ground, together with Schaeffer's "perceiving" and some cases of Chion's casual listening (e.g. more general sound categorisations).

In conclusion, by being aware and recognising the variety of listening intentions, it is possible to notice and appreciate the different impressions a same sound can convey and responses it can generate in the receiver, as well as the amount of information this last one is able to gather.

2.2 THEORIES OF EMOTION

As discussed in the previous sections, humans can emotionally react to sounds, whether they are everyday sounds, sound effects or a musical soundtrack, based on parameters such as pitch, pace, intensity, frequency, regularity and more. But the main question when investigating this topic is: what is an emotion?

According to the Cambridge Online Dictionary[9] an emotion is "a strong feeling such as love or anger, or strong feelings in general", whereas based on the definition given by the Collins Online Dictionary[10], an emotion is "a feeling [...] which can be caused by the situation that you are in or the people you are with". These definitions may be suitable when trying to give a simple and brief explanation of what emotions are, or when trying to understand the

[9] https://dictionary.cambridge.org/
[10] https://www.collinsdictionary.com/

meaning of emotion for everyday purposes. However, when discussing emotions from a scientific point of view, the aforementioned definitions are not satisfactory.

Robert Plutchik, an American psychologist and Professor whose work on emotions is considered to be one of the most valid, estimated that over the course of the 20th century, more than 90 definitions of the word "emotion" have been proposed by researchers, psychologists and other theoreticians (2001), highlighting the little consent around this enigmatic topic.

Several researchers tried to investigate the world of emotions in order to conceptualise and demonstrate a theory on which the scientific world could agree with, classifying emotions as primary/basic or fundamental or along the dimensions of valence (positivity or negativity) and arousal (excitement or relaxation). No set of basic emotions proposed so far has been identified as the right one, since there are different opinions on the number of basic emotions, on their identification and on the reason why they are basic (Ortony & Turner, 1990).

The present work will examine in depth only the first type of categorisation (primary/basic) explained following the theories of Charles Darwin, who tried to investigate the origins of some emotional driven behaviours. In particular, the sets of emotions proposed by Robert Plutchik and Paul Ekman will be examined, inasmuch they offered a list of basic and primary emotions, respectively, that are still taken in consideration today. The theoretical foundations and findings of these last two researches represent the basis in which the experimental research presented in Chapter 3 was made.

To begin the analysis on emotions, the work that has been the most influential for this topic in the last centuries has definitely been *The Expression of the Emotions in Man and Animals* by Charles Darwin (1872) in which the author attempted to explain the origins and development of expressive actions and movements elicited by different stimuli through three principles.

These principles are: the principle of serviceable habits, the principle of antithesis and the principle of the direct action of the excited nervous system on the body.

The first principle states the power of the force of habit, underlining how different movements can be performed, even unconsciously, once they are well established in our lives. Namely, if these movements, that are functional for desire gratification or for sensation ease, are often repeated in the course of an individual's life, they become habits that can be performed every time the same desires or sensations are felt, even if with less intensity. Indeed, actions are associated with other actions and with states of mind, therefore if a presented stimulus alters the present mind state, a chain of different movements will be produced. Moreover, some of these actions may also be willingly partially repressed, causing slight movements that are not under the individual's control and that are recognised as expressive.

A couple of examples of "serviceable habits" reported by Darwin are the eyebrow raise when we are surprised and extending the arms to protect ourselves when falling. In the first case, surprise is caused by something unexpected or unknown happening to us, therefore, we try to find the cause and, in order to do that, we open our eyes as much as we can "so that the field of vision may be increased, and the eyeballs moved easily in any direction" (Darwin, 1872, p. 281). All of these movements were voluntary at first, centuries or millennia ago, then, they became habitual due to their repetition every time the same or a similar stimulus occurred, and finally, they became hereditary to the point that they can also be willingly performed to conceal or deceive a certain emotional state.

Then, the second principle is that of antithesis. The movements produced in this case are the exact opposite of those produced for specific situations and purposes of the first principle. To better understand antithesis, Darwin presented an example from the animal world. He described the different movements and postures of a dog and a cat ready to attack their

enemies: hair erection, mouth open, teeth uncovered, tail raised and a general body bulking to give an intimidating impression and to get ready to fight. These movements belong to the first principle, while their opposite, thus not hostility but affection towards for example their owner, are the antithesis. These actions are therefore performed due to an opposite will and, as their counterpart, they have become habits. If during the serviceable habits process every functional movement performed voluntarily corresponds to the action of a specific set of muscles, the opposite movement, therefore the antithesis, requires an opposite set of muscles, e.g. pushing or pulling, turning right or left.

Darwin also noticed that these kinds of movements connected to a mind state are innate or inherited since they happen identically also in other individuals, regardless of age, race and species.

Lastly, the third principle states that some expressions happen in humans and in animals in order to let the nervous system discharge the excess of excitement (Hess & Thibault, 2009). For instance, laughter is an action done to free ourselves from an excessive presence of nervous energy caused by a physical (e.g. tickle) or psychological (e.g. something funny) tension. Even the smallest excitement of a sensitive nerve causes a reaction to the heart, which then starts an action-reaction process with the brain and it causes us to change appearance or move in a certain way.

One of the easiest ways to identify emotions is through the classification into exciting or depressing. If the organs of the body and mind are in excess of energy, hence they are excited, they perform their actions more rapidly, for example if the individual is happy or angry, while if this last one is sad, they will produce the opposite reaction by working in a slower way. The movements induced by the nervous system are those acquired through habit, as seen in the first principle, so the reaction to a previously experienced exciting or depressing stimulus will be unconsciously reproposed when the same or a similar stimulus is

presented. Taking rage as an example, the heart is subject to a strong acceleration and, as a consequence, the body's appearance changes along with the movements we are led to practice. In the first case, the face turns red, the nostrils dilate, the voice changes and the body often trembles, whereas in the second one, we clench our teeth, the respiration changes and the muscular system tends to act in a violent manner.

In conclusion, for Darwin, emotion expressions are manifestations of an underlying emotional state aimed at survival, they can be visible or not, they can be observed from lower animals to human beings and, most importantly, they are innate or inherited, supporting the thesis that they exist thanks to evolution.

The work of Charles Darwin has been fundamental for understanding the origins and the mechanisms of emotional expressions, however, the scientist did not provide a definition of the term “emotion”.

Robert Plutchik, an American psychologist, had the goals of clarifying what an emotion is and of finding a way to measure it. In trying to do so, he took inspiration from four intellectual traditions on the topic of emotions, which he named: the evolutionary theory of Darwin, seen above, the psychophysiological one of James, the neurological of Cannon and the psychodynamic one of Freud (Plutchik, 1982). By trying to fit these pioneering approaches into a more general theory of emotion, Plutchik constructed a systematic structural model of emotions and outlined eight prototypical patterns of behaviour fundamental in the evolution process. These dimensions, organised in four couples of polar opposites are: destruction and protection, incorporation and rejection, reproduction and deprivation, and orientation and exploration, and they can be observed both in humans as well as in animals. To each of these patterns of behaviour corresponds a series of actions aimed at survival and a cluster of words expressing emotions can be assigned to each one of them. As Darwin also stated, a variety of emotional responses are triggered by events that are

important for the organism, such as being a predator or being a prey, seeing a friend or an enemy (as in the above-mentioned case of the antithesis reactions in cats and dogs), being in front of a novel occurrence or seeing a mate. To start giving a more specific definition of "emotion", the psychologist combined his and Darwin's findings by saying that emotions are body reactions to survival problems, hence they represent a problem solving mechanism aimed at maintaining control over specific triggering events (Plutchik, 1982).

Anyway, an individual rarely experiences emotions in a pure state and he or she can also experience different emotions at the same time, which may even vary in intensity, degree of similarity and polarity. The union of all of these varying factors is what makes the research on emotions difficult to study and to precisely define once for all.

Nevertheless, Plutchik founded his theory on the eight patterns of behaviour and, by associating an emotion term to every behaviour, he theorised that there are eight primary emotions, which are fear, anger, joy, sadness, acceptance, disgust, anticipation and surprise (1982).

To represent his findings in a visual way, Plutchik used a circumplex model, already theorised by Schlosberg, 1941 (as cited in Plutchik, 2001), in which he added a third dimension representing the intensity of an emotion producing, therefore, a model shaped like a upside-down cone. The farther a non-basic emotion is from the top circle, the less intense and recognisable it is.

What is of particular interest for the present study is that Plutchik applied the colour-wheel concept to his model. This analogy, already proposed by McDougall (as cited in Plutchik 1962) consists in considering emotions as hues, which can vary in intensity (saturation) and polarisation. As a matter of fact, just like complementary colours are placed on opposite sides of the wheel, Plutchik's eight basic emotions are arranged in four pairs of polar opposites: ecstasy and grief, rage and terror, vigilance and amazement, and adoration and loathing.

Furthermore, similar emotions are close to each other and their mixture gives life to a set of secondary or tertiary emotions. A further examination on Plutchik's colour-emotion analogy will be discussed in the following section.

Being this model a three dimensional one, the vertical dimension represents arousal and it ranges from a state of total excitement to that of total lack, which, in colour terms, refers to the saturation of the hue.

To summarise, according to Plutchik's psychoevolutionary theory, an emotion is

> "a complex chain of loosely connected events that begins with a stimulus and includes feelings, psychological changes, impulses to action and specific, goal-directed behaviour. [...] They are responses to significant situations in an individual's life [...]" (Plutchick, 2001, pp. 345-346)

and it mainly has two functions: communicating information about one's intentions or actions and increasing one's chances of survival in emergency situations. These functions highlight the strong similarity between emotions and sound, which, as mentioned above, through its characteristics such as pitch, intensity, regularity and frequency, can be an indicator of someone's intentions and emotional state (e.g. a low pitched knock repeated few times with low intensity is a symbol of sadness) or its meaning could be crucial in emergency situations (e.g. an alarm).

Even if according to Plutchik's theory there are a total of eight primary emotions, over the centuries, many researchers, philosophers and psychologists have proposed a different number of basic or primary emotions, indicating that there is not an unique and well established number to refer to (Ortony & Turner, 1990). What is certain is that nearly all the

existing proposals include fear, anger, sadness and many also include joy, love and surprise (Plutchik, 2001; Ortony & Turner, 1990).

Indeed, also Paul Ekman's list of basic emotions includes the first group. In particular, Ekman, the main exponent of emotion studies whose work has been influenced by Darwin, identified six basic emotions: anger, contempt, disgust, enjoyment, fear, sadness and surprise (Ekman, 1999). According to Ekman, the term "basic" has three connotations.

The first one indicates that these emotions differ a lot from one another, not only for valence and arousal, but also for their reception and response. For instance, even if only the negative emotions are considered, their appraisal, cause, behavioural response and other causes or effects will be different depending on if the person is sad, angry or scared. The same concept applies to positive emotions.

