1 2 9 0

UNIVERSIDADE Đ
COIMBRA

Francisco Azevedo Gomes de Almeida e Sousa

Jack-in-the-Mug
An interface to assist in the production of Foley sound effects

Dissertation in the context of the Master in Design and Multimedia advised by Professor Fernando Amílcar Bandeira Cardoso and Professor Tiago José dos Santos Martins da Cruz and presented to the Department of Informatics Engineering of the Faculty of Sciences and Technology of the University of Coimbra

Outubro de 2020

Acknowledgements

Aos professores Amílcar Cardoso e Tiago Cruz, pela dedicação mesmo durante estes longos meses.

Ao Branko Neskov e ao António Porém Pires, pela sua disponibilidade e ajuda imprescindível.

Aos meus amigos, por continuarem presentes em todas as etapas da minha vida.

À Bea, Filipe, Lousada e Zé Pedro, pelo apoio e companhia.

E um especial obrigado à minha família que continua a crescer, pelo apoio e dedicação e por serem responsáveis pela pessoa que sou hoje.

Faculty of Sciences and Technology
Department of Informatics Engineering

Jack-in-the-Mug

An interface to assist in the production of Foley sound effects

Francisco Azevedo Gomes de Almeida e Sousa

Dissertation in the context of the Master in Design and Multimedia advised by Professor Fernando Amílcar Bandeira Cardoso and Professor Tiago José dos Santos Martins da Cruz and presented to the Department of Informatics Engineering of the Faculty of Sciences and Technology of the University of Coimbra

October 2020

1 2 9 0

UNIVERSIDADE Ð
COIMBRA

This page is intentionally left blank.

Abstract

During the observation of a movie or other audiovisual media, the image stands out more than the sound, but the sound contains details that cannot be transmitted by the image. Sound gives relevant information on off-screen actions and has the capability to immerse the audience in the scene's atmosphere. The necessity of having a good sound quality is respected and appreciated by the film industry, being sound professionals recognized through various awards.

The sounds recorded while a movie scene is being filmed are rarely present in the final montage. Instead, the sounds present in a movie, such as the characters sounds and the environment sounds, are created and composed by professionals through various techniques. One of the most predominant techniques is called Foley, homonym to one of the first practitioners - Jack Foley. The method consists on the recreation and synchronization of sounds with visual assistance of the scene by specialized artists in an audio production studio. The artists use a broad collection of props to recreate the intended sound.

Jack-in-the-Mug is an interactive physical interface to support the recreation and synchronization of sound effects in a scene of a movie, an animation in a video game or other purposes in the audiovisual media context. The interface's objective is to facilitate the production of Foley sound effects by allowing the user to generate different sounds with a single object and by emulating the conventional method with the different props.

The present dissertation reports the conceptualization, implementation and experimentation of this interface for Foley sound production.

Keywords

HCI; Physical Interface; Sound; Foley Sound Effects; Sound Design; Art; Cinema

This page is intentionally left blank.

Resumo

Durante a observação de um filme ou outro conteúdo audiovisual, a imagem sobressai mais do que o som, mas este contém detalhes que não podem ser transmitidos através da imagem. O som oferece informação pertinente de ações fora de cena e tem a capacidade de imergir no público a atmosfera da cena. A necessidade de uma boa qualidade de som é bastante apreciada e respeitada pela comunidade cinematográfica, inclusivamente os seus profissionais são reconhecidos através de prémios.

Os sons gravados durante a filmagem de uma cena de um filme raramente estão presentes na montagem final. Contrariamente, os sons presentes num filme, como os sons das personagens e os sons do ambiente, são criados e compostos por profissionais através de várias técnicas. Uma das técnicas mais predominantes é chamada de *Foley*, homónima a um dos primeiros praticantes - Jack Foley. O método consiste na recriação e sincronização de sons por artistas especializados com o apoio visual da cena num estúdio de produção áudio. Os artistas utilizam um leque extenso de adereços para recriar os sons pretendidos.

Jack-in-the-Mug é uma interface física interativa de suporte à recriação e sincronização de efeitos de sons para uma cena de um filme, uma animação de um videojogo ou outros propósitos num contexto audiovisual. O objetivo da interface é facilitar a produção de efeitos sonoros *Foley*. Para tal, a interface permite o utilizador gerar diferentes sons através de um único objeto e emula o método convencional de interação com os diferentes adereços.

A presente dissertação reporta a conceptualização, implementação e experimentação da interface de produção de sons *Foley*.

Palavras-Chave

IHC; Interface Física; Som; Efeitos de som *Foley*; Design de Som; Arte; Cinema

This page is intentionally left blank.

Contents

This page is intentionally left blank.

Acronyms

ADC Analog-to-Digital Converter. 11

DAC Digital-to-Analog Converter. 11

DAW Digital Audio Workstation. 27, 33, 37, 42, 44

IMU Inertial Measurement Unit. 48, 56

MIDI Musical Instrument Digital Interface. 27–29, 41, 44

OSC Open Sound Control. 28, 33, 41, 44, 62

This page is intentionally left blank.

List of Figures

This page is intentionally left blank.

List of Tables

This page is intentionally left blank.

Chapter 1

Introduction

Cinema is one of the most popular arts which moves innumerous fans to the movie theatres. The seventh art, since it was introduced by the Lumière brothers to the world, has undergone major changes due to technological innovation. The technological advancements focused on the enhancement of the recording devices (e.g., cameras and microphones) and the reproduction devices (e.g., television and sound speakers). Also, new possibilities, such as realistic visual and sound effects emerged with the rise of digital methods.

The end of the silent era transformed the sound into an important medium to create quality cinematographic products. The sound started to be applied for different purposes into diverse modes, improving the experience of the audiovisual artifact's viewer. Consequently, the new possibility to integrate sound in cinematographic content caused the creation and exploration of new methods to develop the soundtrack of a movie. The method from which the dissertation lay emphasis is the production of Foley sound effects.

Foley is one of the most predominant techniques used in film production to develop sound effects, consisting in the recreation of sounds with visual assistance of the image where the sound is intended to. The method straightforwardly synchronizes the sound to the image by the action of a Foley artist that reproduces the sound of various events on the movie's scene through several props, such as the characters footsteps, the clothing movement, and more specific sounds.

As new audiovisual artifacts forms appeared, the Foley technique was adopted into areas other than the cinema. In particular, the more recent area that came with the digital world, the video games, appropriated the Foley sound effects.

Considering the necessity of the Foley artists to resort to a diversity of physical objects to recreate a credible soundtrack complemented with the new technological resources and respective progress of the human-computer interaction field of study, the following question occurred:

- Is it possible to create some kind of physical interface with a computer system that could assist in the production of Foley sound effects?

It is upon this question where the present dissertation unfolds.

1.1 Motivation

Sound is an important medium to transmit information in audiovisual or audio artifacts. It allows the listener to perceive details of the environment that surrounds a scene and emotions of a movie's character or event. For instance, the clock's ticking sound in an excerpt of a movie may either represent the existence of a clock in a closed space and/or suggest the feeling of anticipation from the character. The multitude of impacts a sound can have in a movie forced professionals to explore new approaches and techniques to improve the quality of movies.

Foley is a technique that facilitates the synchronization between the sound effects and the image. The technique also allowed to incorporate realistic sounds and give freedom in the selection of sounds limited by the imagination and creativity of the Foley artist. This technique dates back to the 1920s and continues to this day as one of the main methods of sound film production. With the birth of the digital world, new mediums of expression emerged and started to require and adopt this method to improve the quality of the products (e.g., video games). The technologic evolution affected a variety of areas in audiovisual production with the insertion of new digital methods and technological resources. Nevertheless, Foley did not suffer a great impact, only affecting the quality of the recording devices. One of the main incentives to create a physical interface with a computer system that can assist in the production of Foley sound effects is this scarcity of explorations in the area.

Nowadays, the production of Foley sound effects requires access to plenty of materials, such as natural resources leading to their wastage. Also, the Foley artists resort to common physical objects that start to wear out and degrade. The interface we propose aims to reduce the waste of natural resources, such as water, and access to sounds of objects without concerning the Foley artist of the wear of the object to affect the sound quality. In addition, the confinement of the interface into one object reduces the necessity of big storage units for all of the objects that the technique demands.

Extending the topic of technology, the appearance of video sharing web platforms such as the Youtube phenomenon generated countless video creators. The interface provides amateurs with the possibility to improve the sound quality of their content and adopt the technique into their workflow since it does not require a Foley studio full of props.

The main motivation of the project focuses on the exploration of new techniques to incorporate sound effects in audiovisual content, allied with the research on the development of Human–computer interaction interfaces.

1.2 Objectives

The main objective of this dissertation is the development of a sound interface to offer a new method of synchronization and edition of sound for Foley artists, either for cinema, video games or other audiovisual media. The physical interface, Jack-in-the-Mug, was created with the intention to be adopted by Foley artists in their workflow.

The development of Jack-in-the-Mug aims to offer a scientific contribution to the design domain by exploring possibilities and limitations of sound interaction and sound design in the production of sound effects. Additionally, the interface opens opportunities to be explored in the artistic domain, such as sound art, performative art and music.
que, para além da aplicação ao Foley, a interface abre também possibilidades de exploração no campo das artes performativas

1.3 Context

The sound is one of the most powerful mediums for expression of emotions and feelings and for transmission of information. Therefore, sound is a crucial element for the quality of audiovisual media.

New horizons have been discovered with the increasing availability of miniaturized computing resources and wireless technologies. One of the horizons is the new means of interaction with sound including a novelty of instruments, extensions and interfaces dedicated for sonic expression.

The significance of sound allied with the insertion of new technological possibilities originated the concept of the dissertation which is the creation of a sound interface based on instrumented physical objects. This concept was then applied for the Foley method to develop sound effects in audiovisual media, more specifically, the creation of a sound interface to auxiliate the production of Foley sound effects. The purpose and necessity of the Foley technique stick with the improvement of sound quality in cinematographic products or other audiovisual media.

The interface is called Jack-in-the-Mug, a wordplay of the words "Jack", the first name of the inventor of the Foley technique, "Mug", because of the resemblance of the physical aspect of the interface, and Jack-in-the-box, a cultural well-known children's toy.

With this intent, the dissertation derived from the study areas of sound, including its presence in the arts and sound design, and from the study areas of interface design and HCI.

1.4 Document structure

This thesis is organized in seven chapters. The current chapter, **Introduction**, introduces the topic of the thesis, accompanied by its motives, objectives and contextualization.

The second chapter, referring to the **State of the Art** of the addressed areas where the project inserts itself, contains a brief summary of the research conducted on the scientific domains and projects developed related to the project. The chapter introduces the sound with a theoric component of it's different perspectives and of the digital sound. The next section reports the relevancy and example of sound in the artistic domain. The third section contains a brief explanation of the Foley technique, followed by the research conducted on the topics: sound design and sound interfaces. Finally, a survey of projects related to the one presented is displayed and analyzed.

Following the State of the Art, chapter three, **Methodology**, explains in detail the scientific procedure used for this thesis.

In the fourth chapter, **Design Overview**, the essential process of conceptualization of the interface is explained, including the investigation of potential users and their needs, and final definition of the attributes of the interface.

Succeeding this definition, the **Prototyping** of the interface initiates. This chapter describes the implementation of the device in two parts: the electronic component and the physical model. The fifth chapter concludes with the information of the system framework.

The sixth chapter, **Results and Observation**, reports the results obtained from the interviews conducted with professionals and the usability tests, including an implementation process for the graphical interface and sound library required for the tests. The interviews consisted of the explanation of the interface concept followed by a questionnaire. The two professionals interviewed were António Porém Pires and Branko Neskov. The chapter ends with a reflection upon the obtained results in relation to the objective of the thesis.

Lastly, the seventh chapter, **Conclusion**, concludes this thesis by responding to the initial question: "Is it possible to create an interface to assist in the production of Foley sound effects?" and is complemented with some suggestions for future work.

This page is intentionally left blank.

Chapter 2

State of Art

2.1 Sound

In order to develop musical instruments, virtual or non-virtual, it is important to understand the concepts of production, propagation and perception of sound (Henrique, 2002). Listening is the primary interface, a path of information exchange, between people and the environment (Truax, 2001). Even though hearing is the first sense that develops in the womb, by the fifth month of pregnancy (Sonnenschein, 2001), it is startling how limited our vocabulary is when we try to explain a sound we heard (Ballou, 2013).

Sound is characterized by several basic phenomena (Everest & Pohlmann, 2015). When characterizing a sound we must take in count the context at hand since there is a duality of the definition of sound (Everest & Pohlmann, 2015). It can be considered as a wave motion in an elastic medium such as air or water - a stimulus - or as an excitement of the auditory mechanism that results in its perception by the brain - a sensation (Everest & Pohlmann, 2015; Kristjánsson, 2016). These two views differentiate the objective quantities of frequency, level, spectrum, etc., that can be measured through a physics perspective, from subjective quantities of pitch, loudness, timbre, etc., that can't be measured, since they are auditory perceptions in our heads, and a sound is characterized through psychoacoustics perspective (Ballou, 2013; Truax, 2001). This distinction allows us to look for changes in the objective stimulus that may influence a sound's sensation (Truax, 2001) due to the fact that a subjective quantity can be influenced from various objective quantities (Ballou, 2013). The interaction between these physical aspects of sound and our brain's response to them is not identically heard by two different individuals as there are no two identical persons (Ballou, 2013).

2.1.1 Physics Perspective

Sound, in a physics perspective, is a physical phenomenon of energy transmission (Farnell, 2013). The transmission of energy from one place to another is made through waves and move by propagation through a medium (Farnell, 2013). The source provokes vibrations in the surrounding medium and these vibrations spread across the medium, carrying the disturbance in the form of oscillating and propagating pressure waves - the sound wave (Lorinc, 2015). The medium must be elastic, such as gases, liquids or gases (Everest & Pohlmann, 2015), as the elasticity of the medium allows the possibility of sound vibrations (Farnell, 2013), and without it, a sound cannot be propagated (Everest & Pohlmann, 2015)(Kristjánsson, 2016). The wave movement carries energy and quantity of movement, but not matter, the energy is transmitted but the medium itself is not transported (Henrique, 2002).

The medium consists of a material or materials between two points in space, which carries vibrations that are created by the sound wave spreading outwards travelling through the medium (Farnell, 2013). Sound waves originate disturbances in the medium's density caused by pressure, velocity and temperature variations (Henrique, 2002). The medium affects the velocity of the sound wave (Henrique, 2002), for example, the molecular structure of a denser material where the molecules are closer together makes the transmission of sound faster (Farnell, 2013). It also depends on the temperature of the material, for instance, the speed of sound in the air is faster in a warmer environment than in a cooler one (Giancoli, 2016). The sound propagates causing fluctuations of pressure on the material and the waves generated by these fluctuations are called compression and rarefaction waves, or pressure waves (Henrique, 2002).

Waves can be classified according to the relationship between the direction of the vibratory movement of the medium particles and the direction of the propagation (Henrique, 2002). Sound waves, in an acoustic context, are considered as longitudinal in the air or other fluids, and traverse and longitudinal in solids (Farnell, 2013). Sound waves are longitudinal waves of compression and rarefaction that are responsible for causing a hearing sensation (Henrique, 2002). Waves are defined as longitudinal when the particles of the material oscillate according to the direction of the wave propagation that is the same as the impulse generator (Henrique, 2002). In solid materials, the molecules oscillate in the same direction as the wave in a longitudinal form, as it does in liquids and gases, but also move weakly in a transverse direction (Henrique, 2002).

Physical Properties

Cycle:
In oscillatory movements, the cycle is the path taken from where the movement repeats the same characteristics (Henrique, 2002). It represents the motion of a point that is displaced, returning through its rest point, overshooting, and then coming back to its rest point once again (Farnell, 2013). It is an abstract number so it doesn't have a unit (Henrique, 2002).

Wavelength:
The wavelength of a wave is the distance that a wave travels to complete one cycle and is defined by λ (Everest & Pohlmann, 2015) and quantified in meters (Farnell, 2013). It's measured between the closest two points with the same displacement moving in the same direction (compression or rarefaction) (Farnell, 2013).

Period:
The period is the time length that a wave takes to complete one cycle (Ballou, 2013). It's represented by the letter T and is measured in seconds (s) (Farnell, 2013).

Frequency:
Frequency (f) is the number of completed cycles per unit of time, being measured in cycles per second (cycles/s) or Hertz (Hz) (Kristjánsson, 2016). Hertz is defined as one cycle per second (Kristjánsson, 2016). If the frequency is the number of cycles per second, then it depends on the period, being one the inverse of the other (Henrique, 2002). Higher frequency equals lower period and vice-versa (Henrique, 2002).

Human hearing can only detect sounds with frequencies between 20 Hz and 20.000 Hz (Giancoli, 2016). However, this interval can be different from one individual to another (Giancoli, 2016).

Amplitude and Level:
The amplitude of a sound wave is the magnitude calculated on a certain point. It can present negative or positive values relative to the rest point, where the amplitude is equal to 0. On sound waves, positive amplitude corresponds to the compression of the medium and negative amplitude corresponds to the rarefaction of the medium (Farnell, 2013). Compression is the section in a wave cycle where the wave is pushing out against the next

molecules and rarefaction is the opposite (Lorinc, 2015).

Intensity is the quantity of energy transported by a wave (Giancoli, 2016). The sound pressure level or sound intensity level is proportional to the amplitude of the sound (Farnell, 2013). Sound levels are expressed in decibels (dB) (Everest & Pohlmann, 2015). Which is a logarithmic scaling unity calculated according to the amplitude of the sound wave because the response of our ear to sound disturbances is not linear but rather logarithmic (Henrique, 2002).

Waveform:
The waveform is the shape of the wave (Lorinc, 2015), is the graphical representation of the amplitude in the time domain (Cipriani & Giri, 2010). There are different types of shapes like sine waves, triangle waves, square waves and sawtooth waves (Lorinc, 2015).

The waveform differs if the sound is a pure sound or a complex sound. While a pure sound is sinusoidal, represented by a sine function or a cosine function (Cipriani & Giri, 2010), a complex sound is reducible to components of sines (Everest & Pohlmann, 2015). All complex waves can be analyzed as a superposition of sinusoidal waves (Giancoli, 2016).

Spectrum:
The waveform is a representation of amplitude changes through time (Cipriani & Giri, 2010), although there are other representations with different domains that allow observing specific properties in a more efficient way (Henrique, 2002). The sound spectrum is a graphical representation where the amplitude of the components is shown in a frequency domain (Cipriani & Giri, 2010). This spectral representation highlights the frequencies in which the sound contains more energy, showing the relevancy of each frequency (Henrique, 2002).

Envelope:
The energy difference in a wave is characterized by the envelope. The envelope is defined by 4 phases: Attack, Decay, Sustain and Release (Farnell, 2013). These phases can occur in various properties of the sound (Farnell, 2013), thus representing a combination of amplitude and wavelength (Lorinc, 2015). The envelope characterizes if a sound is more impulsive or more reverberant (Sonnenschein, 2001).

The Attack occurs at the beginning of the sound reproduction, where the transmitted energy goes from null to the maximum. Following this spike of energy, the Decay occurs and the energy charge falls until it reaches a constant equilibrium - Sustain phase. Lastly, the Release occurs, being the phase when the wave stops producing energy and just transmits the stored energy. The format of the envelope is not always regular as some of the phases may not manifest (Farnell, 2013).

Speed:
The speed of a sound wave remains constant during its lifetime on the same material and it depends on the properties of the material (Farnell, 2013). The waves can change its speed in case there is a transition of one material to another or a change in the medium's temperature (Farnell, 2013).

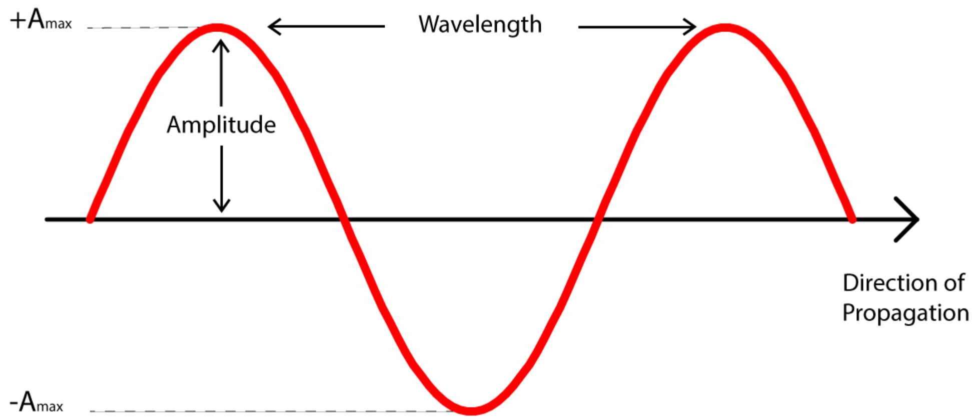

Figure 2.1: Properties of a sound wave.

