論文 / 著書情報
Article / Book Information

題目(和文)	ヒトのマルチモーダル知覚特性に基づく聴触覚刺激の合成
Title(English)	Synthetic audio-tactile stimuli generation based on human multi-modal perception
著者(和文)	アルフォンソ バランドラ
Author(English)	Alfonso Balandra
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第11114号, 授与年月日:2019年3月26日, 学位の種別:課程博士, 審査員:長谷川 晶一,小野 功,中本 高道,三宅 美博,宮下 英三
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第11114号, Conferred date:2019/3/26, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

TOKYO INSTITUTE OF TECHNOLOGY

DOCTORAL THESIS

Synthetic audio-tactile stimuli generation based on human multi-modal perception

Author:
Alfonso BALANDRA

Supervisor:
Asoc. Prof. Shoichi HASEGAWA

Faculty Member Judges:

Prof. Nakamoto Takamichi

Prof. Miyake Yoshihiro

Asoc. Prof. Miyashita Eizo

Asoc. Prof. Ono Isao

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Engineering

Department of Computer Intelligence and Systems Science

February 7, 2019

A mis padres y a mi hermano.

Gracias Gracias Gracias

"If you want to find the secrets of the universe, think in terms of energy, frequency and vibration."

Nikola Tesla

Tokyo Institute of Technology

Abstract

Department of Computer Intelligence and Systems Science

Doctor of Engineering

Synthetic audio-tactile stimuli generation based on human multi-modal perception

by Alfonso BALANDRA

In this document we propose a novel multi-modal listening environment to enrich the music listening though the use of a redundant haptic and visual stimuli. The environment, which we called the Haptic Music Player, uses music score to create a visual animation and a synthetic haptic vibration that resembles the sound of one instrument in the song. To develop the synthetic haptic vibration several psychophysical experiments were performed, where the similarity perception between an audio and haptic signal was evaluated. Also the impact of a redundant audio-tactile and visual stimuli was evaluated. The experiment's results demonstrate that the attack synchrony and resemblance of the envelope shapes, between the audio and haptic signals, are important to be similarly perceived. Also we found that the selective listening of the users improves when using a audio-tactile-visual stimuli. All these findings were applied on the implementation of the Haptic Music Player. Therefore, though the Haptic Music Player the user is able to easily identify and follow any instrument in a song. Also, this environment helps the user to understand the melody, structure, and the role of each instrument in a song and consequently the user could have a deeper understanding of music.

Contents

Chapter 1

Introduction

1.1 Background

When listening to a live concert usually we are able to hear the music though our ears, but we can also feel the vibration from the sound source though our body. This phenomena naturally enhance the listening experience, mostly because sound and vibration are part of the same natural phenomena, so is natural that when we hear a loud sound we expect also to feel its correspondent vibration.

The same happens when we feel or manipulate objects in our daily life. When we haptically manipulate objects the interaction creates vibration on the surface of the manipulated objects, and these vibration can also generate sound. So, is also natural for us to expect sound when we haptically interact with the objects around us. For example when: knocking on a door, writing with a pencil or use the computer keyboard.

We can listen to recorded music in many formats and environments: in the car, though a portable player, in the cinema and of course in home. Nevertheless, the formats used to play music cannot recreate the haptic sensation of been in a concert hall listening to a live concert. The speakers or headphones that we frequently use to play music are able to play the recorded sound with high fidelity. The haptic vibration can also be recorded with a microphone or with an accelerometer, but is difficult to recreate the complete vibration sensation though the complete human body.

An easy solution to recreate that haptic sensation though the complete body, could be just increase the volume of the music by using very big speakers. By these means, the increase on the mechanical vibration of the speakers may also increase the vibration in the air and floor and consequently create a haptic sensation. This technique is simple and if the conditions are favorable the haptic sensation can be pleasant and consistent with the sound.

But, there are some environments and applications where the audio signals cannot be enhanced by using loud or big speakers. In addition if the loud speakers are used as a haptic signal, only the low spectrum of the signal will be perceived as a tactile sensation. Which for some kind of applications may be an appropriate solution for other applications it is not. For example: with this technique is difficult to selectively enhance high pitch instruments, like a guitar or piano, without being masked by low register instruments, like bass or drums. In addition, this specific technique implies that the audio signal must be displayed with a high volume.

Usually handheld devices like: cellphones, tablets and video games incorporate haptic actuators inside them, so they can provide a haptic feedback to the user. But, in most cases these handheld devices do not have good quality haptic vibrators. Usually to reduce costs, these devices had cheap, rotary mass vibrators (RMV) or a

FIGURE 1.1: This figure shows diverse haptic actuators used in modern handheld devices. **a)** The haptic actuator used in Apple cellphones, where the amplitude of the signal can be controlled [45]. **b)** Oculus touch controller, equipped with haptic actuators with a refresh rate of 320Hz and a frequency range of 0~160Hz. **c)** a state of the art lineal actuator with a wide frequency range of 0~450Hz. [11]

pager motors, these devices usually just allow on-off vibration, so the amplitude or frequency of the vibration cannot be controlled.

Nowadays, newer products are provided with better haptic interfaces. Usually haptic transducers are used in order to provide a better haptic feedback. For example the Occulus Rift VR controllers have a haptic actuator with a refresh rate of 320Hz and a frequency range of 0~160Hz. Also the the Apple cellphones, have a new type of mechanical haptic actuator where the amplitude of the signal can be controlled.

Even if the new haptic interfaces provide better and better haptic feedback, these haptic feedback still has limitations. Usually the frequency and the bandwidth of these specific actuators cannot be used to directly display audio as a haptic signal. Also, the maximum amplitude that these devices can display is very low, because these were designed to save battery life. Also, most of the cases the length and strength of the haptic signal is controlled by the device's API (see Figure 1.1). So due to the previously mentioned limitations, we can conclude that the direct usage of an audio signal as a haptic signal in these kind of haptic devices will not have the best results.

Therefore, due to the clear limitations of the haptic actuators in mobile devices, an audio signal cannot be directly displayed on mobile devices, like the ones we previously mentioned (see Figure 1.1). So we consider unfeasible to directly use a audio signal on a hand held device vibrator to create representative haptic vibrations to resemble the audio source. The vibration may be have a weak sensation and/or the vibrations may not be similar to the audio source they want to resemble. So a different strategy must be considered in order to create a resemblant audiotactile sensation for mobile devices.

Another important factor that must be considered is the signal output latency that these haptic devices have. Even for specialized devices for gaming and virtual reality is technically difficult to provide a perfect synchrony between the haptic signal and the visual simulation or video game and audio signal related to the haptic cue. So due to these technical reasons and the hardware limitations of the haptic actuators, seems necessary to perform a perception analysis in order to understand which is the best method to crate a representative haptic vibration of an audio signal.

1.2 Haptic Music Player - Proposal

As mentioned before, directly use a audio signal as a haptic signal, may be simple and straight forward solution to create a haptic signal that resembles an audio signal. But this trivial technique has some limitations: the audio signal must be amplified, it is noisy and in most of the cases only the low frequency of the audio spectrum can be perceived. For handheld devices, if an audio signal is used directly on the device's vibrator, due to the hardware limitations, the signal may be weak and it may not be similar to the sound signal.

Then, for these cases it seem mandatory to modify or enhance the audio signal in order to overcome the hardware limitations, like Chang's did in their proposal of 2005 [9], where she proposed a combination of audio filters and signal amplification techniques to improve the haptic signal sensation for specific hardware. More recently, Hwang [25] also proposed a similar technique to amplify the signal on certain haptic frequency bands, also a solution designed for specific hardware.

In contrast, to Chang's and Hwang's proposal, we believe that is necessary to find a general method and solution to this specific problem. So we propose to study the multimodal audiotactile human perception in order to design a method to create synthethic haptic vibrations that have a high resemblance to their respective audio source and also suitable to display on diverse kinds of haptic hardware. Then we also want to show the results of this specific proposal by developing an specific piece of software, which we called the Haptic Music Player, that enhance the music listening experience of the users though the usage of vibration and score visualization.

In order to find a general method to translate an audio signal into a resemblant haptic vibration, is mandatory to understand and explore the limitation and capabilities of the human perception of synchronized audio tactile signals. In psychophysics, the area responsible of research the human psychological ability to perceive reality, provides many and reliable information about the human auditive and somatosensory sense abilities and limitations. For example, there are some studies that mention that, the volume of an audio signal can be biased though the a synchronized haptic signal [47]. But there is not detailed previous documentation about the human multi-modal perception of audiotactile stimuli. So we need to know more specific information about the human perception of audiotactile signal in order to propose a general method to create a resemblant haptic signal.

Finally, the results from the performed psychophysical evaluation, are going to be directly applied in the development of the Haptic Music Player. The Haptic Music Player is a multimodal enhancement listening environment that is meant to enhance the users music selective listening by displaying a redundant haptic and visual stimuli. Also we propose to evaluate the effectiveness of the proposed environment using another psychophysical experiment meant to evaluate if the redundant visual-tactile stimuli of the Haptic Music Player is indeed effective to enhance the users' selective listening (see Figure 1.2).

1.3 Haptic Music Player - Motivation

The principal motivation for the implementation of this proposal was the realization of the lack of proper psychophysical studies on this specific subject. We consider

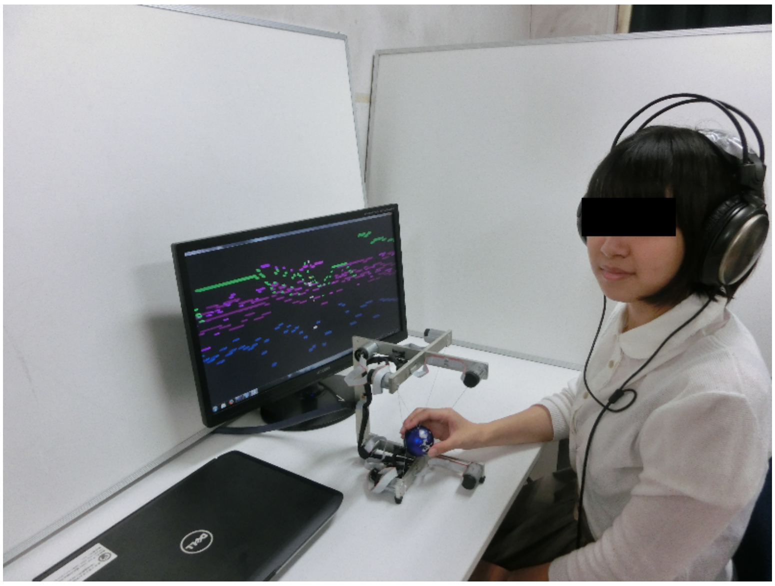


FIGURE 1.2: This figure shows the complete Haptic Music Player environment. The environment plays a MIDI song, along the music, while the user can see a graphical representation of the song melody and also can feel the vibration of an specific instrument of the song. Our hypothesis is that, the complete system helps the users to focus their listening attention to have a closer and deeper understanding of the music.

necessary to explore the human audiotactile perception of synchronized audiotactile signals. In specific, we are interested on explore the characteristics of the signal that have more impact their perceived similarity. So by these means we can propose a general method to optimize the haptic signal characteristics to maximize its subjective resemblance to an specific audio signal, using limited hardware.

In addition, other possible motivation to create a music resemblant haptic vibration was, that some people do not have the enough music education or the capabilities to understand music with the same level as a trained musician would do [34]. Then we propose that by using an specific kind of visual and haptic stimuli, we can help untrained people to focus their listening attention to understand the music melody, structure and rhythm of a song, in the same way a trained musician would do.

Chapter 2

Related Research

2.1 Previous Research

In this chapter we will talk about the previous efforts related to this proposal. The following proposal are related to this research in our final goal, to enhance the music listening experience by using haptics. Most of this proposals, use diverse types of hardware and methodologies to create, enhance, modify or design the audiotactile signal, but the common point between all these articles is the final goal to enrich the audiotactile music listening experience.

2.1.1 Early proposals

One of the erliest proposal that we could find was a proposal by Gunther [19] in 2002. Gunther proposed a wearable design haptic interface a full bodysuit with several types of transducers attached to it. They used the V1220 a coil-based transducer and an an old commercial video game low-frequency transducer called Interactor. While 12 V1220 where attached to several part of the body (wrists, shoulders, ankles, thighs and arms) the Intractor has disassembled and attached to the lower back. Then all the transducer are connected to a Digi001 audio interface, and the complete system is controlled with MIDI while the audio signals are specifically made for each composition.

Essentially, Gunther interface is a experimental music haptic interface that can be used for music performance or experience. Gunther proposal considers in some degree the tactile and auditory human perception to design the signals. Even so, this proposal is very open, Gunther do not clearly specify the goal of the project or the possible applications that his proposal may have, but it is clear that the final goal is to enhance the music listening experience though haptics.

Another of the earliest attempts to improve the audio-haptic signals was made by Chang [9] in 2005. She proposed an audio manipulation technique for an specific multifunction transducer (Multi Function Transducer) . Chang's proposed technique enhance and audio signal to improve the haptic signal quality and perception for her specific hardware the MFT.

To explain Chang's work is necessary to explain what is an MFT. An MFT is an special kind of speaker capable of displaying both and audible signal and a vibrotactile output at the same time from an audio signal. For the MFT the haptic effects of the audio signal can only be displayed between 100 HZ to 300 HZ, frequencies above this threshold only generated and audio signal.
Consequently, Chang focused herself to enhance the audio signal bellow the 300 Hz.

So she used an specific audio manipulation technique divided on several steps, first she separated the haptic signal and the audio signal by a stop band filter. Then if the signal amplification levels are low, then the haptic signal is enhanced. But if not, then no further modification is applied and the haptic signal is remixed with the audio signal and displayed. But if the haptic signal is weak, then the signal is enhanced by matching several combinations of haptic texture or haptic icons to the temporal amplitude characteristics of the signal, after this improvement the signal is remixed with the audio signal and displayed in the MFT.

So, similar to our proposal Chang proposed a new way to create a haptic signal that resembles the the haptic components of an audio signal. Also similar to us, she considered the limitation of the a low-end hardware (MFT) to display such enhanced signal. In addition, she also performed a subjective perception experiment to evaluate the haptic perception of the signal. But, contrary to our proposal instead of enhancing a audio signal we propose to create a completely synthetic haptic signal. This give us the versatility to modify the signal accordingly to the specifications of different kinds of haptic interfaces. Also contrary to Chang's evaluation, were only the final results were evaluated, the purpose of our perception experiments was to find the properties of an audio-tactile signal that enhance it's similarity perception.

2.1.2 Haptic chairs

In the year 2009 two different but similar proposal were published one by Karam [27] and the other by Nanayakkara [35]. These proposal are very similar because both use as a haptic interface a chair with several speakers attached to it (see Figure 2.1).

Karam proposed a system composed of 8 different voice coils attached to the back of the chair. For the haptic signal Karam used a haptic signal methoaphore of the human cochlea to display the signal on the different voice coils placed in the back of the chair. Karam proposed a sensory substitution technique, were to exchange the characteristics between the audio and haptic modalities, for this she separated the audio signal in specific bands, choosing each band motivated by our bias towards the preservation of the original signal and our underlying stimulus organization, fundamentally based on Gunther's work [19]. Finally, she evaluated the system based on it's capability to display emotion though a audio-tactile signal.

Besides, Nanayakkara build a similar char with only 4 speakers, two stereo speakers mounted on the arm-rests, one on the food-rest and the other on the back-rest. The specific purpose of Nanayakkara research was to enhance the music listening experience of deaf people. So, instead of relying on synthetic, enhanced or haptic methaphores, he directly used the amplified audio signal as a haptic signal. Nanayakkara argumented that any additional information delivered though the haptic signal may disrupt the musical experience, also he mentions that there is not enough information about how the central nervous system process any kind of audio-tactile signal. In addition to the haptic-signal, Nanayakkara used a simple music animation so the deaf people will be able to have a better listening experience. Then he used MIDI to create and display a music animation. Finally, the complete system was subjectively evaluated by several deaf people, with positive results.

In contrasts to Karam's proposal, we tried to consider the human perception of audio-tactile signals to design the signal itself, not just to evaluate the final system

FIGURE 2.1: shows the haptic interface proposed by Karam, composed of 8 speakers on the back of the chair. Uses a haptic metaphor to display the signal [27] © [2009] IEEE.

result. Also, Karam proposed an interesting sensory substitution method but it is fundamentally based on roughly comparing the perceptual similarities between the audio and somatosensory senses and gradually improving it by evaluating the emotional characteristics presented on the substitution method.