Then, the second meaning is connected to Darwin's study on emotions. Some emotions are basic because they have evolved in order to deal with fundamental life tasks, e.g. fighting, escaping, reproducing and more, therefore our judgement on a stimulus and the consequent response we are going to carry out, are influenced by our ancestral and personal past and are thus innate. If it is true that basic emotions are a result of evolution in order to deal with life tasks, apart from conveying expressions that help understand other individuals what is happening, they should generate some physiological changes as well.

Experiments found that there are distinctive patterns of nervous systems activities for the emotions of anger, fear, disgust and sadness (Levenson et al., 1991; Levenson, Ekman & Friesen, 1990, 1983; Roberts & Weerts, 1982; Schwartz, Weinberger & Singer, 1981; Graham, 1962; Ax, 1953, all as cited in Ekman 1999). For Ekman, the reason why no specific patterns have been found for surprise and enjoyment lies in the fact that these two emotions do not have a survival function. Other authors claim that surprise is not an emotion since

valence is thought to be a necessary condition for a state to be classified as such and surprise is a state that can also be neutral (Ortony & Turner, 1990).

Finally, the last nuance of the term basic is probably the most common in everyday language. Something basic is something that, if used in combination with another basic element of the same category, can create a more complex element. In the case of emotions, the mixture of two or more basic emotions can generate complex or compound emotions, which, in Plutchik's view, are the secondary ones. However, Ekman does not consider this third meaning since, apart from Plutchik, no other researcher was currently concerned with it.

By shifting from our ancestors and from the animal world to civilised life, emotions can be triggered in humans not only by actions or events, but also by words, sounds and music. Because of this further process, Ekman proposed two appraisal mechanisms. The first one is automatic, meaning that "the interval between stimulus and emotional response is sometimes extraordinarily short [...] the appraisal mechanism is not only quick but it happens without awareness" (Ekman, 1977 as cited in Ekman 1999 p. 51). This is the case of fundamental life tasks, as described in the second meaning of "basic", in Darwin (1872) and in Plutchik's eight patterns of behaviour (1962, 1982). Whereas, if the evaluation of the stimulus takes more time, is then slower and conscious, as in the case of hearing someone's voice or listening to a song, an extended mechanism is activated. This mechanism may automatically arouse or alert the individual, but no specific emotion may be elicited.

In addition, emotions last in the range of seconds and minutes, and whenever they are perceived for a longer period of time, they become moods, which are states of mind overly saturated with emotions. Sometimes, emotions, moods and personality traits are confused with each other, and language is probably the biggest factor of misunderstanding. As Plutchik (2001) and Ekman (1999) point out, language makes the topic of emotions more obscure and hard to investigate, since there probably exist more emotional words than emotions and it is

therefore difficult to describe an emotion in an unequivocal way. Moreover, according to Ortony and Turner (1990) one of the reasons why there is such confusion and disagreement on the definition of a unique set of basic/primary emotions is due to the fact that theorists use different terms to presumably refer to the same emotion, e.g. joy and happiness or anger and rage.

Furthermore, in previous studies, Ekman (2019, 1970) validated Darwin's theory of universal facial expressions of emotions due to evolution purposes. Emotions triggers may vary from person to person and culture to culture, for instance what makes a person angry does not have the same effect on another person and what makes the latter person happy may not make the former one feel equally happy. However, even if the stimuli are different, the set of face muscles activated as a response to the stimulus remains the same.

To sum up Ekman's view on emotions, the author identified six basic emotions (anger, contempt, disgust, enjoyment, fear, sadness and surprise) which, according to Darwin's theory of evolution, are mechanisms used to deal with fundamental life tasks and were adaptive in our past. These emotions are basic since they all present specific characteristics: they have distinctive universal signals and physiology, they generate automatic appraisal, they are present also in other primates, they do not last for a long time and they are spontaneous, not voluntary.

2.3 COLOURS AND EMOTIONS

The relationship between colours and emotions is more present in our everyday life than one may actually think. In some language idioms, this association is quite evident thanks to expressions such as "feel blue", to indicate a feeling of depression or sadness and the

consequent “Blue Monday”, considered to be the saddest day of the year; “see red” and “black with anger” when someone is angry, “green with envy” when someone is jealous, “white with fear” when someone is scared and many more (Jonauskaite et al., 2020; Kurt & Osueke, 2014).

In the scientific field, the leading figure of “coloured emotions” is Robert Plutchik with his emotion-colour wheel. Plutchik was not the first scholar to propose this analogy (see McDougall, 1921 as cited in Plutchik 1962), but he is the first one who proposed a tridimensional model in which he assigned specific colours to the different emotions. As seen above, emotions are classified in the same way that hues are: they can change not only in saturation and brightness, which represent the various intensity and polarisation degrees of an emotion, but they are also arranged in a way for which the most similar ones are close to each other, while the complementary ones are situated at opposite extremes. Furthermore, as it happens with primary colours, their mixture generates a set of other colours, called secondary. In this perspective, red and blue make green, and anger and joy will make pride (according to Plutchik, 1962). Image 1 visually represents the colour-emotion associations proposed by the psychologist.

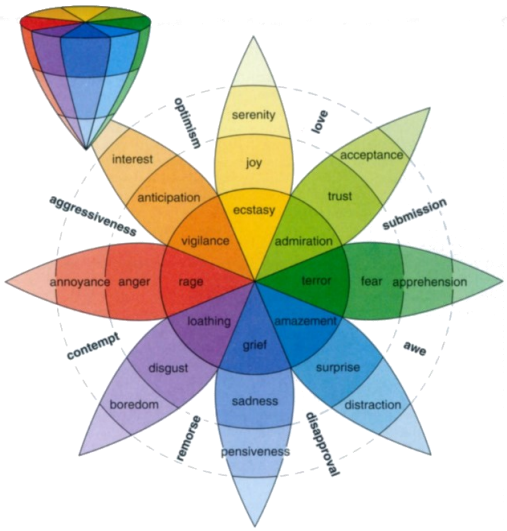

Image 1. Plutchik’s three-dimensional circumplex model describing the analogy between colours, emotions and their relationship. Found in Plutchik, 2001.

However, Plutchik's model is not supported by any empirical evidence that proves that the superimposition of the colour wheel onto the emotion one is correct, nor that green is the colour of terror and yellow of ecstasy. This model is therefore appropriate when explaining, according to Plutchik, how emotions work thanks to the colour-emotions analogy, but its combinations do not represent scientific findings and must for this reason seek validation.

A multitude of studies from across the world have tried to investigate if there is a direct relationship between a specific colour and an emotion, trying to find a consensus solution that could be applied in every future case. Results showed that these kinds of associations are not fixed and universal, but they may change from person to person depending on factors such as gender, age and culture (Güneş & Olguntürk, 2019; Gong et al., 2017; Gao et al., 2007; Manav, 2006; D'Andrade & Egan, 1974). For instance, in Western cultures black is the colour of funerals, whereas in China it is white.

Thus, there is not a consensual result showing that a determined colour is able to elicit one and one only emotion or that people always associate a colour to a given emotion. However, an analysis of different studies carried out in different countries as well as in different time spans, showed how there is a trend and a general good agreement, even if not total, on some emotions. The emotions considered in the analysis are those examined in the experiment in Chapter 3, and they are four among Ekman's basic emotions list, that is to say happiness, anger, sadness and fear.

When choosing the colours for an experiment on emotions, it is very important to take in consideration that colours are not only composed of hues, but saturation and brightness are crucial factors (Wilms & Oberfeld, 2017; Hupka et al., 1997; D'Andrade & Egan, 1974). These two perceptual colour dimensions have proved to be, in some cases, more important than the hue itself when it comes to emotion elicitation. Wilms and Oberfeld (*ibidem*) manipulated all the three dimensions of colour to see their effects on emotions. In doing that,

they used the dimensional approach, therefore they did not divide emotions into primary/basic and secondary, but according to valence, arousal and dominance. Results showed that saturation and changes in hue had a significant impact on arousal, since higher ratings were registered for highly saturated colours. On the other hand, colours with medium saturation and a high brightness had more impact on the dimension of valence.

Therefore, selecting the right colour, as a combination of the right hue, level of saturation and of brightness is crucial, since the same hue but with different parameters might not elicit the same emotional response. For instance, a bright yellow with maximum saturation may evoke happiness, while a dark, low saturated yellow, with some hints of green, can be associated with sickness (Kaya & Epps, 2004; D'Andrade & Egan, 1974).

In the researches carried out by Kaur (2020) [1][11], Demir (2020) [2], Binzak Fugate and Franco (2019) [3], Hanada (2017) [4], Da Pos and Green-Armytage (2007) [5][12], and Kaya and Epps (2004) [6], some interesting results can be found.

First of all, a bright yellow is associated with happiness in every study, together with a light orange. This last colour-emotion association, however, is not present in [1] but it is replaced with black, which is perceived by the participants as happy since it is seen as a symbol of fashion and power. In the following tables (Table 1, Table 2, Table 3 and Table 4), the hexadecimal (hex) codes for each colour are presented. This system has been used to group all the different colour systems used in the mentioned studies in an unique way.

	[1]	[2]	[3]	[4]	[5]	[6]
HEX	#ede939	#ffff00	#ffff33	#faf54e	#f8a723	#ede939

Table 1. Hex codes for happiness.

[11] Numbers in [x] will be used to refer to the studies in the tables as well as in the following pages.

[12] In Table 1, Table 2, Table 3 and Table 4, Da Pos and Green-Armytage values [5] are referred to the mean values of all the colours associated with the different emotions.

Secondly, almost all the studies, except for [2] who does not provide any data, agree that red is associated with anger. However, different degrees of brightness and saturation have been considered. Concerning brightness, a dark and medium saturated red was detected in [1] and [5], whereas a lighter but still medium saturated red has been found in [4] and [6]. In [3], red is still a symbol for anger, but this time the colour is fully saturated and bright.

	[1]	[2]	[3]	[4]	[5]	[6]
HEX	#862410	/	#ff0000	#e14750	#89312e	#d7414b

Table 2. Hex codes for anger.

Regarding sadness, all the studies agree that grey is the colour associated with this emotion. However, other options were also found: black in [6], a more or less saturated and bright light blue in [4] and [5], respectively, whereas in [3] this emotion is also represented through a deep blue.

	[1]	[2]	[3]	[4]	[5]	[6]
HEX	#58473a	#808080	#666666	#818080	#95a7af	#838686

Table 3. Hex codes for "sadness".

Finally, in these studies fear appears to be associated with black in [2], [3] and [6], and with purple in [4] and [5]. [1] did not find any correspondence for this colour.

	[1]	[2]	[3]	[4]	[5]	[6]
HEX	/	#000000	#000000	#a36ab8	#6a2c6d	#000000

Table 4. Hex codes for "fear".

To summarise, research made in the last two decades shows some similarities, especially concerning the medium saturated red and anger, bright saturated yellow and happiness, and grey-sadness combinations.