2.1.2 Psychoacoustics Perspective

Upon the arrival at the ear, the sound becomes the subject of study for psychoacoustics (Truax, 2001). Psychoacoustic is the study of the human perception of sound, covering the physical structure of the human ear including its structure and function, as well as the human perception of the sound (Ballou, 2013). This science investigates the relation of measurable objective quantities of waves with the perception of sound and subjective phenomena (Ballou, 2013)(Farnell, 2013).

In psychoacoustic, sound is a sense, a human experience. Therefore, studying sound takes into account how it is perceived, how we extract features and meaning from it, and how we categorize and memorize it (Farnell, 2013). Since psychoacoustic views sound as perception and no two persons are identical there are large individual differences (Ballou, 2013).

A stimulus wave when reaching the ear triggers a set of phenomenons that result in electrical discharges that are sent to the brain and create the sensation of a sound (Everest & Pohlmann, 2015). Psychoacoustics investigates this process by the auditory system to extract information, the process called hearing (Truax, 2001).

The human ear is usually considered in three parts: the outer ear, the middle ear and the inner ear (Ballou, 2013). The outer ear is composed by the pinna and the auditory canal (Everest & Pohlmann, 2015). The pinna's job is to amplify the ambient sounds and funnel them into the auditory canal (Sonnenschein, 2001). The sound waves travel through the auditory canal to the eardrum, which vibrates as the wave reaches it (Giancoli, 2016). The middle ear is composed by the hammer, anvil, and stirrup which transfer the eardrum's vibrations to the inner ear. As the vibrations reach the oval window, all the components of the inner ear transform the vibrational energy of the sound wave into electric energy, which is then sent to the brain (Giancoli, 2016). The energy is converted into information that the brain can use, creating the sensation of the sound (Sonnenschein, 2001).

Psychoacoustics Properties

Loudness:
Loudness is a subjective perception of a sound's intensity (Sonnenschein, 2001) and varies among listeners, unlike level or intensity, which are physical or objective quantities (Ballou, 2013). Loudness is a sensation that is related not only to the intensity of the sound wave (Giancoli, 2016) but also depends on the frequency and other factors (Farnell, 2013). Since the ear is not sensitive in the same way to all the frequencies, the sensation of equal loudness of sounds with different frequencies requires different intensities between those sounds (Giancoli, 2016).

Some other factors, along with intensity or frequency, affect the perception of loudness. For example, if multiple sound sources reproduce the same sound at the same time, the sound perceived seems louder than if it was only one source (Sonnenschein, 2001). Sonnenschein defines that the intensity of a sound perceived by a person is influenced by how far the source is and the relevancy of the sound in our lives (Sonnenschein, 2001).

Pitch:
The perception of the sound's pitch allows to distinguish a high-pitched sound from a lowed-pitched sound (Cipriani & Giri, 2010). In a more common approach, the sounds we perceive as pitch are musical tones, being a pure tone or a complex tone (Ballou, 2013). If a sound has pitch it can be used to make a melody (Ballou, 2013).

In the same way intensity affects the perception of loudness, the pitch's main influential physical property is frequency (Farnell, 2013), as the parameters of pitch follow the scope of low to high frequencies (Sonnenschein, 2001). The frequencies are examined by a membrane in the inner ear that functions as a frequency analyzer (Ballou, 2013).

However, frequency is not the only physical property of sound that affects pitch (Everest & Pohlmann, 2015). For example, the higher intensity in low-frequency sounds generates a perception of lower tones than in a lower intensity environment, while in high frequency sounds higher intensity generates a perception of higher tones (Everest & Pohlmann, 2015). Also, pitch recognition is a process that integrates the two ears. When two sounds with different pitches are presented to each ear independently the listener will not hear two pitches, but a combination of these two sounds (Ballou, 2013).

Timbre:
According to Farnell, "Timbre describes our perception of sound color" (Farnell, 2013). Timbre is the term used to describe the tonal quality of complex sounds (Everest & Pohlmann, 2015)(Lorinc, 2015), is the quality of a sound that distinguishes different types of sound production (Lorinc, 2015). It allows the listener to detect a difference between two independent musical instruments while playing the same note (Everest & Pohlmann, 2015)(Ballou, 2013). This way, timbre is what makes a musical sound unique, even if it has the same pitch and loudness of another musical sound (Lorinc, 2015), whereas waveform and envelope are the physical attributes that influence the perception of timbre (Lorinc, 2015).

Localization:
The perception of the location of the sound source is captured by the external ear where the information of a sound's direction is encoded and transmitted to the brain for interpretation (Everest & Pohlmann, 2015). Usually, we assume the loudest source to be the closest one (Farnell, 2013).

Rhythm and Organization:
Rhythm is the characterization of sound through time (Sonnenschein, 2001). Sounds can be rhythmic, which leads to the fact that this predictability can provide either tranquillity

and assuredness or pain to the listener, or irregular, where the irregularity of the sound can alert, frighten, confuse or produce laughter on the listener (Sonnenschein, 2001).

Rhythm depends on the listener's physical and biological background, while the organization depends, also on the listener's physical and biological background as well as the listener's social and educational background (Sonnenschein, 2001).

The organization is how orderly the sound signals are to the human ear, being classified as organized or chaotic (Sonnenschein, 2001). A great example is how unintelligible and chaotic a foreign language seems to be until it is learned and becomes organized (Sonnenschein, 2001).

2.1.3 Digital Sound

The digitalization of sound is possible thanks to transducers. Transducers are machines that convert changes in one form of energy to changes in another kind of energy (Farnell, 2013). In the case of sound, transducers convert air pressure into an electrical form, the audio signal, or electrical variations into sound (Farnell, 2013). Some examples of worldwide known transducers are the microphones and the loudspeakers. Microphones are devices that convert acoustical energy into electrical energy (Ballou, 2013). Loudspeakers are devices that perform inversely, converting electrical energy into acoustic energy (Ballou, 2013). This electrical representation of sound is termed analog (Truax, 2001), where the waveform of the electrical signal is analogous to the original sound (Henrique, 2002), since the audio signal is a continuously changing voltage that corresponds to variations in sound pressure (Truax, 2001).

Currently, it is impossible to convert sound from its acoustic form into a digital form, or vice-versa, without the analog signal stage (Truax, 2001). The digital audio is represented by binary numbers that consist of ones and zeros (Farnell, 2013), as these numbers are used as the conventional representation of digital values (Truax, 2001). Usually, the digital representation of sound is codified with the Pulse Code Modulation (PCM) method (Henrique, 2002). Digital sounds can be created from scratch using a computer software (Cipriani & Giri, 2010) or can be digitized from the analogic form (Henrique, 2002)(Truax, 2001). In order to reproduce a digital sound a Digital-to-Analog Converter (DAC) is required (Henrique, 2002)(Truax, 2001). The DAC converts binary numbers to a proportionate voltage (Truax, 2001), so it could be broadcasted by a loudspeaker (Henrique, 2002).

The digitization of an analog sound captured by a microphone is obtained through an Analog-to-Digital Converter (ADC) (Henrique, 2002)(Truax, 2001), which converts points along the continuous analog signal's curve to binary numbers (Truax, 2001). The digital representation of sound is obtained by sampling the analog signal (Truax, 2001), a step by step approximation of the analog wave is performed (Henrique, 2002). The residual difference between the original analog wave and the digital time sampled wave appears as noise (Henrique, 2002). The sampling is typically made at a fixed rate in the order of 20 to 40 kHz (Truax, 2001). The discrete values obtained from this operation are called samples and the greater the number of samples per second (i.e. greater sample frequency), the more accurately small variations can be preserved (Truax, 2001).

The Shannon Theorem, also known as the Sampling Theorem or the Nyquist Theorem, states that when a continuous signal is discretized to obtain a digital signal, the sampling frequency must be at least twice the highest frequency in the original, in order to avoid the aliasing (Henrique, 2002). Aliasing is the phenomenon where a digital sound is reproduced with a lower frequency sound that wasn't present in the original signal (Henrique, 2002). This effect is caused when the original sound is digitized with a frequency lower existing one or more frequencies higher (Henrique, 2002).

2.2 Sound in Artistic Artefacts

The presence of sonic elements in artistic concepts is relatively new (Ciumakova, 2018), although Douglas Kahn defends that "none of the arts is entirely mute" (Kahn, 1999).

The invasive characteristic of sound, how we are constantly being affected by it as almost everything in our daily lives results in a sound, makes it into a powerful mean of artistic expression (Rudi, 2011). Douglas Kahn considers sound as an "artistic raw material" and states that is the artists work to finesse an artistic material into something conceptually, socially, politically, aesthetically, and poetically interesting (Kahn, 1999).

The technological advancements that provided affordable high quality sound production facilitated the exploration of sound in the most various arts (Rudi, 2011) and the desire of artists to pursue new forms of artistic materials, yet to be investigated, allied with the new tools attached to it, evokes challenges to set limits of the use and influence of the new artistic medium (Ciumakova, 2018)(Rudi, 2011). Also, nearly all arts are strongly shifting towards the digital domain (Rudi, 2011).

Simon O'Sullivan argues the central role of the existence of affects in art and defines art, in a deconstructive approach, as a "bundle of affects" waiting to be experienced (O'Sullivan, 2001). Sound has that property and can affect us in four different ways (Kristjánsson, 2016). Firstly it can affect us physiologically, when a sound has a physical impact on us by affecting our hormone secretions, our brain waves, breathing or heart rate. For example, when we are exposed to loud noises we will probably increase our stress and heart rate (Kristjánsson, 2016). It can also affect us cognitively, when a sound affects our concentration or the perception of the surroundings. For example, if two people are talking at the same time we must choose which one we want to listen to because it's difficult to concentrate on more than one sound (Kristjánsson, 2016). The third way refers to our behavior, when a sound affects our actions and we find ourselves avoiding unpleasant sounds or seeking pleasant sounds. For example, our driving is likely to be influenced by the music we are hearing, as we will probably drive more calmly with soft music than an aggressive one (Kristjánsson, 2016). Lastly, sound can affect us psychologically, influencing our emotional state. The most common way is music, as we associate music with feelings and memories, being capable of making us sad or happy. It can also be associated with many more emotions as for instance, most people may sense a bird singing as reassuring, because our genetic memory "tells" us that in the past it meant that our surroundings were safe (Kristjánsson, 2016).

2.2.1 Sound as a principal medium in artistic artefacts

The first examples of sound being used as an artistic material, without in its musical form, was to complement the visual arts (Rudi, 2011) and performative arts (Cancellaro, 2005). One of the first examples in the production of sound was in theatre plays to reinforce natural phenomena and footsteps sounds off stage, the first sound effects (Cancellaro, 2005). Along with the technological advances, recorded sounds were also adopted, i.e., the London theatre used in 1890 a phonograph to play a baby's cry (Dakic, 2009). In the 1920s, these sound effects were adopted by radio drama, using the same machinery developed by the theatrical arts. As new technologies advanced these machines were replaced by the sound library full of multiple recorded sounds with different natures. Finally, in the late 1920s, sound is introduced to the moving picture (Cancellaro, 2005).

The technological advancements created new tools that allowed the disjunction of sound and image. This separation of sound from the image generated a new artistic form, Sound Art (Licht, 2009). Due to the diversity of sound as an artistic form of expression on its own or accompanied by other mediums, there is a lot of disagreement and debates when defining what is sound art (Ciumakova, 2018)(Licht, 2009). In the early beginnings, sound as a principal medium of an art artifact was used in a variety of forms of expression (Ciumakova, 2018)(Licht, 2009)(Rudi, 2011). These forms could consist of audio montages, where the artists explored the aesthetic of sound. Some of these innovative examples are the Weekend by Walter Ruttman (1929), the sintesi radiofoniche by F.T. Marinetti (1933) and the origin of the genre of musique concrète by Pierre Schaeffer and Pierre Henry (Licht, 2009). There were also physical artifacts constructed as sculptures with a sound producing capability, often produced for a gallery or a museum installation - the sound sculptures and the sound installations (Licht, 2009). Some of the most prominent first works are from the artists Bill Fontana, Luigi Russolo, Max Neuhauss and John Cage (Licht, 2009).

Figure 2.2: Sound effects machine for theatre. Theatrical Meteorology – The Graphic January 23 1875

Retrieved from: `http://www.theatrecrafts.com/pages/home/topics/sound/sound-effects-for-the-stage/`. Accessed online: 9th August 2020.

Contemporary Art

The intent of this section is to study the importance of sound as a means of expression and its relevance as a principal medium for an artistic artifact. To this effect, this section is a survey of some artistic sound works in contemporary art, excluding sound in more recognizable forms of art, like cinema and conventional music.

resono (2015) by Rui Penha

resono is an interactive installation of an ecosystem with fifteen elements. The artist created this sound installation based on the idea that the cornerstone of autonomy is the capacity to say: no!. It seeks to answer the questions: can there be interaction without choice? Can we build true empathy between beings that are not free?

The ecosystem's elements are permanently listening to their surroundings. They wait, with different balances between anxiety and curiosity, for a new visitor. The elements react to the sound they hear trying to enter in a dialogue with the other elements and the visitor. The elements do not produce sound in an autonomous way, being necessary for the visitor of the exposition to start the dialogue. Each element responds to the message dependent on the approach taken by the visitor and the characteristics of the elements. These elements can react with fear or timidly at first, although as the visitor establishes and encourages the interaction, they become more confident and louder. Each successful interaction will contribute to the development of their unique personalities, and the artist expects the same element of empathy will happen to the visitors. The dialogue can expand between the elements and start a collective conversation. This performance is unique to each visitor, as it depends on the initial stimulus and memory of each being (Penha, 2015).

Collage I (2015) by Fernando Fadigas

Collage I is a 30 minute sound collage produced by Fernando Fadigas for the Festival Silêncio and publicized in Lisbon at The National Museum of Contemporary Art and at the Museum of Chiado. This piece is composed of various sound fragments retrieved from radio, cinema, voice, field recordings, sound poetry, excerpts of punk music and intervention music (Fadigas, 2015).

Metálica (2018) by Alexandre Estrela

Although Metálica is a video projection, we believe this work of Alexandre Estrela is still relevant to the use and exploration of sound in artistic artifacts. Metálica is composed by a projection of a photo in a copper screen. The projected image is affected and distorted as a response to the sound of knocks in a metal surface, consequently creating the illusion that the screen vibrates with the sound of metal, as if the screen really acquired vital phenomena of the knocks, accordingly expelling noises or gaining movement. The sound component of this artistic composition develops an element of tension between the material and immaterial, where the sound deceives the mind as it seems to affect the image (Castelo Branco, 2019).

Nocturne (2015) by Samson Young

Nocturne is a sound performance by Samson Young. The performance revolves on the reaction of a collective of night bombing videos stripped away of the soundtrack. The videos collected by the artist on youtube are played on a monitor during the sound performance. The artist plays live Foley performance, as he attempts to restore the sound of the silent video. The performer has 23 props which he uses to recreate the visualized sounds. These series of props have a vast nature, going from rice to a FM transmitter, or an airsoft pistol to a laptop (Young, 2015).

Lowlands Away (2010) by Susan Philipsz

The 2010 Turner Prize winner Lowlands Away, by Susan Philipsz, incorporated installed recordings, in three bridges in the center of Glasgow of the artist singing different versions of the same Scottish song over the water of the River Clyde. The exhibit of this piece consisted of her voice sounding in an empty gallery conceived to reverberate in the river waters. Susan Philipsz was the first sound artist to win a Turner Prize (Botella, n.d.).

435 prepared dc-motors, 2030 cardboard boxes 35x35x35cm (2017) by ZIMOUN

435 prepared dc-motors, 2030 cardboard boxes 35x35x35cm, is an architectural mechanism built with dc motors, cardboard and wires. This mechanism consists of a wall of cardboard boxes, hanging from the ceiling, which are constantly forced to move by the mechanical system. The sound is produced by the soft collision between the boxes. The Swiss author of this piece explores the complexity of sound and movement and the mechanical rhythm and flow of a mechanical system, creating an organic orchestra of rustling noises using minimalist materials (Koblyakova, 2017).

Figure 2.3: 435 prepared dc-motors, 2030 cardboard boxes 35x35x35cm by ZIMOUN Retrieved: `https://www.metalocus.es/en/news/435-prepared-dc-motors-2030-cardboard-boxes-35x35x35cm-zimoun`. Accessed online: 10th September 2020.

Diachrony - Transforming Movement Into Sound and Lights (2016) by Arash Azadi and Anahit Ter-Sargsyan

Diachrony - Transforming Movement Into Sound and Lights is an experimental performance developed by Arash Azadi in collaboration with Anahit Ter-Sargsyan to seek new alternative ways of artistic expression. In this performance the movements of the dancer are translated into abstract sound and visual entities using the motion capturing interface Kinect. The performer movements are influenced by the public reactions, which consequently will be transformed into sounds and visual effects (Azadi, 2016, August 1).

2.2.2 Sound and Image in the Arts

In an artistic context sound and image are very close together. There are a lot of artistic forms that use both as means of expression, being present in older performative arts, like theatre and dance, as well as present in movie theaters, like cinema and animation, or even arising through the new possibilities of the digital world, like audiovisual art and video-games.

This section documents the relationship between our auditory and vision perceptions and the effectiveness of the combination of the two to transmit information and emotions.

Relation between sound and image

Images can originate sounds from our own minds (Sonnenschein, 2001), or when imagining a situation or a place we can "hear" the soundscape of that environment through our life's experiences (Farnell, 2013). Another example is, when an alarming sound stands out, our brain instantly tries to draw a visual reference with which to attach the sound without a visual context (Cancellaro, 2005).

There is a definitive connection between sound and image since vision and hearing are senses constantly present and influencing each other in our day to day experiences (Cancellaro, 2005). Although they are more valuable together than on its own when retrieving information (Cancellaro, 2005), these two senses are very different. Our eyes focus is more narrowed and the primary brain processing is done intellectually, using reason. On the opposite, our ears focus is broad and its information processing is emotional, using intuition (Sonnenschein, 2001). The imposition of sound as an integral part of film production and how some actors failed to make this transition from the silent era, along with the rise of television, when almost all radio talent was able to move, is an interesting fact that proves this divergence between image and sound perceptions. This proves how familiar voices predominated over any mental visual image (Ballou, 2013).

Sound in the audiovisual chain

Unification:
Sound unifies or binds the image flow. In temporal terms, sound unifies visual breaks by overlapping it and also brings unity by establishing atmosphere and provides unity through nondiegetic music (Chion, 1990).

Punctuation:
Sound allows punctuation in scenes and dialogues which can be introduced subtly in a scene through discrete ways to emphasize a word, digitize a dialogue or close the scene (Chion, 1990).

Anticipation:
Considering sound and image are not uniformed elements, they can induce different sentiments in the public by indicating directions, following patterns of change and repetition. The viewer's expectations can be fulfilled or not (Chion, 1990).

Synchresis:
This phenomenon is the spontaneity and inevitability between a particular auditory phenomenon and a visual phenomenon when they occur simultaneously, being organized according to the gestalt principles (Chion, 1990).

Gestalt Principles

Although our perception of sound and image are very different, there are some similarities. One of the examples are the Gestalt Principles, a theory developed in terms of visual perception which is also applied in sound perception. These principles conceptualize the mechanisms that our brain uses to analyze information in certain situations (Sonnenschein, 2001).

Figure and Ground:
We often separate an image in two segments, the figure, where we can define an object or a character, and the ground, that is less perceptual and appears as a background (Todorovic, 2008).

It also happens in the perception of sound, a higher intensity or higher importance sound becomes the focus, where the background ambience is muffled (Sonnenschein, 2001).

Good continuation and closure:
Visual elements that are oriented and aligned with each other, even if blocked by other elements, tend to be perceived as a group, refering to the continuity principle. The same happens to a closed figure alluding to the closure principle (Todorovic, 2008)(Sonnenschein, 2001).

In audio, good continuation relates to the principle that changes of a single sound source tend to be smooth and continuous rather than abrupt. If an abrupt change occurs, a new sound source is assumed (Sonnenschein, 2001)(Dakic, 2009). The closure principle states that if a sound is overlapped by another or other event that makes the sound less relevant, we perceive it as one sound if it persists after that disconnection (Sonnenschein, 2001).

Proximity and Similarity:
When observing various objects closer to one another we assume them as a group or when there is a similarity of the physical aspect of the objects. We associate objects into groups by how similar or closer they are (Todorovic, 2008).

In audio, the proximity principle occurs when we perceive sounds near in time, and we assume they came from the same object (Sonnenschein, 2001). Likewise, sounds are assumed as connected if their psychoacoustics characteristics are similar (Dakic, 2009), even if they are separated in time (Sonnenschein, 2001).