Also Nanayakkara's haptic solution is similar to ours because Nanayakkara also proposed a music animation to enhance the music listening experience for deaf users. But opposite to Nanayakkara we believe that is necessary to evaluate the subjective human perception of visual-tactile signals, in order to understand how deaf-people perceive a visual-tactile stimulus. We consider that Nanayakkara proposal can be significantly improved if these psychophysical studies are performed.

2.2 State of the Art

2.2.1 Dual band Audiotactile music player

In 2013 Hwang proposed a novel method to crate a resemblant haptic vibration from music [25]. Hwang prosed a novel real-time vibration generation algorithm to extract the vibration commands from music, something similar to a "haptic equalizer" where the two haptic signals were extracted one that resembles the treble of the song and the other that resembles the bass of the song. This vibration generation technique was specifically designed for precise dual band haptic actuator called, DMA (Dual Mode Actuator). Compared to other vibrators the DMA is capable of displaying two principal frequencies displaying a superimposed vibration.

First the audio signal is segmented and then separated into two different bands (treble and bass) using the FFT (Fast Fourier Transform) and using some frequency threshold values. Then the treble band is further subdivided into five different subbands in a log scale, from 200Hz to 6400Hz. Then a weight is assigned to each band, where each weight determines the amplification gain for each sub-band. Then for

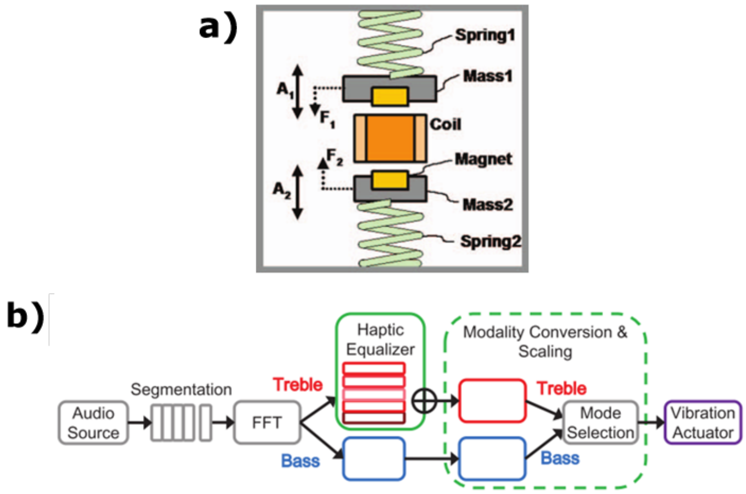


FIGURE 2.2: a) shows an internal structure diagram of the DMA (Dual Mode Actuator) b) shows a flowchart of the algorithm proposed by Hwang [25] © [2013] IEEE.

different types of music the the amplification weights are preset. After that both signals are then adjusted and scaled to the human perception frequency ranges, adjusting the range to the previously know haptic psychophysical magnitude functions. Finally, a subjective study was conducted by comparing the proposed method with a lineal and dual band modes (see Figure 2.2).

Hwangs work has a very strong foundation on audiotactile perception and his goal is also very similar to our proposal goal; to enhance the music listening experience though haptics. Even so, it is evident that Hwang specifically designed his algorithm to fit the characteristics of this specific device the (DMA), to apply the same algorithm on a different type of actuator could produce bad results or be impossible. Also Hwang algorithm has a strong haptic psychophysical foundation, but he overlooks the crossmodal audio-tactile perception relying just on the magnitude haptic perception. Contrary to Hwang's research we first want to generally analyse the cross-modal perception of simultaneous audio-tactile signals to then apply those findings to create a general and basic strategy that can be applied and modified to be used on diverse kinds of hardware.

2.2.2 HapTONE

A very recent proposal by Ogawa *et.al.*[37] is named HapTONE. This is an entertainment system that uses auditory, tactile and visual cues to enrich the players experience. Ogawa developed a haptic piano, in which each piano key an structured vibration and a distance sensor was placed. When the keyboard is played, it will reproduce the respective sound of the note and also a haptic vibration. In addition a visuals cues and interactive animations can also be projected on the keyboard. This entertainment environment is very similar to our proposal, because audio, haptics and visuals are used to enrich the experience.

In detail, Ogawa's work is outstanding. For haptics, he placed a vibration and a distance sensor on each key of the keyboard. So, the key distance and velocity are considered to accurately play and syncronize the audio and haptic signals. Also, Ogawa uses pushing action velocity to change the velocity of the audio and haptic signals accordingly. In addition, the keyboard is very versatile, because it can be programmed to play string, wind, percussion and even non-musical instruments.

In general, Ogawa's proposal is also similar to our proposal because Ogawa used audio, visual and haptics to enrich the music playing experience. But due to the different hardware used by Ogawa, his proposal is oriented to musicians instead of music listeners. In contrast, we propose to enrich the music listening experience and provide the user a closer and deeper experience with music.

In addition, Ogawa used decaying sinusoidal vibrations to present differences between instrument's material, like a wood and metal xylophone and glockenspiel. For percussions a recording and play back vibration data is displayed directly to the user. The precise approach and haptic interface used by HapTone is very straight forward, so a deeper audiotactile perception analysis was not required. Basically, because HapTONE is meant to be an playing music entrainment environment. On the other hand, in the Haptic Music Player the user cannot interfere with the music, the user just listens while receiving the vibration. In this particular case the interaction with the haptic signal is passive and it is leaded by the audio signal itself, consequently for us it was imperative to perform a deeper research and analysis about the human perception of synchronous audio-tactile signals.

Chapter 3

Human Perception

3.1 Auditive Sense

The auditory system in human and other mammals is located on the sides of the head. This helps us to easily locate horizontal location of the audio signal source. By their depth in the skull and function the human auditory system can be separated in: outer ear, middle ear and inner ear.

The outer ear consist of the auricle (also called pinna) and the ear canal. The function of the auricle and pinna is to focus the sound vibration into the tympanic membrane. Many mammals, like dogs and cats, have special muscles in their auricles to move them in certain direction to focus their hearing in different directions; but humans do not have this ability.

The middle ear consist diverse structures, bones and muscles located in the tympanic cavity. When the tympanic membrane moves or vibrates it also moves three small bones called: malleus, incus and stapes. The movement of these 3 bones transfer the sound energy from the ear drum to the oval window. These bones and muscles are not just responsible of the energy transmission between the outer and inner ear, they also protect the inner ear from loud noises by a reflex movement of the stapedius and tensory tympani muscles.

The inner ear consist of cochlea and the vestibular system. The vestibular system is just dedicated to coordinating movement and balance, while the cochlea is responsible of convert the sound patters from the outer ear into electrochemical impulses. Finally, these electric impulses will be transmitted to the brain though the auditory nerve (see Figure 3.1).

Humans have an outstanding performance to analyses the frequency and the harmonics of an audio signal, which lets us identify voices and the origin of certain sounds. This process starts in the cochlea. The human cochlea is an spiral-shaped structure located bellow the vestibular system and it surround the final portion of the auditory nerve. Inside, it can be split into 3 different sections the scala vestibuli, scala media and the scala tympani. Separating these cavities there are the Reissner's membrane and the basilar membrane, when the cochlea is stimulated by the middle ear, the fluid between the cavities vibrates and stimulates both membranes.

Attached to the basilar membrane the are two different types of cells, called the inner hair cells and the outer hair cells. These cells are arranged in rows accordingly to the frequency of sound they detect and every cell has very tiny and long hair like projections called sterocylia. And above the hair cells the tectorial membrane is attached to the Reissner's membrane. So, when the cochlea cavities vibrate also the membranes between the cavities vibrate causing the hair cells to brush along the tectorial membrane. Finally, the brushing of the sterocylia creates a electrochemical

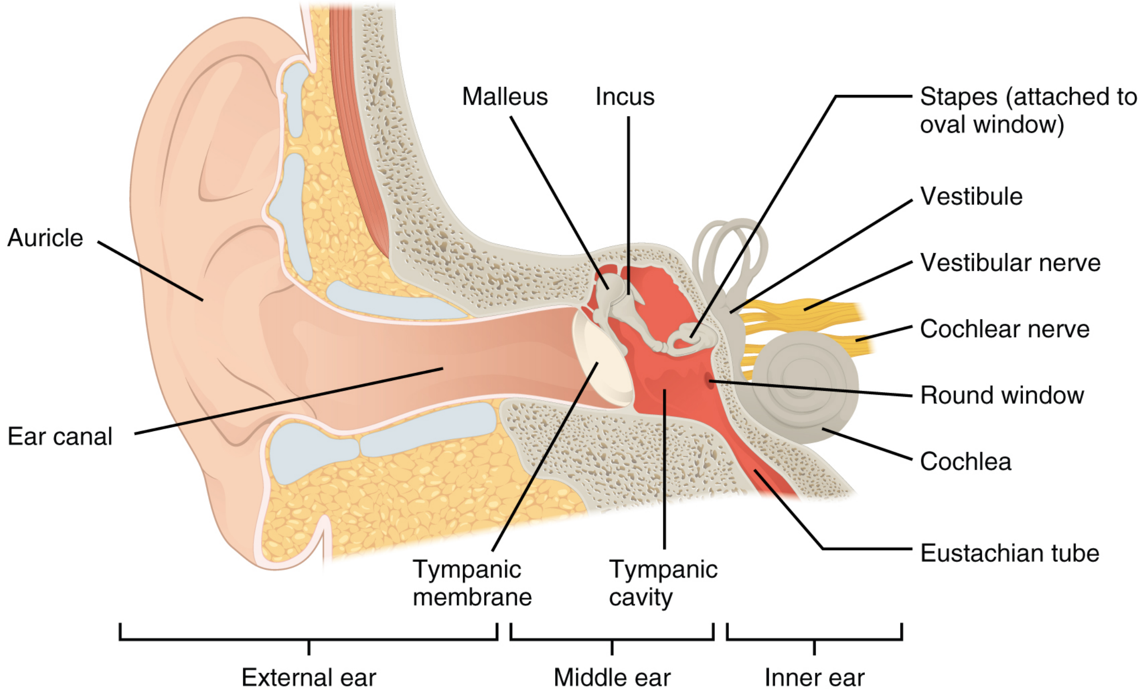


FIGURE 3.1: The figure shows a cross-section of the human ear. Here we can see the structures of the outer, middle and internal ear. (CC BY-SA 4.0) [38]

reaction on the hair cells and then the created electrochemical stimuli creating a neural impulse that travels along the auditory nerve into the brain. And this creates the perception of sounds that we hear on everyday life (see Figure 3.2).

3.2 Touch Sense

The somatosensory sense is the located all over the human body inside the skin. The human skin can have different external qualities like: hairy, hairless, thicker some parts, thinner in others. But just like our skin is not the same in every place, also the somatosensory sense is not the same all over skin. Mostly because the special receptors of the somatosensory sense are not evenly distributed inside the skin, these are distributed on different concentrations in our body. Evidently this is something easy to understand just by touching ourselves in different parts of our own body.

These receptors are also called mechanoreceptors because they respond to the mechanical stimulation and pressure. These receptors are located on the epidermis (outer layer) and the dermis (underlying layer) and consist of a "nerve fiber" and an associated expanded ending. On the hairless skin of the hand palm there are four different populations of tactile receptors, connected to different nerve fibers, as shown in the Figure 3.3. The four different groups of mechanoreceptors found in the hand palm are named after the anatomists who first described them: Meissner corpuscles, Merkel cell neurite complexes, Pacinian corpuscles and Ruffini endings.

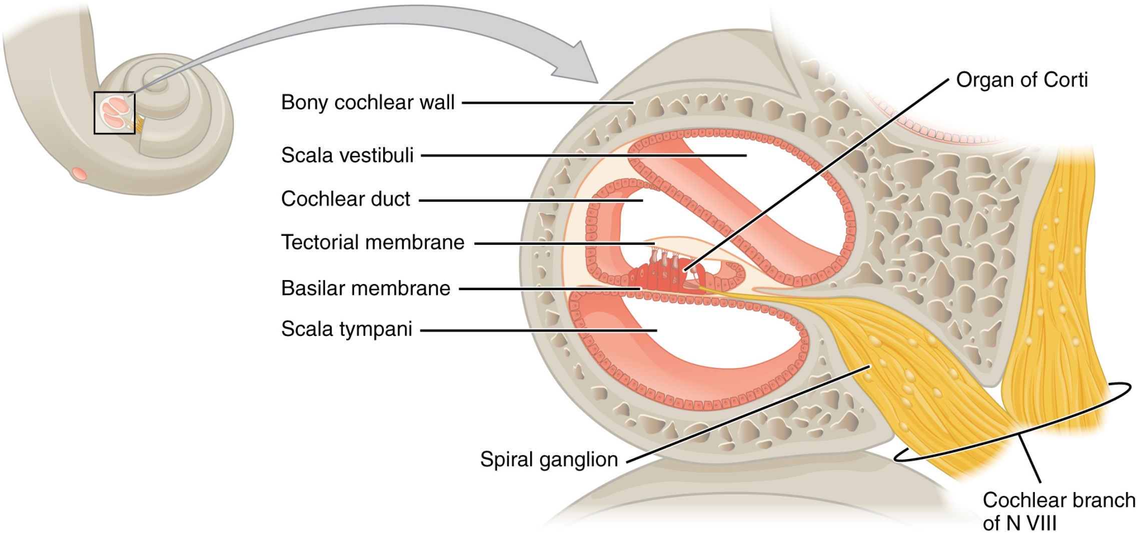


FIGURE 3.2: The figure shows an illustration inner human ear. Also the figure shows a diagram of the human cochlea and its internal structures. (CC BY-SA 4.0) [38]

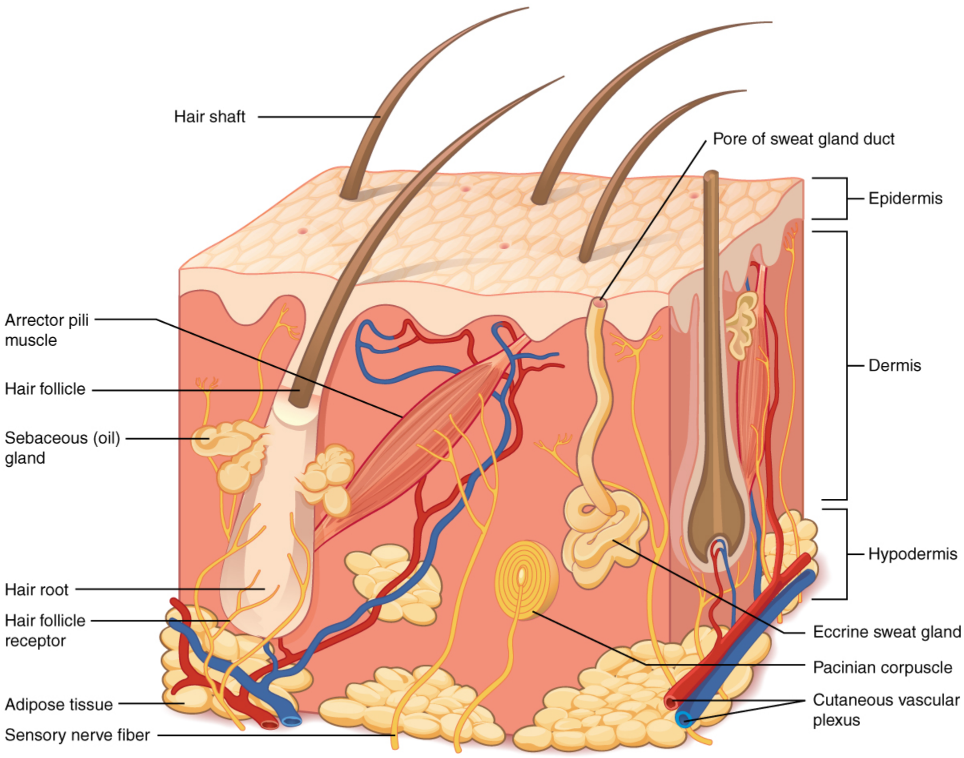


FIGURE 3.3: The figure shows a cross section of the hairless skin of the human hand. The figure shows the 4 types of different mechanorecpetors of the human hairless skin. Also the image shows, how deep in the human skin the different receptors are located. (CC BY-SA 4.0) [38]

Receptor	Feature Sensitivity	Primary Functions
SA I	Sustained pressure very low frequency (< ~5Hz) Spatial deformation	Texture perception, pattern and form perception
FA I	Temporal changes in skin deformation (~5Hz-50Hz)	Low frequency vibration detection
FA II	Temporal changes in skin deformation (~50Hz-700Hz)	High frequency vibration detection
SA II	Sustained downward pressure, lateral skin stretch skin slip (low sensitivity to vibration across frequencies)	Finger position, stable grasp

TABLE 3.1: Put some table.