Numerous other studies have been made throughout the years. However, a multitude of them did not provide colorimetric specification, failed to consider saturation and brightness, or used colour words, such as “YELLOW”, “BLUE”, instead of providing a visually coloured stimulus (Wilms, Oberfeld, 2017).

Some of these studies include Jonauskaite et al. (2020) [7], Gilbert et al. (2016) [8], Sutton, Altarriba (2015) [9], Hupka et al. (1997) [10] and D’Andrade and Egan (1974) [11]. Nevertheless, even if they did not provide the specific colour chosen to the reader, their findings are in line with the others mentioned above, concerning saturation and brightness in the case of [11] and hues for the others. More specifically, [7], [8] and [9] agree with the previous studies that yellow is the colour mostly associated with happiness. [10] does not provide data, while [11] agrees that happiness is mostly represented through highly saturated colours. Then, red is also the colour that mainly evokes anger, with [9] and [11] agreeing that it is a low saturated colour. Concerning sadness, grey [7], black [7], [8] and blue [8], [9] are once again the colours that people associate more with this emotion. [11] agrees that an unsaturated colour represents sadness. Finally, fear is connected to black in [7], [9] and [10], with [11] confirming that unsaturated dark colours mostly elicit a sentiment of fear. [8] did not provide any data.

This analysis will turn out to be useful in the choice of the colour for the stimuli for the experiment described in Chapter 3.

In broad terms, these findings partially support the associations made by Plutchik, with the exceptions of sadness, which is reported to be mainly grey, or an unsaturated colour in general, and fear who is mostly perceived as black, with some cases of purple.

These associations can be applied to many fields such as interior design (Güneş & Olguntürk, 2019; Savavibool et al., 2016; Kurt & Osueke, 2014; Manav, 2006), web design, artificial intelligence (Nijdam, 2005), marketing, colour perception and visual effects with colour correcting.

Concerning interior design, especially in shop architecture, colour is undoubtedly an element that enhances the personality of the brand, as well as the aesthetic qualities and overall branding (Savavibool et al., 2016). A careful colour choice can also be useful to understand the feelings of consumers. As demonstrated by Gong et al. (2017), colour preferences have a direct effect on the feelings that people have towards colours, meaning that the most preferred ones tend to evoke more positive emotions, while disliked or hated ones are more likely to generate negative emotions. It is therefore important to also take into account the target consumer's colour preferences and expectations when designing products (Gilbert et al., 2016) and, in general, all the brand touchpoints with the public. However, customers are not the only ones who will benefit from the emotional effects of colour. Indeed, the colour of the workspace is able to influence workers too, both psychologically and physiologically, for example by impacting their mood and emotional states, heart rate, productivity and creativity (Savavibool et al., 2016).

It is also worth mentioning another example of how colour-emotion links can be used in marketing, as a way to define a brand's strategy. In marketing there are the so-called "brand archetypes", which are syntheses of a brand's personality and image traits operated through different means. Some use fairy tales characters such as "the joker", "the hero", "the king" and "the wise", others prefer personas like "the guardian", "the warrior", "the patriarch" and "the earth mother", while others refer to brands as colours. This is the case of Needscope by Kantar[13], a data management company. With the Needscope model, a company uses

[13] https://www.kantar.com/expertise/brand-growth/brand-purpose-and-positioning/needscope

emotions and colours to define a brand's placement, with each colour-emotion representing a set of characteristics that must be applied in the brand's strategy as well as in every touchpoint with the public in order to be coherent with the overall image. In total there are six colours: red, purple, blue, brown, orange and yellow, and they symbolise energy, danger, passion and desire, nobility, power and luxury, strength, serenity, loyalty and authority, stability, comfort and wholesomeness, happiness and warmth, hope, fun and optimism, respectively. Nonetheless, even if the company refers to six "emotions", after the analysis done in section § 2.2, it is clear that the ones presented by Needscope are not emotions, or at least not primary or basic ones, but rather feelings and personality traits. It was still interesting to allude to an application of colour-emotion links outside of the most common one of interior design.

The last application of colour-emotions links is in the world of media production and, most specifically, the interaction between sounds, colours and emotions. It is this last concept the main focus of the present study.

As seen in Chapter 1, depending on its perceptual and physical parameters, sound has the ability to evoke a number of different feelings and emotions to the listener, whether it is presented alone or combined with an image. Since it has been proved that colours have the same power, it is interesting to examine what happens if the two elements are combined.

Pauletto and Iop (2021) and McDonald et al. (2022) investigated this process. It is important to notice how these two studies tried to investigate the same phenomenon, but, in doing so, they used two very different methods. Four main differences can be found in these studies:

- Pauletto and Iop used four of Ekman's primary emotions, while McDonald et al. used a dimensional approach along arousal, valence, together with familiarity and enjoyment

- in the first case, emotional everyday sounds were used (“knocking”) whereas the stimuli of the second one consisted of piano performances
- five coloured doors were used in Pauletto and Iop (red for anger, yellow for fear, blue for happiness and brown for sadness), while two hue filters (red and blue) were added to McDonald et al.’s videos
- and finally, the first study used still images, whereas the second one used videos.

Even if these two studies presented methodological differences, they both showed that hues had little effect on the overall perception of emotions or arousal ratings. This task was mainly carried out by sound alone, bringing the aural modality to dominate over the visual one. A possible explanation for these results could be that both these studies did not consider changes in saturation and brightness, Pauletto and Iop considered different colour-emotion combinations then the ones presented above.

A last example of music and colour associations is the study conducted by Bresin (2005). Here, participants were asked to listen to twelve musical performances expressing different emotions and performed with three different instruments, and rate them using colours. Results showed that people used different colours to classify the same musical piece and that these differences depend mostly on the performance and the instrument. In particular, this study is in line with the ones presented above, registering yellow as the colour used to describe happiness and red to describe anger.

To conclude, this section showed how, even if there are some differences, there is a certain degree of agreement on the colour-emotion links found and how these relationships can have several fields of application.

3. THE EXPERIMENT

As seen in the sections above, both sounds and colours have the ability to elicit emotions. Concerning sounds, it has been shown how both music and sound effects, as well as everyday sounds and artefacts, have this power. On the other hand, concerning colours, it has been proved how, even if there is not a consensual solution, associations between specific colours and a related emotion have been found. These findings can have different fields of application, from media production to marketing and interior design.

The present section illustrates and discusses the experiment through which we tried to merge these two aspects by further investigating the influence of colours in emotion perception expressed via sound, more specifically, knocking sounds.

Findings may be applied to all the various communication fields concerned with emotion elicitation: from bigger products such as movies, to video games, music videos, video advertisements, personal or professional videos in platforms such as YouTube and TikTok, and more immersive solutions such as virtual reality.

For the development of the questionnaire and other methodological matters, we collaborated with Prof. Emery Schubert and Diana Zhang of the University of New South Wales Sydney, and Dr. Anthony Chmiel of the Western Sydney University.

The experiment was conducted in June 2022.

3.1 INTRODUCTION

Everyday sounds are very important means of information: from the identification of the source of an event, e.g. a car crash, to the recognition of specific emotional states, e.g. an angry person knocking on our door. As exhaustively discussed in Chapter 1, the combination and recognition of different sound parameters, such as intensity, pitch, timbre, duration and regularity, are of fundamental importance both when designing the sound for an artefact and for a media product, as well as in everyday life. By the perception of such aural stimuli, a person can initiate a set of different responses as a reaction to the sound sources, depending on the emotional state elicited (see Chapter 2). For instance, an alarm will automatically activate a flight mechanism, whereas hearing someone saying something funny will unleash nervous energy in the form of laughter.

Furthermore, as it has been examined in Chapter 2, emotions can not only be elicited by sound, but they can be associated with colours as well.

Knocking sounds are among the impact sounds that compose everyday sounds, and, as seen above, they have the power to elicit emotions in the listener. The impact of these kinds of sounds on emotions has been investigated by Barahona-Ríos and Pauletto (2020), Bedenko et al. (2020), Houel, Arun et al. (2020), and Vitale and Bresin (2008). The influence of colours on these types of sounds has then been investigated by Pauletto and Iop (2021).

The present experiment aims at investigating the influence of certain colours, proved to be associated with specific emotions, on emotionally performed knocking sounds.

The work's hypothesis is that the association between knocking sounds and emotions, already validated by Pauletto and Iop (2021), changes if the knocking sound is performed on a door of a different colour. This study draws inspiration from the above mentioned one since audio stimuli from the same dataset, used also in Barahona-Ríos and Pauletto (2020), have been used and, as a consequence, the same set of basic emotions has been selected. However, the association between colours and emotions is different and the present work does not investigate the influence of the door material on emotion perception.

The starting point for the research has been the work conducted by Sandra Pauletto and Alessandro Iop (2021) in which they investigated the perception of emotions in multimodal stimuli, namely emotional knocking actions presented together with images of doors of different colours and materials. Their results showed a prevalence of the aural modality on the visual one.

Another study that has been of inspiration for the present one, from the methodological point of view, has been the one conducted by McDonald et al. (2022). Also in this study, results showed a dominance of the aural modality on the visual one.

Nonetheless, even if the present experiment draws inspiration from the two mentioned ones, the preparation of stimuli and some methodological matters have been carried out in a different way.

To briefly summarise, the common ground between the present study and the previous ones is the choice of the emotions and use of knocking sounds extracted from the same dataset in Pauletto and Iop (2021), whereas the use of a 0 to 10 scale and the follow-up questions on synesthesia have been taken from McDonald et al. (2022).

The differences, on the other hand, concern the use of video stimuli, and not still photographs as in Pauletto and Iop (2021), different colour-emotion associations than in *ibidem*, and the use of sliders built on a continuous scale, and not on an integers one as in McDonald et al. (2022).

Therefore, our hypothesis of colour-emotion associations are:

- Yellow = happiness
- Red = anger
- Purple = fear
- Grey = sadness
- White = neutral

These associations have been made according to the analysis discussed in § 2.3.

For these reasons, we expect to obtain different results from Pauletto and Iop (2021).

3.2 METHOD

3.2.1 STIMULI

For this experiment, audio stimuli have been selected from Barahona-Ríos and Pauletto's dataset[14] of emotional knocking actions (2020), which has also been used in Pauletto and Iop (2021). Knocking actions have been executed by Ulf Olausson, a professional Foley artist. This dataset has been chosen because it is already validated and also because it is professionally performed, therefore the quality of the sound, as well as the overall performance should be better for the analysis than knocking actions performed by non-trained volunteers. Five audio recordings, one for each emotion plus a neutral one have been chosen among the proposed ones.

[14] https://zenodo.org/record/3668503#.YqcHbqhBy5c

To construct the visual stimuli, videos of a hand knocking on a white door have been recorded. In order to do that, audios were listened to multiple times to render the synchronisation between the knocking action and the sound in the best way possible. The videos have been recorded using an iPhone 8 external camera and have then been edited using HitFilm Express to add the audio and trim them so that the length was of 3 seconds each and later with Adobe Premiere Pro to change the colour of the door.