Common fate and belongingness:
If visuals objects move together they tend to be integrated into perceptual wholes (Todorovic, 2008).

The same happens in sound, if two sound components undergo the same changes in time are perceived as part of the same source. A single component of sound can only be assigned to one source at a time, which is known as belongingness. (Sonnenschein, 2001)(Dakic, 2009).

2.2.3 Cinematography

A movie is composed of two components. The image that appears in the screen, image track, and the sound emitted by the sound speakers, the soundtrack (Buhler, Neumeyer, & Deemer, 2010). As George Lucas states, in an interview to the New York Times in 1992, "Sound is half the experience in seeing a film," (Fantel, 1992). Sound is present in motion pictures since the silent era (Buhler et al., 2010) but sound practices and roles in film production and exhibition constantly changed through the history of cinema, mostly affected by the technological improvements over the last decades (Buhler et al., 2010). The book “Hearing the Movies” by James Buhler, David Neumeyer and Rob Deemer reports a very detailed history of sound in the motion picture's world (Buhler et al., 2010).

In the early beginnings of the movie industry, the silent era, sound was different between venues or even different depending on the time of day and usually performed live by singers, small orchestras, or pianists. Sound was not developed by the filmmakers but exhibition theaters would accompany movies with sound performances (Buhler et al., 2010). Since the origin of motion pictures, there were multiple attempts to unite sight and sound. This idea started before the invention of the motion picture. One of the first examples of this attempt was the Kinetophone, developed by Edison and Dickson and introduced to the public in 1913 (Dakic, 2009). The Kinetophone had technical imperfections that fomented disapproval from the audience, causing the idea of sound movies to be abandoned until 1915 (Dakic, 2009).

The technological advancements of the 1920 decade allowed experiments in sound reproduction and inclusion of recorded and edited optical sound track, but technicians and artists had problems with the synchronization and the fidelity of the sounds and mixing audio track elements (Buhler et al., 2010). Lights of New York, released in 1928 by Warner Brothers, was the first all-dialogue film (Dakic, 2009). In the same year, Steamboat Willie, a Disney's cartoon, was the first film in which the sound track was made in post production and included sound effects, music and dialogue (Dakic, 2009). These experiments led to the classical era where microphones and speakers were more advanced, allowing sound recording and reproduction to be more faithful, and rerecording was adopted as a process of film production. Rerecording permitted to focus on capturing dialogue on set, and sound effects and music to be added in post production to the sound track (Buhler et al., 2010). The first film that started sound design history was King Kong (1933). The sound developer of the movie, Murray Spivak, was the first to manipulate sound in a creative way (Dakic, 2009).

As time passed, technologies related to sound continued to improve from 1950 to 2000. Some of the more relevant are the adoption of magnetic tape in post production, the stereo sound, followed by the shift to dolby stereo, and the emergence of sound designers and digital sound (Buhler et al., 2010). Additionally, sound recording and reproducing systems were constantly developing and improving (Dakic, 2009). The technological, commercial and cultural changes from the 50s to the 90s contributed to how distinct the movie industry in the classical era is to the movie industry in the present day (Buhler et al., 2010).

Nowadays, a movie is divided into two components, the image track and the sound track. This division requires to sync both tracks (Buhler et al., 2010). Sound and picture are perceived by different mechanisms although together they can become better than the sum of the parts. Sound is integrated along the picture into a complete whole (Dakic, 2009). The objective of the process of synchronization is to connect a sound event with the visual event, creating a virtual reality of everyday events (Cancellaro, 2005).

Similarly to the image track, the sound track has a distinction of foreground and background, being their relationship fundamental to transmit the narrative. While the foreground is the main focus, where the attention of the public must be, the background provides the sense of presence unifying sound and image as a physical place (Buhler et al., 2010).

Sound plays a narrative role in a movie. It can be used for story telling, telling the story directly, mainly from speech and as a subliminal narrative method, where the sound supports the narrative of the movie and integrates some elements, in a subtle way to not captivate the complete attention of the public (Dakic, 2009). It also allows the viewer to know details that cannot be transmitted with only the image. The sensation of dimension, velocity or weight of a character or another element in a movie can't be perceived through the image in many occasions (Sonnenschein, 2001). Furthermore, it can tell if a scene falls within a large or a small area, the type of material of an object and give perception of distance and proximity (Cancellaro, 2005). The psychoacoustical properties of a sound allow the listener to characterize something with the sound footprint (Sonnenschein, 2001). Sound is defined by textures and characteristics which can be recorded or created and are crucial to emulate an environment or visual object (Cancellaro, 2005).

Moreover, sound plays a grammatical role in a film, providing a form of continuity or connective tissue. Sound in film is normally exaggerated from the reality in isolation, being overhyped, but in conjunction with the picture assumes a more natural form. Although sound influences the story perception, it also has an impact on the film's rhythm. For example scenes where sound is absent feel longer, whereas scenes with background music feel faster (Dakic, 2009).

Sound design is the technique creative field applied on the production of sound in movies, videogames and other artistic fields and multimedia (Dakic, 2009). Normally the sound in movies and television shows is produced in post-production the interconnection of sound artifacts to the soundtrack is established (Dakic, 2009).

Movie audio production steps

The process of sound in a movie or television show consists of four phases: recording, mixing, synchronizing and editing (Dakic, 2009).

Recording:
This phase has as its main objective to record all sounds that can be potentially used, with the highest quality possible during filming (Dakic, 2009).

Mixing:
When the levels of microphones used during the recording phase are manipulated. This manipulation intends to give more emphasis at certain sounds(Dakic, 2009).

Synchronizing:
The synchronization of the image with the sounds that were not recorded during the filming of a scene which can be dialogues and sound effects, such as Foley (Dakic, 2009).

Editing:
The last phase, where all sounds are arranged and combined to obtain the final product (Dakic, 2009).

Soundtrack components

Speech:
Speech, or dialogue, is essential to inform the plot of a movie (Dakic, 2009) as it allows the observer to organize and interpret the image (Buhler et al., 2010). Through the dialogue it is possible to obtain the speaker's emotion complementing what is observed at the image track (Buhler et al., 2010). Generally, speech is in the sonic foreground and music and sound effects are part of the background (Dakic, 2009).

Music:
In the beginnings of cinema, music had 2 concrete objectives: cover up the noise from the projector and offer emotional context (Dakic, 2009). Nowadays there is a wide range of applications of music in movies. Music can offer information to the narrative, define the emotional state of a character or a moment and influence the perception of the time and continuity, filling empty moments and smoothing transitions (Buhler et al., 2010)(Sonnenschein, 2001). The music used can be specifically composed for the film or can be a commercial music (Dakic, 2009).

Sound Effects:
Everything in the image track is expected to produce sound as in the quotidian of our lives (Dakic, 2009). Sound effects are almost all sounds that aren't speech or music (Dakic, 2009). These sounds are relevant to the narrative and help to extend the physical environment (Buhler et al., 2010). The sound effects track is compiled of many tracks layered together and mixed to form an organic soundtrack (Cancellaro, 2005). The sound effects can be different types of sound and created in different ways (Cancellaro, 2005). Usually, they permit the introduction of objects, animals, or characters that are not present in the scene (Buhler et al., 2010).

The sound effects are composed of diverse types of sound artifacts (Cancellaro, 2005)(Dakic, 2009).

- Hard or cut effects, which are sound effects obtained from a source, and not recorded on the set. These effects are normally obtained in sound-effects libraries and are cut in or combined with other effects by an editor to match the picture (Dakic, 2009). Hard or cut effects make up a large part of sounds of objects and actions in the image track (Cancellaro, 2005);

- Ambient sounds that represent background noises present on the place where the scene occurs. These sounds give an "artificial" presence of the space and continuity to a scene without interfering with the sonic foreground (Dakic, 2009);

- Designed sound effects that are all non-natural sounds developed by a sound designer (Cancellaro, 2005). Normally used for futuristic technology or to transmit emotions in a musical context (Dakic, 2009);

- Foley effects, which are recorded outside of the filming process (Buhler et al., 2010). Foley sound effects are performed by a specialist artist with the use of various tools and visual support of the recorded scene (Ament, 2014).

2.3 Foley

The Foley sound effects are responsible for the realism of sound in movies (Dakic, 2009). The most common Foley sounds are footsteps of the characters which are always developed through this method (Dakic, 2009). The name Foley comes from a professional in the field, Jack Foley (1891-1967), who invented the method used currently (Cancellaro, 2005)(Ament, 2014). Foley recording was invented in the early beginnings of the sound film era. At the time sound in movies was mainly composed of dialogue and music but lacked in noises of movement. The sounds of the characters movements, like footsteps, were composed by sound cuts. The editors would cut a sound of a character, using constantly the same edited sound through the movie. Foley proposed to project the moving image in a sound stage where a person would recreate the character's sound movement and record the sounds in sync with the character while the film runs. This new method allowed the recording of multiple sounds in sync with the picture just by performing the action (Dakic, 2009), resembling the original way of recording sound effects (Cancellaro, 2005). Jack Foley had the opinion that the person performing the Foley sound effects had to get into the role and act accordingly, and consequently improving the quality and realism of the sound produced (Cancellaro, 2005).

Nowadays the method remains the same, the sound effects are produced and recorded while the image track of the scene is running (Dakic, 2009), and was adopted by video games, animation, TV shows and other audiovisual art forms. The sounds are produced in a specific studio, a Foley studio. The Foley studio has a sound workstation, where the recorded sound is edited, and a Foley stage, where the sound is recorded (Ament, 2014).

The Foley stage is equipped with recording devices and contains multiple surfaces and props which the artist interacts with to recreate the sounds (Ament, 2014)(Cancellaro, 2005). The Foley stages are usually already developed for this practice intent, but it's not mandatory. For example, the sound designer of Godfather II (1974), Walter Murch, recorded the sounds in a replica of the real space of the movie scenes (Dakic, 2009).

The production team of Foley sound effects mainly consists of a Foley artist, a Foley mixer and a Foley editor (Ament, 2014). The Foley artist recreates steps, impacts and countless other sounds synchronizing them with the image (Cancellaro, 2005)(Ament, 2014). The sounds are recreated in an exaggerated format so they would be more audible (Dakic, 2009). The Foley mixer is responsible to record the sounds recreated. And the Foley editor processes the sounds recorded (Ament, 2014)

During the investigation of the production of Foley sound effects technique, We came across documentaries that we believe are significantly relevant to be mentioned.

- The Secret World of Foley (2014)(Jewel, 2014)
- Actors of Sound (2016)(Molina, 2016)
- The Magic of Making Sound (2017)(Great Big Story, 2017)
- Inside the Pinewood Foley Studio (2018)(BAFTA Guru, 2018)
- Sound Effects - Batman: Arkham City Behind-the-Scenes Video (2012)(GameSpot Trailers, 2012)

The following table presents a list of techniques observed and mentioned in the documentaries. The table has two columns: one with the tool used (left) and the other with the sound that intends to recreate (right). There are no sounds that aren't recreated in their real context.

Sound created	**Foley Technique**
Birds flying	Feather duster
Dog footsteps	Gloves with clips on the tip of the fingers
Waterfall	Bucket of water
Storm	Tree branches
Horse running	Plunger with coated with glue tape
Walk on snow	Walk on sand
Walk on leafs	Film tape
Ice breaking	Pine cone
Boat	Tank with water
Wind	Fur cloth
Bones cracking	Leek

Table 2.1: Survey of Foley techniques to produce specific sounds.

Figure 2.4: Foley Artists Ken Dufva and David Fein recreating sounds of footsteps. Photo of Steve Lee, Hollywood Lost and Found. (*The Foley Grail*, 2014, p. 9) (Ament, 2014)

Figure 2.5: Foley Studio of Pinewood Studios
Frame retrieved from the documentary: Inside the Pinewood Foley Studio (3:52) (Jewel, 2014)

2.4 Sound Design

Sound design is a creative technical field used in several areas (Dakic, 2009). It's a discipline that includes different sound productions techniques, such as recording, mixing, sampling, editing, synthesizing, and also specific techniques such as Foley (Hahn, 2020). Presently, the production and composition of sound is used for very different products, such as sound art, theatrical productions, computer software, mobile applications and even greeting cards (Licht, 2009). The multi-functionality of sound design is enabled by the computers that facilitates the ability to model sound in every way, opposed to the lack of freedom in acoustic instruments (Cipriani & Giri, 2010).

Nonetheless, in the 1970s the term Sound Design became the denomination of this practice (Buhler et al., 2010). Sound design was present in theatre production in India and China around 3000BC through the accompaniment with music and sound. Likewise, the renaissance comedia dellàrte, born in the 16th century in Italy, used music and sound effects to make the theatrical pieces more ridiculous (Dakic, 2009). As previously mentioned in the segment 2. Sound in Artistic Artifacts, sound continuously became more commonly used in theatre plays, and later advanced to the worlds of radio, cinema and television. New innovations for sound to explore rose with the evolution of the digital environment. For instance, the 20th century novelty, the video games, started with very basic digital sounds, as the industry progressed sounds became an integral part. The sound design process in video games is similar to the film industry, although there is essential programming that is required (Cancellaro, 2005).

The principal task of a sound designer in the audiovisual media, is to develop an overall sound character for the project, as sound is an integral part of the aesthetic. Sound designers are informed of the intent of the project in the pre-production phase to produce the sound landscape. The soundtrack is organically designed to support the thematic material present in the visual side (Cancellaro, 2005).

In the book "Designing Sound" by Andy Farnell (Farnell, 2013), the author defines sound design as a structure supported by three pillars, the three bodies of knowledge. These bodies of knowledge are: the Physical, sound as a physical phenomenon; the Mathematical, sound as a digital audio signal; and the Psychological, sound as a human experience. These three subjects support the audio processing techniques used to design sound (Farnell, 2013). Sound designers use different types of sound production techniques and tools to develop the sonic component of a project, such as mixing, recording, editing a sound or make sound from scratch (Hahn, 2020).

2.4.1 Sound Synthesis

Sound synthesis consists on the production of sounds without using acoustic instruments (De Poli, 1983).The wide availability of high-quality technological tools for recording and synthesis facilitated the generation of sounds for designers and musicians (Dubnov, 2000). Sound synthesis is typically performed to facilitate some interaction or control of a sound, to facilitate the sound designer's search for a suitable sound or to create non-natural sound (Moffat, Selfridge, & Reiss, 2019). The production of sounds can be accomplished by reusing existing sounds; adopting mathematics, physics or biology theories and practices; generating sound mechanically or electronically; and many other methods. The sound synthesis techniques can be divided into two types: the Analog Synthesis, which consists of applying oscillators, filters and amplifiers into the audio analogue signals; and the Digital Synthesis, which consists of using a computer to process the numeric representations of sound (Dubnov, 2000). Our project lies in this digital space. Synthesis techniques can produce a sound directly from given data or generate, and consequently modify, signals (De Poli, 1983).

Fixed-Waveform Synthesis

The fixed-waveform synthesis is a method that attempts to reproduce a periodic repetition of the waveform without variations, similar to the characteristic of pitched sounds. In this technique, an envelope for the amplitude and frequency fluctuations can be added to increase the quality of the sounds generated, as the sound becomes less static in time (De Poli, 1983).

Subtractive Synthesis

The sound produced by a subtractive synthesis technique is a filtered complex waveform (De Poli, 1983) removed from specific harmonic content (Dubnov, 2000). This method that started the concept of sound synthesis (O'Sullivan, 2012, February 02), uses filters to selectively change specific frequency components, where the undesired components are attenuated and the remaining components are eventually amplified (De Poli, 1983).

Additive Synthesis

The additive synthesis takes a different approach from the subtractive synthesis method. Rather than removing components to obtain the harmonic structured desired, the additive synthesis constructs harmonic structures by combining together multiple sine waves of varying levels and frequencies (O'Sullivan, 2012, February 02). Through this technique complex sounds are produced by the superimposition of elementary sounds (De Poli, 1983).

Wavetable Synthesis

The wavetable synthesis resorts to a table with various switchable frequencies played in specific orders, the wavetables, allowing it to produce sound that can evolve really quickly and smoothly. The waveform does not change spontaneously, but smoothly changes into the various waves in the table (O'Sullivan, 2012, February 02). This method is often used to join the initial part of one sound onto a sustained part of another (Dubnov, 2000) and is an effective way to create pads or harsh-sounding tones (O'Sullivan, 2012, February 02).

S&S (Samples and Synthesis)

S&S applies the combination of wavetable and sample replay techniques, and adds in the filtering and shaping of subtractive synthesis. The sample replay technique replays complete samples of sounds, with a loop for the sustained part of the sound (Dubnov, 2000).

Granular Synthesis

The granular synthesis technique develops sounds from short sound fragments, the "grains" (Dubnov, 2000). It works on the same principle as wavetable synthesis, except it uses a multitude of sounds (O'Sullivan, 2012, February 02). The grains are played with a very short duration close to the timing of which the human hearing system assumes one sonic event instead of two (Dubnov, 2000). There are two ways to implement a granular synthesis: organize the grains into frames, in each frame the parameters of all the grains are updated, or scatter the grains within a mask, which bounds a particular parameter region (De Poli, 1983). The control parameters are the number of grains in a given time period, their frequency content and their amplitude (Dubnov, 2000).

Reverberation

Reverberation attempts to simulate an acoustic environment in a sound. This method can be obtained using two filters: the Comb filter, where the signal is delayed and attenuated obtaining an exponentially decaying repeated echo, and the All-pass filter where the input signal is attenuated and subtracted from the delayed signal obtaining a compensated feedback effect and maintaining the echoes (De Poli, 1983).

Physical Modeling Synthesis

The physical modeling synthesis uses mathematical equations to imitate how an instrument works, resulting in realistic sounds, synthetic sounds or a mixture of both (Dubnov, 2000). The method programs a set of algorithms to define the characteristic of the intended sound (O'Sullivan, 2012, February 02).

The digital sound synthesis methods are more varied than the analog synthesis ones and there are new ones with specific objectives emerging continually (Dubnov, 2000).

The Karplus-Strong algorithm (Karplus & Strong, 1983)

The Karplus-Strong synthesis aims to simulate the sound of a plucked string or drum timbres. Alex Strong invented in 1978 the synthesis method based on the wavetable synthesis method by adding a modification to average two successive samples producing a slow decay of the waveform. The modification resulting in a tone sounds similar to the decay of a plucked string, as the sound derives from differing decay rates from different harmonics. Being necessary to start the note with a lot of high harmonics to produce a realistic sound the algorithm fills the wavetable with random values at the beginning of each note and this randomness generates a slightly different harmonic for each note with the same pitch. In 1979, Kevin Karplus added to the algorithm a drum variant using a probabilistic recurrence relation

Physically Informed Sonic Modeling (PhISM) (Cook, 1997)

Physical informed Sonic Modeling is based on many synthesis techniques, such as physical modelling, Fourier synthesis, wavelet synthesis and granular synthesis. Perry R. Cook developed two different synthesis algorithms from the PhISM family to recreate sounds of percussion instruments.

The Physical Informed Spectral Additive Modelling (PhISAM) was developed to recreate sounds of instruments exhibiting exponentially decaying resonant behaviors. The PhISAM algorithm was implemented using resonant filters. The modal synthesis or oscillator parameters are driven and controlled by rules derived from predetermined Fourier boundary methods, and/or from analysis data extracted from recorded sounds.

The Physical Informed Stochastic Event Modelling (PhISEM) is suitable for instruments characterized by random interactions of sound-producing component objects. The PhISEM algorithm is based on pseudo-random overlapping and adding of small grains of sound according to predetermined physical rules and parameters. Perry R. Cook also adopted the PhISEM to resynthesize sounds of footsteps, "Modeling Bill's Gait: Analysis and Parametric Synthesis of Walking Sounds". The synthesis of walking sounds involved an analysis and extraction of information of recorded sounds. A particle based synthesis model was created and parameterized with the information obtained (Cook, 2002).

Physics-based sound synthesis for crushing, walking and running (Fontana & Bresin, 2003)

In this research project, two sound synthesis models were developed following a physics-based approach to recreate sounds of crushing, walking and running based on crumpling sounds. A crumpling sound occurs when our auditory system interprets a superposition of microscopic sonic events from a unique source.

The synthesis algorithms implemented by influence of an existing physically-based impacts model, where the model was superimposed to the dynamic and temporal stochastic characteristics present in crushing events.

Sounding Liquids: Automatic Sound Synthesis from Fluid Simulation (Moss, Yeh, Hong, Lin, & Manocha, 2010)

In the project Sounding Liquids a sound synthesis was developed to create liquid sounds directly from visual simulation of fluid dynamics. The source of the sound produced by fluids is assumed as the bubbles within the fluid. Each bubble produces a different sound depending on their size and format (the combination of spheres). The sound synthesis is based on fluid dynamics. In the project different scenarios were developed to test different types of simulations: objects falling into the water, a simulation of a creek flowing through a meadow, duck moving in a bathtub, pouring water and a simulation of a wall of water being released creating turbulent waves and sound as the water reflects off the far wall.