The touch sense can be classified accordingly to their adaptation rates and the size of their receptive field of its nerve fibers. The receptive field of a nerve fiber is defined as the extend of the body area that triggers a nerve response and the adaptation rate of a nerver fiber is defined as the velocity of reaction of the receptor, that can be classified on fast-adapting fibers (FA) and slow-adapting fibers (SA). The fast-adapting fibers (FA) quickly respond to the stimulus with burst of nerve signals, these receptors are triggered when the stimulus is applied and with it is removed. On the other hand, the slow-adapting fibers (SA) remain active throughout the period during the stimulation. Then by using these specific mechanorector properties we can classified the touch sense fibers as shown in the Table 3.1.

If we analyse the data of the Table 3.1 we can see the difference between the fibers populations. The change in their behaviour is mostly caused by the type of mechanoreceptor the fiber is connected with. The fast-adapting fibers are connected to Meissner corpuscles (FA I) or with Pacinian corpuscles (FA II), while the slow-adapting fibers are connected to Merkel complexes (SA I) or with Ruffini endings (SA II). These to different types of mechanoreceptors work together to inform us about every individual object we touch or manipulate. Almost every simple task involves all the previously mentioned fibers and receptors. For example: to open a lock with a key we need to SA 1 to feel the key in the pocket, the SA II are used to grasp the key, FA I allows to dynamically adjust the grasping force to avoid the key to slip and finally the FA II allows to feel when the key had hit the end of the keyhole.

Along the previously mentioned receptors, the touch sense also has thermoreceptors and kinesthetic receptors. The kinesthetic receptors are special muscle fibers that are located along the muscles fibers, these special receptors allow use to sense the position of our limbs and the different kinds of movements that we are making. While, the thermoreceptors are special nerve fibers located along the epidermal and dermal layers of the skin that let us know temperature of the things we touch relative to our own temperature.

3.3 Multimodal Perception

To create a representative haptic signal from an audio signal for the entertainment environment(see Chapter 5). We plan to use the MIDI notes' pitch numeric values as input parameters to create a synthetic haptic signal with a distinguishable frequency and amplitude. Therefore it is necessary to map the frequency and amplitude of each note into a perceivable haptic range, so the different frequencies and amplitudes between the haptic stimuli of each note can create a melodic sensation. This strategy is similar as the method purposed by McLean and Enriquez [33]. In their study McLean and Enriquez used the term *haptic icons* to define brief haptic signals that have an implicit abstract message that can be used to convey to an specific object's event, function or state [33]. Though a multidimensional scaling analysis they suggested that frequency, wave shape and amplitude are the most dominant parameters that affect the differentiability between their *haptic icons*. Then from the reported results of McLean and Enriquez we can estimate that it also a good idea to use the amplitude changes of the haptic signal to represent different kinds of sounds in a haptic signal. Even so, we consider mandatory to understand and compare the perception abilities of the auditory of the somatosensory sense, to design an optimal strategy to be create a resemblant audiotactile signal.

3.3.1 Multimodal Timbre definition

In order to define a haptic signal that is similar to an auditive signal, first is very important to define what are the properties that make a audio signals distinguishable from each other, in order to emulate the same phenomena into haptics. So we have to clearly understand which are the properties of an sound that make it different from other similar auditive signal. These set of properties is also known as timbre.

The American National Standards Institute defines timbre as: "Timbre is that attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar. Timbre depends primarily upon the spectrum of the stimulus, but it also depends upon the waveform, the sound pressure, the frequency location of the spectrum, and the temporal characteristics of the stimulus." [4]. So if this definition is considered, the timbre of an auditory sensation is given by:

1. Pitch
2. Spectrum
3. Waveform
4. Sound pressure
5. Frequency spectrum
6. Amplitude temporal characteristics.

In order to simplify this definition, we decided to limit the timbre's definition to only 3 physical audio properties: pitch (fundamental frequency), amplitude or time envelope (temporal amplitude envelope) and the frequency spectrum (harmonics). So we can focus on these audio signal properties to understand the limitations on the somatosensory sense compared to the auditory sense. Therefore, we consulted the previous haptic psychophysical perception literature to compare the human perception of frequency and amplitude in both senses, in order to evaluate how this these properties could be mapped into a haptic signal.

3.3.2 Frequency perception analysis

We define pitch as the human subjective perception of distinguish the frequency of auditive stimuli that allows ordering the signals on a frequency-related scale. On this specific, the auditory system performance is magnificent, it can distinguish different kinds of frequencies with very small variations. For pure signals, the frequency JND (Just Noticeable Difference) in humans is very small, Riez reported a differential sensitivity value ($\triangle E/E$) of 0.05 to 0.15 depending and the frequency, with the minimum value around 2.5 kHz [42]. Another study reported a frequency JND of 3Hz frequencies below 500Hz and about 0.6% for frequencies above 1000Hz [6].

On the other hand, the frequency JND for haptics on the finger has been reported to be 18% for a sinusoidal vibration and it is independent of the amplitude or acceleration of the applied stimuli [40]. Other studies report a smaller frequency JND value of 10% in the index finger [2]. If both senses are compared only on the terms of their frequency JND, the ear can make frequency discriminations of 0.6% while the somatosensory system can only discriminate frequency variations larger than 10%. So on this terms the auditory systems is far superior than somatosensory system.

Then, if the frequency perception range for both senses is also compared, the reported somatosensory frequency range, for pure tone vibrotactile signals, goes from 5Hz to 400 Hz [32] while other studies confirmed that people can detect vibrations up to 700Hz [49]. Now for audio, the absolute auditory frequency perception range, for pure tone signals, goes from the 0.032 kHz to around the 16 kHz. And the best auditory frequency perception range from 0.250 kHz to 8 kHz [22]. And again the auditory system is also far superior with a wider frequency range compared to the somatosensory system.

Now, if the JND frequency values and the reported frequencies ranges of both senses are considered, then is evident the impossibility of directly map all the distinguishable audio frequencies into the narrow haptic frequency perceivable range.

For and informal estimation we are going to overlook the fact that the frequency JND variation accordingly to the base frequency and we are just going to consider the smallest JND for both senses with a a base stimuli frequency of 250Hz (where the haptic sense has a smaller freq JND) [42]. Then, if the best perception frequency range for audio has a length of 7.750 kHz and the complete haptic perception range has a 700Hz lenght, then the JND in Hz for a 250Hz base signal will be: 3Hz for audio and 25 Hz for haptics. Then if these quantities are divided, we obtain $7750/3 = 2583.3$ for audio and $250/25 = 10$ for haptics. So this rough calculations tell us that we can perceive 2583 different frequencies for audio and only 10 for haptics.

This rough comparative shows the impossibility of correctly displaying all the perceivable different audio frequencies into haptics. If this technique is used, then different audio frequencies will be represented with different haptic frequencies below the minimum haptic perception threshold. Therefore, we tried to tackle this problem from a different perspective (see Chapter 5).

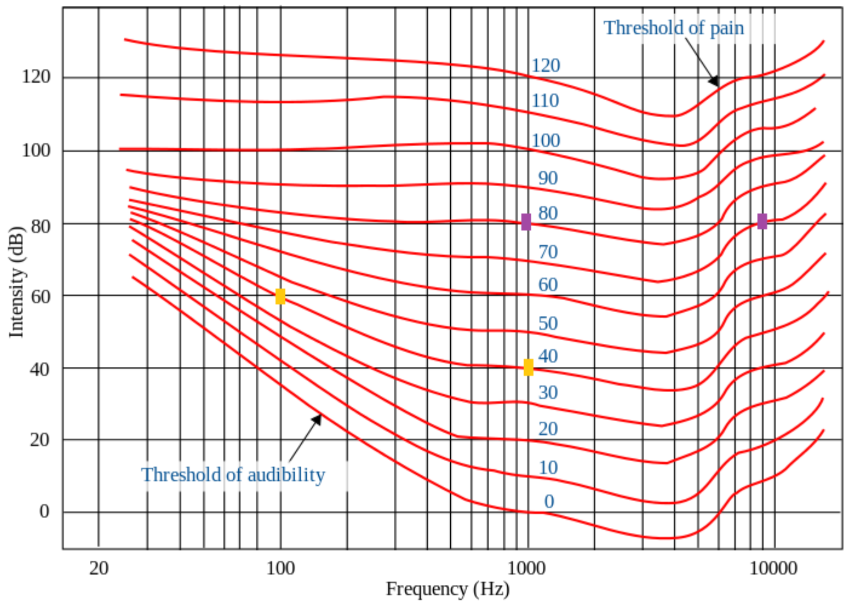


FIGURE 3.4: The figure shows the equal-loudness curves for the human auditory sense. The audibility threshold is show in red, while the yellow and purple squares represent equal-loudness levels on different frequencies. (CC BY-SA 4.0) [36]

3.3.3 Amplitude perception analysis

As the work of McLean and Enriquez [33] suggest, to crate representative haptic signals with different kinds of sensations it seems important not just to consider the frequency of the signal, is also necessary to use their amplitude and waveform. Due to the somatosensory sense limitations to perceive frequency variations, it seems a good idea to rely on the haptic signal amplitude to create a vibration that resembles an specific audio signal. To do so, it is important to understand the differences between the amplitude perception in both senses.

The amplitude perception in sound is usually associated with the perceptual quality of loudness, so the more amplitude the sound signal has the louder it will sound. But the frequency is also related to the perception of this specific sound property. As mentioned before the human frequency hearing range is wide (20~20000Hz), but the reported lowest sound pressure level perceived in humans is between 2000 to 6000 Hz. So this means that, the absolute detection threshold is affected by the audio signal frequency; as we can see in the Figure 3.4. Also in the Figure 3.4 we can see that the lowest amplitude detection in humans rises around when the frequency is low and and also it rises again when the signal frequency is very high.

In the Figure 3.4 we can also see several red lines, these lines show the equal-loudness curves. To obtain these equal-loudness curves several subjects were asked to equate the loudness of sounds with different frequencies. So if we follow the equal-loudness line we will find all the possible amplitude/frequencies combinations that make a sound to sound at the same intensity. For example, the Figure 3.4 shows two orange squares, one at 50dB SPL and the other at 40dB SPL, also we can see that both orange squares are on the 50dB equal-loudness level. So this means that for a 100Hz frequency to sound as loud as 1kHz, the 100Hz frequency is necessary to present them at 50dB SPL and 60dB SPL respectively.

These for the purple squares, show in the Figure 3.4, these indicate that 9kHz tone presented at 80dB will sound as loud as 1000Hz tone presented the same amplitude level [53].

In addition there is another interesting phenomena that must be considered in auditive amplitude perception. The perception of loudness can also be altered if an sound is being heard for a long period of time. Due to the design of our perceptual mechanisms, that depend on the summation of energy, this will cause that the sound will gradually be heard with more and more intensity. This specific process is called temporal integration. Temporal integration will also happen if two auditive signals with the same frequency and amplitude are presented in less than 100ms between one to the other [53], then the second signal will have a little difference in loudness perception [53]. So for intensity perception experiments the temporal integration effect must be considered in the design of any auditive psychophysical experiment.

For the somatosensory sense the same kind of amplitude perception studies had been performed [50]. This and other studies [15], shown that the somatosensory sense intensity perception also depends on the frequency of the presented signal. As we can see in the Figure 3.5 the absolute threshold curve has is lower around the 250Hz and the minimum value of that curve is around -20dB at exactly 250Hz. Also by performing similar experiment for the auditory sense, where the participants were asked to equate the intensity of signals with different frequency, the haptic equal-intensity curves were also found. The tactile equal-intensity curves can be used to determine which frequency and amplitude two signals should have to be equally perceived. For example, in the Figure 3.5, we can see that 100 Hz haptic vibration should be displayed at 24dB to be have the same perception intensity as a 500Hz haptic vibration presented at 10dB.

Also, the same as the auditory sense, the somatosensory sense also presents temporal integration, because also the perception mechanisms of the somatosensory sense are also design to integrate the energy over brief periods of time. So, if two haptic vibrations with the same frequency and amplitude are presented between each other in a less than 500ms [51] then the amplitude of the second signal will be perceived with a greater intensity. And also, this intensity summation will increase proportional to the vibrations' length. And of course the temporal integration must be considered on any psychophysical study involving haptic signals.

3.3.4 Harmonics perception analysis

Harmonics are complex signals composed of two or more pure tone vibratory signals. The lowest frequency contained in a complex signal with harmonics is know as the fundamental frequency, while the rest of the harmonics in the signal are known as overtones. Almost all the sounds in nature have harmonics, like: human voice, birds, cars, etc. On the other hand the objects that can produce pure tone signals are more rare, like: tuning forks and some flutes.

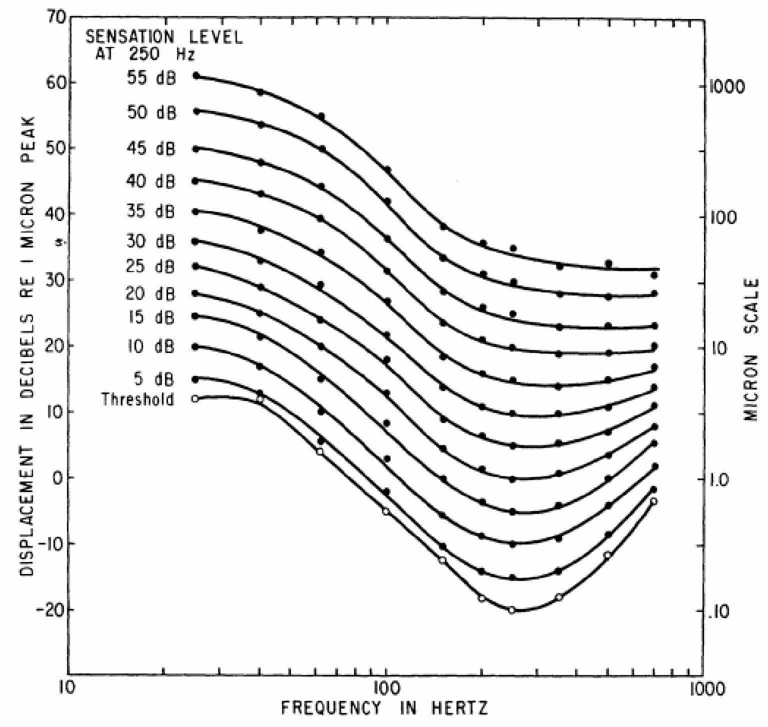


FIGURE 3.5: The figure shows the equal-sensation curves of the human somatosensory sense. The data was taken using the direct-scaling method and the sensation levels refer to 250HZ. [©1928 American Physical Society] [50]

We can clearly identify if an audio signal is a pure tone or a complex signal by performing an spectrum analysis, with a mathematical process called: Fourier analysis. With this specific mathematical procedure we can observe the harmonic spectrum of the audio signal. Or in simple words we can see the energy intensity of all the fundamental frequencies present in a complex audio signal.

To create a haptic signal that resembles an audio signal is also necessary to evaluate both senses in terms of their perception of harmonics. For audio it is clear that our perception of harmonics is outstanding, because in our daily life we can easily identify harmonic spectrum of complex signals all the time. For example: we can identify female and male voices, different types of vehicles and different types of musical instruments. It has been reported that humans can distinguish up to 5 harmonics and this has been reported on the harmonic perception on 7-month-old infants [10]. While for adults is has been reported [23] the we can perceived harmonic complex tones containing over up to 11 successive harmonics. So, it is evident that the auditive sense has an outstanding performance on discriminating harmonics.