After the analysis presented in § 2.3, the following colour-emotion associations were chosen:

- Happiness and yellow (hex code: #ede939)
- Anger and red (hex code: #d7414b)
- Fear and purple (hex code: #a36ab8)
- Sadness and grey (hex code: #838686)
- Neutral and white (natural door colour, average hex code: #e6dfdc)

In total, twenty-five audiovisual materials were produced (5 audio stimuli x 5 colours).

Concerning fear, even if black was the colour mostly associated with it, we selected purple for two main reasons: the first one being that black was also found to be a quite consistent association also with sadness, therefore choosing this colour could have created some confusion and interference. Secondly, and consequently to the first observation, purple is the colour that refers to fear in a more unique way and therefore is less likely to be associated with other emotions.

3.2.2 PROCEDURE

An online questionnaire was created using the online platform PsyToolkit[15] (Stoet, 2010, 2017). After a brief presentation of the study aim and of the privacy policy, participants were informed on the structure of the survey. Twenty-five questions composed of a 3-second-video followed by four sliders were presented to them. The videos and the sliders were positioned on the same page. After watching the video, which could be seen as many times as they preferred, participants had to judge the emotion portrayed in it by moving four sliders constructed on a continuous scale ranging from 0 (total disagreement) to 10 (maximum agreement). Before each slider, the following statements were reported:

- The emotion expressed in the video is **HAPPINESS**
- The emotion expressed in the video is **SADNESS**
- The emotion expressed in the video is **FEAR**
- The emotion expressed in the video is **ANGER.**

The order of presentation of the stimuli has been randomised for each participant.

Before starting the experiment, participants had to listen to two audio recordings, one high in intensity and the other one low, in order to adjust the volume of their device's speakers or headphones and to not change it throughout the course of the experiment. These two audio samples have been randomly extracted from the same main dataset used for the experiment. Different audios were used in order to avoid creating bias on the actual sounds proposed for the research.

At the end of the twenty-five questions, some personal data was collected (gender, age, country of origin) as well as information about the participants' sight and hearing state. Furthermore, since the experiment involves the union of two senses, they were presented with a set of questions investigating synesthesia, taken from McDonald et al. (2022).

[15] https://www.psytoolkit.org/

3.2.3 PARTICIPANTS

Sixty-six participants completed the questionnaire. However, six of them have been removed from the final analysis because they reported having hearing (1) and sight (1) disorders, such as colour blindness, an essential factor for the success of the experiment. Moreover, 3 people did not watch all the videos and 1 reported only hearing them and not paying attention to the visual part. Therefore, a total of 60 participants have correctly completed the experiment.

33 participants are females (55%), 25 males (41%), 1 non-binary (2%) and 1 preferred not specify their gender (2%). Concerning age, the range is 18-66, with 32 people in the 18-25 range (54%), 17 in the 26-34 (29%), 4 in the 35-42 (7%), 2 in the 43-50 (3%), 3 in the 51-59 (5%) and 1 above 60 years old (2%). One person wrote ">45" and was therefore not taken into account in this specific section of analysis.

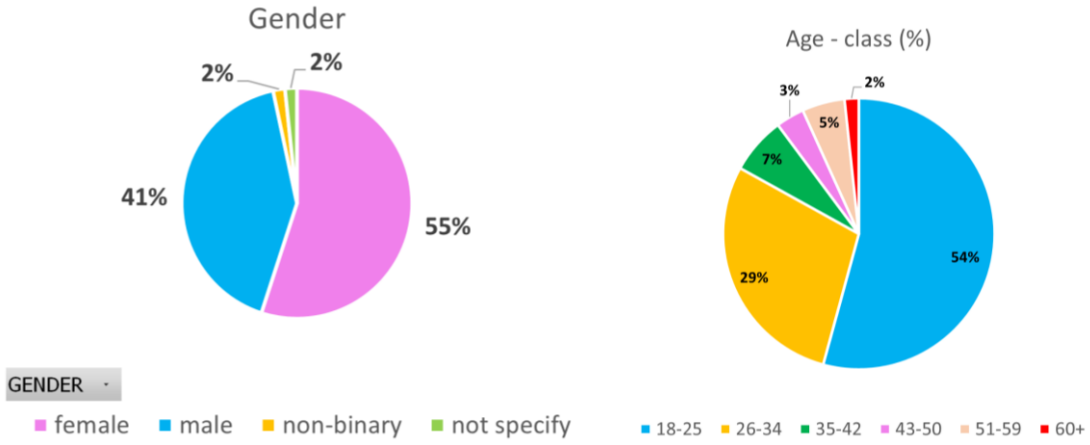

Figure 1. Distribution of participants based on gender (left) and age-range (right).

This study was conducted cross-culturally and Figure 2 shows the distribution of participants based on their country of origin and frequency.

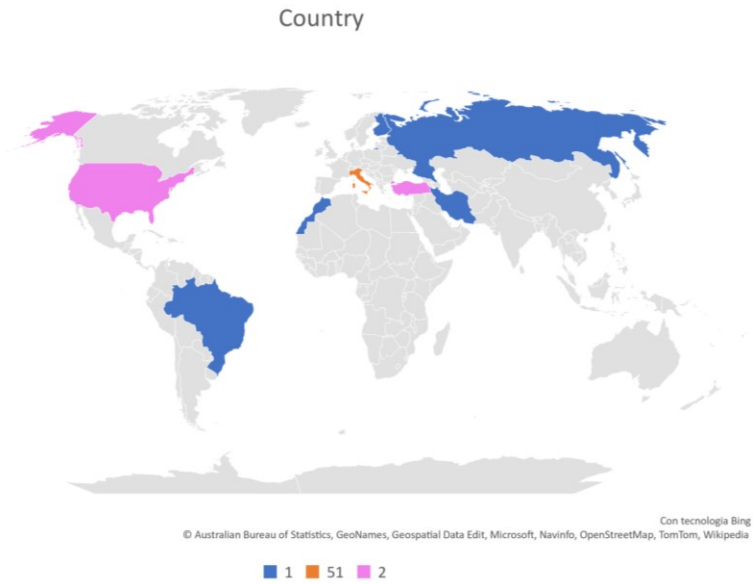

Figure 2. Distribution of participants based on their country of origin.

The majority of participants comes from Italy (51), 2 come from the United States of America, 2 from Turkey, 1 from Brazil, 1 from Finland, 1 from Iran, 1 from Morocco and 1 from Russia.

48% of them (29) reported having normal sight, while 52% (31) reported having corrected-to-normal sight, for instance with glasses or contact lenses. Furthermore, 98% of them (59) reported having normal hearing, while 2% (1) reported to have corrected-to-normal hearing.

Concerning the device in which the participants watched the videos, 62% (37) used their smartphones, 20% (12) their laptop and 18% (11) their computer monitor.

To complete the whole experiment, participants spent an average of 19 minutes.

In addition, since this experiment aims at investigating the effects of the union of two senses, we also asked participants some questions regarding synesthesia. The following questions were taken from McDonald et al. (2022). Each one of them was followed by a pertinent example to better explain the concept of synesthesia to participants who were not familiar with it.

1. Do numbers or letters cause you to have a colour experience?
2. Do weekdays and months have specific colours?
3. Do you imagine or visualise weekdays, months and/or years as having a particular location in space around you?
4. Does hearing a sound make you perceive a colour?
5. Do certain words trigger a taste in your mouth?
6. Do you feel a sense of touch when you smell things?
7. Do you suspect that you experience an unusual blending that other people do not have (other than the one listed above)?

The following tables (Fig. 3 1-7) show the answers for each question, highlighting how females report being more sensitive than males to synesthesia especially when it comes to words triggering a taste in the mouth (question 5), a sense of touch triggered by smell (question 6) and other cases of sense blending not presented in the questions (question 7). The sense fusion most important for the present experiment, the so-called “coloured hearing” (question 4) is once again predominant in women (13 out of 33) than in men (3 out of 25).

	Total
female	33
no	24
yes	9
male	25
no	21
yes	4
non-binary	1
no	1
not specify	1
yes	1
Total	60

1

	Total
female	33
no	18
yes	15
male	25
no	19
yes	6
non-binary	1
yes	1
not specify	1
yes	1
Total	60

2

	Total
female	33
no	26
yes	7
male	25
no	22
yes	3
non-binary	1
no	1
not specify	1
no	1
Total	60

3

	Total
female	33
no	20
yes	13
male	25
no	22
yes	3
non-binary	1
no	1
not specify	1
yes	1
Total	60

4

	Total
female	33
no	19
yes	14
male	25
no	21
yes	4
non-binary	1
no	1
not specify	1
no	1
Total	60

5

	Total
female	33
no	16
yes	17
male	25
no	18
yes	7
non-binary	1
no	1
not specify	1
no	1
Total	60

6

	Total
female	33
no	14
yes	19
male	25
no	18
yes	7
non-binary	1
yes	1
not specify	1
no	1
Total	60

7

Figure 3. Answers participants gave to the questions on synesthesia, grouped by gender.

3.3 RESULTS

Mean values were calculated for the ratings of each video's perceived emotion.

By looking at the average values, it is possible to notice that both happiness and sadness confirm our initial hypothesis, with an average of 7,8 points for the yellow-happiness association and of 7,1 for the grey-sadness one (Fig. 4 C and D).

On the other hand, anger seems to surprisingly have been more associated with the colour white, with an average value of 8,1 points, whereas the association with red comes in second place, equally to the association with purple, with an average of 7,8 points each.

Then, fear tends to be confused with anger, with this last one receiving higher average values in each colour-emotion association (Fig. 4 B). This result is consistent with that of Pauletto and Iop (2021), who found in their study the difficulty in correctly identifying these two emotions and therefore there was no statistical significant difference between them.

However, in general fear was the audio mostly confused with anger, rather than vice versa, and red and purple are the two colours that make the perception of this emotion more evident, 6,3 and 6,1 points, respectively.

Figure 4 shows all the average values, along with the standard deviation (σ) for each intended emotion. In general, the standard deviation values highlight a great variability among the data. This aspect is more clearly shown in the box plots in Figure 5.