Foley became prominent because it allowed a sound designer to perform, act or create the desired sound in sync with the action and gave almost total control to the artist over the sound. Similar to Foley, sound synthesis can allow for control over digital sounds in real-time, which can be obtained from physical parameters. The public demand is increasing for instantaneous and realistic interactions with machines, notably in the video game industry. David Moffat et al. state that "Sound Synthesis can be considered digital Foley" (Moffat et al., 2019).

2.4.2 Software Tools

Digital Audio Workstation (DAW)

A DAW is a digital system designed for recording and editing digital audio. Early DAWs, "the integrated DAWs", were hardware units that included a mixing console, a data storage division and an analog to digital converter. Nowadays, audio editing and post-production is performed primarily with software, the DAW software (Christensson, 2012). The book "The Foley Grail" by Vanessa Theme Ament (Ament, 2014) contains multiple references of Pro Tools, but there are a variety of other DAWs such as Ableton Live, WaveForm and Reaper.

A DAW has access to a sequencer, sampler, analogue and digital synthesizers, effects units, mixer, samples of real instruments and other plugins used by audio engineers and sound designers to develop a sound track or a music (Dubnov, 2000).

Pure Data

Pure data (Pd)[1] is a visual programming environment developed by Miller Puckette that runs in different types of devices. Pd can be used to process and generate sound and interface sensors, input devices and Musical Instrument Digital Interface (MIDI) allowing developers such as musicians and researchers, to create software graphically without writing lines of code.

2.5 Sound Interfaces

The exponential evolution of the technologic advances opens, continually, new opportunities and possibilities to sound control and sound synthesis, where new interfaces and changes emerge regularly. The principal motivation of the pioneers of electronic music and sound interfaces was the exploration of unusual canals of expression and musical control (Paradiso & O'modhrain, 2003). Common examples of these practices are the Telharmonium and the Theremin. The Telharmonium by Thaddeus Cahill in 1906, was one of the first musical instruments to generate sound from electricity. The Theremin by Leon Theremin created in 1920, in turn, was one of the first sound interfaces of free gestural control (Henrique, 2002).

The introduction of the modular analog synthesis and controlled by tension allowed the separation of the synthesizer controller, resulting in the appearance of a variety of musical and sound interactive interfaces despite being incapable to generate audio independently. The interfaces developed at the time range from the portable keyboard to the controllers that are responsive to brain waves. The appearance of the MIDI in the 80s, separated, more clearly, the interface from the generation of sound (Paradiso & O'modhrain, 2003).

Musical Instrument Digital Interface (MIDI)

MIDI is a connectivity standard for transferring digital instrument data. It is supported by multiple platforms, such as computers, synthesizers and electronic instruments. MIDI data includes the information of the note played, velocity and time active, and other types of information, such as the instrument ID, sustain pedal timings, and controller information (for example: pitch bend and vibrato). This data is displayed in a digital format. If several notes are played at the same time, the MIDI data is transmitted for all the notes (Christensson, 2017).

[1]Pure Data [Visual Programming Language]. (1996). Retrieved from `https://puredata.info/`

Open Sound Control (OSC)

OSC[2] is a communication protocol optimized for modern networking technology. The protocol provides real-time control of sound and other media processing. OSC is flexible and easy to implement and allows for multiple implementations, such as real-time sound and media processing environments, software synthesizers, and hardware devices for sensor measurement.

2.5.1 Developed Projects

Nowadays, the research for new forms of sound interfaces for musical and sound expression continues. The International Conference on New Interfaces for Musical Expression (NIME)[3] is an annual conference that brings together researchers and musicians to share their knowledge and their latest work on new musical interface design. The dutch research center Studio for Electro Instrumental Music (STEIM)[4] created in 1969 develops and experiments new sound interfaces and other artifacts for sound art and live performances.

During the research phase of the project, a survey of sound or musical interfaces was necessary to learn and understand the recent techniques and explorations accomplished in this area.

Sibilim (de Souza Nunes, Visi, Coelho, & Schramm, 2019)

The Sibilim is a low-cost musical interface. The interface is composed of a resonance box made of cardboard, the source of the sound, containing buttons, whose position can be configured by the user using a smartphone's camera. The buttons are mapped to a set of MIDI notes or control parameters. The sound is generated through synthesis or sample playback.

The Sibilim was built for music education, to develop the consciousness of music perception and to stimulate creativity.

Jam-O-Drum (Blaine & Perkis, 2000)

The Jam-O-Drum is a multi-user interactive musical system. The interface allows the participation of six users simultaneously on an interactive surface with the intent to make the group take a collaborative approach to musical improvisation, to make collective musical experiences available to novices, to explore music and motion in graphics and to experiment different musical and visual styles.

The system architecture consists of two separate computers, one for music and other for graphics, connected by the MIDI data protocol, a sampler to play the musical notes and also a drum and percussion sound module, which receives signals from the drum pads embedded in the table. The graphic computer was connected to an LCD video projector, that projected the image received down onto the table. The Jam-O-Drum table was built with 10" drum pads mounted directly onto the frame. Multiple software programs were used during the development of the various prototypes, for either sound processing and production or graphical components.

[2] Open Sound Control [Communication Protocol]. (2002). Retrieved from http://opensoundcontrol.org/

[3] https://www.nime.org/

[4] https://steim.org

Shaping Sounds (Walther, Müller, Brade, & Groh, 2013)

Shaping sounds is composed of an elastic surface and a collection of objects. Each object is represented by a specific shape, so it can be identified, and is associated with a continual repetition musical content.

The user is able to grasp the objects, raising and lowering them, due to the elastic capacities of the surface. When the user lifts an object the reproduction of the sound sample associated with it is activated, while doing the opposite causes it to stop. Other functionalities were implemented in order to apply changes in the sound samples while the object is up, such as the possibility to change the volume and speed of the sample.

Cave of Sounds (Murray-Browne et al., 2014)

The Cave of Sounds is an interactive sound installation formed by eight musical instruments. Each instrument was developed with the intent to belong in this ensemble.

Sonicsphere (Panagiotis Tigas)
Sonicsphere is a sphere with the size of a hand whose orientation is mapped into a space of tones.

Joker (Wallace Hobbes)
The user using a mask with conductive tape activates different percussion instruments when touching the mask.

The Animal Kingdom (Daniel Lopez)
Interface where the user interacts with it imposing shadows in a translucid surface. The interface has the capacity to identify hands and fingers and detects common shadows of animals, like a dog or a bird. On detecting these animals, it produces a sound related to the animal.

Generative Net Sampler (Tadeo Sendon)
On entering the invisible cylindric activation zones the user triggers sound samples collected from the internet.

Lightefface (Kacper Ziemianin)
This instrument contains 24 light-dependent resistors that modulate the intensity of a harmonic with a fundamental frequency. The sound played is affected by the illumination on the object's surface captured by the resistors.

Campanology (Dom Aversano)
Generation of rhythms through mathematical patterns controlled by hand movement.

Rockmore (Sus Garcia)
A theremin which the proximity of the hand of the user affects the sound.

Wind (Tim Murray-Browne)
The movement of the user's hands on his body is captured using a 3D camera. The body is then mapped into a grid of harmonious notes that play flute sounds.

The Mutantrumpet (Neill, 2017)

The Mutantrumpet is a hybrid electroacoustic instrument, designed to erase the boundaries between acoustic and electronic musical creating and performance.

This expanded acoustic instrument and electronic controller contains an acoustic trumpet, in which the sound is converted into MIDI data, and multiple other controllers, such as joysticks, and a STEIM Junxion board, that maps the controllers on the board to a variety of routings and allows multiple configurations. The Mutantrumpet can be incorporated with multiple software applications, including Junxion, LiSaXC (the STEIM live sampling program), Ableton Live, Jack Router and Resolume. The Resolume is used for the live video interaction.

Mutantrumpet is capable of interacting with audio and video simultaneously. The controllers are mapped to video parameters using Resolume, creating a true synthesis of the two media in performance. The audio is played by synthesizers in Ableton Live, and its live sampled sounds are used to animate the visual material. This visual feedback allows the audience to perceive the interactivity of the performance.

Figure 2.6: Exposition of the project "The Cave of Sounds", London. Retrieved: Sound Synthesis and Evaluation of Interactive Footsteps for Virtual Reality Applications.(Murray-Browne et al., 2014)

SculpTon (Boem, 2014)

SculTon is a malleable tangible interface for sound sculpting oriented to live performance. The interface, which was developed within the concept of shaping sound, is an equivalent of physically shaping an object. The organically shaped object, the physical interface, is capable of detecting subtle and detailed physical manipulations when the user handles and modifies it. The information detected is then used in a real-time formant synthesis.

SculpTon consists of a physical interface, a software and a stompbox. The physical controller is composed of an array of sensors (analog, digital and acoustic) and 4 LED for visual feedback embedded in a malleable structure covered with an opaque latex hull. The software is a framework for Pure Data, so it is possible to manage different routines, data processing, mapping and sound synthesis. The stompbox connects the physical interface and the software using two serial plugs for the interface and a USB for the serial communication with the laptop. The box contains an Arduino Mega board, a low-power audio amplifier, a footswitch, and a potentiometer.

2.6 Related Projects

This section introduces three projects related to ours, containing a characterization of each project where both technicalities and functionalities are addressed. Finally, a reflection and analysis of the projects and how they shaped the conceptualization of our project. Some aspects of the project will be mentioned in a general context, since they will be detailed in the next chapter.

DIGIT: A Digital Foley System to Generate Footstep Sounds (Aly, Penha, & Bernardes, 2017)

This project consists of a system composed of a tangible interface surface and a software tool. The DIGItal sTeps (DIGIT) was created with the intention to assist the development of footsteps sounds in a Foley context. The sound produced is obtained using the acoustic gestures captured on a control surface to retrieve similar matches from a large database containing annotated footsteps sounds.

The sounds stored in the database follow a criteria of attributes using Max's descriptors object, which characterize temporal and spectral attributes of the footsteps sound. The database includes a unique type of show on four different types of floor, being wood, cement, gravel and water. The sounds were divided into three categories of speed (slow, average and fast), each type of speed containing 300 footsteps sounds on each one of the floors. The sounds were collected from the AudioStepsPro database. This use of a high number of sounds is related to the interest of minimizing repeated sounds and allowing subtle nuances captured by the physical interface to be recognized.

The physical interface consists of a rectangular metal plate where a contact microphone is attached to capture the performed gestures. The input data is analyzed to create a representing acoustic profile of the gesture. These gestures are detected using the Max object bonk, to trigger the real-time extraction using the Max object descriptors. The data obtained is then used to search a footstep on the annotated database that best corresponds with it. To avoid repeating the same footstep and to promote timbral variation, the four most recent sounds used are excluded by the search algorithm. The selected footsteps sounds are reproduced using a granular sound synthesis engine and post processing effects can be applied.

Physically-based sound synthesis and control of footsteps sounds (Turchet, Serafin, Dimitrov, & Nordahl, 2010)

This project is one of the more relevant as it's an experiment more focused on developing an interface to assist Foley artists than the DIGIT, who focused on the system as a whole. Three interfaces were developed for the synchronization of footsteps, producing the sound of footsteps in different types of surfaces.

In this project, a sound synthesis mechanism based on physical phenomenons was developed. The mechanism, implemented with Pure Data, pretends a sound as a result of physical interaction between objects, and the details as a sound result of the material's properties and shape of the objects. In the case of this interface, the sound of the footsteps is the result of multiple impacts of the shoe on the surface. To this process a distinction between solid surfaces and granular surfaces (e.g., sand) was made.

The modulation of solid surfaces was decomposed in an exciter and a resonator, being the exciter, the foot, and the resonator, the floor. There were also developed models of impact and friction, so the interface could allow continuous interaction, like dragging the foot, and discrete, like one step.

In the granular surface, the synthesis was implemented with algorithms following a physically informed sonic model. This model follows probabilistic rules for the simulation of the interaction of the various particles, meaning that each particle has a certain probability to produce sound.

The interfaces developed allow the reproduction of footsteps on the following surfaces: metal, wood, snow, gravel, sand, grass, dry leaves, vegetation pothole and ground with rocks.

- **Microphones on the floor**

  The interface used 4 microphones on the floor to capture the sound of the user's footsteps to extract the information of the reaction forces on the floor used to control the sound synthesis of the footsteps. The sound recreated was transmitted to the user's headphones.

- **Tangible acoustic interface**

  The interface consists of a plate with 4 accelerometers, positioned in the middle of each side of the plate so it can calculate the position of the feet. Since it was developed to not be influenced by other external sounds as the first did.

- **Shoes with pressure-sensitive sensors**

  The physical interface is a pair of sandals with two pressure-sensitive sensors placed in the extremities of the sandal's sole. The sensors detect the pressures force exercised during the user's movement.

Scrubber: An Interface for Friction-induced Sounds (Essl & O'Modhrain, 2005)

The Scrubber is an interface for friction-induced sound, where the user engages in familiar gestures, such as wiping or scrubbing, to reproduce sound. The developers emulated the natural tactile sensation experience of friction united with a friction-related sound controlled by the performer.

The physical interface consists of a gutted white-board eraser with it's interior filled with silicone and sensors that are used to detect the user's actions. The collective of sensors are two actively powered microphones embedded on a cavity in the silicone and a force sensing resistor. The microphones are oriented downward towards the surface to detect the interactions with it. The irrelevant actions, such as grabbing the interface, are prevented from being detected by the silicone. The force sensing resistor detects the contact force applied on the surface.

The interface can control a variety of sound synthesis algorithms, using the data detected by the sensors. In this project some experiences were made with a granular synthesis, a wave-table synthesis and a physically informed modeling.

De-MO: Designing Action-Sound Relationships with the MO Interfaces (Bevilacqua et al., 2013)

The Modular Musical Objects (MO) are an ensemble of tangible interfaces and software modules. This ensemble of hardware and software components was developed with the purpose to create new musical instruments or to augment objects with sound and to support the exploration and experimentation of action and sound.

The MO focused on the relationship of action and sound. The relevancy of this interface to the project is due to the free movement of the user when interacting with the interface resulting in multiple scenarios, being these the motion-sound relationships developed for the interface.

Some of the scenarios are:

- **The Grain Stick** is a scenario where the MO hardware is combined with an object. When tilting the object, a cascade of small sound events occurs.
- **The Sound Shaker** is a scenario where the object creates rhythmic patterns when it is being shook, recalling a shaker or maracas.
- **The Sounding Surface** is where the user plays on a surface. The different interactions, like caressing or scratching, are detected and mapped to corresponding sound textures.
- **The Throw a Ball** is where the resulting sound illustrates a ball trajectory.

The physical wireless interface contains a 3D accelerometer and 3-axis gyroscope, two buttons and 9 LEDs allowing additional sensors to be added using a I2C Bus. The wireless transmission is performed using the IEEE 802.15.4 protocol to a receptor that communicates via Open Sound Control with a computer running the software modules.

The MO software modules are integrated into Max, using the MuBu toolbox. These execute multiple algorithms for the real-time processing of motion data and sound. The processing techniques can be standard signal for low level features or more sophisticated analysis techniques, using machine learning techniques, digital signal processing and simple physical models.

2.6.1 Analysis and Reflection

The approach taken in the initial conceptualization of our project was to develop an interface for Foley footsteps, since it is so predominantly used for this effect. During the research we stumbled upon the second project, changing the concept to broader use of this technique so that the interface could be used to generate any sound the artist or performer wanted to, since the objective of our project strived to be an exploration of a new type of interface. To this purpose, a collective of different types of interaction were added to the concept.

The first two projects are both focused on the recreation of footsteps sounds, but both focused on different aspects of the system. While the first revolves around the development of the sound library, the second revolves around testing different types of interfaces. Both of the developed systems had a mandatory sound library and each of the sound libraries has different methods for producing sound. The DIGIT library is composed of annotated sounds of footsteps with different descriptors that are matched with the input data captured. The second project uses a physical model based on sound synthesis developed with Pure Data where the calculated values of the forces applied on the three different physical interfaces are used to recreate the sounds. To avoid restricting the interface to a single sound library we adopted to our project the necessity of the user to use the interface with different types of sound production procedures, for example with a DAW or a Pure Data patch.

The third project, the Scrubber, is very similar to one of the interactions used in our project and contains multiple experiences on sound synthesis techniques applied on the recreation of sounds produced by wiping or scrubbing. This interface is very similar to our interface, since the intention is to control sound synthesis algorithms without forcing the user to use a specific library.

The last project focuses on the relationship of sound and gesture, not necessarily for Foley. The ensemble of interfaces resembles our project since they promote different types of interaction being up to the user to choose. For different scenarios of interaction additional sensors or other resources may be added using an I2C Bus. The Sounding Surface scenario was added to our project as an interaction. The difference between our project and the MO project is the more restrictive interactivity of our project, as the MO project allows the user to add sensors promoting the possibility of new interactions.

This page is intentionally left blank.

Chapter 3

Methodology

The methodology adopted in the project was the Design Science Research (DSR). The methodology revolves around an iterative cycle between the fundamental activities to build and and the evaluation of the design artefacts. The DSR's three pillars are the gathering of the maximum knowledge regarding the scientific areas coincident to the project, the oriented experimentation that leads to the development of the artefacts and the analysis of the outcomes at each experimental phase. Thus, the methodology is based on the search of solutions for the identified problems.

3.1 Objectives

This dissertation proposed to accomplish the following objectives:

1. Understand the relevancy of sound as a medium of art, specifically in cinema.
2. Study the Foley technique for production of sound effects, including current practices.
3. Explore ways to replicate some of the traditional Foley practices based on the interaction with the props into a single interface.
4. Research and implement a collection of sensors to gather information from the user's actions.
5. Research sound synthesis and sound editing techniques suited for the data obtained from the user's interaction.
6. Create an intuitive, useful, responsive and coherent interface capable of assisting the production of sound effects by mimicking the conventional Foley method.

The last objective is accomplished through the collection of the other five objectives and is the final outcome of this dissertation. The development of Jack-in-the-Mug intends to explore new ways of sound effects production and sound expression.

3.2 Process

The project's work process consisted of three phases. The iterative cycle of the Design Science Research approach was applied in each phase, represented in Fig. 3.1.

The first phase was divided in two sub-processes: the contextualization and the definition of the concept. The contextualization was a general study of topics that influence the project and its theoric investigation was made thoroughly: sound and its presence and relevancy in artistic mediums; sound interfaces; sound design; and the Foley technique for sound effects production. The concept's definition occured in parallel, focusing on the analysis of a collection of projects and practices related to the theme to comprehend the current techniques on the different scientific areas. Thereafter, the conceptualization and definition of the project were conducted. This phase outlined the State of Art and defined the final objective of the dissertation. It also included the investigation of promising sensors for the interface's intent.

The second phase focused on the analysis and extraction of data and tools. The analysis comprised the study and experimentation of different sensors to capture the information of the user's interaction with the interface, including the collection of libraries and algorithms developed with similar intent, and the study and experiment of different sound synthesis and sound manipulation techniques to be manipulated by the data collected from the sensors. In this phase, an experimental study was conducted to evaluate different tools to transmit the information collected by the interface to software programs used by sound designers, such as Digital Audio Workstations and Pure Data.

The last phase consisted of the creation of the interface, by refinishing the collected information into the final product. This phase revolved around the evaluation of the results and consequently adjustments.

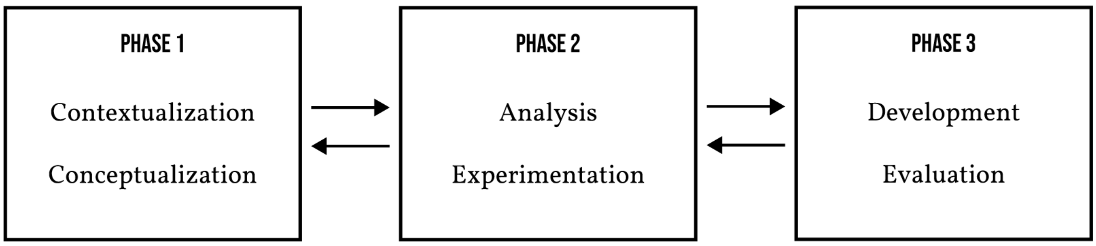

Figure 3.1: Representative scheme of the process.

This page is intentionally left blank.

Chapter 4

Design Overview

This chapter details the steps taken to define the conceptual identity of the project, containing an essay about the concept of the system and how the ideas evolved into the definitive concept, including a brief explanation of the features and motives of their selection. The process initialized with the development of user profiles to acknowledge needs of potential users for our interface. The information gathered was then applied during the definition of the interactions, the sound library and the sound effects production process.