But on the other hand the somatosensory sense do not have the same ability to discriminate harmonics as the auditory sense. The study performed by Toshiharu reveals that the harmonic discrimination in humans may be different depending on the contact zone. Also reveals that we can relatively accurately discriminate a second harmonic in a complex signal with a fundamental frequency of 40Hz. So we suspect that the touch sense ability to perceive harmonics is low, because the mechanisms and purpose of the touch sense are very different to the auditory sense.

For the harmonics multimodal audiotactile analysis, we could mention the work by Picinali [39]. In this study Picianali evaluated the discrimination harmonics threshold in humans by comparing the audio and somatosensory senses. He reported that for both modalities the spectral difference between different stimuli could be perceived, with difference of 28.7 dB between modalities. In the best case scenario, this difference was reduced, to 5 dB, when the two tones composing the stimuli were not in harmonic relation. Even if we are able to perceive differences in the harmonics in an audio-tactile signal, the difference between the modalities is still very big 28.7dB. This study leads us to thing that is not a good idea to rely on the haptic signal harmonics to create a resemblant audiotactile vibration. Mostly, because almost all the audio signals in nature are complex signals composed of several different harmonics.

3.4 Multimodal Perception of Music

Music is an abstract representation of sound that creates specific responses in the brain triggering different kind emotions, which highly depend on the music interpretation, the listener and the music itself. It is evident that music is a complex multidimensional stimuli. And since the final purpose of this research is to create a resemblant haptic signal of a single instrument in a song. It is also necessary to understand the perception mechanism involved in the multimodal perception of music.

In order to propose an effective way to improve music listening though a haptic stimulus or a multimodal stimuli. It is mandatory to understand if there are intrinsic properties of music or special perceptual mechanisms in the human brain that help us to focus our listening attention into an specific instrument or sound.

In psychology there is a music perception theory called the psychoacoustic theory of "auditory stream segregation" which proposes that humans can effortlessly segregate tones sequences into individual streams of melodies based on just on the tone of the melody [7]. Additionally, it has been demonstrated that our perception of melodic structures can also be manipulated if the piece is cleverly composed, such as in many of Bach's contrapuntal compositions [54]. Along to this findings, the temporal relation between the voices in a polyphonic song also are an important perceptual clue for stream segregation [41]. Therefore, for the selective listening of music the structure of the melody and its registry seem to be influential factor that must be considered.

Also other factors as the instruments timbre may also have an impact on our selective perception of music. In has been demonstrated that, in polyphonic music two voices are more recognizable if their respective melodies have a closer relation in pitch. In cases when their notes' onsets have the same tempo participants begin to rely on timbre to distinguish them [18]. In addition, the instruments timbre seems to play an important role as well. For instance a very homogeneous timbre can be used to obfuscate the listeners' abilities to identify the concurrent number of voices present in a polyphonic song [24].

There is plenty of literature in music psychology that indicates that the timbre, pitch or the melodic content of a polyphonic song can influence the selective perception of a polyphonic song. Therefore, it is mandatory to evaluate if the multimodal stimuli in the entertainment environment proposed in the the Chapter 5 is in fact effective to improve the individual selective music listening.

Also due to the multidimensional properties of music, it is not possible to generalize if a haptic or visual or a combination of both stimuli would be more or less effective to improve the individual perception of music. As mentioned in the introduction, there is plenty of evidence that for simple search tasks using multimodal stimuli (visual-tactile or visual-auditive) the visual stimuli seems to be dominant, this specific phenomena is already known as the Colavita effect [46]. The Colavita effect indicates that humans are more likely to respond to a visual stimulus when they are exposed to a combinations of auditory and visual signals or combinations of tactile and visual signals. Now, the Colavita effect was only be tested using simple pure tone audio signals, therefore due to the multidimensionality and complexity of music is irresponsible to blindly assume that the Colavita effect will hold when an individual uses a bi-modal or tri-modal stimulus to listen to music.

On the other hand, there is evidence that indicates that when a tri-modal stimuli (visual, tactile and auditive) is used the Colavita effect disappears. Accordingly, Hecht & Reiner, it seems that when a tri-sensory stimulus is given, the probability of missing two signal is much smaller that missing only one signal, therefore the visual dominance disappears. Also, Hech and Reiner findings were primarily found on searching task using simple audio, visual and tactile stimuli, therefore is also irresponsible to assume that the Colavita effect will also disappear, when using a music instead of a simple audio stimuli.

Chapter 4

Evaluation on Multimodal Envelope and Frequency Perception

In this Chapter we will talk about the all the perceptual experiments performed to understand the human ability to judge the similarity between an audio and a haptic signal. In the section 4.2, we discuss how we tried to understand if we are able to perceive the envelope similarity between a haptic signal and a specific instrument audio signal in a MIDI song. In section 4.3.1 we discuss a different set of experiments. These experiments were focused on understanding which are the envelope characteristics that have more impact on the similarity perception between an pure tone audio signal and a pure tone haptic signal.

The main objective of the presented experiments, in this section, is to find some general guidelines that can be directly applied to design haptic signals with a high resemblance to an specific audio signal. This knowledge can be finally applied into the design of the haptic signal in the proposed multimodal listening enhancement environment (see Chapter 5). Also we consider that, the results of the presented experiments can be directly applied to improve the tactile sensation of diverse kinds of haptic or multimedia application applications like: video games, virtual reality and even haptic manipulation of simulators.

4.1 General Description

Before performing and designing the psychophysical experiments, we consult the different psychophysical documentation in order to generally compare the limitations and attributes of both senses. From these general analysis we understood that due to their specific purpose and functions, the auditory and touch senses were very different. In audio humans are usually have a outstanding performance to perceive small changes in frequency, but compared to haptics the human performance to perceive frequency variations or overtones is poor. On the other hand, the human auditive and tactile senses are really good to perceive subtle differences in amplitude. Also other related research like the work of Adelstein [1], Tikka [47] and Caclin [8] gave support to our idea that there should be a relation between the waveform of the signals and their perceived similarity. So we started deeply explore the previous psychophysical research about the relation between the signals envelope shape and their perceived similarity.

Tikka [47] reported that the audio stimuli intensity perception can be biased if a synchronized haptic stimulus was also presented. Also Tikka determined that the best corresponding physical parameter to be perceived as the acceleration of the haptic stimulus pulse. But Tikka did these observations by using a pulse signal modelled with a Gaussian bell function. So, further analysis may be needed to generalize these results for signals with diverse envelope shapes.

Also, Caclin [8] reported a relation between the auditory localization judgement and audio tactile stimuli. She demonstrated that the apparent location of a sound can be biased towards a synchronized tactile stimulation. This research also showed us that simultaneous audiotactile stimulus can impact human perception.

Even so, we want to specifically evaluate if the envelope shape characteristics of both signals could impact the similarity perception of them. And to our knowledge there isn't any reported information on this specific subject.

The lack of specific documentation about the subjective similarity of audiotactile stimuli, pushed us to conduct our own perceptual experiments. In order to understand the human perception of synchronized audio and tactile signals. In specific we focus our experiments on trying to find a good strategy to create a syntethic haptic vibration with a high resemblance to an specific audio signal. Then, these experiments evaluate which characteristics of the haptic signal had more impact on the perceived similarity between both signals. For example: We suspect that the attack synchronization between a haptic and audio signals was a crucial characteristic to increase the perceived similarity between them. We suspected this, based on the reported audiotactile asynchrony threshold values reported by Adelstein [1]. Adelstein reported that humans can detect asynchrony between a haptic and audio signals, if their asynchrony is bigger than 24ms. For this study Adelstein used short pulses for the audio and haptic signal, but we suspect that Adelstein results can be generalized on the attack asynchrony detection of signals with a similar waveform shape.

So we believe it was necessary to perform further research, and understand in detail the characteristics of both signals that have more impact on the perceived similarity of both.

Due to the previously mentioned research results, we suspected that the attack synchronization and the waveform of both signal may impact their perceived similarity. So we formulated different hypothesis about the envelope correlation perception and we designed several perception experiments accordingly these possible hypothesis:

1. A similarity between pure tone haptic and audio signal envelope shapes can be easily perceived and discriminated.

2. The perceived similarity of the signals may be positively affected by their attack synchronization and vice-versa.

3. The perceived similarity of the signals may be positively affected by a precise amplitude correlation of their envelopes.

4. For a complex audio signal, like music, the amplitude and frequency variations between the signals, may mask the perception their waveshape correlation.

The previously mentioned assumptions were made in base of the previously reported results of related previous research. The assumptions 1 and 3 are based on the reported results of Tikka [47], where she prove that the relative intensity of a haptic signal can bias the intensity perception of an audio signal. The assumption 2 is based on the generalization and application of the reported human asynchrony audiotactile threshold by Adelstein [1]. Finally, the assumption 4 is founded on several informal experiments, performed during the early experimentation phases of this research.

Also, these hypothesis were also proposed to elucidate a strategy to optimize the perceived similarity of a haptic signal with its audio counterpart. So if these hypothesis are rejected or accepted this will guide us to define an haptic signal with good audiotactile resemblance.

Finally, after performing the perception experiments we were able to propose a multimodal listening environment, where the users are able to listen to the music, while looking to a self explanatory representation of the music structure and while feeling a representative haptic vibration of one instrument in the song. The goal of Haptic Music Player was to create an accessible multimodal listening environment to have a deeper understanding of music by focusing the listener attention to specific instrument in a song by using a combination of haptic and visual clues.

4.1.1 Similarity Definition

To clarify, when we refer to audio-haptic similarity perception, we allude to the human ability of perceive the same characteristics between two different stimuli. In this specific case, the ability of perceive the same envelope characteristics between simultaneous audio and haptic signals.

4.2 Music & Vibration Perception

Two psychophysical experiments were performed to evaluate if the envelope correlation of an audio-tactile stimuli can improve the subjective perceived similarity between both signals. In both experiments, the users ranked the similarity between an specific instrument in a song and several haptic signals build with different envelope shapes. On the first experiment, the frequency and amplitude of the haptic signals were changed accordingly to the notes' key (as described in the Chapter 5), while in the second experiment the notes' key was ignored and the haptic signal used only constant frequency and amplitude. Also is important to mention that direct the results of this studies were consider to the design the synthetic haptic vibration used in the proposed entertainment environment: The Haptic Music Player.

4.2.1 Envelope Masking Evaluation

Experiment #1 - Variable Amplitude & Frequency

In the first experiment the participants' task was to rank the perceived similarity between 3 different haptic stimuli and the sound of 4 different instruments in a song. The haptic signals were build using different types of envelopes: a triangular envelope, a square envelope and the actual sound envelope. The haptic signal with the same envelope characteristics as the audio signal was defined as the *Analogue*

Envelope. The haptic signals' frequency and amplitude was defined by the technique mentioned in the Section 5.2. So, the amplitude haptic range was set between: $a_{min} = 7.5$dB and $a_{max} = 30dB$, while the frequency was set between: $f_{min} = 50$Hz and $f_{max} = 250$Hz. This particular frequency range was used in order to avoid aliasing in the haptic signal, due the haptic device refreshing rate (1000Hz).

The experiment was performed using the virtual environment, mentioned in Chapter 5. This environment let the users listen to the music see the animation and feel the instrument's vibration. For this experiment, the Haptic Music Player was slightly modified to let the user rank and change between different haptic vibrations. The particular music piece used for the experiment was a MIDI rendition of Bach's 1079 Sonata - Largo movement [31]. This particular song used 4 difference instruments: harpsichord, violin, contrabass and flute. So, we presented 3 different envelopes for every instrument, then in total every user had to rank 12 different audio-tactile stimuli. While listening to the music, the user was able to change between the 3 different envelopes at any time, and rank them using an A,B,C scale. Also, the user was instructed to rank the 3 haptic envelopes before continue to the next instrument. The 3 haptic envelopes were presented in random order. The experiment finished after the user ranked the 12 different audio-tactile stimuli presented in the song.

In order to clarify the similarity concept among the users, without bias the their particular preference, 2 rounds of practice were performed before the main experiment. For the practice rounds the isolated tracks of violin and contrabass from Bach's BWV 1079 Sonata - Allegro movement [31] were used.

For the main experiment, we randomly presented 3 different haptic signals built with different envelopes and then we asked the participants to rank the presented vibration accordingly to the similarity between the instrument's sound and the vibration. After finishing every practice round the analogue envelope position was reported to the user, so the participant could understand the similarity between both signals by his own perceptional means. Also these practice rounds helped the users to familiarize with the keystrokes used to: change the vibration (*0~1*), rank the vibration (*A*,*B*,*C*) and change the instrument (*t*).

Experiment #2 - Constant Amplitude & Frequency

The second experiment was performed with the same conditions, methodology and participants as in the first experiment. But contrary to the first experiment, for this particular experiment the haptic signal was displayed with a constant haptic frequency and amplitude. So, the notes' pitch was not considered to compute the haptic vibration, instead and constant frequency (f_h) of 250Hz and a peak amplitude (a_{max}) of 1mm was used for every haptic signal of every note. This experiment was performed to evaluate the cross-modal similarity perception of the signals' envelopes under more controlled circumstances. By these means, we evaluated if the amplitude and frequency variability affect the users' cross-modal envelope perception.

4.2.2 Experimental Conditions

Again a special version of the Haptic Music player was used (see Chapter 5). A computer with an Intel i7-3770S, Windows 7 and a Realtek ALC662 sound card was used to perform both experiments. While, the haptic stimuli was displayed with Spidar-G6 [29] haptic interface, and a 16 bit Realtek chipset sound card was used to display the sound with a 48000 Hz sound quality.

Both experiments were performed with the same conditions and participants, so the same 11 participants, 5 females and 6 males, made both experiments. All of them healthy adults between 23 to 30 years old.

Perceptual experiments are clearly very difficult to perform, because the specific variable to be measured (the individual perception) can be easily biased by several factors like: the participants' mood or the given explanation.

To avoid biased lectures, special measurements were taken to avoid biasing the subject own subjective perception. So before performing any of the experiments, every subject made several rounds of practice before taking the real experiment. With these practice rounds we tried to evaluate if the subject was able to understand by their own perceptual means the difference or similarities between both signals. Consequently, the examiner only asked the users to grade, select and order the signals by their similarity. Is important to point out that in none of the mentioned experiments the similarity concept between the signals was explained to the participants.

To isolate the user's auditive and tactile senses during the experiment, several precautions were taken. The experiment was taken on a quiet room. Also, the participants used earplugs (Sennheiser MX 475) to listen to the audio stimuli, while using passive noise cancel earmuffs.

In addition, to minimize any inadvertent vibration coming from the haptic device, the user rested his right forearm on a armrest at the same height as the haptic device, and the armrest was separated from the table which the haptic device was placed. Additionally, the user index finger was attached to the haptic device pointer using a velcro strap, and the subject was not allowed to grab the pointer with another finger or touch the haptic device strings. Furthermore, the haptic device was placed over urethane foam on a solid 1.5cm thick iron plate to eliminate any vibration transmission from the haptic device into the the user's left hand, positioned on the laptop keyboard.

Also, to avoid the potential visual cue from the haptic device mechanisms movement, the user passed his right arm though a tall barrier (see Figure 4.1). And to prevent any problem with any enhancement perception [51] the audio-tactile stimuli were displayed with a 500 ms time interval between them.

Due to the random latency by the audio drivers, a random delay on the synchronization between the audio and haptic signals was always present. By measuring the delay between the signals and by using low latency drivers (ASIO v2.13) [48]. The latency between the signals was always monitored and controlled to be between 0ms~2ms. By these means, this random delay was always below the human audio-tactile asynchrony sensitivity threshold [1] (~24ms), so this should not affect the participants performance.