Intended emotion
ANGER

Door colour	HAPPINESS	SADNESS	FEAR	ANGER
WHITE	1,8	2,3	4,8	**8,1**
	2,5	2,8	3,5	2,3
YELLOW	1,7	1,8	4,1	7,5
	2,5	2,6	3,6	2,4
GREY	1,5	1,9	4,0	7,3
	2,4	2,5	3,2	2,9
RED	1,6	2,0	4,1	**7,8**
	2,2	2,5	3,2	2,6
PURPLE	2,2	1,9	4,2	7,8
	2,8	2,6	3,4	2,4

A

Intended emotion
FEAR

Door colour	HAPPINESS	SADNESS	FEAR	ANGER
WHITE	2,4	1,5	5,7	7,1
	2,9	2,3	3,4	2,8
YELLOW	2,9	1,2	6,0	6,9
	3,3	2,0	3,5	2,9
GREY	2,7	1,6	5,9	6,8
	3,1	2,3	3,6	3,0
RED	3,0	1,0	6,3	7,2
	3,3	1,6	3,2	3,1
PURPLE	2,2	1,4	**6,1**	**7,2**
	3,1	2,3	3,6	2,9

B

Intended emotion
HAPPINESS

Door colour	HAPPINESS	SADNESS	FEAR	ANGER
WHITE	7,3	2,1	2,0	1,6
	2,3	2,6	2,7	2,3
YELLOW	**7,8**	1,6	1,5	1,4
	2,1	2,4	2,3	2,3
GREY	6,9	2,8	1,9	1,5
	2,5	2,8	2,6	2,5
RED	6,9	1,9	2,0	2,2
	2,6	2,5	2,6	2,7
PURPLE	7,2	1,9	1,6	1,6
	2,4	2,5	2,5	2,4

C

Intended emotion
SADNESS

Door colour	HAPPINESS	SADNESS	FEAR	ANGER
WHITE	2,1	6,1	4,1	1,7
	2,6	3,4	3,5	2,5
YELLOW	1,6	6,7	3,9	1,6
	2,1	2,9	3,5	2,3
GREY	2,2	**7,1**	3,9	1,5
	2,8	2,9	3,3	2,6
RED	1,8	6,7	4,1	1,7
	2,5	3,5	3,5	2,6
PURPLE	1,5	6,7	3,5	1,2
	2,3	3,2	3,6	2,3

D

**values refer to the 0-10 continuous scale used in the sliders.*

Figure 4 . Tables A to D show the mean values of the perception for each intended emotion. White rows symbolise mean values; light blue rows represent the standard deviation; yellow cells show the highest value, whereas green cells show the starting hypothesis. If there are no yellow cells, the highest mean value corresponds to the hypothesis.

To better understand how the intended emotion is perceived through the different colours, results have been mapped in box plots (Figure 5).

Regarding anger (Figure 5 E), it is possible to notice how the median is the highest for the colour white and the lowest for the colour yellow. Another very clear aspect of this boxplot is the dispersion of data: the colour white presents a narrow range, with values starting from >4 (outliers excluded), meaning that participants had a quite similar opinion on this colour-emotion association. The same reasoning can be applied to the colour purple, which has a small range with values starting from 5. On the contrary, the colour grey presents the widest range, with values from 0 to 10, showing the high dispersion of data and the consequent disagreement of participants when judging this association. Concerning the hypothesised colour-emotion association, red-anger, the dispersion of data is wider, with values starting from 2, meaning that this colour-emotion association did not

produce a stronger perception of anger than the other ones. In general, the five batches of data present some similarities among them: they all present a left asymmetry, underlining how the data dispersion is more concentrated on lower values and a general agreement on the medium-high ones.

Moving on to fear (Fig. 5 F), this emotion is the one with the widest data dispersion in every case: the full range of possible evaluations (0-10) has been used for every colour. Results suggest that participants perceived this emotion differently. However, red and purple present a narrower interquartile range. Even in this case, all the batches present a left asymmetry.

Concerning happiness (Fig. 5 G) the yellow box is the one with a higher median, mean value and less dispersion than the other ones. Once again, all the five batches present a left asymmetry, underlining how the dispersion is more concentrated on the lower values, and that participants tended to give medium to high values to this emotion. Nonetheless, the batches present a rather similar symmetry among them, suggesting that the stimulus was correctly identified as happy, independently of the colour.

Finally, the sadness box plots suggest that participants evaluated the sad audiovisual stimuli using values mainly belonging to the upper range, showing once again a left skewed distribution. Data appear to be more concentrated on the grey box, which also presents the higher mean value, median, first quartile and lower limit, confirming once again our initial hypothesis.

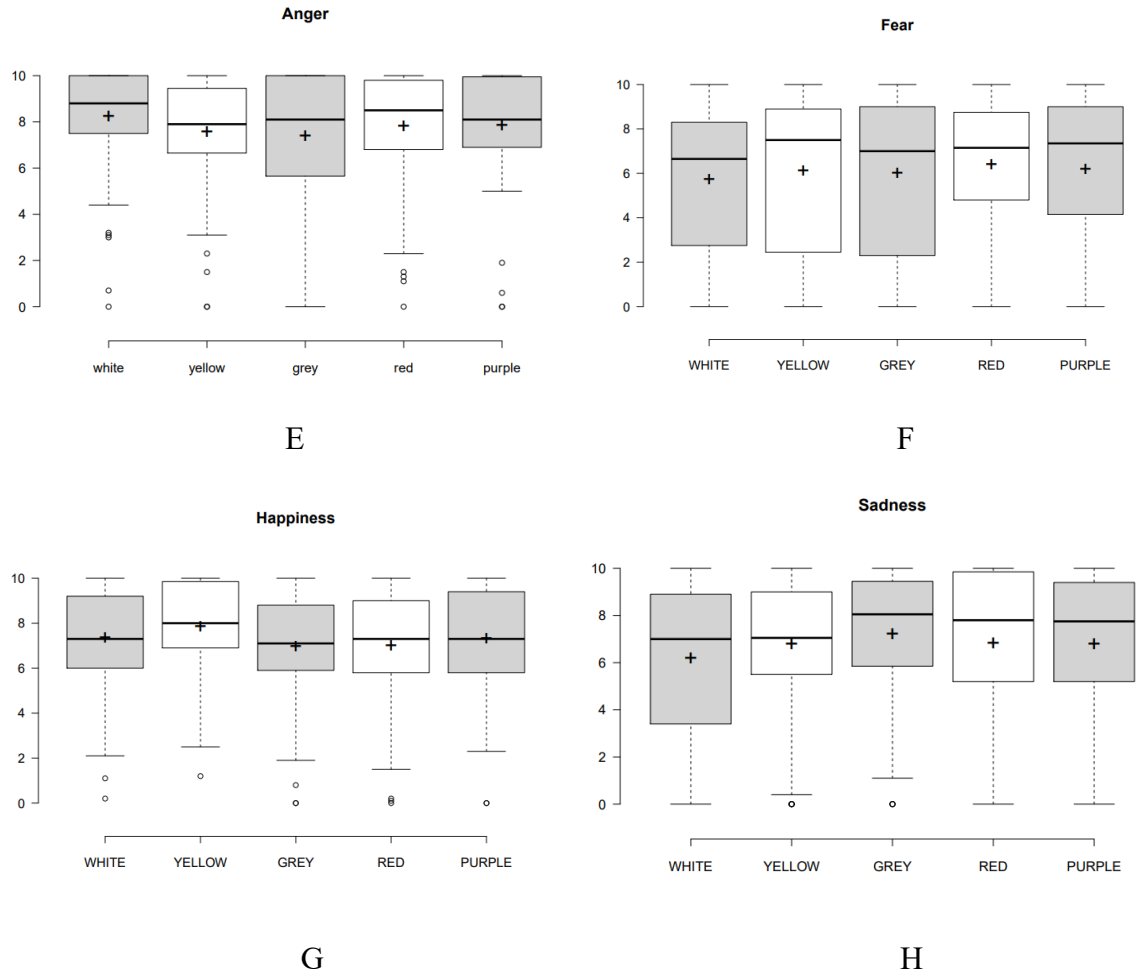

Figure 5. Box plots representing how the intended emotion is perceived through the different colours.

Furthermore, to better see the expected influence of colours in the perception of emotions, we analysed the mean values of the scores given to the neutral audio stimulus. As seen in Table 1. the neutral audio presented together with the yellow door, elicited once again a higher perception of happiness (4,7). The same emotion was mostly associated with red and purple, with an average value of 4,2 each. Moreover, the neutral audio but presented with a white door was mostly associated with sadness, with a mean value of 4,4. The same emotion received the highest average score for the colour grey, reinforcing our initial hypothesis.

On the contrary, fear and anger did not receive the highest average score for any colour. However, fear was mostly averagely associated with grey (3,8) and anger with red and purple (2,5 each).

Nevertheless, all the average scores found in neutral are below 5, showing that no emotion prevailed on the others thanks to the visual channel. This could indicate that the audio, together with the hand movements, were correctly identified as none of the emotions proposed to participants.

Intended emotion
NEUTRAL

Door colour	HAPPINESS	SADNESS	FEAR	ANGER
WHITE	3,8	**4,4**	3,7	2,0
	2,9	3,4	3,2	2,8
YELLOW	**4,7**	3,4	3,0	2,1
	3,0	3,1	3,2	2,6
GREY	3,6	**4,0**	3,8	2,2
	2,7	3,1	2,9	2,8
RED	**4,2**	3,8	3,0	2,5
	3,1	2,9	2,9	2,6
PURPLE	**4,2**	3,6	3,5	2,5
	2,9	3,3	3,1	3,0

Table 1. Mean values (white rows) and standard deviation (light blue rows) for the neutral audio stimulus. Yellow cells represent the highest values related to the colour, whereas red numbers represent the highest score for the emotion.

To have a better visual representation of the distribution of data, five sets of box plots were created for each colour and perceived emotion when using the neutral stimulus (Fig. 6).

At first glance, it is clear that this time all the box plots present a right skewed distribution, meaning that participants concentrated their responses on the lower part of the 0-10 range. The most symmetrical box plot is yellow-happiness.

Almost in every case, the full range has been used, with the exception of white-happiness, white-anger, yellow-anger, grey-fear, red-sadness and red-anger.

The interquartile distributions are reasonably similar, with the exceptions of white-anger, which is more concentrated on 0-3 values and yellow-happiness, which is in the >2 - <7 range.