The project started with the intent to design a musical interface based on instrumented physical objects. As already mentioned in the State of Art, new musical interfaces projects appear every year, as in laboratories such as Steim or as proposals for conferences such as NIME. To steer away from this overcrowded theme, we conducted a search for new purposes to design our sonic interface. During this research, the definitive conceptual idea occurred - to develop an interface to assist in the production of Foley sound effects. The initial concept focused on the production of sound effects of footsteps, where the Foley artist would walk in a carpet, incorporated with multiple sensors, to recreate the character's walk. Since it was similar to the project of Luca Turchet et. al (Turchet et al., 2010), we decided to change the concept to an instrumented physical object to auxiliate the production of sounds of multiple natures, including footsteps.

Hence, the definitive main concept was determined as a physical object interface for Foley sound effects production. A study on the usual work process for producing Foley sound effects was conducted to determine which features our interface needed. The main purpose of this study was to define how the user would interact with the interface, what sounds should the sound library contain and how the interface would insert itself in the framework of the Foley sound effects production process. An approach similar to contextual design was adopted in the design process to conceptualize Jack-in-the-Mug.

4.1 User Profiles

The first step of the conceptualization of our interface focused on the research of our user's needs to create profiles for each type of user. These profiles would define what implications we must take in count to produce our interface. The primary users of our interface are the Foley Sound Effects Production Team and the Amateur Audiovisual Content Creators, which want to use the Foley technique to develop sound effects.

The Table 4.1 and the Table 4.2 contain the user profiles of the Foley Sound Effects Production Team and Amateur Audiovisual Content Creators, respectively.

User	Foley Sound Effects Production Team (Primary)	
Characteristic	**Detail**	**Design Implications**
Access to multiple props	The different props allow to recreate different sounds	Require access to an great variety of sounds
Different gestures with the props	Each prop to reproduce sound may be used in different manners	May require multiple forms of interaction based on the most common ways to handle the props
Freedom of movement	The Foley artist is in a studio room when performing with the different props	May require to be wireless
Is a work based on experimentation	The Foley artists experiment with different props to accomplish a specific sound	May require freedom of sound production
Access to an auxiliary image	The Foley artists perform and synchronize the sounds auxiliated by the scene image running	Require to the possibility to synchronize the sounds with the movie image
Has experience on producing sound effects with the Foley technique	The Foley technique is very old and its professionals already develop systematic processes of work	May require to be inserted in the conventional method of production of Foley sound effects without changing the method drastically

Table 4.1: User Profile of a Foley Sound Effects Production Team

User	Amateur Audiovisual Content Producer (Primary)	
Characteristic	**Detail**	**Design Implications**
Has difficulties to apply the Foley method to produce sound effects	Wants to produce sound effects using the Foley method	May require the same needs of a Foley sound production team
Doesn't have the same resources of a Foley Studio	Wants to add sounds in the project but don't have access to the same resources of a Foley Studio	Must be compact

Table 4.2: User Profile of a Amateur Audiovisual Content Producer

Although the interface was built for Foley, the freedom of expression required to allow the application of such techniques, potentiates the use of the interface in other artistic mediums. Sound Artists and Musicians were introduced as secondary users. A research of sound in artistic artefacts was conducted and afterwards documented in section 2.2, *Sound in Artistic Artefacts* of the chapter 2, *State of Art*. Also, the sound control capabilities that Jack-in-the-Mug required may be useful for Sound Designers or Sound Producers.

The Table 4.3 and the Table 4.4 contain the user profiles of the Sound Artists and Musicians and Sound Designers or Sound Producers, respectively.

User	Sound Artist and Musician (Secondary)	
Characteristic	**Detail**	**Design Implications**
Experimentation	Some artists are eager to test and experiment new forms of art	May require the same needs of a Foley sound production team
Freedom of Expression	The artists job revolves around expression	Must allow the artist to express himself or be used in a transformative way; Must allow the artist to change between interactions freely; May require it to be used as a MIDI interface

Table 4.3: User Profile of a Sound Artist or Musician

User	Sound Designer and Sound Producer (Secondary)	
Characteristic	**Detail**	**Design Implications**
Manipulate Digital Samples	Sound Designers and Sound Producers job involve the manipulation and control of sound samples	Must allow to control and manipulate sound samples himself. May require OSC protocol or MIDI
Create sound effects	Sound Designers create sounds by using synthesis methods or by applying effects	Must allow the creation of sound effects

Table 4.4: User Profile of a Sound Designer or Sound Producer

4.1.1 Scenarios and Storyboards

After the research and development of the users profiles, two scenarios for each primary user were created to make an analysis exercise about circumstances in which the interactive system will be used. Also, storyboards were drawn to get a visual representation of the scenarios.

Foley Sound Effects Production Team

"During the production of Foley sound effects, a Foley artist is using a hammer to recreate the sonic event of the character nailing nails.

The Foley artist breaks the overused hammer.

Instead of searching for a new hammer, the Foley artist proceeds to grab the Jack-in-the-Mug.

The Foley artist interacts with the interface in a similar manner as he would with the hammer to produce the sounds.

The sounds produced are received and verified by the Foley editor."

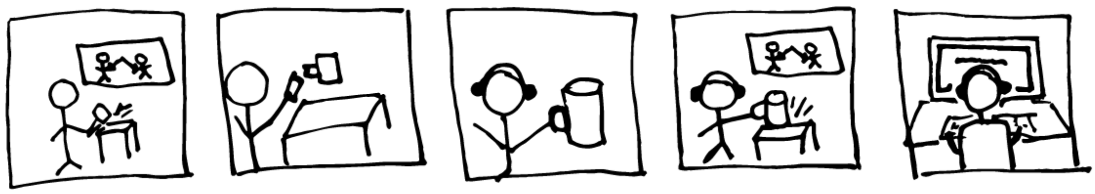

Figure 4.1: Storyboard: Foley Sound Effects Production Team.

Amateur Audiovisual Content Producer

"Joe is an aspiring movie director. He is developing a short film to test and evolve his capabilities.

Joe is currently in the audio production phase and wants to use the Foley technique to produce sound effects for his short film.

Joe doesn't have the same resources of a Foley studio.

Joe proceeds to grab the Jack-in-the-Mug.

He then interacts with the interface to produce sound effects and sync them to the picture.

The sounds produced are present in the DAW used by Joe."

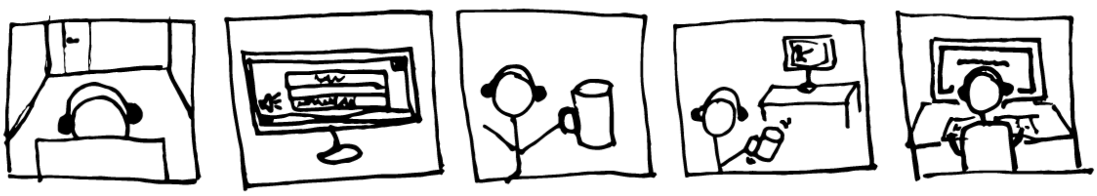

Figure 4.2: Storyboard: Amateur Audiovisual Content Producer.

4.2 Definition of the Interactions

The study evidenced the freedom of expression of Foley artists to recreate the sound, as in the use of multiple props and in the experimentation of new methods. With this freedom in mind we took the approach to try to minimize the constraints of our interface, while allowing a reasonable number of different movements to be produced and sensored. Each of these movements should be easily performed by an average person, without much training. Using an interface that could resemble a familiar object sounded us as a good way of proceeding. A mug served as inspiration for our object. Jack-in-the-Mug, the name of the interface, is a wordplay with the words Jack Foley, Mug and Jack-in-the-Box. The table 4.5 contains the details of the four interactions selected that are also visually represented in Fig. 4.3. The third column of the table 4.5 contains references from where we based on when defining the interaction and demonstrates examples of how the the Foley artist can use this collective for other purposes. The need for diversity of interactions is related to the intention of offering the Foley artist the freedom to use the device according to his needs and preferences.

The interface contains multiple buttons to offer the possibility to change between interactions freely during a performance. This concept was defined mainly to satisfy the expressional needs of a sound artist or a musician.

Interaction	**User's Action**	**Sounds Recreated**
Collision	The user hits the object bottom into a surface. The intensity of the impact is applied in the sound processing	The interaction was adopted to recreate sounds of impacts. For example: - Footsteps - Punch - Glass breaking
Scrub	The user moves the object on a surface. The velocity of the movement is applied in the sound processing	The interaction was adopted to recreate sounds of something being dragged on a floor or scrubbed on a surface. For example: - Rake a garden - Table cleaning - Grate cheese
Shake	The user shakes the object. The velocity of the movement and the event where the movement's direction changes are applied in the sound processing	The interaction was adopted to recreate sounds of hollow objects containing materials or props used with this method by Foley artists. For example: - Water Bottle being shaken - Maracas - Key-Chain being shaken
Rotation	The user rotates the object in the different axis. The degrees are applied in the sound processing	The interaction was adopted to recreate sound of liquids or continuous sounds. For example: - Jar of water filling a glass - Door squeaking - Control wind and other ambient sounds

Table 4.5: Interactions defined for the interface

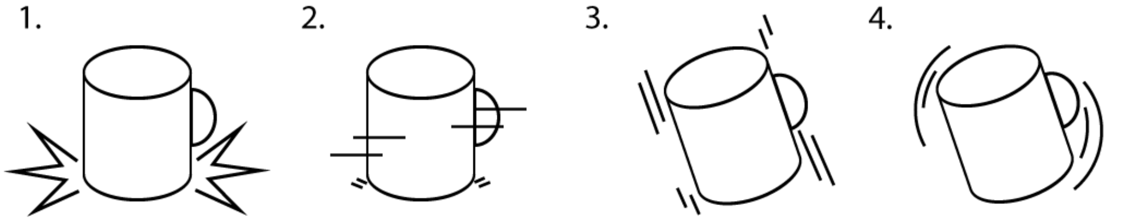

Figure 4.3: Sketch of the interface and interactions.
1.: Collision 2.: Scrub
3.: Shake 4.: Rotation

4.3 Definition of the Sound Library

As already mentioned, the Foley technique is based on experiments and can be used to create an infinite number of sounds. For this reason we decided not to focus on creating a sound library that would confine the user in a specific set of sounds, but to give the necessary resources for the creation of different sound effects. The user would have total freedom on the sound he intended to reproduce. Also, there is a variety experiments made on sound synthesis techniques to produce sound effects.

DAWs are commonly used in film production and other areas where a soundtrack is composed. DAWs can receive data from recording devices or by OSC and MIDI messages. There is a variety of this type of software on the market, such as Pro Tools and Ableton, so our interface was developed to be used on the user's prefered program offering and testing the possibility of using different DAWs. These software tools are used for producing, recording, mixing and editing audio. Therefore, the system's data stops when it reaches the digital interfaces and is, consequently, processed.

Pure Data is a visual programming language, often used for sound design. The language allows to receive, control, manipulate and transmit data that can be applied for audio, video and graphical processing. The programs implemented by Pure Data are called patches. In our system, Pure Data patches are used to receive and process the data received by the Jack-in-the-Mug and send to the intended DAW. The patches can create sounds resulted from sound synthesis algorithms or other methods. The audio is transmitted using an audio routing software that connects the Pure Data output to the DAW's input. A patch can send MIDI and OSC information using a virtual MIDI device connected to the DAW.

In brief, the sound producer develops a Pure Data patch that receives and processes the information detected by our physical interface. Afterwards, it sends MIDI or OSC messages or a digital sound to a DAW. Although our project did not focus on the creation of a sound library, we compiled a set of sounds for the usability tests and, also, for the experiments while implementing the interface.

4.4 Definition of the Foley Sound Effects Production Process

The production of Foley sound effects typically requires three professionals: a Foley artist, a Foley mixer and a Foley editor. The Foley artist has the role to recreate and synchronize the intended sounds while the image is running, being in contact with a diversity of props. The Foley mixer is in charge of the microphone's positioning. Lastly, the Foley editor is responsible to verify the audio quality and if the sound is in sync with the image.

Jack-in-the-Mug was developed with the traditional method of production in mind, applying small changes to the process. The Foley artist continues to have the same role but, instead of using a diversity of props, he only uses one prop - the interface. The interface detects the movement and intensity applied to it and the data collected is then used for processing the intended sound. The Foley artist continues to proceed with the conventional method when mimicking the character's behaviour, i.e., if the character of the movie effects does an intense movement the Foley artist also has to do an intense movement with the interface. In summation, the Foley artist must apply the same intensity present on the image to be mapped into the sound to be produced. The process does not require the presence of a Foley mixer when using the interface due to the absence of microphones. The Foley editor performs the usual job task.

The exploration of props to reproduce the planned sounds is substituted by the exploration of sound synthesis and other methods for digital sound production. Sound designers or sound producers are inserted by the interface in the workflow to implement the sound processing methods.

The Flow Model in Fig. 4.4 is based on the workflow of the process to produce Foley sound effects by a Foley Sound Effects Production Team using the interface. The model focused only on this primary user but the compactness of the interface provides the possibility for amateur artists to develop their own sound effects, as the interface eliminates the need of a Foley studio. In this case the amateur producer can assume the role of all subjects. Likewise, due to the possibility of other types of users, such as the secondary users, some tasks may be removed and substituted, as well as, the subjects can be different. For example, a sound artist may use the interface in a performative artistic medium, such as dance, where the final output is a loudspeaker that transmits the sound to the viewers.

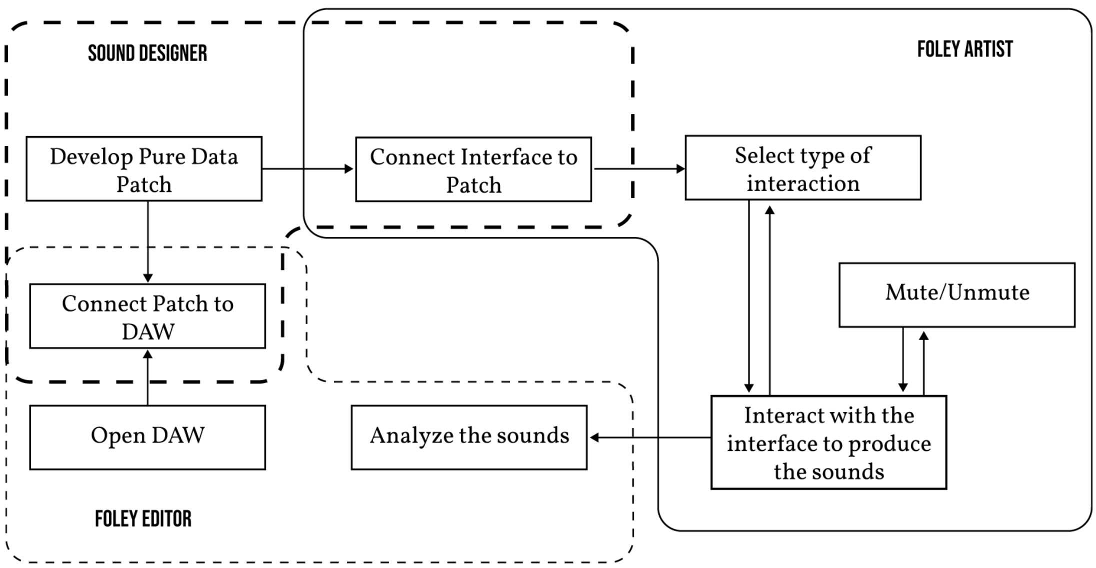

Figure 4.4: Flow Model of the Foley Sound Effects Production Process with the interface.

This page is intentionally left blank.

Chapter 5

Prototyping

The establishment of fundamental requirements during the conceptualization of the interface initiated the development of a prototype. The requirements in conjunction with the information collected and analyzed promoted the initial prototyping experiments.

The prototype was created to put into practice the interface's theoretic concept. The current chapter contains a description of how the prototype was implemented and of the final system framework.

5.1 Implementation

The implementation of the prototype was divided into two parts: the electronic component and the physical model.

5.1.1 Electronic Component

The electronic component consists of the electronic equipment used to detect the user's interactions with the device. Initially, an investigation, and consequently programming and testing of possible equipment, was conducted to determine which microcontroller board and sensors should be embedded into the physical interface.

Microcontroller board

The main feature required from the microcontroller board was the possibility to transmit data wirelessly so cables wouldn't interfere with the user performance. The first experiments to obtain data from the multiple sensors were conducted using an Arduino UNO board, which is unable to transmit data via wireless on its own, and programmed using the Arduino IDE. The inability of the Arduino UNO board to transmit data wirelessly was ignored due to the possibility of incorporating a Bluetooth Module or a WIFi Module onto the final prototype. Even though the Arduino UNO board was used for most experiments, instead of incorporating a Bluetooth Module or WIFi Module, the board was replaced with an ESP32 with Bluetooth Module and a WIFi Module already incorporated in it. This replacement did not interfere with the experiments already conducted since the ESP32 board can be programmed using the Arduino IDE. The communication protocol selected to transmit data was the Bluetooth.

In this first prototype, the processing capacity of the board and latency of response was not taken into consideration since the final objective is to evaluate the theoretical concept of the interface. However, considering our interface's main focus is the synchronization of sound and image, this specification of the board should be taken into consideration for the final artefact. Andrew P. McPherson et al. examined multiple microcontroller and wireless devices with a computer-based sound generator to test the latency response of these devices. This experiment proved Bela to be a possible platform for the final artefact (McPherson, Jack, Moro, et al., 2016). Bela [1] provides a latency under 1ms by replacing the microcontroller and the computer with a high-performance embedded board.

Sensors

The predominant data required by all four interactions is the the interface's movement. The first step was to define which Inertial Measurement Unit (IMU) would be employed to capture the movements of the user while using the interface. The sensor selected was the MPU-6050, that contains a 3-axis gyroscope and a 3-axis accelerometer allowing to detect the angular velocity and the acceleration forces acting on the device, respectively.

The values obtained from the MPU-6050 are the analog voltage read converted into digital data. These, most commonly known as raw values, vary between 0 and 32750. The acquired raw values are transformed into interpretable data using the TinyMPU6050 library[2]. The library offers functions that return the angular data in radians and the acceleration data in m/s^2, both in the 3 axis. Although, the acceleration's data doesn't attend to the accelerometer influence by the gravity.

To withdraw the effects of the force of gravity, the values of the acceleration are subtracted to the values of the acceleration of the last iteration of the processing loop. In summation, the acceleration in each axis is equal to the current values obtained minus the last values obtained. Succeeding the determination of the acceleration vector, the acceleration magnitude is calculated using the formula:

$$||a|| = \sqrt{a_x^2 + a_y^2 + a_z^2}$$

.

In order to facilitate the sound design process, the magnitude calculated is transformed into a scale of 0 to a 100. The maximum value of acceleration was set at $2.5m/s^2$, after multiple experiments.

The next segment is divided into the four defined interactions, in order to clarify the explanation of the multiple sensors as they were implemented and tested according to each data the specific interaction required.

Collision

- Impact intensity:
  - The intensity is acquired using the magnitude of the acceleration in the moment of impact offered by the MPU-6050.
- Moment of Collision:
  - The first sensor tested to determine the moment when the interface hit a surface was a condenser microphone, which detects the shock waves of the impact, but it was discarded since it would also detect claps and other sounds in the room. During the experiment phase, other sensors were used such as the tilt sensor, discarded due to the detection of other events, and the tap module, which was discarded since it didn't recognize every impact.

    Finally, the idea of using custom push buttons on the base of the interface to detect the moment of impact emerged. The custom buttons are composed of foam, foil and two jump wires. One of the jump wires is connected to the

[1] Bela [Embedded Computing Platform]. (2014). Retrieved from `https://bela.io/`

[2] TinyMPU6050 (Milan, G) [Arduino library]. (2019). Retrieved from `https://github.com/gabriel-milan/TinyMPU6050`

5 Volts and the other to an analog pin. When the user hits the interface on a surface the foam shrinks making the foil disc connect the two jump wires. The data collected from the digital pin is the electric current, so when it turns to 4095, which corresponds to 5 Volts, is assumed that an impact occurred.

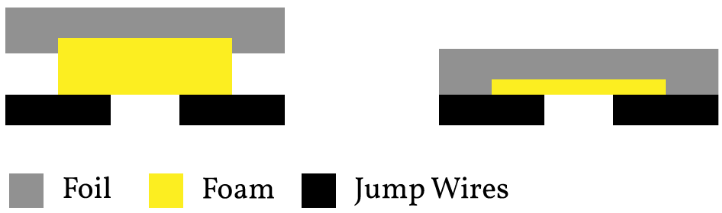

Figure 5.1: Sketch of the push button's

Scrub

- Acceleration Magnitude:
  - The acceleration of the movement is determined from the data captured by the MPU-6050.
- If the interface is superimposed on an object:
  - The Scrub interaction, apart from the acceleration of the movement, needs a sensor to verify if the interface is superimposed on an object. The initial, and last, experiment was to use an Infrared Obstacle Avoidance Sensor. This sensor is composed by an infrared receiver and an infrared LED. The infrared LED emits infrared signals, which are reflected back to the receiver if an obstacle is obstructing them. If the receiver doesn't detect infrared signals, it is assumed that no obstacle is in front of the sensor.