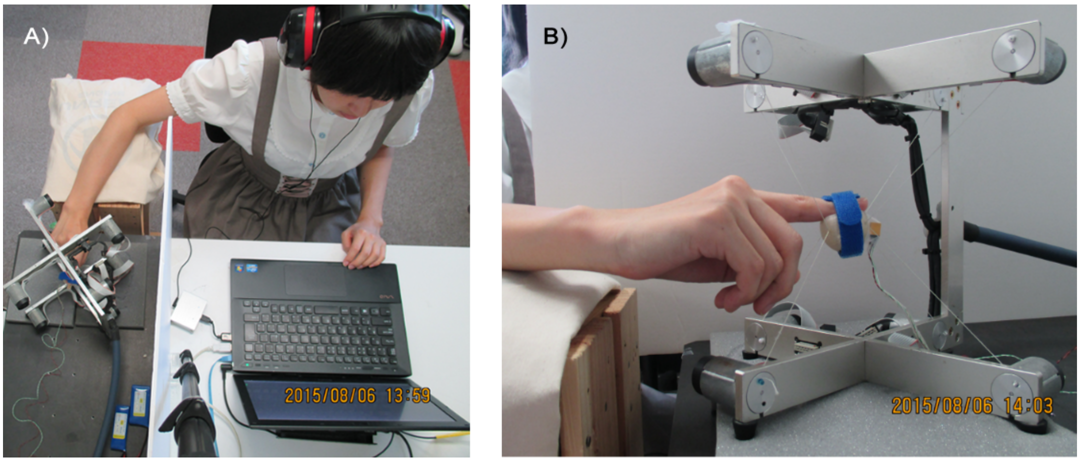


FIGURE 4.1: The figure A) shows the overall experiment configuration. In this picture, the user uses in-ear headphones and earmuffs to listen to the audio and passed his arm though a tall barrier to feel the haptic signal. The image B) shows how subjects' finger attached to the haptic device grasp, in order to normalize amplitude between the subjects.

4.2.3 Envelope Masking - Results

If the Copeland's method is applied to the results of the first experiment, then the Copeland's favorites of each instrument are: the square envelope for violin and flute, the triangle envelope for the harpsichord and the actual sound envelope for the contrabass (see Figure 4.2). These results suggest, that the participants preferred haptic signals with simpler envelopes (square and triangular) over the analogue envelope, when the frequency and amplitude of the haptic signal was variable.

On the other hand, in the second experiment the Copeland's method results show, that the analogue haptic vibration was the best raked haptic audio-tactile stimuli for all the instruments. So these results show a clear preference to the analogue envelope for all instruments. In addition the same preference is clear by counting the number of votes given to the best raked envelope.

As mentioned before both experiments were performed under the same conditions and with the same participants. So, the contrast in the results suggest that, the frequency and amplitude variability between the notes, created by the presented mapping technique, masked the haptic envelope perception. Consequently, we consider that the participants were not able to perceived the envelope similarities of both signals with the same accuracy as in the second experiment.

The results of the first experiment show that even if the participants preferred the simpler haptic envelopes, however the participants chose simpler envelopes, who had more amplitude similarities to the audio envelope. For example, for the flute and violin the square envelope were preferred over the triangular, so in this case we suppose that the steady sustain of violin and flute caused a preference of the square envelope. Also we suppose that the similar decay rate between the harpsichord and triangular envelope caused the preference of the triangular envelope over the square envelope. Even so, for the contrabass the users preferred the analogue envelope. So,

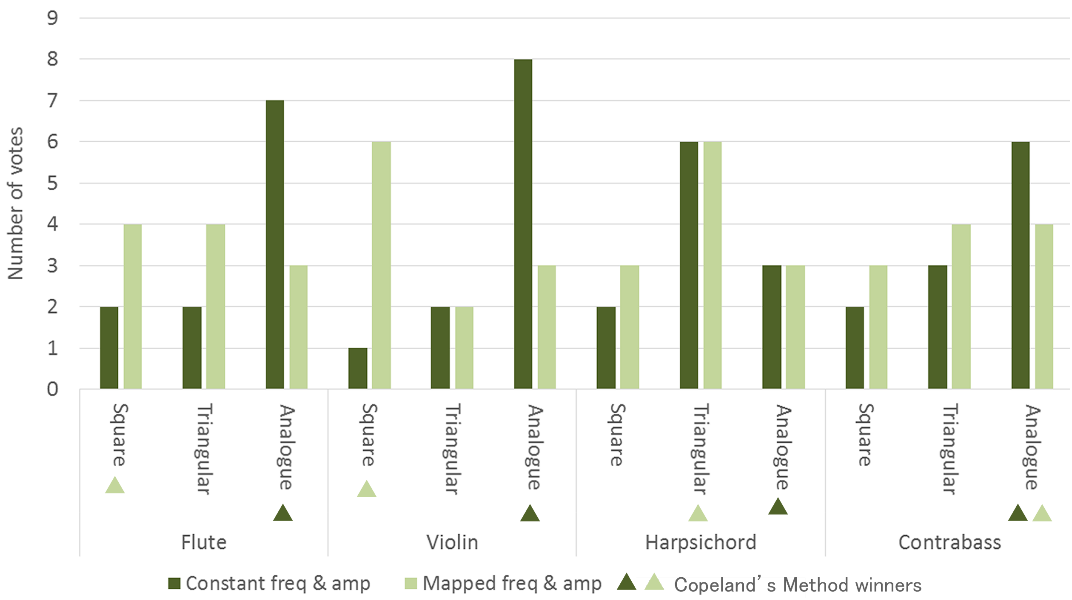


FIGURE 4.2: The figure shows the number of votes received by the best ranked audio-tactile stimuli. The votes of the experiment with a constant amplitude and frequency for haptics are shown in light-green, while the results in dark green are for the constant frequency and amplitude experiment. And the small triangles indicate the Copeland's winners for every instrument.

we suppose that the particular amplitude fluctuations in the contrabass envelope shape helped the users to perceive the haptic envelope shape. In any case it is necessary to perform further studies to clarify these observations.

4.3 Pure Tone Audio & Vibration Perception

In this Section we will talk about a set of 3 different experiments, performed to understand the impact of the envelope shape on the perceived similarity of an audio-tactile signal. On this experiments are focused on finding which specific phases of the envelope shape had more impact on the similarity perception of the signals. To elucidate this, we performed 3 different experiments focused on different aspects of the envelope shape: The first one, was focused on study the correlation perception of the waveform of both signals. The second, was focused on study the impact of the asynchrony of both signals on different signal phases. And the third one, was focused on improve the similarity perception of a audiotactile signal by changing some characteristics of the haptic signal.

The results of these studies, were directly applied to the Haptic Music Player haptic signal. So, the temporal amplitude characteristics of the haptic signal in the Haptic Music Player were defined accordingly to the results obtained on these perception studies.

4.3.1 Experimental Conditions

The experimental conditions for this set of experiments was almost the same as the described in the Section 4.2. The same precautions to isolate the auditive and tactile senses of the participants were taken: usage of strong earmuffs on the earplugs (Sennheiser MX 475), the haptic device was placed over iron and urethane foam to avoid vibration transfer to the table and a white screen was placed between the user face and right arm. Also to avoid amplitude differences between subjects, the right index finger of all the participants was attached to the haptic device grip with a Velcro strip; just as shown in the Figure 4.1.

To perform the perceptual studies an special software was developed. The frequency, amplitude and the envelope of the haptic signal was modeled using an ADSR filter as mentioned in the Section 5.2.

This set of experiments were performed by 8 subjects: 5 males and 3 females, between 23 to 40 years old. All the participants were healthy and didn't report any audition problems. The developed experimental software used a laptop (VAIO model SVS13AD11N) with Windows 7 and an Intel Core i7-3520M processor. The audiotactile stimuli was displayed with Spidar-G6 [29] haptic interface, while a 16 bit Realtek chipset sound card a was used to display the sound with a 48000 Hz sound rate.

4.3.2 Experiment #3 - Correlation Perception

The goal of this psychophysical experiment was to evaluate the correlated perception of diverse audio-tactile signals, by using haptic envelopes with different temporal amplitude properties. In the experiment several subjects rated the perceived similarity between different audio-tactile stimuli, built with a combination of 6 different sounds and 4 different vibrotactile signals.

The waveforms for the sound signals were designed based on musical instruments envelopes: Violin, Trumpet, Organ, Flute, Guitar and Harpsichord. While the vibrotactile signals had simple envelope shapes like: Square, Triangular and Logaritmic, except for last haptic stimulus that had the same envelope characteristics as the sound signal, which we called the "*Analogue Vibrotactile Stimulus*" or "*Analogue Stimulus*" (see Fig 4.3).

The different audio-tactile stimuli were build by combining 6 sound signals and 3 haptic signals with different envelopes (Square, Triangular, Logarithmic) and 1 haptic signal with the same envelope as the audio signal (*Analogue Stimulus*); giving 24 combinations in total (see Figure 4.3).

A simple computer software was developed to create and present the simultaneous audio-haptic stimuli. This software randomly presented every signal combination in 6 rounds. On every round a different sound was presented in combination with 4 different vibrations (Square, Triangular, Logaritmic and Analogue). Then on the next round another randomly picked sound was presented with 4 different vibrations, until finishing all 24 combinations. On each round, with a numeric scale from 0 to 7 (0 for the worst similarity and 7 for the best one), the subject graded the similarity among the 4 different vibrotactile stimuli. Then, with the keyboard, the user could freely play and grade any of the 4 different presented audio-tactile stimuli. After grading all the stimuli presented in each round the user continued with the next one. Also, to avoid the identification of the stimuli by its order, the stimuli

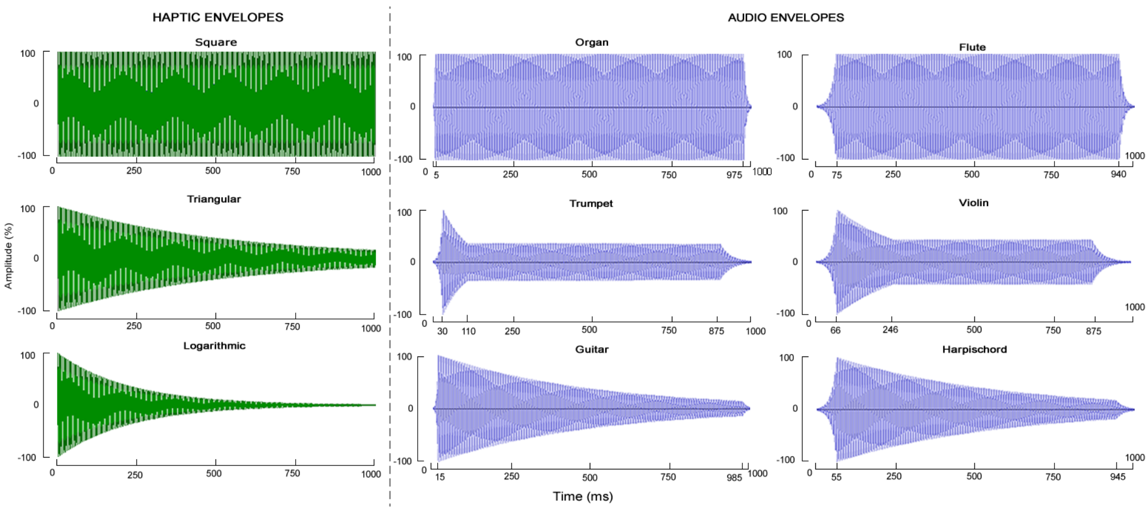


FIGURE 4.3: **Experiment #3 envelope shapes.** The different kinds of envelopes used to build the audio-tactile signal for the perception experiment #3. The blue waveforms are used for the audio signal and the green waveforms are used for the haptic signal. The blue envelopes are similar to the waveforms of different musical instruments. while the green haptic waveforms have simple envelope shapes. Finally, to create the identical audio-tactile stimuli *(analogue stimuli)* the same audio waveform (blue) is used for audio and haptics.

and the rounds order were randomly shuffled for each participant.

An ADSR envelope was used to verify if the amplitude changes were above the reported minimum amplitude perception levels. The amplitude changes, at the attack, sustain and release stages, of the audio and haptic signals were measured and compared with the minimum thresholds for audio and haptics. So in all the changes between the phases would be perceivable by the participants.

The maximum amplitude and sustain of both signals were measured. For haptics we measured a max amplitude (a_h) of 23dB and a sustain (s) of 10dB (with 1μ REF value). While, for audio a max amplitude (a_h) of 65.6dB SPL and a sustain (s) of 55 dB SPL were measured. If the considered minimum perception threshold values are: 10 dB SPL (for a 250Hz sine wave) for audio [26] and -20 dB (for 250Hz and 1μ reference value) for haptics [50]. Then, for audio and haptics the attack, sustain and release stages are above their respective thresholds.

Correlation Perception - Results

The average similarity ratings, provided by the participants, to every stimuli are shown in the Figure 4.4. These results show a preference for the *analogue stimuli* (the stimuli with the same haptic and audio waveform). But also in some cases a simpler haptic stimuli received similar grades or better grades than analogue stimuli, like: organ-square and the guitar-logarithmic. These specific audio-tactile stimuli had a high envelope shape correlation between the haptic and audio signals (see Fig. 4.3). This suggests that, for these specific cases most users were not able to discriminate

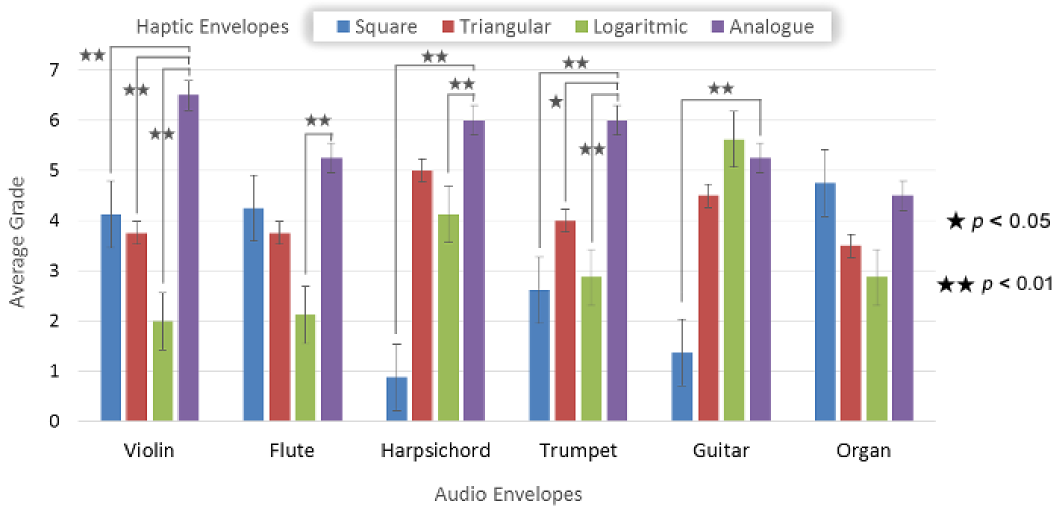


FIGURE 4.4: **Experiment #3 results.** This graphic shows the user's ratings average given to a different combinations of audio-tactile stimuli build with different haptic envelope shapes. The average grade is shown on the y axis. On the x-axis, the audio-tactile stimuli are grouped by instrument (audio envelope) and the haptic envelope shape is indicted by the color of each bar. The p-values of the t-test showed and statistical difference between the violin-analogue signal and the other audio-tactile signals for this instrument. Also, the trumpet results showed the similar statistical results.

small temporal amplitude differences between the simple haptic signals and audio signal, so the participants were confused only by simple envelope with higher resemblance with the audio waveform.

Also, for the organ-square and guitar-logarithmic cases, if we observe the time lag in their attack-decay inflexion points (t_1), between the audio and haptic signals, it was relatively small: $\Delta t_1 = 15$ms for the guitar-logarithmic stimulus and $\Delta t_1 = 5$ms for the organ-square stimulus. Therefore, we consider that the high envelope shape correlation and the relative small Δt_1 between the audio and haptic signals was enough to produce a rise in the similarity ratings for the organ-square and the guitar-logarithmic cases. Even if their envelope correlation was not perfect and their attack were not exactly synchronized.

To compare the results, a one tailed t-test was performed between the analogue signals grades and the other simple haptic signals ratings (see Figure 4.4). For the violin and trumpet, the p-values between the analogue signal and all the simple haptic signals were smaller than the chosen significance level ($\alpha = 0.05$). Therefore, the flute and violin analogue stimulus grades and the other audio-tactile stimuli grades were statistically different. So for these specific instruments, the users' preference for the analogue stimuli was clear. Then, this means that the participants were able to clearly identify the haptic analogue stimuli among the other simpler haptic signals.

To explain these results, we consider this two possibilities: First, any of the presented simpler haptic signals had enough envelope resemblance to the audio signal envelope characteristics to create confusion. Or Second, the attack difference (Δt_1)

between the audio and haptic signals was significantly large: $\Delta t_1 = 66$ms for the violin and $\Delta t_1 = 30$ms for the trumpet. A little bit larger than the reported audio-tactile asynchrony threshold (24ms) [1]. So we suspect that any of this two possibilities or the combination of both may caused the characteristic results for the violin and trumpet.