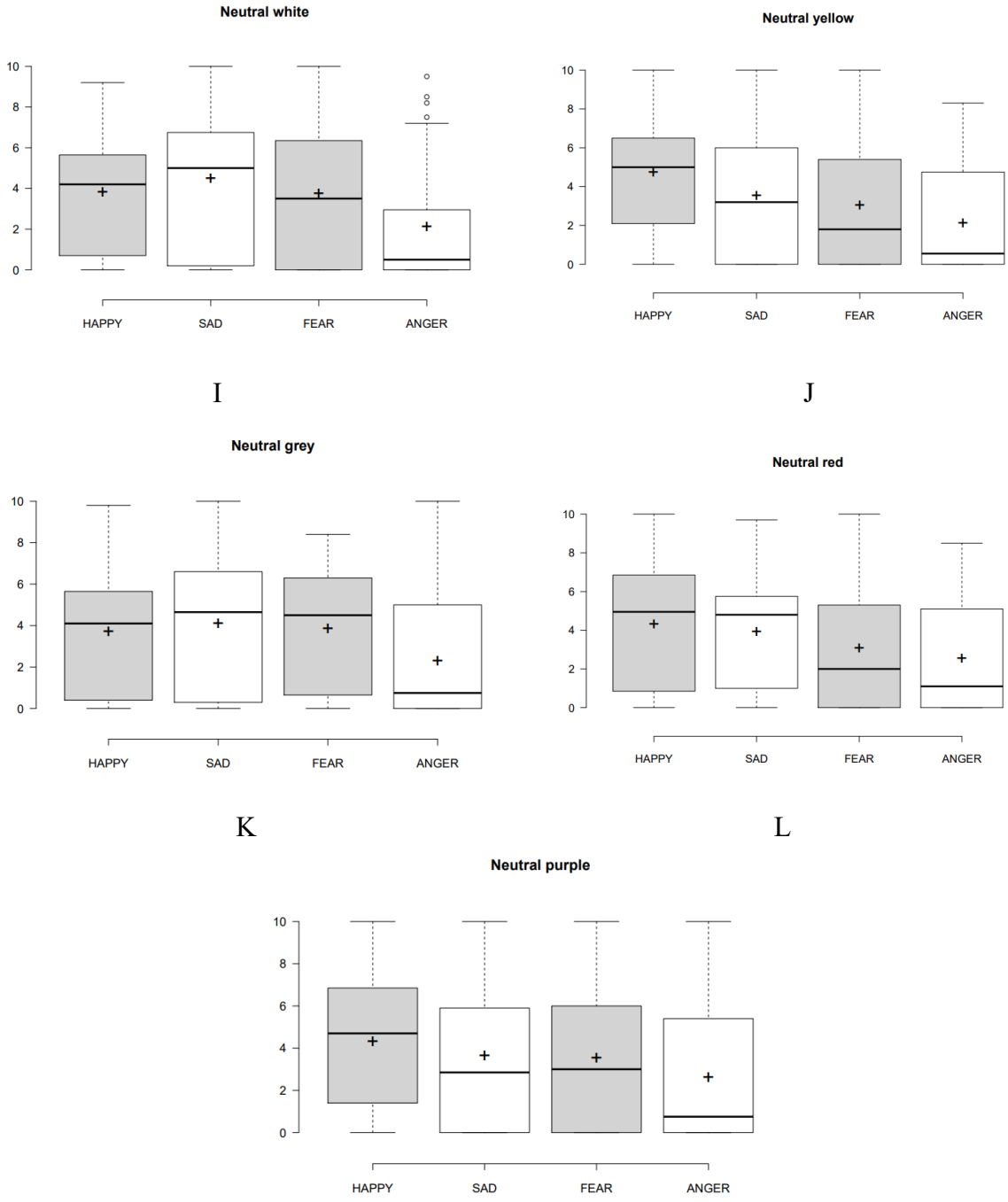

Figure 6. Box plots representing the distribution of data for each colour when using the neutral audio stimulus.

3.4 DISCUSSION

We can therefore conclude that, independently from the visual modality, that is colour, happiness, sadness and anger are the emotions that were mostly correctly identified. On average, also the neutral audio stimuli were correctly identified, since participants gave medium to lower scores to the videos, meaning that they did not identify a specific emotion among the ones presented.

On the other hand, fear tended to be confused with anger. This result is in line with that of Pauletto and Iop (2021).

Nevertheless, colours influenced the perception of the emotions, especially of happiness and sadness with yellow and grey, respectively, even if not in a clear way. Therefore, even in this case, there has been a predominance of the aural modality on the visual one, as found in McDonald et al. (2022) and Pauletto and Iop (2021). However, the visual channel still performs a weak effect on the overall perception of emotion. This can be especially seen in the already mentioned yellow-happiness and grey-sadness couples when presented with the respective intended emotional stimuli, but in the neutral ones as well. Indeed, even in this case, yellow was mostly associated with happiness and grey with sadness, confirming even further these colour-emotion associations.

By looking at the data, it is also possible to notice a remarkable variance among them. This aspect might depend on factors such as the mistaken understanding of the task to do due to the use of the English language for every participant and the gradual decrease of the attention threshold when doing the experiment.

Besides the problem of fear being confused with anger, not all the initial colour-emotion associations have been proved to be right. Indeed, anger surprisingly resulted to be more averagely associated with white.

Lastly, even if the study involved people coming from different countries, a statistical comparison of the data grouped by country has not been done due to the different distribution of the participants' nationality.

CONCLUSIONS

This work presented an in-depth focus on sound, starting from its definition to an analysis of its functions depending on its physical and perceptual parameters and on how, thanks to the different effects caused by a careful manipulation of such characteristics, sound can be applied to many different fields. This manipulation is aimed at eliciting a set of different responses and emotions in the listeners.

Among the various areas in which sound can be employed, particular attention has been paid to the world of cinema. Indeed, this field can, in a way, summarise all the other occasions in which the design of a sound is essential: from the use of more cinematic effects such as special sound effects that do not exist in real life and the design of a musical soundtrack, to the recording of everyday and artefacts sounds. For an innovative and accurate performance of these kinds of sound, the role of some sound workers has been highlighted. Among them, it was noteworthy to mention sound designers and Foley artists, who, together with other sound workers such as the sound editor and sound mixer, have completely changed the history of cinema. Through their work, they make the audience immerse themselves in the atmosphere of the movie they depict and, to do that, they need to design sounds that will elicit a set of different emotions.

Since the connection between sound and emotion is so tight, the present work examined more in depth the theories of emotions proposed in the past years. Indeed, to create the right atmosphere or to elicit a specific emotion through sound in the best way possible, it is of fundamental importance to understand how our auditory perception works and how different physiological but also physical responses are generated in our bodies when we hear a certain sound.

Connected to the theories of emotions, previous studies that analysed the relationship between colour and emotions have been reported. As discussed, even if those researches did not find a

consensual solution to the colour-emotion associations dilemma, they still proved that some emotions are better represented by some colours.

With these analyses on sounds, emotions and colours, it was possible to undertake the experiment presented in Chapter 3.

Our study drew inspiration from the one conducted by Pauletto and Iop (2021), but, since we used different audiovisual stimuli, different colour-emotion associations and a different emotion-scoring method, we expected to obtain different results from the mentioned study.

Our hypothesis was that the association between knocking sounds and emotions, already validated by that study, changes if the knocking sound is performed on a door of a different colour. More specifically, our initial hypothetical connections were happiness and yellow, red and anger, grey and sadness, and purple and fear. After the analysis of our data, only two of them were confirmed: happiness and sadness. Anger and fear, on the contrary, did not meet expectations. In the first case, the colour that received a higher score was white, followed by red and purple. Whereas in the second case, the audio stimuli were averagely confused with anger, independently of the colour. This last result is in line with that obtained by Pauletto and Iop (2021). However, red and purple are the colours that were mainly associated with fear.

Furthermore, for all the emotions investigated, the aural channel prevailed on the visual one, yet this last one still had an effect on the perception of the emotion, even if weakly. As seen, this aspect is more evident in the yellow-happiness and grey-sadness associations.

Future studies on this topic may include different methodological aspects. Firstly, to bypass the possible misunderstandings due to the use of a different native language, it could be useful to use facial expressions, proved to be universal by Ekman (2019, 1970). In alternative, since pictures of real people may influence the participants decision based on the gender depicted in the photo, black and white emojis could be used. In the current digital age, people are getting more used to express their feelings on text messages through emojis, so using these images on the experiment, might help

to even out results and create less data dispersion. In the case of emojis, removing colour is fundamental so that they do not interfere with the experiment target visual stimuli.

Secondly, we conducted the study cross-culturally. However the different nationalities were not equally distributed enough to proceed with a statistical analysis. For this reason, future studies might further investigate this aspect by gathering results from more homogeneous nationalities distributions.

Finally, since we selected the colour purple for fear and not black, which was reported to be the colour that participants of previous studies tended to associate mostly with this emotion, it could be interesting to see if, paired with this colour, fear becomes more recognisable, and therefore less confused with anger.

ACKNOWLEDGMENTS

I would like to thank Professor Rodà for introducing me to the world of sound with his class, for supervising this work and for introducing me to his international team of researchers.

I would then like to thank my family for always supporting me during my academic studies, making sure I always had everything I needed to pursue them.

Lastly, I would like to thank my classmates (and friends) for the times spent together and for making university life return to normality after the pandemic.

Lo studio rende liberi

Asia

REFERENCES

1. Abelin Å, (2004). Emotional McGurk effect in Swedish, *Speech, Music and Hearing, KTH, Stockholm, Sweden TMH-QPSR*, 46.
2. Abelin Å., (2007). Emotional McGurk effect - an experiment, *Proceedings of the 7th International Conference on Epigenetic Robotics*.
3. Alsius A., Paré M., Munhall K.G., (2018). Forty Years After Hearing Lips and Seeing Voices: the McGurk Effect Revisited, *Multisensory Research,* 31, pp. 111-144.
4. Ament V.T., (2009). The Foley Grail. The Art Of Performing Sound For Film, Games, And Animation, *Focal Press*.
5. Barahona-Rìos A., Pauletto S., (2020). Synthesising Knocking Sound Effects Using Conditional WaveGAN*, Proceedings of the 17th Sound and Music Computing Conference,* pp. 450-456.
6. Bedenko D., Kotanova A., Huidobro P.P., supervised by Pauletto S., (2020). Emotions And Sound Design: Exploring Emotions In Knocking Sounds, *n/a.*
7. Binzak Fugate J.M., Franco C.L., (2019). What color is your anger? Assessing color-emotion pairings in English speakers, *Frontiers in psychology*, pp. 1-17.
8. Bovo R., Ciorba A., Prosser S., Martini A., (2009). The McGurk phenomenon in Italian listeners, *ACTA Otorhinolaryngologica Italica*, pp. 203-208.
9. Bresin R., (2005). What Is The Color Of That Music Performance?, *Proceedings of the International Computer Music Conference -ICME 2005*, pp. 367-370.
10. Chion M., (1994). Audio-Vision. Sound on Screen, *Columbia University Press*, (translated by Claudia Gorbman, with preface by Walter Munch).
11. Crook T., (2001). Radio Drama. Theory and Practice, *Taylor and Francis Group.*

12. Cytowic R.E., (1989). Synesthesia And Mapping Of Subjective Sensory Dimensions, *Neurology 39*, pp. 849- 850.
13. Darwin C., (1872). The Expression of the Emotions in Man and Animals, *John Murray*.
14. D'Andrade R., Egan M., (1974). The Colors Of Emotions, *American American Anthropological Association*, 1(1), pp. 49-63.
15. Da Pos O., Green-Armytage P., (2007). Facial Expressions, Colours and Basic Emotions*, Colour: Design & Creativity*, 1(1), pp. 1-20.
16. De Gelder B., Vroomen J., (2000). The perception of emotions by ear and by eye, *Cognition and Emotion*, 14 (3), pp. 289-311.
17. De Götzen A., Trento S., (2011). Foley Sounds vs Real Sounds, *Sound and Music Computing Conference 2011*.
18. De Götzen A., Sikström E., Grani F., Serafin S., (2013). Real, Foley Or Synthetic? An Evaluation Of Everyday Walking Sounds, *Sound and Music Computing Conference*.
19. Demir Ü., (2020). Investigation Of Color-Emotion Associations Of The University Students, *Color Res Appl*., 47, pp. 871-884.
20. Donaldson L.F., (2014). The Work of an Invisible Body: The Contribution of Foley Artists to On-Screen Effort, *Alphaville Journal of Film and Screen Media*.
21. Doyle J., (2013). Subtlety of Sound: A Study of Foley Art, *Senior Honors Projects*, paper 333.
22. Ekman P., (1970). Universal Facial Expressions of Emotions, *California Mental Health Research Digest*, 8(4), pp. 151-158.
23. Ekman P., (1999). Handbook of Cognition and Emotion, *John Wiley & Sons Ltd,* pp. 45-60.
24. Ekman P., (2019). Universal Emotions, *Paul Ekman Group*, https://www.paulekman.com/universal-emotions/ .
25. Esposito E., Sini S., & Castagneto, M., (2018). *R*oman Jakobson, linguistica e poetica, *Ledizioni*.