Shake

- Acceleration Magnitude:
  - The acceleration of the movement is determined from the data captured by the MPU-6050.
- If the interface is on movement:
  - The acceleration data obtained from the MPU-6050 informs if the interface is moving or if it's static. The interface is assumed in motion when the magnitude of the acceleration surpasses a specific limit. The interface's movement is considered to stop when the interface remains more than 500 milliseconds above the limit imposed. The value $0.6m/s^2$ defined for the limit was established through multiple experiments of trial and error by movementing the sensor.
- Moment when the movement changes to the opposite direction:
  - The first experiments to detect the event of the movement's change of direction used tilt sensors, which were discarded since it didn't recognize every event.

Since we already had access to the acceleration of the movement, a theory emerged. The premise was that the acceleration values should be 0 when the interface changes to the opposite direction, based on Newton's second law of motion where force equals mass times acceleration. The moment when the interface changes its direction the sum of the forces applied onto the interface is equal to 0. So, if force equals mass times acceleration, when the acceleration is 0, the sum of the forces is, also, 0. This theory was tested but the values obtained from the acceleration never reached 0.

After some research and analysis of the data collected from different movements, a pattern was discovered between the values of the sum of all accelerations and the moment of directional change. The value of the sum of the accelerations changed from positive to negative and from negative to positive when the interface changed its course, which proves the initial premise, since, in order for these variation events to occur the acceleration must have been 0 at some point during the delay of the process loop iteration of the microcontroller board.

Rotation

- Angular deviation in each axis:
  - All the data necessary for the Rotation interaction is obtained from the MPU6050 and accessed using the TinyMPU6050 library. The only adjustment added from the data obtained from the library was the subtraction of the angular deviation when the interface became connected from the current angular deviation. This way the user would be aware that the angles would be considered according to the position of the interface as it turned on. The values obtained were later converted to degrees, since the library returned the angles in radians.
- Rotation speed:
  - The angular speed is acquired using the same values from the acceleration magnitude.

5.1.2 Physical Model

The physical modelling process started after the necessary electronic components were determined, since the first task of the process was to define the size of the physical object which was affected by the size of these components. The definition of the dimensions of the object was accomplished through the creation of multiple cardboard box models to verify if the electronic components would fit and how to position them. The cardboard box models also allowed us to get a more realistic sense of the volume of the object. The final cardboard box model is present in Fig. 5.2, whose format is a quadrangular prism, with 12x12 cm base and 18 cm of height, containing a shelf at 8.5 cm from the base.

Figure 5.2: Photo of the final cardboard box model.

The next phase consisted on the construction of a 3D model using 3D modelling software. The software selected was Rhinoceros 3D [3]. The approach taken for modelling the physical object was to initially recreate the cardboard box model selected with the electronic components positioned in the Rhinoceros 3D to obtain a basic foundation model. The cardboard box was modelled with one surface curved to resemble a mug, since during the conceptualization of the interface the object was most of the times mentioned as "caneca", which means mug in Portuguese. Boxes with dimensions based on measurements made on the electronic equipment were positioned according to the position of the equipment in the interface.

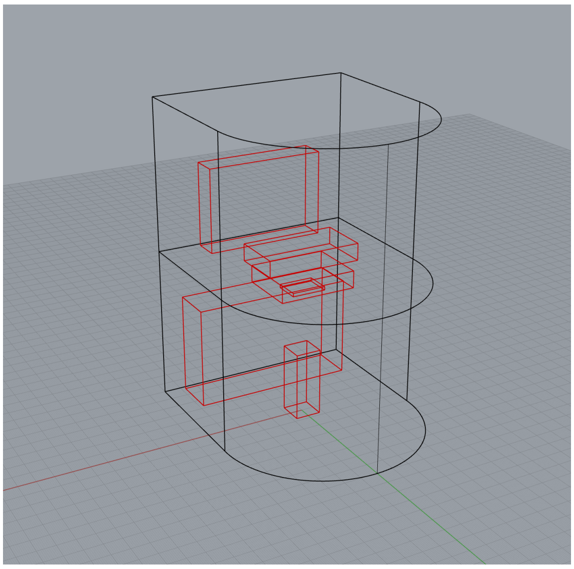

Figure 5.3: Screenshot of the wireframe of the cardboard box model and the sensor boxes in Rhinoceros 3D.

[3]Rhinoceros 3D [Computer Software]. (1998). Retrieved from `https://www.rhino3d.com/`

The model of the real object was developed according to the material and printing method selected, MDF with 5mm of thickness and laser cutting machine respectively, being divided into two pieces, represented in Fig 5.4. The piece on the center dovetails with the main piece (left), so it would be easier to access the electronic equipment by dislodging the pieces. The main piece contains all the equipment where the other piece is only used to hide them.

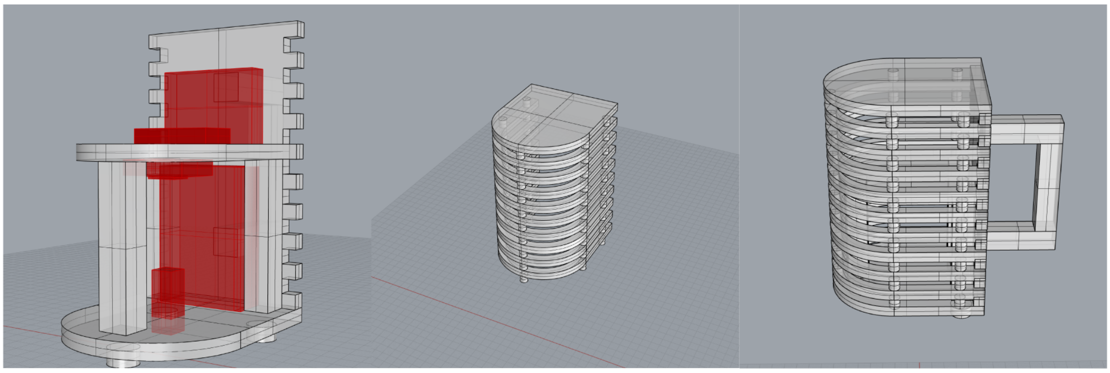

Figure 5.4: Three screenshots of the interface model during production in Rhinoceros 3D.

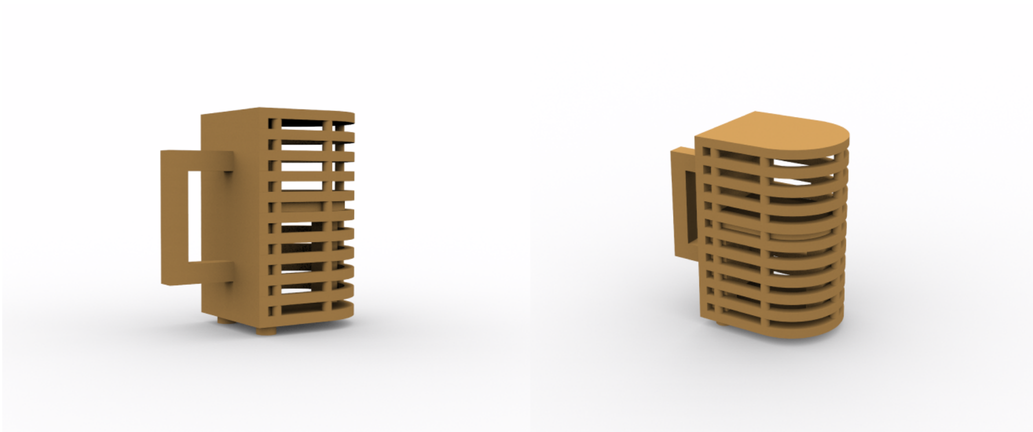

Figure 5.5: Render image of the interface model.

The proceeding phase intended to transform the model into 2D pieces for the laser cutting machine. The pieces were developed to fit with each other with special attention to the thickness of the MDF. The printed 2D pieces, show in Fig. 5.6, were constructed into the physical object using polyvinyl acetate glue. The two pieces assembled can be observed in Fig. 5.7 and Fig. 5.8. Finally, the electronic components were attached to the main piece, Fig.5.9.

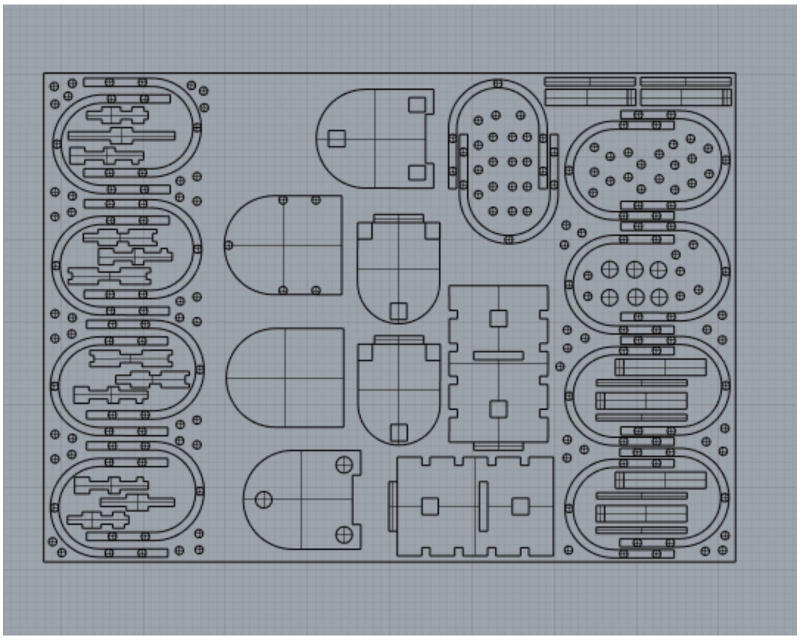

Figure 5.6: Render image of the interface model.

Figure 5.7: Photo of the main piece of the interface assembled without the electronic equipment.

Figure 5.8: Photo of both pieces of the interface assembled.

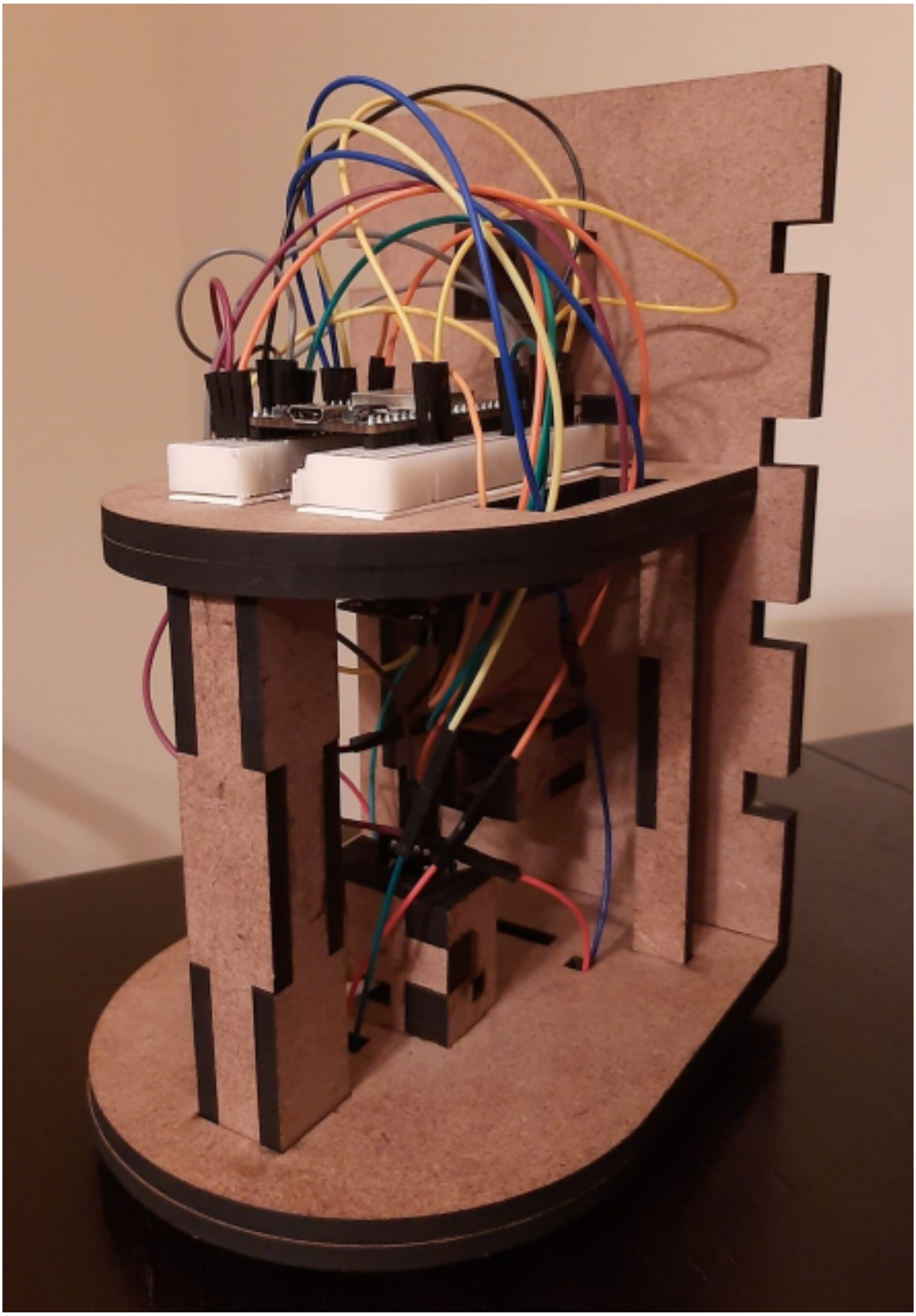

Figure 5.9: Photo of the main piece of the interface assembled with the electronic equipment.

5.2 System Framework

5.2.1 List of equipment

- ESP32-WROVER-B (Microcontroller Board)
- MPU-6050 (IMU)
- Infrared Obstacle Avoidance (Obstacle Sensor)
- 3x Custom Push-buttons (Collision Sensor)
  - Foam
  - Foil
  - 2x Jump Wires
- Jump Wires
- 2x Breadboards
- Two piece structure (Physical Model)
- Powerbank

5.2.2 Circuit Diagram

The circuit diagram of the prototype is presented in Fig. 5.10. The three custom push buttons are represented by a single push button.

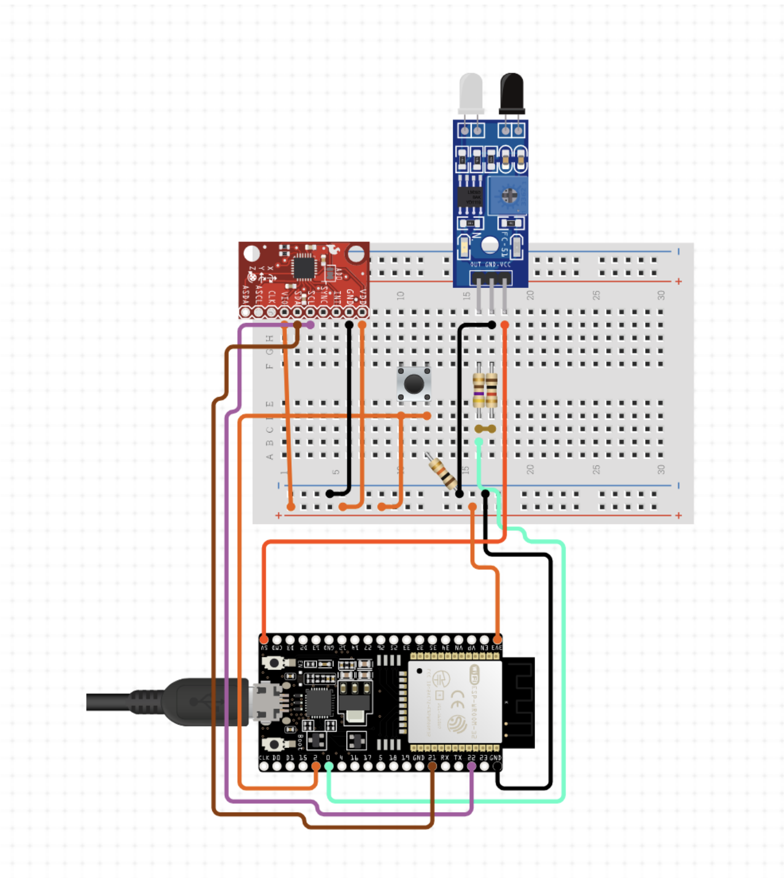

Figure 5.10: Circuit Diagram of the prototype.

5.2.3 Data Transmitted

This section contains the data transmitted in every iteration of the processing cycle, except for the Shake interaction that only transmits when the interface is on movement. The acceleration vector is transmitted in all interactions to offer the sound designer more access to potential useful information.

Collision

- Acceleration Vector (Values in m/s^2)
- Acceleration Magnitude (Values between 0 and 100, where 100 equals to $2.5m/s^2$)
- Value of Collision (The value transmitted is constantly 0 and only turns 1 when an impact is detected)

Scrub

- Acceleration Vector (Values in m/s^2)
- Acceleration Magnitude (Values between 0 and 100, where 100 equals to $2.5m/s^2$)
- Value of Superimposition (The value is 0 when the bottom of the interface is not blocked by an object and 1 if it is blocked by an object)

Shake

- Acceleration Vector (Values in m/s^2)
- Acceleration Magnitude (Values between 0 and 100, where 100 equals to $2.5m/s^2$)
- Value of Movement (The value transmitted is 1 when a directional change of the movement occurs, -1 when the movement stops and 0 when neither happens)

Rotation

- Rotation Vector (Values in degrees)
- Acceleration Magnitude (Values between 0 and 100, where 100 equals to $2.5m/s^2$)

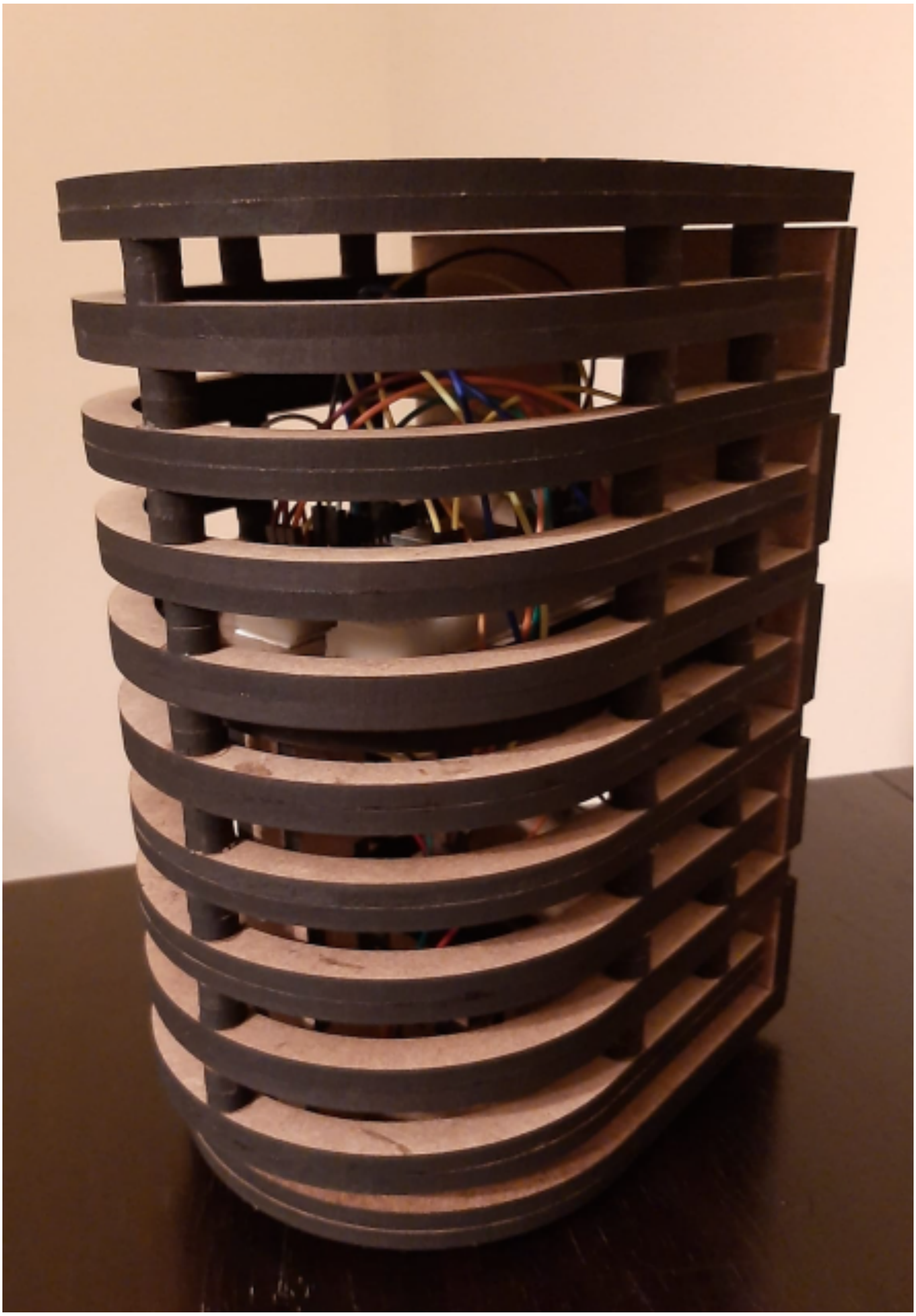

Figure 5.11: Photo of the main piece of the interface assembled with the electronic equipment.