On the other hand for the flute case, the flute-square and the flute-analogue did not have similar ratings, even if the square and the flute envelopes had a high envelope shape correlation. However, the delay between the attack inflexion points (Δt_1) among the square and flute signals was 75ms. Therefore, it seems that the asynchrony between the audio and haptic signals at the attack could be perceived by most of the participants, so consequently the flute-square did not had high similarity ratings.

4.3.3 Experiment #4 - ADSR Asynchrony Perception

The goal of this experiment is to evaluate which parts of the ADSR envelope are more helpful to the user in order to discriminate the similarity between both signals. For this experiment, different audio-tactile stimuli, with different inflexion points, were built. The audio-tactile stimuli was built using the same ADSR technique as in the Experiment #3. For the audio signal only the violin envelope was used. While for the haptic signal the violin envelope inflexion points (Δt_1, Δt_2, Δt_3) were modified. Thereby, the audio-tactile stimuli had the same audio envelope, however the inflexion points' timing of the haptic signal (Δt_1, Δt_2, Δt_3) where modified. In order to evaluate if the participants were able to distinguish the attack, delay and release timing differences between the audio and haptic signals, and consequently understand which ADSR phases have more impact on creating a similar perception on both senses.

This experiment consisted of 3 rounds, where 7 different audio-tactile stimuli were presented. For every round the violin sound was displayed to the user in synchrony with the 7 different haptic stimuli. The haptic stimuli had the same amplitude-temporal properties as the violin audio envelope, however the inflexion points at attack, sustain and release were modified on every respective round. So, for the attack round the attack inflexion point (t_1) time values were: 0ms, 36ms, 56ms, 66ms, 76ms, 96ms, 136 ms. For the decay round the decay inflexion point (t_2) time values were: 66ms, 216ms ,236ms, 246ms, 256ms, 276ms, 346ms. And finally for the release round the release inflexion points (t_3) time values were: 825ms, 845ms, 865ms, 875ms, 885ms, 905ms, 1000ms (see Figure 4.5). To have a constant decay length, during the attack round, when the attack inflexion point (t_1) was modified, the decay inflexion point (t_2) was also modified accordingly to the forward or delay of the attack inflexion point (t_1).

By these means, 19 different audio-tactile stimuli and 3 analogue stimuli, were displayed during the complete experiment. As in experiment #3, for every participant, the rounds and stimuli order were randomly shuffled. But for this experiment, only the violin envelope was selected, mostly due the previously obtained results on the experiment #3. Because the violin was the instrument with the lowest p-values between the analogue stimuli and the simple haptic envelopes.

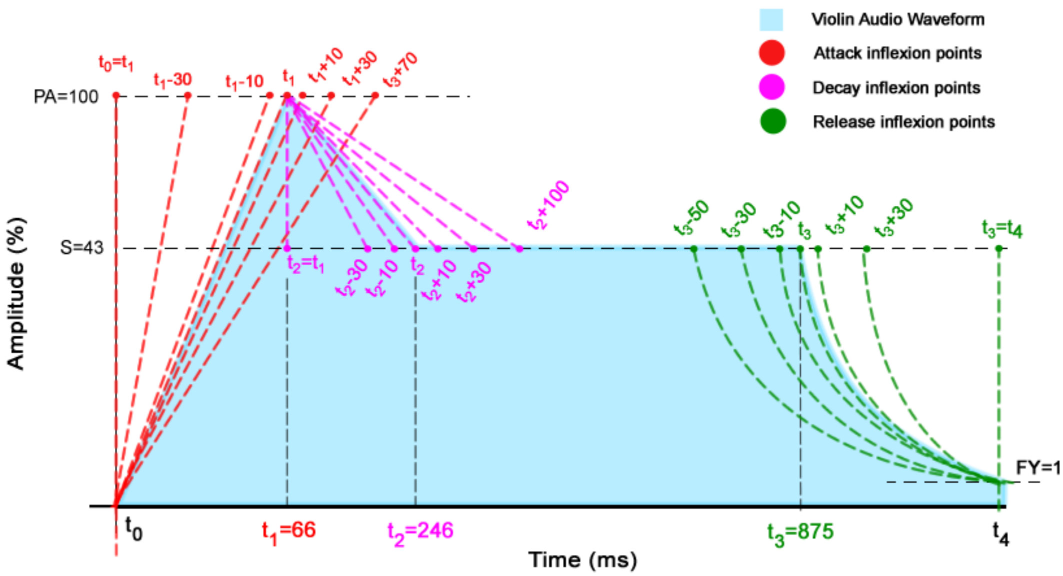


FIGURE 4.5: **Experiment #4 stimuli description.** The figure shows the audio-tactile signals timing used in Experiment #4. The audio waveform used for every audio stimuli is shown in clear blue. The haptic signal inflexion points are shown in red (attack round), pink (decay round), and green (release round). So for the analogue audio-tactile signal the haptic inflexion points are: t_1 for attack, t_2 for decay, and t_3 for release. The delayed inflexion points, show their delay in milliseconds, using the analogue signal timing as reference.

As in the Experiment #3 the users were instructed to grade the similarity between the audio and tactile stimuli, by giving a number between 0 (worst similarity) and 7 (best similarity). And also, before the main experiment, the similarity concept was introduced to the user by doing two practice rounds. After each practice round the users were notified about the position of the analogue stimuli, in order to let them perceive the amplitude similarity between the envelopes by themselves.

Due to the high similarity between the haptic signals, most of the users reported difficulties to identify the differences among the haptic stimuli. So, only for this experiment, the users were notified about the specific ADSR phase where the haptic signals were modified, by displaying on the computer screen these words: "beginning" for the attack round, "middle" for the decay round and "end" for the release round.

Also, the participants were notified that, the sound used was the same for all rounds and cases. Besides that, it was also notified that, all the haptic stimuli were different between the each other. Thereby, we attempt to focus the attention of the participants into the haptic signal.

ADSR Asynchrony Perception - Results

The results indicate the users inability to find the differences between the haptic signals for the decay and release rounds, even if their phase inflexion points were beyond the asynchrony haptic-audio threshold [1]. In the decay and release rounds

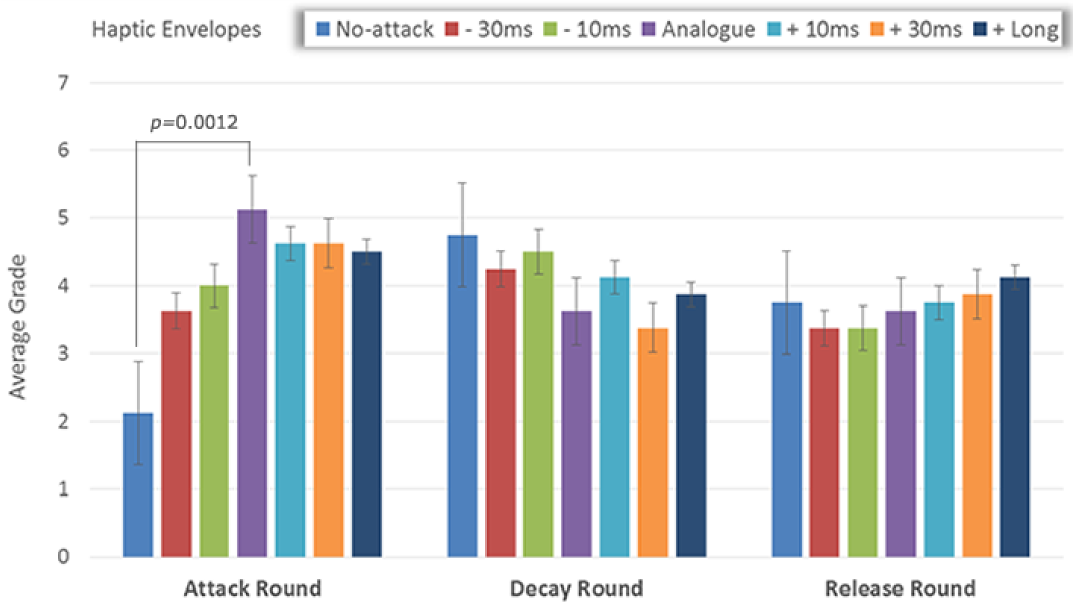


FIGURE 4.6: **Experiment #4 results.** The average ratings given to the different audio-tactile stimuli in the Experiment #4. The y axis shows the user's rating average. On the x-axis, the audio-tactile stimuli are grouped by round and the color of each bar indicate the haptic signal delay. The analogue stimuli of every round is shown in purple.

there isn't a clear difference between the 7 different audio-tactile stimuli combinations in each round (see Figure 4.6). Also, in both rounds the average grade error bars overlap, indicating a lack of significant statistical difference between the rankings of the 7 audio-tactile stimuli in these rounds. Consequently, the p-values between the analogue stimuli and the other displayed stimuli, on both rounds, didn't show any statistical difference ($p <= 0.05$). Hence, this indicate that the participants were not able to perceive if the decay and release amplitude changes were synchronized, delayed or forwarded compared to the decay and release of the audio signal.

In contrast, the attack round results showed an small difference only when the haptic attack inflexion point was forward to the sound. Specifically, the cases: -0ms ($\Delta t_1 = 0$ms), -30ms ($\Delta t_1 = 36$ms) and -10ms ($\Delta t_1 = 55$ms) had different results. The -30ms and -10ms audio-tactile stimuli received similar average ratings, and both cases received slightly lower grades compared with the analogue and delayed stimuli (+10ms, +30ms, +Long). Also, the no-attack (-0ms) stimulus received the lowest ratings compared to the other stimuli in the attack round (see Figure 4.6).

The no-attack case (-0ms) received the lowest grade compared with all the other stimuli in the round. Also, the no-attack (-0ms) stimuli ratings have a clear statistical difference if compared to analogue signal grades ($\Delta t_1 = 66$ms) of the attack round. If a single t-test between the none-attack and the analogue signal grades was computed, then a p-value of 0.0012 was obtained (see Figure 4.6). This suggests that, the participants were able to clearly perceive the attack amplitude asynchrony, of 66ms between the none-attack haptic signal and the audio signal. Consequently this stimuli received the worst ratings in the attack round.

The low ratings that the attack stimuli −0ms,−30ms, −10ms received, indicate that the participants were capable to perceive the asynchrony between the haptic and audio signals only when the haptic attack was displayed before the audio attack. In contrast, the signals with a haptic stimuli displayed after the audio attack (+10ms,+30ms, +Long), received similar ratings, without any statistical difference compared with the analogue stimulus rate (see Figure 4.6). Therefore, we consider that the users are not able to discriminate the asynchrony of the attack when the haptic attack is displayed after the audio attack.

Due to the unclear results at the decay and release rounds is impossible to suggest that the timing differences between the signals at their decay or attack may impact on the similarity perception. But, the attack round and the experiment #3 results show that the attack can impact the perceived similarity between the signals. So, we consider that, the participants were able to perceive the overall envelope shape of both signals, better than comparing small timing difference between them. So, another experiment was necessary to evaluate the role of the general envelope shape of both signals on their perceived similarity.

4.3.4 Experiment #5 - Overall Envelope Perception

The experiment #4 results showed the participants capability to detect audio-tactile asynchrony at the attack. Besides, the experiment #3 results also showed that the temporal amplitude changes (envelope shape), between the audio and haptic signals, can affect the similarity perception. Therefore, this experiment is focus on clarify the results found on Experiments #3 & #4.

In this experiment the haptic signal diverse envelopes were carefully design in order to resemble just an specific stage of the audio envelope, while the rest amplitude properties of the audio signal were not represented on the rest of the haptic envelope shape. In this experiment all the audio-tactile stimuli were build using the violin audio envelope for all the stimuli.
The Figure 4.7 shows all the haptic envelopes used in the experiment: The envelope A is the analogue envelope used as a control. The envelope B has the same attack as the audio signal, however the envelope shape during the decay, sustain and release do not resemble the audio signal envelope. The envelope C has no attack phase, however the rest of the envelope shape matches the audio envelope perfectly. The envelope R1 has a high correlation with the analogue signal, however all its inflexion points are unsynchronized. The envelope R2 has a long decay with no sustain and also all their inflexion points are not exactly synchronized with the audio envelope. Finally the R3 envelope has the same attack, however the signals doesn't have decay or release only sustain.

Every haptic envelope in this experiment have specific envelope properties that resemble certain properties of the violin audio envelope. For example, (see Figure 4.7) the haptic envelopes B and R3 have the same attack as the audio envelope, however their envelopes have a low correlation with the audio signal. Therefore, envelopes B and R3 have been designed to check the subjective similarity perception of the attack resemblance. In addition, the envelope C hasn't any attack, however the rest of the envelope is the same as the audio signal. Then the envelope C is designed to evaluate the impact of the envelope shape alone. Additionally, the R1 and

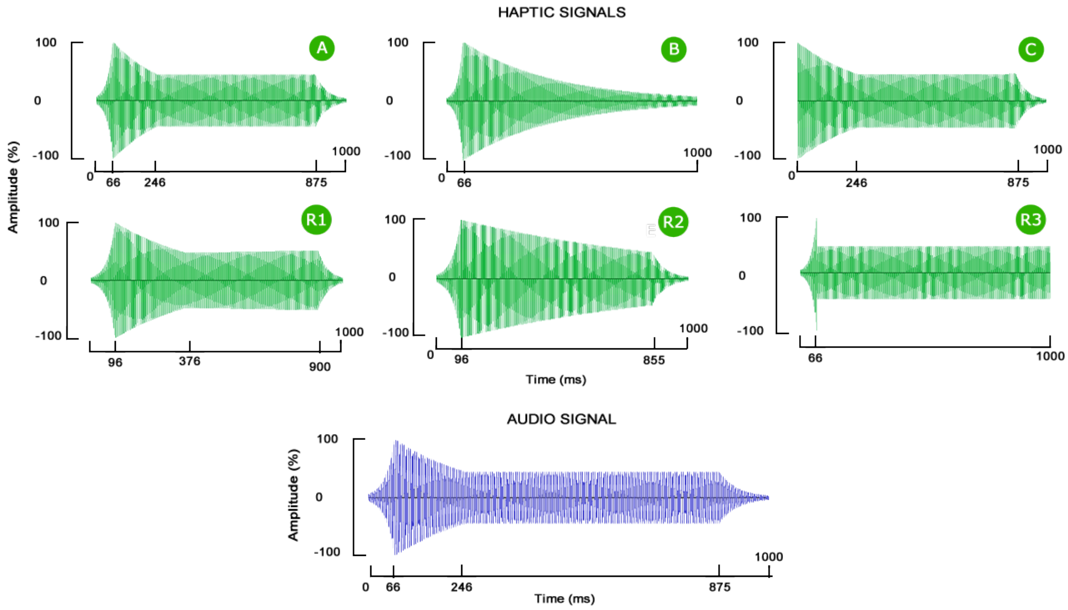


FIGURE 4.7: **Experiment #5 signal description.** In Experiment #5 several haptic signals (green) were designed to resemble specific phases of the audio signal (blue). If compared to the audio signal: (A) is the analogue stimuli, (B) has a sync attack with uncorrelated envelope, (C) has correlated envelope with no attack, (R1,R2) have sync attack and similar envelope and (R3) have sync attack but no decay and release.

R2 envelopes have a similar attack and similar envelopes compared with the audio, however any of their ADSR inflexion points is exactly the same as the audio signal it self. Even so, these envelopes were designed to have the same or better rankings than the analogue stimulus (A). By the specific design of these haptic envelopes, we pretend to study impact of the attack and envelope correlation on the participants subjective amplitude similarity perception.

This experiment was performed under the same conditions and proceedings used in the experiments #3 & #4. The haptic stimuli were displayed in 3 different rounds: For the round #1 A,B,C and R1 audio-tactile stimutli were displayed, for the round #2 the audio-tactile stimuli A,B,C and R2 were displayed and for the round #3 A,B,C and R3 were displayed (see Figure 4.7). The haptic signals were placed in that order, to place diverse audio-tactile stimuli with a high correlation on different envelope phases in every round. By these means, we tried to display contrasting audio-tactile stimuli to minimize any user confusion caused by comparing very similar audio-tactile stimuli.