26. Fagel S., (2006). Emotional McGurk Effect, *Speech Prosody*.
27. Fahlenbrach K., (2008). Emotions in Sound: Audiovisual Metaphors in the Sound Design of Narrative Films, *Projections*, 2(2), pp. 85-103.
28. Fernay L., Reby D., Ward J., (2012). Visualized Voices: A Case Study Of Audio-Visual Synesthesia, *Neurocase*, 18(1), pp- 50-56.
29. Garwood I., (2010). Review Essay: The Foley Grail. The Art Of Performing Sound For Film, Games, And Animation, *Animation: an interdisciplinary journal,* 5(3), pp. 369-374.
30. Gassner S., Wilson D., (nd). A Study in Synchresis, *Dept. of Applied Information Technology*, Göteborg University and Chalmers University of Technology.
31. Gilbert A.N., Fridlund A.J., Lucchina L.A., (2016). The color of emotion: A metric for implicit color associations, *Food Quality and Preference*, 52, pp. 203-210.
32. Goller A.I., Otten L.J., Ward J., (2008). Seeing Sounds And Hearing Colors: An Event-Related Potential Study Of Auditory-Visual Synesthesia, *Journal of Cognitive Neuroscience*, 21(10), pp. 1869-1881.
33. Görne T., (2019). The Emotional Impact Of Sound: A Short Theory Of Film Sound Design, *EPiC Series in Technology*, 1, pp. 17-30.
34. Gong R., Wang Q., Hai Y., Shao X., (2017). Investigation On Factors To Influence Color Emotion And Color Preference Responses, *Optik*, 136, pp. 71-78.
35. Güneş E., Olguntürk N., (2019). Color-Emotion Associations In Interiors, *Color Res Appl*., pp. 1-13.
36. Hanada M., (2017). Correspondence Analysis Of Color-Emotion Associations, *Color Res Appl.*, pp- 1-16.
37. Hess U., Thibault P., (2009). Darwin and Emotion Expression, *American Psychologist*, 64(2), pp. 120-128.

38. Hietanen J., Leppänen J., Illi M., Surakka V., (2004). Evidence For The Integration Of Audiovisual Emotional Information At The Perceptual Level Of Processing, *European Journal of Cognitive Psychology*, 16(6), pp. 769-790.
39. Hillman N., Pauletto S., (2014). The Craftsman: The Use Of Sound To Elicit Emotions, *The Soundtrack*, 7(1), pp. 5-23.
40. Houel M., Berg A., Arun A., Iop A., Barahona-Rìos A., Pauletto S., (2020) Perception Of Emotions In Knocking Sounds: An Evaluation Study, *Proceedings of the 17th Sound and Music Computing Conference*, pp. 419-425.
41. Hupka R.B., Zaleski Z., Otto J., Reidl L., Tabarina N.V., (1997). The Colors of Anger, Envy, Fear, and Jealousy: A Cross-Cultural Study, *Journal of Cross-Cultural Psychology*, 28, pp. 156-171
42. Jackson B., (2007). Fights, Footsteps and Thrills. Foley Adds Real-Life Drama to The Bourne Ultimatum, *MIX,* pp. 58-62.
43. Jonauskaite D., Abu-Akel A., Dael N., Oberfeld D., Abdel-Khalek A., Al-rasheed A., Antonietti J. Bogushevskaya V., Chamseddine A., Chkonia E., Corona V., Fonseca-Pedrero E., Griber Y., Grimshaw G., Hasan A., Jelena H., Hirnstein M., Karlsson B., Laurent E., Mohr C., (2020). Universal Patterns in Color-Emotion Associations Are Further Shaped by Linguistic and Geographic Proximity, *Psychological Science*, 31(10), pp. 1245-1260.
44. Kaya N., Epps H.H., (2004). Relationship Between Color And Emotion: A Study Of College Students, *College Student Journal*, 38(3), pp. 396–405.
45. Kaur A., (2020). A Link Between Colors And Emotions; A Study Of Undergraduate Females, *International Journal of Engineering Research & Technology (IJERT)*, 9(9), pp. 553-557.
46. Knight-Hill A., (2019). Sonic Diegesis: Reality and the Expressive Potential of Sound in Narrative Film, *Quarterly Review of Film and Video*.

47. Kurt S., Osueke K.K., (2014). The Effects of Color on the Moods of College Students, *SAGE Open*, pp. 1-12.
48. Li, A., Fang, Q., Jia, Y., Dang, J. (2013). Emotional McGurk Effect? A Cross-Cultural Investigation on Emotion Expression under Vocal and Facial Conflict, *Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data*.
49. Li X.F., Logan RJ., Pastore RE., (1991). Perception Of Acoustic Source Characteristics: Walking Sounds, *The Journal of the Acoustical Society of America 90,* pp. 3036-3049.
50. MacDonald J., (2018). Hearing Lips and Seeing Voices: the origins and development of the "McGurk effect" and reflections on audio-visual speech perception over the last 40 years, *Multisensory Research*, 31, pp. 7-18.
51. Manav B., (2007). Color-Emotion Associations and Color Preferences: A Case Study for Residences, *Wiley Periodicals*, 32(2), pp. 144-151.
52. Massaro D.W., (1998). Perceiving Talking Faces: from speech perception to a behavioural principle*, Book Reviews*, Cambridge, MA: MIT Press, pp. 130-135.
53. Massaro D.W., Egan P.B., (1996). Perceiving Affect From The Voice And The Face, *Psychonomic Bulletin & Review*, 3(2), pp. 215-221.
54. McDonald J., Canazza S., Chmiel A., De Poli G., Houbert E., Murari M., Rodà A., Schubert E., Zhang J.D., (2022). Illuminating Music: Impact of Color Hue for Background Lighting on Emotional Arousal in Piano Performance Videos, *Frontiers in Psychology*, 13.
55. McGurk H., MacDonald J., (1976). Hearing Lips and Seeing Voices, *Nature*, 264, pp. 746-748.
56. McGurk H., MacDonald J., (1978) Visual Influences On Speech Perception Processes, *Perception and Psychophysics*, 24(3), pp. 253-257.
57. Munhall K.G., Gribble P., Sacco L., Ward M., (1996). Temporal constraints on the McGurk effect, *Perception & Psychophysics*, 58(3), pp. 351-362.

58. Murray Schafer R., (1969). The New Soundscape. A Handbook for the Modern Music Teacher, *BMI Canada Limited*.
59. Nicholls M.E.R., Searle D.A., Bradshaw J.L., (2004). Read My Lips. Asymmetries in the Visual Expression and Perception of Speech Revealed through the McGurk Effect, *American Psychological Science*, (15), pp. 138-141.
60. Nijdam N.A., (2005). Mapping Emotion To Color, *n/a*.
61. Ortony A., Turner J.T., (1990). What's Basic About Basic Emotions?, *Psychological Review*, 97(3), pp. 315-331.
62. Pauletto S., Iop A., (2021). Perception of Emotions in Multimodal Stimuli: the Case of Knocking on a Door*, 18th Sound and Music Computing Conference*, pp. 233-237.
63. Plutchik R., (1962). The Emotions: Facts, Theories and a New Model, *Random House*.
64. Plutchik R., (1982). A Psychoevolutionary Theory Of Emotions, *Social Science Information*, 21(4,5), pp. 529-553.
65. Plutchik R., (2001). The Nature of Emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice, *American Scientist*, 89(4), pp. 344-350.
66. Rapan E., (2018). Shepard Tones and Production of Meaning in Recent Films: Lucrecia Martel's Zama and Cristopher Nolan's Dunkirk, *The New Soundtrack*, pp. 135-144.
67. Savavibool N., Gatersleben B., Moorapun C., (2016). The Effects Of Colour In Work Environment: A Systematic Review, *7th AicE-Bs2016Edinburgh*, 1(4), pp. 262-270.
68. Selfridge R., Pauletto S., (2021). Investigating the Sound Design Process, *KTH Royal Institute of Technology*.
69. Southworth M.F., (1967). The Sonic Environment Of Cities, *MIT*.
70. Stoet, G. (2010). PsyToolkit - A software package for programming psychological experiments using Linux, *Behavior Research Methods*, 42(4), pp. 1096-1104.

71. Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments, *Teaching of Psychology*, 44(1), pp. 24-31.
72. Suied C., Susini P., McAdams S., (2008). Evaluating Warning Sound Urgency With Reaction Times, *Journal of Experimental Psychology Applied*, 14(3), pp. 201-212.
73. Susini P., Houix O., Misdariis N., (2014). *S*ound Design: An Applied, Experimental Framework To Study The Perception Of Everyday Sounds, *The New Soundtrack*, 4(2), pp. 103-121.
74. Sutton. T.M., Altarriba J., (2015). Color associations to emotion and emotion-laden words: A collection of norms for stimulus construction and selection, *Behav Res*.
75. Tajadura-Jiménez A, Basia M., Deroy O., Fairhurst T.M., Marquardt N., Bianchi-Berthouze N., (2015). As Light As Your Footsteps: Altering Walking Sounds To Change Perceived Body Weight, Emotional State And Gate, *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems,* pp. 2943-2952.
76. Turchet L., Rodà A., (2017). Emotion Rendering in Auditory Simulations of Imagined Walking Styles, *IEEE Transactions on Affective Computing*, 8(2), pp. 241-253.
77. Valle A., (2011). C'era (Ancora) Una Volta Il West. Note A Partire Dal Suono In Una Sequenza Celeberrima, *La Valle dell'Eden*, 25-26, pp. 101-113.
78. Vernon P.E., (1934). Auditory Perception, *PE British Journal of Psychology*, 25(2), pp. 123-283.
79. Vitale R., Bresin R., (2008). Emotional Cues In Knocking Sounds, *Proceedings of the 10th International Conference on Music Perception and Cognition*.
80. Visell Y., Fontana F., Giordano B.L., Nordahl R., Serafin., Bresin., (2009). Sound Design And Perception In Walking Interactions, *J. Human-Computer Studies*, 67, pp. 947-95.
81. Whissell C., (1999). Phonosymbolism And The Emotional Nature Of Sounds: Evidence Of The Preferential Use Of Particular Phonemes In Texts Of Differing Emotional Tone, *Perceptual and Motor Skills*, 89, pp. 19-48.