This page is intentionally left blank.

Chapter 6

Results and Observations

In order to complete our study, two approaches were taken to obtain the necessary information about the project so that we could analyze and reflect upon it. The first approach was to conduct interviews with professionals in sound production, in order to obtain an analysis of the concept from the point of view of someone inside the industry. The second approach consisted of usability tests to examine the interface on its practicality, responsiveness and intuitiveness, as well as discover possible implementation errors.

The final section of this chapter contains an analysis and reflection of the information collected from both approaches.

6.1 Professional Opinion

The main purpose of the interviews was to acquire a critical opinion of the interface's concept from sound professionals and potential users. The interviews were performed remotely and started with a brief explanation of the concept of the interface with visual assistance, although all the interviewees commented on the necessity of being able to experiment the interface to offer more convict opinions, which was the initial idea. However, it was discarded due to the current situation of world pandemic. After the conceptual explanation of the interface, a previously prepared questionnaire related to the project was conducted. During the conversation, the interviewees offered new insights about the project, aside from the previously prepared questions. The professionals interviewed were António Porém Pires (Sound Designer) and Branko Neskov (Sound Designer).

6.1.1 Data collection and analysis

The following section contains the questionnaire and respective answers by the professionals:

1. On a scale of 1 to 5, where 1 means "very low" and 5 "very high", evaluate the Foley's method regarding its relevancy in the audiovisual and sonic worlds.

   Answers: 5 | 4.

   It is a consensual opinion that the Foley sound effects are very relevant to both worlds, but not always required to create interesting sonic artistic artefacts.

2. On a scale of 1 to 5, where 1 means "very low" and 5 "very high", evaluate the applicability of the physical interface in the productional method of Foley sound effects.

   Answers: 3 | 3.

This was the most relevant question to validate the concept of our interface, mentioned by the interviewees as the most difficult answer as they couldn't give a convict answer, since they didn't had access to the interface. It was also a consensual opinion that the interface was likely to not insert in the Foley sound effects production, but could be used in projects with no access to a Foley artist.

3. Indicate up to three positive aspects.

   Answers:

   - It would be an optimal solution to restrict the production to one single device.
   - The lack of need for microphones which, despite detecting some interesting type of noise, they can also detect undesirable sounds.
   - The multiple possibilities to control the sound synthesis.
   - Offer a new possibility for projects with no access to a Foley artist. For example, an independent movie or videogame.
   - The wireless capability.
   - Future potential for other types of purposes, aside from Foley.

4. Indicate up to three negative aspects.

   Answers:

   - The loss of the artistic side of Foley. Branko Neskov complemented with "Foley Artist is called an artist for a reason. They perform art".
   - The possibility to create trends, multiple movies would use the same type of audio by using the same library.
   - The process of developing the sound synthesis for the giant amount of environmental sounds of a movie would take too much time.
   - The fact that is a very rigid object.
   - The unrealistic sound artefacts resulting from the current sound synthesis algorithms.
   - The latency of using the Bluetooth protocol is equal to 1 frame of difference.

5. Suggest a different scenario from Foley in which it might make sense to use this interface.

   Some of the scenarios mentioned were already thought during the conceptual design process of the interface. The scenarios mentioned were:

   - Music
   - Sound Controller for other purposes than Foley,

6. On a scale of 1 to 5, where 1 means "very low" and 5 "very high", evaluate the relevancy in the exploration of new interfaces for artistic purposes, regardless of the concept of the interfaces being based upon traditional artistic production methods.

   Answers: 5 | 5

   It is consensual between the interviewees that the exploration of new interfaces is necessary to find new forms of expression, even if they are not adopted by the industry.

6.2 Usability Tests

The usability tests were structured in two parts: the first, to introduce the testers to the interface and verify possible implementation errors, and the second, to test the interface in a Foley sound effects production environment. This way the tester was more familiarized with the interface when performing the Foley technique. At the end of each part, the testers answered questions dependending on the objective of the part.

A graphical interface (Fig. 6.1) and a sound library were developed to perform the tests.

The graphical interface was developed in Processing[1]. It allowed the tester to change between interaction, to mute and unmute the sound, to play and pause the auxiliary video. Another important function of the interface was to allow the user to start recording, which would initiate a countdown. Upon completion, the scene would start and the user would begin using the interface. Once the scene reached the end, the audio recording stopped. These multiple buttons were considered for the physical interface during the conceptualization but were not implemented on the developed prototype. The graphical interface was also a communication bridge between the physical interface and the sound library.

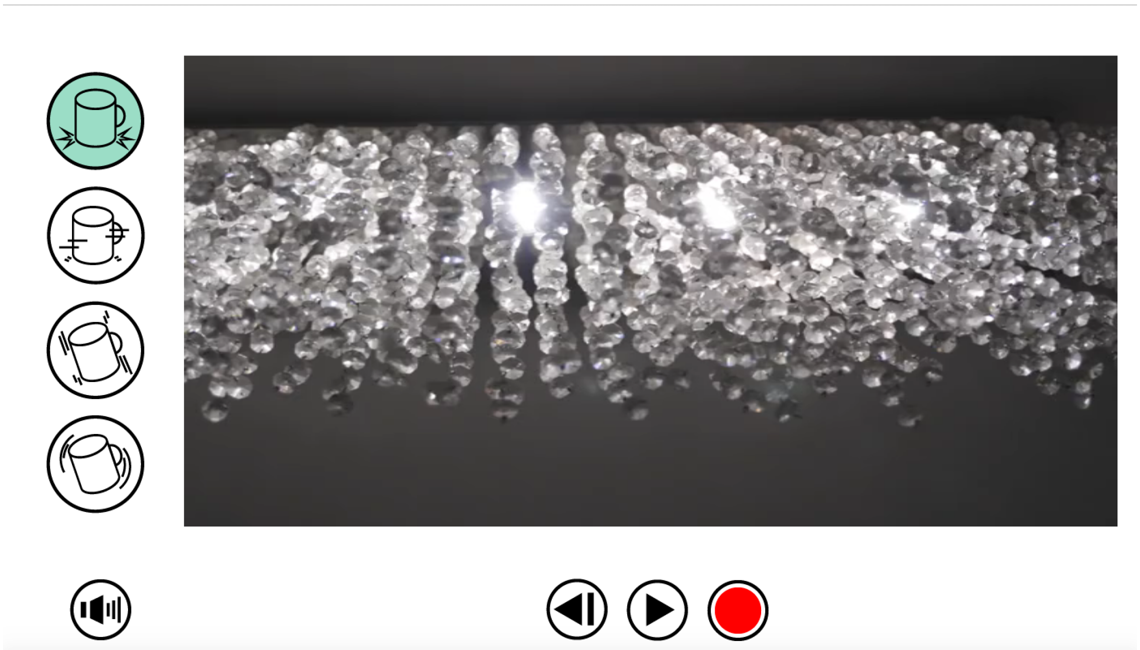

Figure 6.1: Screenshot from the Graphical Interface with the movie selected for the Shake interaction.

6.2.1 Sound Library

The sound library consists of five Pure Data patches that receive the data via Open Sound Control from the graphical interface. One of the Pure Data patches is used in the first part of the testing experiment and each of the other four in each interaction in the second part of the experiment.

[1] Processing [Programming Language]. (2001). Retrieved from `https://processing.org/`

Part I

The sounds produced by the first patch are all non-natural, with the aim to make the user more aware of the characteristic of each interaction.

In the Collision interaction, the volume and the sustain of the sound reproduced are affected according to the intensity of the impact, so a more intense impact results in a loud and long sound, and a less intense impact results in a quiet and short sound. The sound is produced by an oscillator with 440Hz.

The Scrub and Shake interactions produce similar results. The Scrub only produces sound when the base of the interface is obstructed by an object. The sound produced is an oscillator in which the frequency changes according to the acceleration of the movement. High values of acceleration correspond to higher pitched sounds and lower values correspond to lower pitched sounds. The same occurs in the Shake interaction, the only difference being that when the interface's movement changes to the opposite direction a "pop" sound arises.

The Rotation interaction is composed of three oscillators sounds each with different frequency. The three sounds are affected by each axis of rotation, where the volume of the sounds are controlled according to the rotation of the interface.

Part II - Collision

The video selected for the Collision interaction is a scene of a character hammering multiple nails. The idea was for the testers to mimic the character's action with the interface as the hammer.

The sound produced when the interface hits the surface are excerpts of a bigger sound file, containing multiple hammering sounds. The excerpts were cut from the original sound and categorized as low, medium and high according to the intensity of the impact assumed. Each category contains three different sounds, so the final sound artefact wouldn't be monotonous.

The same categories are applied to the intensity of the impact of the interface. Additionally, more intensity equals to higher volume.

Video:
cottonbro. (2020, August 13). Free stock video of antique, carved, closeup [Video file]. Pexels. Retrieved from: `https://www.pexels.com/video/5095290/`. Accessed online: 10th October 2020

Sound:
InspectorJ. (2017, October 27). Hammering Nails, Close, A.wav [Audio file]. Freesound. Retrieved from: `https://freesound.org/people/InspectorJ/sounds/406048/`. Accessed online: 10th October 2020

Part II - Scrub

The video selected for the Scrub interaction is a scene of a character cleaning a floor. The idea was for the testers to mimic the character's action with the interface as the scouring pad.

A sound of a single scrub with a scouring pad was recorded and then transformed into a seamless loop. The creation of the loop was done by cutting a segment of the audio recorded. The final seconds of the segments were interpolated with the beginning of the segment and later applied fade-out and fade-in effect, respectively.

The pitch of the looping sound is affected by the acceleration of the movement, using the "phasor" object from Pure Data.

Video:
Martina Tomšič. (2020, February 24). Close Up Footage Of Someone Cleaing The Floor [Video file]. Pexels. Retrieved from: `https://www.pexels.com/video/close-up-footage-of-someone-cleaning-the-floor-3801162/`. Accessed online: 10th October 2020

Part II - Shake

The video selected for the Shake interaction is a more abstract use of the interface. The scene contains multiple glass filaments hanging from a ceiling and moving slightly. The idea was for the testers to shake the object as the multiple glasses move and collide with each other.

For the production of this sound, five sounds of glass hits were collected. The Pure Data patch reproduces these sounds randomly while the interface is moving. The intensity of the movement defines the delay from one sound to another. Higher values of acceleration equals to a lesser time interval between sounds.

These are also randomly modified using a reverb which receives random values to create a less monotonous sound and give a more interesting sonic atmosphere to the sound. Furthermore, the volume of the sound produced is defined randomly.

Video:
Engin Akyurt. (2020, September 16). [Video file]. Free stock video of beautiful, bright, chandelier Pexels. Retrieved from: `https://www.pexels.com/video/5372702/`. Accessed online: 10th October 2020

Sounds:
HonorHunter. (2015, April 25). Clanking glass.wav [Audio file]. Freesound. Retrieved from: `https://freesound.org/people/HonorHunter/sounds/271655/`. Accessed online: 10th October 2020

13GPanska_Gorbusinova_Anna. (2017, January 25). Tinkling the glasses.WAV [Audio file]. Freesound. Retrieved from: `https://freesound.org/people/13GPanska_Gorbusinova_Anna/sounds/377989/#`. Accessed online: 10th October 2020

Unicornaphobist. (2015, February 4). Glass Tap.wav [Audio file]. Freesound. Retrieved from: `https://freesound.org/people/Unicornaphobist/sounds/262958/`. Accessed online: 10th October 2020

Part II - Rotation

The video selected for the Rotation interaction is a scene of a character filling a glass of water. The idea was for the testers to mimic the character's action with the interface as the jar of water.
Three sounds were collected and transformed into seamless loops using the same technique used for the Scrub interaction.
The volume of each sound elevates as the interface is more inclined. The three sounds were then mapped into different angle values. The first sound starts producing when the interface is at 10º according to the floor and reaches the maximum volume at 20º, the second one starts at 20º and reaches the maximum volume at 40º, and the third and final sound starts at 40º and reaches the maximum volume at 60º.

Video:
Free Videos. (2016, July 09). [Video file]. Pexels. Retrieved from: `https://www.pexels.com/video/pouring-water-in-drinking-glass-853752/`. Accessed online: 10th October 2020

Sounds:
AdamJordaan140087. (2017, September 7). pouring water [Audio file]. Freesound. Retrieved from: `https://freesound.org/people/AdamJordaan140087/sounds/401672/`. Accessed online: 10th October 2020

delaneyrose13. (2013, November 17). Water Being Poured into Glass [Audio file]. Freesound. Retrieved from: `https://freesound.org/people/delaneyrose13/sounds/207361/`. Accessed online: 10th October 2020

vataaa. (2017, July 27). Water in a glass [Audio file]. Freesound. Retrieved from: `https://freesound.org/people/vataaa/sounds/398330/`. Accessed online: 10th October 2020

6.2.2 Process of the experiment

The experiment began with a simple explanation of the project and purpose of the test. Afterwards, the paper with questions prepared for the experiment was offered to the tester to write his personal information: age and gender. The test coordinator had access to a manual to guide the tester through the different tasks of the parts. In both parts, the tests were conducted in the following order: Collision, Scrub, Shake and Rotation.

The first part, the introduction to Jack-in-the-Mug, consisted of multiple tasks to inform the testers about the different interactions and how different inputs resulted in different consequences. The user was asked to use the interface with the different interactions and test the functions offered by the graphical interface. The practical component of this part finished with the tester using the interface without restraints, so he would be more comfortable for the second part. Following the practical component, the tester answered questions relative to the first part that were specific for each interaction. The tester graded from 1 to 5, being 5 the highest grade, the interface in terms of its response and functionality. The purpose of these questions was to validate the implementation of the prototype, to ascertain whether the implementation errors would affect the success of the second part of the experiment.

Before the second part, a brief summary of the Foley sound effects technique was explained. This part, the Foley sound effects production, consisted of multiple tasks whose final goal was to create part of a soundtrack of a scene for each interaction. Initially, the testers were asked to experiment with the interaction freely and to start training the synchronization of the sound effects by playing the video when they became more familiarized with the sounds produced. The interaction experiment ends when the tester presses record to perform the final soundtrack. At the end of each interaction, the testers had access to the audiovisual content they created and answered questions relative to the performance they experienced with the interface. The questions, being the same for each interaction, were about the relation between visual support and the interaction, the sound quality and the final artefact recreated.

The usability test ended with the tester responding to five general questions about the interface. The first three consisted on grading questions from 1 to 5 and the last two on pointing out negative and positive aspects of the experiment.

6.2.3 Data collection and analysis

The total population of the experiment is composed of a collective of 3 individuals aged 24 years (2 male and 1 female).

In the group of answers relative to the first part of the experiment with the purpose to validate the implementation of the prototype, an average of 4.25 was recorded. The score of the multiple grades from the 24 answers can be observed in table 6.1.

During the experiments an error was detected, as it can be verified in the second column of the table 6.1 where the testers graded the value 1 a total of three times. This corresponds to the question: "Evaluate, on a scale of 1 to 5, the responsiveness of the interface in relation to the moment when the movement change its direction and reproduction of the "pop" sound" , which verifies if the directional change of the movement in the Shake interaction was detected.

After the completion of all the experiments, an analysis for the possible motive verified the error was in the Pure Data patch instead of the prototype.

Additionally, if the error was in fact from the prototype it would not affect the rest of the experiment, since the sound used for the Shake interaction did not require the detection of the change in the movement's direction. By observing the data collected from the table 6.1 and by discarding the implementation error for the motives already mentioned, we can deduce that the prototype implementation did not affect negatively the answers in the second part of the experiment.

Grade	1	2	3	4	5
Score	3	0	1	4	16

Table 6.1: The score of the grades from the answers relative to the first part of the experiment.

The answers related to the second part of the experiment are divided into 4 groups.

The questions for the first group of answers intended to validate the selection of the interactions, where the testers graded the interaction according to the context of the visual support. This group of answers had an average of 4.75, corresponding 10 (83%) of them to the grade 5. The score of the multiple grades can be observed in the table 6.2. According to these values we can observe that the testers felt comfortable with the interactions for the different videos, including more abstract concepts such as the sound selected for the Shake interaction.

Grade	1	2	3	4	5
Score	0	0	1	1	10

Table 6.2: The score of the grades from the answers relative to the first group of the second part of the experiment.

The second group refers to the quality and realism of the audio produced. The purpose of the questions on this group was to evaluate the sound library. The average result was 3.83, which proves the sound library was satisfactory for the testers. The table 6.3 contains the values for each grade. The least appreciated sounds were from the Scrub and Shake interactions, and the most appreciated sound was from the Rotation interaction, being graded 5 by all the participants.

Grade	1	2	3	4	5
Score	0	0	5	4	3

Table 6.3: The score of the grades from the answers relative to the second group of the second part of the experiment.

The testers were able to visualize and hear the result of their performance in an audiovisual artefact. The third group of questions evaluates the sound-image synchronization of these artefacts from the perspective of the tester. The score of each grade can be observed in the table 6.4, scoring an average of 3.66.

Grade	1	2	3	4	5
Score	0	1	5	3	3

Table 6.4: The score of the grades from the answers relative to the third group of the second part of the experiment.

The last group of answers, relative to the satisfaction of the tester towards the artefact produced, had an average of 4. The table 6.5 contains the scores for each grade, proving in most of the cases the tester was satisfied with the final product.

Grade	1	2	3	4	5
Score	0	1	2	5	4

Table 6.5: The score of the grades from the answers relative to the forth group of the second part of the experiment.

The experiment ended with five general questions about the interface and experiment. The first three questions are represented in table 6.6. Analyzing the data collected we can deduce the interface is accessible to the common user, being intuitive and entertaining. The difficult level of sound-image synchronization with the interface obtained interesting positive values for the first prototype, which didn't have all the resources defined for the final product.

In the last two questions, the tester was asked to provide up to three negative and positive aspects of the experiment. The most common negatives aspects were related to the sound of the Shake interaction, commenting the sound had a low quality and was not apropriated to the scene, and the handle of the prototype, being loose and not ergonomic.

	1	2	3	4	5
Interface Intuitiveness	0	0	0	0	3
Ease of sound-image synchronization	0	0	0	2	1
Interface entertainment	0	0	0	0	3

Table 6.6: The score of the grades from the answers relative to the forth group of the second part of the experiment.

6.3 Analysis and Reflection

The purpose of our thesis was to evaluate the concept of a physical interface to assist in the production of Foley sound effects. The results obtained from the interviews with professionals related to the area are quite significant for this evaluation, since it offered an insight from the point of view of the industry. The usability tests allowed to evaluate the interface's interactive module, as in its intuitivity and characteristics based on the Foley technique.

The most significant question during the interviews with the professionals in relation to the purpose of the thesis was the evaluation of the practicality of the interface in the process of the Foley technique where both professionals graded with 3. During both interviews, the interviewees reported the same opinion, that it was improbable for the interface to be accepted/adopted in a Foley studio environment, since the function of the studio is to enable the artistic performance of the Foley artist through the various props and infrastructures offered by the studio. Another potential problem of the interface is the lack of realism of the sounds produced using the sound synthesis algorithms in today's world, since the interface concept is, in part, dependent on the evolution of this scientific domain.

Nevertheless, both sound designers agreed that the interface had interesting attributes and could be used transformatively in the creation of sonic artefacts for audiovisual media, such as movies and games, apart of the Foley process. António Porém Pires also referred the utility of the interface in independent projects with no access to a Foley studio.

However, the small sample of participants can't provide a convict statement. The usability tests, according to the answers obtained, show some evidence that the the prototype to be intuitive and clear for the production of sound effects and its use to be entertaining. The testers evaluated positively the difficult level of sound-image synchronization and the different video support in correlation to the sounds produced from the interaction.

In conclusion, according with the results obtained, an interface to assist in the production of Foley sound effects would unlikely be adopted in the Foley sound effects production. However, an interactive sound interface whose concept is based on the Foley sound effects method could be useful to create sound artefacts of an audiovisual media or other means of expression.