Also, before taking the experiment, the participants were introduced to the similarity concept with the same method as in the experiment #3. In addition, as in the previous experiments, the same computer, haptic interface and software were used for this experiment. And as in the previous experiments, the amplitude changes at the attack, sustain and release were measured and compared to the the minimum perception threshold values [26] [50], for both audio and haptics.

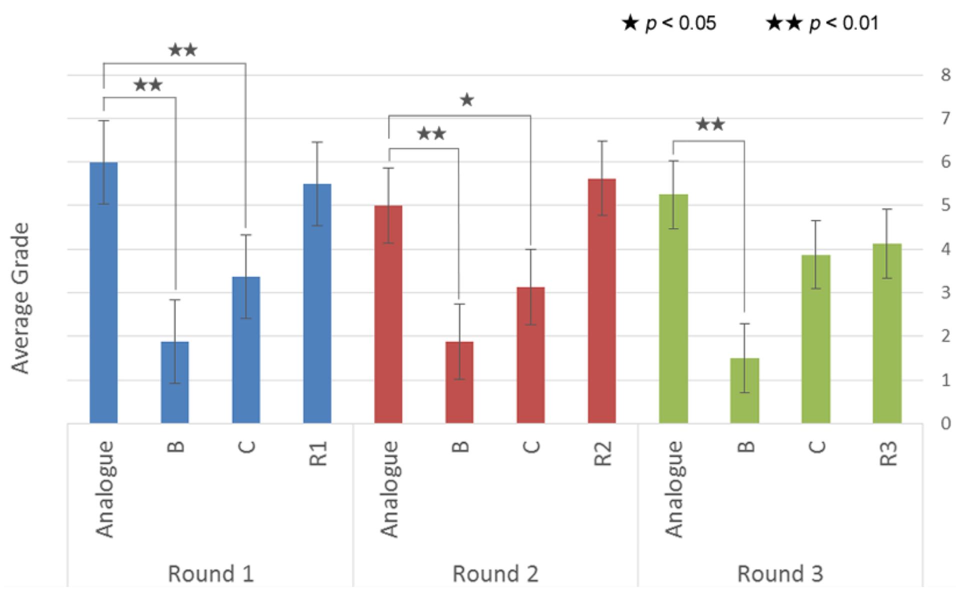


FIGURE 4.8: The figure shows the user's average rating given to the audio-tactile stimuli in the Experiment #5. The y axis shows the rating average. On the x axis, the stimuli are grouped by rounds and each bar has the name of the haptic signal used to create the audio-tactile stimuli. Also, the p-values are shown only for the stimuli with a statistical significance difference ($p < 0.05$) from the analogue stimuli of their respective round.

4.3.5 Overall Envelope Perception - Results

Again as in the experiment #3 the analogue signal preference ratings were high. However for rounds 1 and 2, the R1 and R2 stimuli received similar or sightly better grades compared to analogue stimulus grades (see Figure 4.8). In specific the R1 signal had a similar attack and envelope shape compared with the analogue envelope, however R1's inflexion points were unsynchronized if compared with the violin envelope inflexion points. But even so, R1's average rating was similar as the analogue signal, Consequently, this indicates that most of the users were not able to distinguish the inflexion points asynchrony between the audio and haptic signal.

Also, the R2 audio-tactile stimuli had slightly higher average similarity ratings than the analogue stimuli ratings (see Figure 4.8). The R2's attack was similar to the violin's attack, but the R2's envelope shape didn't resemble the amplitude properties of the audio envelope. And also R2's inflexion points were not perfectly synchronized with the audio signal. Even so, the R2 similarity rating were higher the the rating given to the analogue signal. So, also for this case, the participants were not capable to perceive the amplitude asynchrony of the audio and haptic signals.

Also, we consider that the R2 high average grade was caused by the larger amplitude of the R2 envelope compared with the analogue signal. Therefore, we consider that the users were able to perceive the R2's stronger vibration improving R2's similarity ratings. Also, as in the experiment #2 results, it seems that the users are not able to clearly perceive the timing difference on the decay between the R2 signal and the audio signal.

In addition, the signals B and C received lower ratings compared to the analogue ratings in every round. Due the experiments #3 and #4 results and the specific envelope properties of these stimuli; we expected this behaviour (see Figure 4.8).

The envelope B had the same attack but a different overall amplitude if compared to the audio signal. While, the envelope C lacked of attack but it had the same envelope shape at decay, sustain and release. Therefore this suggests that, the combination of an synchronized attack and a good envelope correlation at the decay-sustain-release phases, increases the perceived similarity. If the audio-tactile signal lacks of any of these properties this will decrease the perceived similarity between the signals, as happened with the envelope B and C.

Also, the one tail t-test performed between the stimulus ratings and the other signal's ratings. shows that, the B and C envelopes ratings, for the rounds 1 and 2, have statistical difference between the analogue signal ratings. This give a statistical foundation to our assumption.

In the round #3 there wasn't any other haptic signal that received similar or better ratings that the analogue stimulus (see Figure 4.8). In this case the R3 signal lacked of decay and sustain, but it's attack was the same as the violin envelope, even so the R3 signal received lower grades compared to the analogue stimulus. This indicates that the users were able to differentiate the amplitude differences during the decay-sustain-release phases. Suggesting one more time, the importance of the overall envelope shape for the amplitude similarity perception.

Additionally, in round #3 the C and R3 envelopes received similar grades, slightly lower than the analogue ratings, but the t-test results didn't show any statistical difference between C or R3 grades and the analogue signal grades. We suspect that, these results may be caused by the temporal amplitude changes of C and R3 were not similar enough to the audio signal. Specifically, C didn't have a synchronized attack, and R3 didn't have any envelope correlation during the decay-sustain-release phases. On the other hand, the envelope B rankings showed statistical difference from the analogue signal grades. Mostly because, after the attack, the envelope shape of B was totally different if compared to the audio signal envelope.

4.4 Results and Discussion

The experiments #1 and #2 helped us to understand how a haptic signal can be used to resemble one instrument in a musical piece. The results of these experiments show that the users clearly preferred the analogue envelope, when the amplitude and frequency of the haptic signal were constant. But, when the amplitude and frequency change between notes the users preferred haptic signals that didn't resemble the envelope shape of the audio signal. So at variable frequency and amplitude the users tend to overlook the envelope shape similarity between the signals.

Therefore, we suppose that the dynamic amplitude detection range of the haptic receptors is more narrow if compared to the same range in audio. So, it seems that displaying the notes' key though amplitude and frequency variations in addition to the instruments' envelope through a haptic signals saturates the haptic mechanoreceptors. Consequently, it seems necessary to omit the frequency and amplitude variations, so the user could be able to perceive the envelope cross-modal similarity.

After performing all the described perception experiments #3, #4 and #5. We can conclude that is not necessary that the haptic signal had a perfect envelope correlation compared to the audio signal, to be similarly perceived. This mostly due, to the differences perception abilities between our auditive and somatosensory senses. This may difficult the perception of the envelope shapes of both signals during time. Even if we are not able to perfectly perceive the envelope correlations in an audiotactile signal, the performed experiments prove, that the envelope shape is important the similarity perception of both signals.

The experiment #5 proves that is not necessary to use the exact same envelope on both signals to create a highly resemblant audiotacilte signal. Because in this particular experiment we created several syntethic haptic signals without the same envelope characteristics as their audio signal, but even so their perceived similarity ranking were as high as the analogue audiotacile signal rankings. We noticed that this increase on the perception similarity may be caused by two specific properties of the audio signal: Its synchronized attack and their overall envelope decay rate.

In a closer detail, the experiments #4 and #5 results show that is important to synchronize the attack of the audio-tactile signal, in order to increase the stimuli perceived similarity. In addition, the experiment #4 shows that, the perceived similarity of the signals decrease if the a haptic attack is presented before the audio attack. We suspect that this maybe related with the reported audio-tactile asynchrony perception [1], so the participants are able to detect the attack asynchrony between the signals. Even so, we cannot mention and specific threshold for this phenomena, due that the experiments were not meant to find this specific threshold. In contrast, the similarity perception doesn't decrease when the haptic attack is displayed after the audio attack.

The experimental results, in the experiment #4 suggest that, for the similarity perception between the audio and tactile signals is important to consider the envelope shape of both signals. This results show that the participants were unable to detect the changes between the different haptic signals during the decay and release stages, even if the participants know where the changes on the haptic signal were made. In addition, the experiment #4 also confirmed that is not necessary that both signals have the exact same envelope shape during the decay-sustain-release phase in order to be similarly perceived. Therefore, we suppose that humans are not able to detect small variations in the decay of the haptic signal because though the decay the tactile mechanoreceptors do not have enough time to be depolarize in order to perceive another stimulus.

Even if the participants were unable to distinguish the differences in the decay and release of the haptic signal, as shown in the experiment #4. The experiment #5 shows that, the participants were able use the general envelope shape during the decay-sustain-release phase to judge if the audio and haptic signals were similar or not. So, these results suggest an inability to distinguish small amplitude changes during the decay, sustain or release phases. But on the other hand, the same results show the participants used the overall envelope shape between the audio and haptic signals, to judge if these were similar or not.

From the results all the performed perceptual experiments. We can summarise the specific finding of each experiment into 3 simple guidelines:

1. To improve the similarity perception, both signals should be synchronized at the attack and their general envelope shape should be similar.

2. The users are not able to distinguish the temporal amplitude characteristics of the signal during the decay or release.

3. For music, the frequency and amplitude variability can mask the envelope correlation perception between the audio and haptic signals.

We consider that these guidelines could be used to design audiotactile signals with a better perceived similarity between the audio and haptic signals. Also these guidelines may help us to design simpler haptic signals that can also resemble an specific audio signal, this can be helpful in cases were haptic devices with limited capabilities are being used. And finally the last guideline can be helpfull to design better strategies to create haptic signals that resemble music.

As mentioned these guidelines were successfully applied to design the Haptic Music Player haptic module. Showing that we can use a haptic device with limited sampling rate capabilities to display a haptic signal that resembles an instrument in a MIDI song. By following the guidelines 1 and 2 we were able to create synthethic resemblant audiotactile signals without using audio itself as input.

Chapter 5

Haptic Music Player

In this chapter we are going to explain the implementation details of all the different modules that compose the Haptic Music Player. The Haptic Music Player is a virtual environment designed to improve the music listening experience. To achieve this the Haptic Music Player creates synthetic haptic vibrations that are designed to resemble the audio of one specific instrument in the song, in addition the user can also see a simple animation of the music. By these means, we tried to create a virtual environment where the user can naturally have a deeper understanding of the music that is listening, just like a trained musician could have. In specific words, we tried to focus the music listening attention to the user on an specific instrument in the song. So, the user can naturally understand the role of that specific instrument in the song and consequently develop a more critical music hearing.

We called this entertainment environment the Haptic Music Player, because it is basically that, a MIDI music player that simultaneously play a haptic signal with a highly resemblance to the audio of one instrument in the song. In addition to the haptic signal we created a synchronized visual stimuli, based on the work of Stephen Malinowski [28]. The relation between the visual stimuli and the music is easy to understand and self explanatory. So, with the combination of the synchronized visual and haptic stimuli we tried to catch the user listening attention into the song itself. So the user can have a pleasant, interactive and closer listening music experience (see Figure 5.1).

It may seem evident that, if a redundant visual-tactile-audio stimulus may be more effective to catch the participants attention compared to a bimodal stimulus, like a: visual-tactile or a audio-tactile stimuli. Although, previous literature on bimodal stimuli perception indicates that humans rely more on the visual sense when they are exposed to a redundant visual-tactile or visual-audio stimuli. This effect is known as the Colavita visual dominance effect [46]. On the other hand, more recent research indicates that when a tri-modal stimuli is presented the Colavita effect disappears, so the participants effectiveness on the simple detection and search tasks increases when using a tri-modal stimuli, compared to their performance using only a visual stimulus [21].

Therefore, due to the Collavita effect and the fact that music is a complex stimulus it was mandatory to evaluate if the stimuli used in the Haptic Music Player in fact improved the selective listening of the users when using a tri-modal stimuli. Therefore, the effectiveness of the proposed entertainment environment was tested on controlled conditions, using specific music and stimuli. In this psychophysical experiment the users selective listening was evaluated using bi-modal and tri-modal conditions in order to understand if a synchronized visual-tactile-auditive stimulus was in fact more effective fix the selective listening of the participants into an specific instrument in a song, compared to bi-modal stimuli (audio-tactile or visual-tactile).

5.1 Implementation

To build a synthetic haptic signal for this particular virtual environment we applied the results from the performed perception analysis. In order to maximize the perceivable resemblance between the haptic and audio signals. The particular methodology used to create the haptic signal was almost totally based in the findings of the performed psychophysical experiments. In addition we performed a subjective study to evaluate if the high resemblance between the auditive and haptic stimuli may allow the users to pay easily pay attention to the music.

Is important to notice that the haptic vibration and the music animation, where designed to be easily understandable by any user, even if the user was a small child or a person with no musical education. To achive this goal the Haptic Music Player was designed to take advantage of the inherent human ability to understand music to naturally focus the listener attention into an specific instrument or melody. So without any explanation any user can use the the system and naturally appreciate and enjoy music with a deeper understanding.

The Haptic Music Player is composed of three main modules: the music visualization, the haptic module and the music module. The audio module is the only responsible of playing the music to the user. The music visualization helps the user to see a simplistic representation of the score of each instrument, while the haptic module is responsible of creating a synthetic haptic vibration of one instrument in the song. To display the correctly haptic signal the complete environment had to run on real time. Consequently, to optimize the environment performance the mentioned modules had to be implemented on independent threads using different refresh rates (see Figure 5.1.

5.1.1 Audio - MIDI Player

The audio module is the responsible of playing the music to the user. For simplicity we decided to implement a MIDI player. Because, the discrete format of the MIDI messages provide us with the necessary data to create the synthetic haptic signals. Data like: the individual notes' pitch, duration and instrument type. Otherwise, if a different audio encoding format like MP3 or WAV had been used, then the audio signals had to be analyzed in real time in order to localize and separate the instruments, to then estimate the pitch and timing of each individual note. Due to scope of this research, we opted to use the MIDI file format instead.

MIDI messages

MIDI or (Musical Instrument Digital Interface) is a technical standard that describes how digital instruments can be connected with computers. This standard used simple MIDI messages to specify the notation, pitch, play velocity, volume, vibration, panning, etc. It also provides with a set of instructions to control the tempo. It is mostly used to communicate one or multiple digital instruments with a computer.

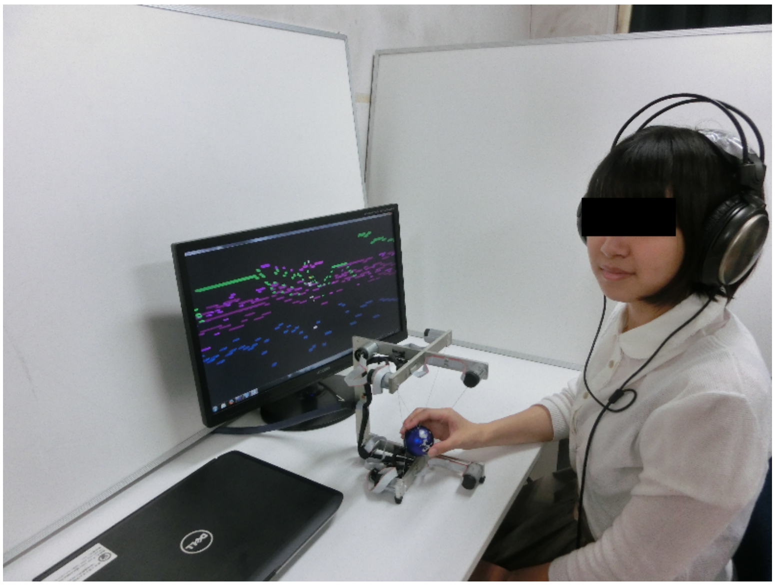


FIGURE 5.1: This figure shows the complete Haptic Music Player environment. The environment plays MIDI song, along the music. The user can also see a graphical representation of the song melody, while they can feel the vibration of an specific instrument of the song. The complete system helps the users to focus their listening attention to have a closer and deeper understanding of the music.