82. Whitelaw M., (2008). Synesthesia And Cross-Modality In Contemporary Audiovisuals, *Senses & Society*, 3(3), pp. 259-276.
83. Wilms L., Oberfeld D., (2017). Color And Emotion: Effects Of Hue, Saturation And Brightness, *Psychological Research*.
84. Wright B., (2014). Footsteps With The Character: The Art And Craft Of Foley, *Screen*, 55(2), pp. 204-220.

VERSIONE ITALIANA (RIASSUNTO)

Questa tesi è un progetto sperimentale sul sound design, con l'obiettivo di esplorare il mondo del suono e le sue applicazioni artistiche e comunicative impegnate nella sollecitazione di emozioni. Quando si progetta una qualsiasi forma di comunicazione, che sia essa un contenuto informativo, una campagna di marketing o un prodotto multimediale, è importante essere a conoscenza di come esprimere determinati concetti, di quali effetti questi possono avere sull'audience e come questi possono essere attuati e rinforzati tramite gli strumenti della comunicazione. Perciò, questo studio mira a fornire qualche indicazione sul design di prodotti multimediali volti a generare reazioni emotive nell'audience, che può essere applicata in campi più artistici come l'industria cinematografica e dei videogiochi, così come nelle campagne di marketing.

Il suono e le emozioni sono apparentemente due concetti molto diversi tra loro.

Il suono è l'elemento che ha completamente rivoluzionato l'industria del cinema negli anni '20. Attraverso il suono è possibile non solo dare vita agli attori e agli oggetti sullo schermo, ma anche esprimere numerose altre informazioni. Infatti, porgendo particolare attenzione alla costruzione del suono, quest'ultimo può diventare un personaggio della storia, uno strumento per unire scene e periodi temporali diversi, così come uno strumento per creare il tono complessivo del film.

Tuttavia, i suoni nei film, come gli effetti speciali e la colonna sonora, non sono gli unici a raggiungere questo obiettivo. I suoni comuni, di tutti i giorni, sono mezzi di informazione più potenti di quanto uno possa immaginare. È infatti quest'ultima tipologia di suoni ad essere l'oggetto della presente ricerca. In particolare, viene discusso il potenziale che il suono ha, specialmente quando relazionato alla sollecitazione di emozioni.

Il campo delle emozioni è una delle più complicate, ma anche studiate, aree di ricerca scientifica. Non solo è difficile specificare cosa sono le emozioni, ma anche identificarle e classificarle è un

compito arduo. Molti studi recenti che indagano il tema delle emozioni si rifanno al set di emozioni base di Ekman (Ekman, 1999), anche se un accordo comune su questo argomento non è ancora stato trovato tra i membri della comunità scientifica.

Le emozioni possono essere sollecitate da stimoli diversi, che siano tattili, olfattivi, gustativi, visivi o uditivi. In particolare, l'interazione tra questi due ultimi sensi è l'oggetto di ricerca dell'esperimento presentato nel Capitolo 3. In ricerche precedenti, sia i colori che i suoni di tutti i giorni hanno dimostrato essere elementi capaci di sollecitare un set di diverse emozioni a seconda dei loro parametri: tinta, saturazione e luminosità da un lato, e tono, intensità, frequenza, timbro, durata, ecc., dall'altro.

La percezione delle emozioni attraverso questi due canali, visivo e uditivo, viene studiato tramite un esperimento che ha l'obiettivo di indagare l'influenza di certi colori, che si sono dimostrati essere associati a specifiche emozioni, su suoni del "bussare" performati emotivamente.

Prima di presentare l'esperimento (Capitolo 3), un'attenta analisi sul suono, sulla percezione uditiva e sulle emozioni è proposta nei primi due capitoli.

Più nello specifico, il primo capitolo esamina la definizione e le caratteristiche del suono, e come queste possono essere utilizzate per raggiungere diversi obiettivi: dal fornire informazioni e costruire una storia, alla sollecitazione di emozioni. Diversi esempi di come i suoni sono usati nell'industria del cinema (*A Quiet Place I, II, Dunkirk* e *C'era una volta il West*), così come su come sono percepiti nella vita di tutti i giorni, sono presentati all'interno del capitolo. Inoltre, viene raccontata l'importanza di figure lavorative nell'ambito del suono, il sound designer e l'artista Foley, assieme alla funzione fondamentale che queste hanno svolto, e svolgono tutt'ora, nello sviluppo degli elementi sonori di un film. Durante la presentazione di questi argomenti, particolare attenzione viene posta alla relazione tra suoni ed emozioni.

Infine, vengono riportati alcuni esempi di illusioni percettive generate dalla combinazione del canale visivo e di quello uditivo, come la sincresi teorizzata da Michel Chion (1994) e l'effetto McGurk.

Il Capitolo 2, invece, presenta la percezione uditiva ed emotiva da un punto di vista più psicologico. Le teorie di Charles Darwin, Robert Plutchik e Paul Ekman vengono presentate, dal momento che fungono da fondamento teorico per l'esperimento del Capitolo 3. Il primo studioso ha indagato l'origine delle risposte fisiologiche e motorie in risposta ad uno stimolo emotivo; il secondo ha proposto una lista di emozioni primarie insieme ad un'analogia tra emozioni e colori; mentre il terzo ha proposto una lista di emozioni base, alcune delle quali sono oggetto di studio della presente ricerca.

Dopodiché, partendo dall'analogia adoperata da Plutchik, il tema delle emozioni viene messo in relazione a quello dei colori, presentando diversi studi che hanno indagato queste associazioni. Questi studi sono stati di grande importanza per la scelta dei colori utilizzati in questo esperimento.

Infine, il Capitolo 3 presenta l'esperimento condotto. La definizione della struttura del questionario e altre questioni metodologiche sono state fatte in collaborazione con il Professor Emery Schubert e Diana Zhang dell'Università del New South Wales Sydney e del Dottor Anthony Chmiel della Western Sydney University.

Questo esperimento prende ispirazione dalla ricerca condotta da Pauletto e Iop (2021), in cui viene investigata la percezione delle emozioni attraverso stimoli multimodali, ovvero suoni del "bussare" presentati insieme ad immagini di porte di diversi colori e diversi materiali. Un altro studio che è servito di ispirazione per quello presente, dal punto di vista metodologico, è quello condotto da McDonald et al. (2022). Entrambi questi studi hanno riportato una prevalenza degli stimoli uditivi rispetto a quelli visivi.

Tuttavia, lo studio presente si differenzia dai due sopracitati per la costruzione degli stimoli audiovisivi e per il criterio di valutazione delle emozioni.

Le nostre ipotesi di associazioni colore-emozione sono:

- Giallo = felicità
- Rosso = rabbia
- Viola = paura
- Grigio = tristezza

- Bianco = neutro.

In totale, sono stati creati 25 stimoli audiovisivi che consistono in un video di 3 secondi di una mano che bussa su una porta di colori diversi. Gli stimoli audio sono stati presi da un dataset professionalmente registrato e già validato (Pauletto & Iop, 2021; Barahona-Ríos & Pauletto, 2020) e i dati sono stati raccolti tramite la piattaforma online PsyToolkit (Stoet, 2010, 2017).

Ogni stimolo audiovisivo era accompagnato da 4 slider, ognuno dedicato ad un'emozione (felicità, tristezza, paura e rabbia), costruiti su una scala continua da 0 (per niente d'accordo) a 10 (molto d'accordo). I partecipanti, una volta guardato il video, dovevano muovere gli slider attribuendo così un punteggio all'emozione percepita.

In totale, 60 partecipanti di età compresa tra i 18 e i 66 anni e provenienti da Paesi diversi hanno correttamente completato il questionario (33 donne, 25 uomini, 1 non-binario e 1 ha preferito non specificare il genere; 51 dall'Italia, 2 dagli USA, 2 dalla Turchia, 1 dalla Finlandia, 1 dall'Iran, 1 dal Marocco e 1 dalla Russia).

Alla fine del questionario sono state poste alcune domande riguardanti il fenomeno della sinestesia ed è emerso che le donne sono più soggette a questo tipo di fusioni sensoriali.

Dall'analisi dei dati degli stimoli audiovisivi è emersa una gran dispersione dei dati, probabilmente dovuta ad un'interpretazione errata del compito o all'abbassamento della soglia d'attenzione durante lo svolgimento dell'esperimento. Solo due delle ipotesi iniziali sono state confermate: felicità-giallo e tristezza-grigio. A sorpresa, rabbia è stata associata maggiormente al colore bianco e paura è stata spesso confusa con rabbia, indipendentemente dal colore. Quest'ultimo risultato è in linea con quello ottenuto da Pauletto e Iop (2021). Tuttavia, rosso e viola sono i colori maggiormente associati a questa emozione.

Inoltre, come nei due studi presentati in precedenza, la modalità uditiva ha prevalso su quella visiva, la quale ha comunque avuto un effetto sulla percezione delle emozioni, anche se debole, specialmente visibile in felicità-giallo e tristezza-grigio. Queste due coppie di colori-emozioni sono

state confermate anche dall'uso degli stimoli neutri. Questi stimoli hanno ricevuto un basso punteggio per ogni emozione, suggerendo che i partecipanti hanno correttamente individuato questa emozione come non appartenente a quelle loro presentate negli slider. Tuttavia, gli stimoli gialli e grigi hanno ricevuto punteggi maggiori per le emozioni rispettivamente ipotizzate.

Alcuni studi futuri possono includere aspetti metodologici differenti. Innanzitutto, il questionario è stato redatto in lingua inglese per coinvolgere la maggior parte di nazionalità possibili, ma questo può aver creato qualche problema di comprensione del compito da svolgere. Per aggirare questa questione, al posto di usare stimoli verbali come "anger", "happiness", "fear" e "sadness", si potrebbero utilizzare le facce di Ekman (2019, 1970) o, in alternativa, usare emojis in bianco e nero rappresentanti queste quattro emozioni.

Dopodiché, anche se questo studio è stato condotto cross-culturalmente, la diversa distribuzione delle nazionalità dei partecipanti non ha reso possibile un confronto statistico. Perciò, studi futuri possono mirare ad ottenere una distribuzione più omogenea per indagare queste associazioni anche attraverso nazionalità diverse.

Infine, in questo studio, una delle ipotesi era paura-viola, nonostante nero fosse il colore che ricerche precedenti hanno maggiormente sottolineato come associato a questa emozione. La scelta del viola è stata fatta in quanto nero era spesso associato anche alla tristezza, fattore che avrebbe potuto creare ulteriore confusione, e perché viola risultava essere un colore unicamente correlato a questa emozione. Studi futuri possono dunque provare ad ipotizzare la combinazione paura-nero per vedere se questa emozione viene confusa in maniera minore con rabbia.