This page is intentionally left blank.

Chapter 7

Conclusion

The long process during the development of this thesis intended to verify the possibility to integrate an interactive physical interface in the production of Foley Sound effects, founded on the information collected from an investigation of the Foley technique and scientific areas where the project inserts itself, as well as the results acquired with the conclusion of the interface's design.

The conception of this interface intended to explore a new method of synchronization and edition of sound, as well as determine limitations and new possibilities of sound interaction and sound design, coupled with the increasing availability of miniaturized electronic equipment and wireless technologies and with the relevancy of sound in the experience of the audiovisual artefacts. The focus of our thesis was the Foley sound effects, due to the pertinence of this method in the sound production for audiovisual media, such as videogames and cinema, and lack of research of the new technological potentialities in the area. The principle of the interface allows the reduction of the use of natural resources and the wear and tear of props, giving access to a similar concept of the Foley technique to the increasing number of new amateur and independent audiovisual creators, due to the appearance of sharing platforms for audiovisual media, such as Youtube, which don't have access to the multiple props and big storage of a Foley studio.

The research on the Foley sound effects production modelled the interface's concept and characteristics, meaning the definition of the multiple interactions of the physical interface were based on the Foley artist interaction with the props and in the adoption of the mentality to not limit the Foley artist into a specific set of sounds. Furthermore, this diversity of interactions and sounds promotes the possibility of the application of the interface in transformative direction for other sonic artistic purposes, such as performance, sound art and music.

The present dissertation proposed to develop an interactive interface and, consequently, evaluate and analyze the concept. During the development of the project, the interface's design was defined and an initial functional prototype was created, including a small sound library used in the usability tests. However, the method of evaluation and analysis initially intended didn't progress due to the small sample of participants in the usability tests and the inability to demonstrate the interface to the professionals interviewed. As a consequence, the results obtained can't be interpreted as proof of concept, due to the premise's subjectivity and the early-stage of the project. Nevertheless, they provided possible guides for potential future work, such as the deviation from an interface for Foley to an interface based on Foley and the validation of the electronic equipment.

As already mentioned, the theme of this dissertation can still be developed. Therefore, future work suggestions are the investigation of realistic sound synthesis techniques to be used by the interface, the construction of a more refined prototype and the experimentation of creating the soundtrack of an artistically relevant audiovisual artefact.

In conclusion, the answer to the question: “Is it possible to create some kind of physical interface with a computer system that could assist in the production of Foley sound effects?” is still unknown, but the first step was taken.

This page is intentionally left blank.

References

Aly, L., Penha, R., & Bernardes, G. (2017). Digit: A digital foley system to generate footstep sounds. In *International symposium on computer music multidisciplinary research* (pp. 429–441).

Ament, V. T. (2014). *The foley grail: The art of performing sound for film, games, and animation.* CRC Press.

Azadi, A. (2016, August 1). *Diachrony - transforming movement into sound and light (2016) (premiere).* [Video]. Youtube. `https://www.youtube.com/watch?v=qdjSOnU2TQg`.

Ballou, G. (2013). *Handbook for sound engineers.* Taylor & Francis.

Bevilacqua, F., Schnell, N., Rasamimanana, N., Bloit, J., Flety, E., Caramiaux, B., ... Boyer, E. (2013). De-mo: designing action-sound relationships with the mo interfaces. In *Chi'13 extended abstracts on human factors in computing systems* (pp. 2907–2910).

Blaine, T., & Perkis, T. (2000). The jam-o-drum interactive music system: a study in interaction design. In *Proceedings of the 3rd conference on designing interactive systems: processes, practices, methods, and techniques* (pp. 165–173).

Boem, A. (2014, 09). Sculpton: A malleable tangible interface for sound sculpting..

Botella, C. (n.d.). Susan philipsz: Sound as invisible sculpture. *ARTPULSE.* (Retrieved Aug 22, 2020, from `http://artpulsemagazine.com`)

Buhler, J., Neumeyer, D., & Deemer, R. (2010). *Hearing the movies: music and sound in film history.* Oxford University Press New York.

Cancellaro, J. (2005). *Exploring sound design for interactive media (design exploration series).* Thomson Delmar Learning.

Castelo Branco, S. (2019, July 06). Alexandre estrela: Métal hurlant. *Contemporânea.* (Retrieved Jul 13, 2020, from `https://contemporanea.pt/`)

Chion, M. (1990). *Audio-vision: Sound on screen.* Columbia University Press.

Christensson, P. (2012). *Daw definition.* Retrieved 2020, Aug 20, from `https://techterms.com`.

Christensson, P. (2017). *Midi definition.* Retrieved 2020, Aug 21, from `https://techterms.com`.

Cipriani, A., & Giri, M. (2010). *Electronic music and sound design* (Vol. 1). Contemponet.

Ciumakova, I. (2018). *Sound art as media art legacy: The documentation of endangered practice* (Doctoral dissertation). doi: 10.13140/RG.2.2.36040.78080

Cook, P. R. (1997). Physically informed sonic modeling (phism): Synthesis of percussive sounds. *Computer Music Journal*, *21*(3), 38–49.

Cook, P. R. (2002). Modeling bill's gait: analysis and parametric synthesis of walking sounds. In *Audio engineering society conference: 22nd international conference: Virtual, synthetic, and entertainment audio.*

Dakic, V. (2009). *Sound design for film and television.* Grin Verlag.

De Poli, G. (1983). A tutorial on digital sound synthesis techniques. *Computer Music Journal*, *7*(4), 8–26.

de Souza Nunes, H., Visi, F., Coelho, L. H. W., & Schramm, R. (2019). Sibilim: A low-cost customizable wireless musical interface. In *Nime* (pp. 15–20).

Dubnov, S. (2000). Martin russ: Sound synthesis and sampling. *Computer Music Journal*, *24*(3), 71–71.

Essl, G., & O'Modhrain, S. (2005, January). Scrubber: An interface for friction-induced sounds. In (p. 70-75).

Everest, F. A., & Pohlmann, K. C. (2015). *Master handbook of acoustics.*

Fadigas, F. (2015, July 09). Collage i [audio file]. (Retrieved Jul 13, 2020, from `https://soundcloud.com/fernandofadigas/collage-i-festival-silencio-2015`)

Fantel, H. (1992, May 03). Home entertainment; in the action with 'star wars' sound. *The New York Times*. (Retrieved January 13, 2020, from `https://www.nytimes.com`)

Farnell, A. J. (2013). Designing sound: procedural audio research based on the book by andy farnell.

Fontana, F., & Bresin, R. (2003). Physics-based sound synthesis and control: crushing, walking and running by crumpling sounds. In *Proc. colloquium on musical informatics* (pp. 109–114).

Giancoli, D. C. (2016). *Physics: principles with applications.* Boston: Pearson,.

Hahn, M. (2020, February 05). Sound design: 7 ways to create your perfect sound. *LANDR*. (`https://blog.landr.com/sound-design/`)

Henrique, L. L. (2002). *Acústica musical.*

Jewel, D. (2014). *The secret world of foley.* [Documentary]. Pinewood Studios.

Kahn, D. (1999). *Noise, water, meat: A history of sound in the arts.* MIT press.

Karplus, K., & Strong, A. (1983). Digital synthesis of plucked-string and drum timbres. *Computer Music Journal*, *7*(2), 43–55.

Koblyakova, M. (2017, March 21). Hear 435 cardboard boxes create a minimalist sound symphony. *Vice*. (Retrieved Jul 16, 2020, from `https://www.vice.com/`)

Kristjánsson, S. (2016). *Sound: effects* (Unpublished doctoral dissertation). Iceland Academy of the Arts.

Licht, A. (2009). Sound art: Origins, development and ambiguities. *Organised Sound*, *14*(1), 3–10.

Lorinc, B. M. (2015). *Sound module for interactive transport simulators* (Unpublished doctoral dissertation). University of Applied Sciences Technikum Wien.

McPherson, A. P., Jack, R. H., Moro, G., et al. (2016). Action-sound latency: Are our tools fast enough?

Moffat, D., Selfridge, R., & Reiss, J. D. (2019). Sound effect synthesis. *Foundations in Sound Design for Interactive Media: A Multidisciplinary Approach. Routledge*.

Molina, L. (2016). *Actors of sound.* [Documentary]. Freestyle digital media.

Moss, W., Yeh, H., Hong, J.-M., Lin, M. C., & Manocha, D. (2010). Sounding liquids: Automatic sound synthesis from fluid simulation. *ACM Transactions on Graphics (TOG)*, *29*(3), 1–13.

Murray-Browne, T., Aversano, D., Garcia, S., Hobbes, W., Lopez, D., Sendon, T., ... Chapman, D. (2014). The cave of sounds: An interactive installation exploring how we create music together. In *Proceedings of the international conference on new interfaces for musical expression* (pp. 307–310).

Neill, B. (2017). The mutantrumpet. In *Nime.*

O'Sullivan, S. (2001). The aesthetics of affect: Thinking art beyond representation. *Angelaki: journal of theoretical humanities*, *6*(3), 125–135.

O'Sullivan, S. (2012, February 02). *The basics of sound synthesis.* Retrieved 2020, Aug 13, from `https://theproaudiofiles.com/`.

Paradiso, J. A., & O'modhrain, S. (2003). Current trends in electronic music interfaces. guest editors' introduction. *Journal of New Music Research*, *32*(4), 345–349.

Penha, R. (2015). Resono. `http://ruipenha.pt/works/resono/`. ([Online; accessed 2020-01-15])

BAFTA Guru. (2018). *Inside the pinewood foley studio | bafta guru.* [Video]. Youtube. `https://www.youtube.com/watch?v=tQl_-MghIjo`.

GameSpot Trailers. (2012). *Sound effects - batman: Arkham city behind-the-scenes video.* [Video]. Youtube. `https://www.youtube.com/watch?v=g5IMQ12hI3w`.

Great Big Story. (2017). *The magic of making sound.* [Video]. Youtube. `https://www.youtube.com/watch?v=UO3N_PRIgX0`.

Rudi, J. (2011). Soundscape in the arts. *Oslo: NOTAM*.

Sonnenschein, D. (2001). *Sound design: The expressive power of music, voice, and sound effects in cinema.* Michael Wiese Productions Studio City.

Todorovic, D. (2008). *Gestalt principles', scholarpedia, 3 (12): 5345.* Online.

Truax, B. (2001). *Acoustic communication.* Greenwood Publishing Group.

Turchet, L., Serafin, S., Dimitrov, S., & Nordahl, R. (2010). Physically based sound synthesis and control of footsteps sounds. In *Proceedings of digital audio effects conference* (Vol. 11).

Walther, S., Müller, M., Brade, M., & Groh, R. (2013). Shaping sounds-a vision for tangible music interaction.

Young, S. (2015). Nocturne. `https://www.thismusicisfalse.com/nocturne`. ([Online; accessed 2020-04-07])

Appendices

This page is intentionally left blank.

Appendix A

Manual used for the Usability Tests.

Jack-in-the-Mug

Manual - Testes de Usabilidade

I Parte: Introdução à interface

1. - Bater com a base da interface física sobre a mesa. Reproduzir o mesmo movimento aplicando diferentes intensidades. Analisar o som reproduzido.

2. - Alterar para interação "Esfregar" usando a interface gráfica.

3. - Coloque a interface física sobre a mesa e faça movimentos sem a levantar. Teste movimentos lentos e rápidos. Analisar o som reproduzido.

4. - Levante a interface física da mesa e faça movimentos sem a aproximar do tampo da mesa. Teste movimentos lentos e rápidos. Analisar o som reproduzido.

5. - Teste os vários tipos de movimentos alternando entre as tarefas 3 e 4. Analisar o som reproduzido.

6. - Alterar para interação "Agitar" usando a interface gráfica.

7. - Realize movimentos aleatórios com a interface física afastada do tampo da mesa fazendo interrupções. Teste movimentos lentos e rápidos. Analisar o som reproduzido.

8. - Alterar para interação "Rodar" usando a interface gráfica

9. - Coloque a interface física a 90º em relação ao chão. Incline a mesma para a frente. Analisar o som reroduzido.

10. - Coloque a interface física a 90º em relação ao chão. Incline a mesma para os lados. Analisar o som reproduzido.

11. - Coloque a interface física a 90º em relação ao chão. Rode a mesma para dentro e para fora. Analisar o som reproduzido.

12. - Teste os vários tipos de movimentos entre as tarefas 9, 10 e 11. Analisar o som reproduzido.

13. - Alterar para interação "Bater" usando a interface gráfica.

14. - Desativar o som usando a interface gráfica. Realize movimentos com a interface física. Analisar se existiu reprodução do som.

15. - Ativar o som usando a interface gráfica. Realize movimentos com a interface física. Analisar se existiu reprodução do som.

16. - Pressione "Iniciar gravação" usando a interface gráfica. Enquanto o vídeo decorre, realize movimentos com a interface física. Verifique se a gravação terminou no mesmo momento que o vídeo.

II Parte: Produção Foley

Bater:

1. - Bater com a base da interface física sobre a mesa. Reproduzir o mesmo movimento aplicando diferentes intensidades. Analisar o som reproduzido.

2. - Pressione “Reproduzir” usando a interface gráfica. Aplique a técnica de Foley usando a interface física. Repetir até se familiarizar com o técnica e vídeo.

3. - Pressione “Iniciar gravação” usando a interface gráfica. Aplique a técnica de Foley usando a interface física.

4. - Analisar o conteúdo audiovisual produzido.

Esfregar:

1. - Esfregar com a base da interface física sobre a mesa. Reproduzir o mesmo movimento aplicando diferentes intensidades. Analisar o som reproduzido.

2. - Pressione “Reproduzir” usando a interface gráfica. Aplique a técnica de Foley usando a interface física. Repetir até se familiarizar com o técnica e vídeo.

3. - Pressione “Iniciar gravação” usando a interface gráfica. Aplique a técnica de Foley usando a interface física.

4. - Analisar o conteúdo audiovisual produzido.

Agitar:

1. - Agite a interface física afastada da mesa. Reproduzir o mesmo movimento aplicando diferentes intensidades. Analisar o som reproduzido.

2. - Pressione “Reproduzir” usando a interface gráfica. Aplique a técnica de Foley usando a interface física. Repetir até se familiarizar com o técnica e vídeo.

3. - Pressione “Iniciar gravação” usando a interface gráfica. Aplique a técnica de Foley usando a interface física.

4. - Analisar o conteúdo audiovisual produzido.

Rodar:

1. - Rodar/Inclinar a interface física. Reproduzir o mesmo movimento aplicando diferentes intensidades. Analisar o som reproduzido.

2. - Pressione “Reproduzir” usando a interface gráfica. Aplique a técnica de Foley usando a interface física. Repetir até se familiarizar com o técnica e vídeo.

3. - Pressione “Iniciar gravação” usando a interface gráfica. Aplique a técnica de Foley usando a interface física.

4. - Analisar o conteúdo audiovisual produzido.

This page is intentionally left blank.

Appendix B

Questionnaire used for the Usability Tests.

1 2 90 UNIVERSIDADE D COIMBRA

Faculdade de Ciências e Tecnologia
Mestrado em Design e Multimédia
Dissertação 2019/2020

Jack-in-the-Mug

Questionário - Testes de Usabilidade

Dados Pessoais:
Idade: _____ Sexo ___

Avaliação: 1 - Negativo | 5 - Excelente

I Parte: Introdução à interface

Bater:

1. - Avalie, numa escala de 1 a 5, a responsividade da interface em relação ao momento de choque e início de reprodução sonora. ___

2. - Avalie, numa escala de 1 a 5, o volume do som reproduzido em relação à intensidade incutida durante o momento de colisão. ___

3. - Encontrou algum erro durante o uso desta interação? Se sim, o que ocorreu?

Esfregar:

1. - Avalie, numa escala de 1 a 5, a responsividade da interface em relação ao momento de sobreposição sobre a mesa e início de reprodução sonora. ___

2. - Avalie, numa escala de 1 a 5, o tom do som reproduzido em relação à intensidade do movimento. Movimento lento deve corresponder a sons graves e movimentos rápidos deve corresponder a sons agudos. ___

3. - Encontrou algum erro durante o uso desta interação? Se sim, o que ocorreu?

Agitar:

1. - Avalie, numa escala de 1 a 5, a responsividade da interface em relação ao ínicio do movimento e início de reprodução sonora. ___

2. - Avalie, numa escala de 1 a 5, o tom do som reproduzido em relação à intensidade do movimento. Movimento lento deve corresponder a sons graves e movimentos rápidos deve corresponder a sons agudos. ___

3. - Avalie, numa escala de 1 a 5, a responsividade da interface em relação ao momento de mudança de direção de movimento e início de reprodução do som “pop”. ___

4. - Encontrou algum erro durante o uso desta interação? Se sim, o que ocorreu?

Rodar:

1. - Avalie, numa escala de 1 a 5, a responsividade da interface em relação ao volume dos sons produzidos perante a rotação da caneca nos 3 eixos. ___

2. - Encontrou algum erro durante o uso desta interação? Se sim, o que ocorreu?

II Parte: Produção Foley

Bater:

1. - Avalie, numa escala de 1 a 5, a interação escolhida em relação ao apoio visual. ___

2. - Avalie, numa escala de 1 a 5, o grau de qualidade/realismo do som produzido. ___

3. - Avalie, numa escala de 1 a 5, o grau de sincronização som-imagem do produto obtido. ___

4. - Avalie, numa escala de 1 a 5, o grau de satisfação em relação ao produto obtido. ___

Esfregar:

1. - Avalie, numa escala de 1 a 5, a interação escolhida em relação ao apoio visual. ___

2. - Avalie, numa escala de 1 a 5, o grau de qualidade/realismo do som produzido. ___

3. - Avalie, numa escala de 1 a 5, o grau de sincronização som-imagem do produto obtido. ___

4. - Avalie, numa escala de 1 a 5, o grau de satisfação em relação ao produto obtido. ___

Agitar:

1. - Avalie, numa escala de 1 a 5, a interação escolhida em relação ao apoio visual. ___

2. - Avalie, numa escala de 1 a 5, o grau de qualidade/realismo do som produzido. ___

3. - Avalie, numa escala de 1 a 5, o grau de sincronização som-imagem do produto obtido. ___

4. - Avalie, numa escala de 1 a 5, o grau de satisfação em relação ao produto obtido. ___

Rodar:

1. - Avalie, numa escala de 1 a 5, a interação escolhida em relação ao apoio visual. ___

2. - Avalie, numa escala de 1 a 5, o grau de qualidade/realismo do som produzido. ___

3. - Avalie, numa escala de 1 a 5, o grau de sincronização som-imagem do produto obtido. ___

4. - Avalie, numa escala de 1 a 5, o grau de satisfação em relação ao produto obtido. ___

Geral

1. - Avalie, numa escala de 1 a 5, a intuitividade da interface física. ___

2. - Avalie, numa escala de 1 a 5, o grau de facilidade de sincronização do som com imagem na Parte II - Produção Foley. ___

3. - Avalie, numa escala de 1 a 5, o grau de diversão que sentiu ao utilizar a interface. ___

4. - Indique até três aspetos negativos da experiência.

5. - Indique até três aspetos positivos da experiência.

This page is intentionally left blank.

Appendix C

Answers to the Usability Test questionnaire from the testers.

	DADOS PESSOAIS		PERGUNTAS GERAIS		
ID Tester	Age	Gender	1	2	3
1	24	M	5	4	5
2	24	F	5	5	5
3	24	M	5	4	5

	I PARTE															
	Bater				Esfregar				Agitar				Rodar			
ID Tester	1	2	3		1	2	3		1	2	3	4	1	2		
1	5	3	N		5	5	N		5	3	1	S	5	N		
2	5	5	N		5	5	N		4	5	1	S	5	N		
3	5	4	N		5	5	N		5	4	1	S	5	N		

	II PARTE															
	Bater				Esfregar				Agitar				Rodar			
ID Tester	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	5	3	4	4	5	3	3	3	3	4	2	2	4	5	5	5
2	5	4	4	5	5	3	3	4	5	3	3	3	5	5	5	5
3	5	4	4	4	5	4	3	4	5	3	3	4	5	5	5	5

COMENTÁRIOS NEGATIVOS

Tester 1
- Som dos vidros não tem muita qualidade
- Os vídeos começaram abruptamente com a ação - dificultando a sincronização

Tester 2
- Som dos videos não realista
- Esfregar som não para c/ movimento
- Pega não ergonómica

Tester 3
- Pega solta
- Som vidros não apropriado
- Esfregar só funciona rápido

COMENTÁRIOS POSITIVOS

Tester 1
- Interessante
- Inovador

Tester 2
- Imperativo
- Util (muitos objetos num só)
- Intuitivo

Tester 3
- Prático
- Fácil utilização
- Multifuncional, versátil