But it can also be used to compose music in MIDI file format and it could be played using virtual MIDI synthesizers. This protocol was standardized in 1993 by music industry representatives and currently is maintained by the MMA (MIDI Manufactures Association).

The MIDI protocol constitutes of a set of standardized instructions. These instructions are very compact and usually can be defined within 4 to 6 bytes, depending of the type of instruction. The most important an usual instructions are the MIDI Note ON and MIDI Note OFF instructions. The MIDI ON and MIDI OFF are specified with 3 bytes. The first part of the first byte, the 3 higher bits, specifies the MIDI command type, the second part of the first byte, the 4 lower bits, specify the channel where that message belongs to (a value from 0-15). Then for the Note ON or OFF instructions are followed by two more bytes, the first byte is used to specify the key of the note (pitch) and the second is used to specify the note's velocity. It must be consider that Note ON and Note OFF message should be always be in pairs. Because the Note ON message is used to start playing an specific instrument with an specific pitch and velocity, while the Note OFF turns off the Note ON event with an specific velocity.

To control the timing, usually all the MIDI files contain a tempo and a time signature messages at the beginning of the MIDI file. The tempo messages indicate how many microseconds a quarter notes should have or in music terms how many (Beats per Minute (BPM)) the song should have, while the time signature messages indicate how many demisemiquaver a quarter note should have. In addition, to the resolution of the midi file (ticks per quarter note) is also used to define the tempo of the song. Also, while the song is playing, the tempo of the song can dynamically be adjusted by adding MIDI tempo messages. Logically, any adjustments to the previously mentioned commands will only change the tempo of the song, it will not change the melody indicated in the MIDI On and MIDI Off events.

MIDI ticks are a mechanism created to control how fast or slow the MIDI messages are played. So it is like an digital metronome. By default, the MIDI clock ticks are defined by the standard to tick 24 times for every quarter note. So even if the MIDI player is running on a multimedia timer (a timer with 1000Hz update rate) the MIDI ticks provide an steady tick to process all the rest of the MIDI messages included in the file on their appropriate timing.

As mentioned before the tempo of in a MIDI file can be changed dynamically by using special MIDI messages called MIDI Tempo messages. As their name says, when the MIDI Tempo messages are parsed the BPM (Beats per Minute) or PPQN (Pulses Per Quarter Note) value is adjusted to the value specified by the message. The MIDI tempo messages have a similar message format as the Note OFF or Note ON MIDI messages. The first 2 bytes specify the type message, while other 3 bytes are used to indicate the tempo value given in micro seconds per quarter note (PPQN). So when the value of the PPQN (Pulses Per Quarter Note) is adjusted then the MIDI clock ticks change rate, and consequently also the MIDI message processing rate also changes.

Audio Module Implementation

We implemented the audio player by using an open source library called OpenMIDIProject [30]. With this library we can parse the content of any MIDI file and play them. To play the MIDI file, first all the MIDI messages contained in the MIDI file are parsed and saved on a list. Then with a multimedia timer each of the MIDI messages saved in the list is send to the Windows Synthesizer and finally the Windows Synthesizer produce the audio signal. Also this library provides specific examples to control the MIDI clock showing the user how it can be inserted inside a high refresh rate multimedia timer.

The dynamic changes of tempo present on most of the MIDI song files, makes mandatory to synchronize the visual and haptic modules by using MIDI clock ticks. Otherwise, all the different modules running on different threads will not be able to synchronize properly. So for our specific implementation the MIDI clock ticks where used to synchronize all the other modules. It was not necessary to change the threads update rate, the processing rate of all the threads was adjusted automatically every time a new tempo message was found. So the audio module played a key role on the implementation of the Haptic Music Player, because the visual and the haptic modules were synchronized by using the MIDI clock timer ticks.

In addition, we took advantage of the MIDI specific and compact command structure, to know the instrument type, note's pitch, note's duration to build the synthetic haptic signal on real time. So, while the message MIDI messages were processed, the haptic module created the haptic signal using the MIDI clock ticks and the note's information the haptic module create and synchronize it's output with the audio signal generated by the MIDI synthesizer.

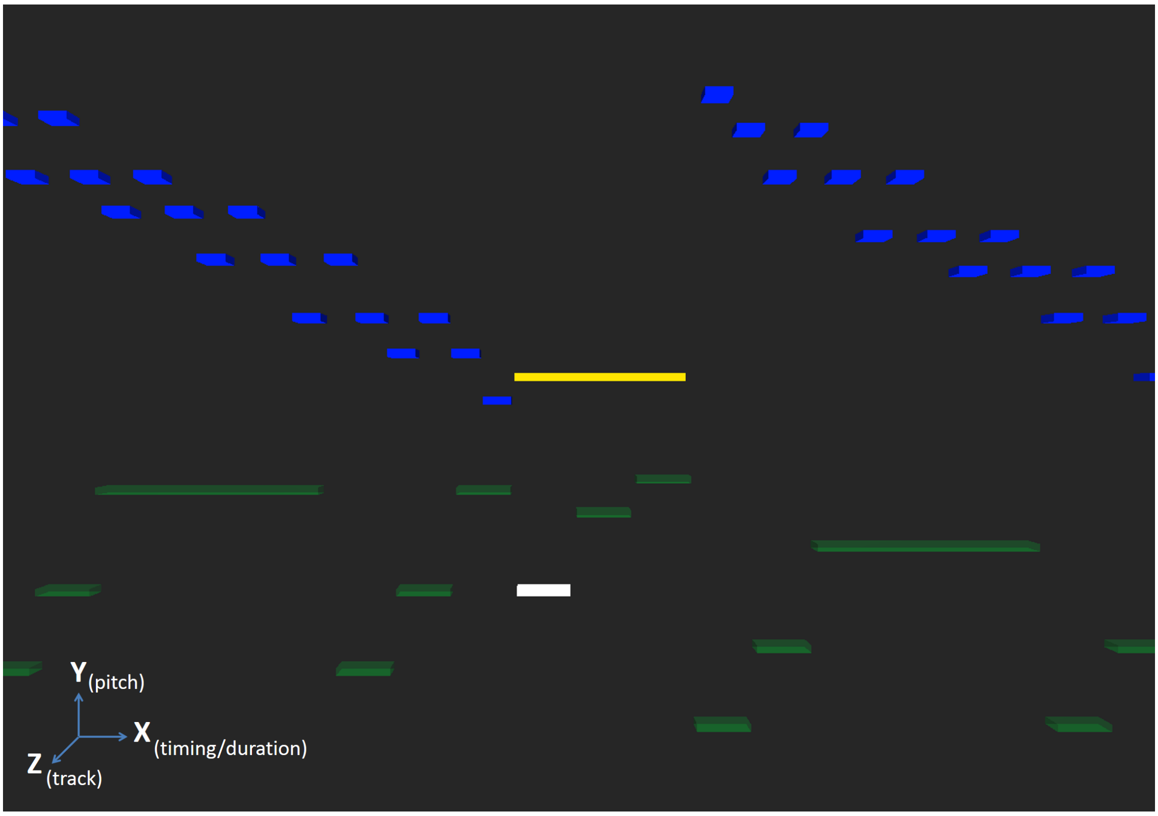


FIGURE 5.2: The visualization of a song with 2 tracks, shown in blue and green. The active haptic track is shown in blue. The current haptic played notes are shown in yellow, while the current played notes of the other tracks are shown in white. The height of each rectangle represents the note's pitch, the rectangles' length represent the notes duration and the timing of each note is represented with the position of each rectangle along the X-axis. Finally the relative position of each track on the Z-axis is used to point out the current active track.

5.1.2 Visual - Music Visualization

The Haptic Music Player visual module is responsible of displaying a simplistic but representative animation of the music melody. The principal purpose of the graphics is to provide some information to the user about the song that the participant is listening, so the participant could naturally have a deeper understanding of the composition. Also this module, along the haptic vibration, helps the user to identify and follow an specific instrument in the song.

The presented animation was build using OpenGL, and it is based on Kevin Kelly's Music Animation Machine project [28]. In the animation every individual note is represented using 3D rectangles. The rectangles' length, position and color represent different properties of every note. The notes' length is represented by using the rectangle length. The rectangles' position in X-axis represents the note's start timing. The rectangle's position in the Y-axis represents the note's pitch. So the notes with a higher pitch are placed higher that the notes with a lower pitch. The rectangles' Z-axis position and color are used to order the different instruments (MIDI tracks) of the song. The rectangles of different tracks are ordered with different depth over the Z-axis. Then the track that is being haptically displayed, by the haptic module, is always shown on the top. So the tracks are rearranged every time a different haptic track is selected (see Figure 5.2).

Additionally, the rectangles move around the screen from right to left with the same tempo as the music. Consequently, the notes that are going to be played are on the right of the screen, while the current notes that are being played are in the middle and the notes that had been played are on the left side of the screen. Also when the rectangles' respective notes are played their color change to identify them. Then, after the notes are being played, the rectangles continue their way though the screen from left to right until they disappear from the screen. Then new rectangles come from the right of the screen and the cycle repeats until the song plays completely (see Figure 5.2).

This module also heavily relies on the MIDI messages and the MIDI Clock synchronization to display the graphics. In contrast to the haptic module, the animation module doesn't work on real time. All the graphics are displayed every time a new MIDI file is selected to be played. So, when the MIDI file is open during the variable initialization process the complete MIDI file is parsed to generate an special list of objects that define every one of the rectangles that represent each instrument note.

During this pre-process phase all the MIDI Note ON and Note OFF messages are parsed for every individual track. During the parsing, when a combination of Note ON and OFF commands was found, the note's properties like the pitch, duration, instrument type and start timing and end timing where used to define the rectangular object used for the graphics. So, all the rectangles properties are defined in advance before the MIDI file is even being played.

After the rectangle objects are created then these are displayed using a OpenGL function to scale the rectangles size and position accordingly to camera frustum. So it doesn't manner the number of instruments or notes contained in the MIDI file, the camera frustum and the rectangles size will be adjusted accordingly to display all of them.

Once the notes are scaled and displayed and the camera frustum is adjusted, the MIDI file starts to play. While it plays the position of the camera in the X axis is moved accordingly to the MIDI clock ticks. By these means the camera will pan though all the displayed rectangles synchronized with the MIDI clock tick.

In addition, if a MIDI Tempo message changes the rate of the MIDI ticks the camera will automatically adjust and its velocity will change. Consequently the camera position and its panning will be always aligned with the current MIDI timing. This camera paning, creates the illusion that the rectangles are moving from right of the screen to the left of the screen and the currently playing notes are always in the center of the screen.

5.1.3 Haptics - Synthetic Haptic Vibration

This module creates the haptic signal by considering: the notes' pitch, notes' duration, the note's timing and the instrument envelope characteristics. And the resulting haptic signal is then displayed in synchrony to the haptic signal.

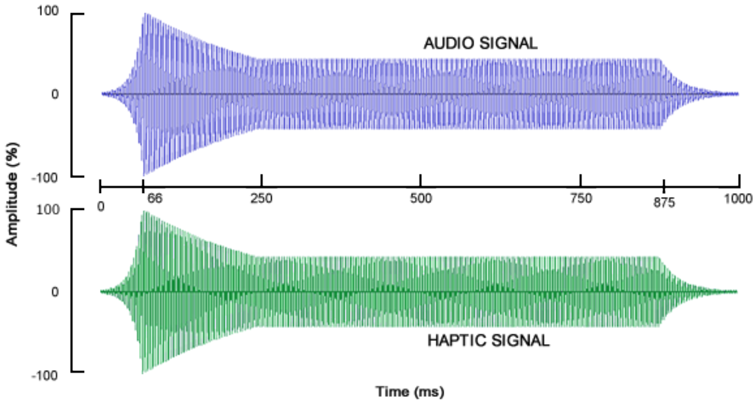


FIGURE 5.3: This image shows the audio waveform of a MIDI violin in blue. The correspondent haptic signal, shown in green, has the same temporal amplitude characteristics as the audio one. So the envelope of both signals are the same.

5.2 Vibrotactile Signal Modelling Technique

As previously mentioned before in real time the haptic module creates a synthetic haptic vibration by using the instruments information and the note's pitch, duration and timing, all of them provided by the MIDI instructions.

Also, as discussed in the Chapter 2, we can notice that there isn't any straight forward or general method to map audio into haptics. Most of previous efforts on this matter defined his own mapping method depending on the hardware used to display the haptic signal [25], [9] or the specific purpose of the audiotactile signal [35]. That is why, we want to provide a more general solution to this problem, so no matter the hardware or system purpose, this method should create a synthetic haptic vibration with a high perceived similarity to its correspondent audio signal.

To provide a more general solution to the audio to haptic mapping problem, crossmodal perception studies must be considered or performed. Due to the fact that the audiotactile crossmodal literature is scarce and do not specifically addressed the similarity perception between audio and tactile signals. Then we were performed those studies, to gradually to improve the method used to translate an audio signal into the haptic perceivable domain. The performed studies specific details are mention in the Chapter 3, but the general results of this studies show us that our ability to compare the similarity between an audio and a haptic signal is very poor. Nevertheless, the same performed studies show us 3 important things to be considered:

First, to improve the similarity perception both signals should be synchronized at the attack and their general envelope shape should be similar. Second, the users are not able to distinguish the temporal amplitude characteristics of the signal during the decay or release. Third, the frequency and amplitude variability can mask the envelope perception (see Figure 5.3).

The previously mention results, are mostly applied to design of the envelope shape of the signal, as mentioned in detail in the Section 5.2.3. On the other hand, to find a general strategy to map the frequency and amplitude of the MIDI notes into a haptic signal, only previous audio and haptic perception literature was used, as

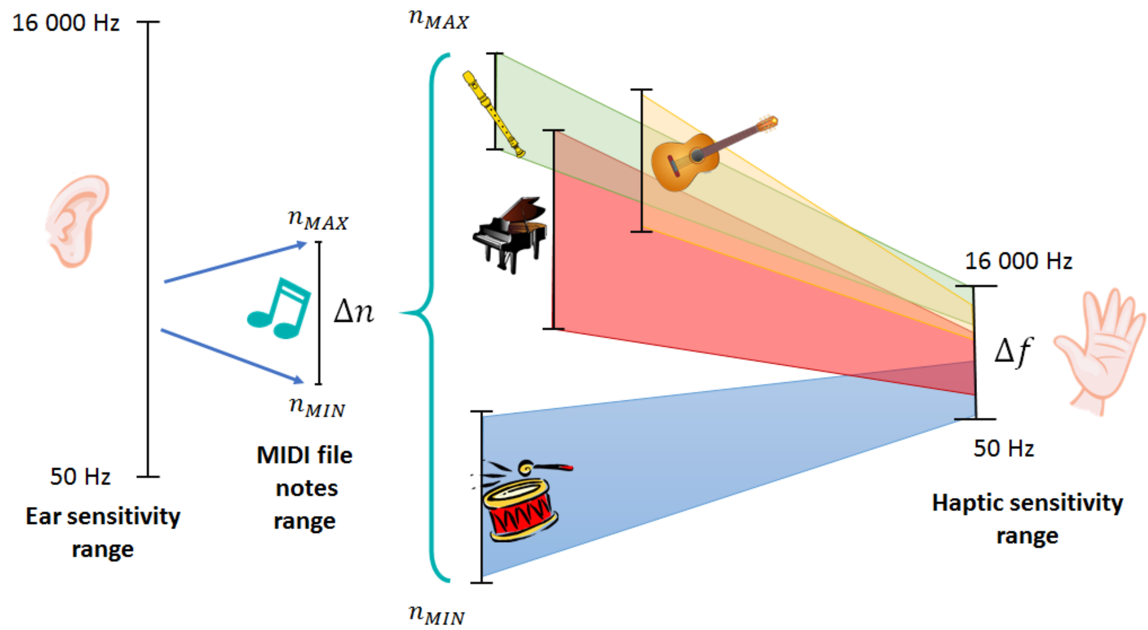


FIGURE 5.4: The figure shows how the notes frequencies presented in the MIDI file ($\triangle n$) are mapped into the human frequency perception range ($\triangle f$). Notice that not all the possible earing frequencies are directly mapped. Also notice that depending on the instrument registry the notes are mapped into the frequency perception range ($\triangle f$).

mentioned in detail in the Section 5.2.2.

5.2.1 Frequency Mapping

If the human and auditory senses are roughly compared in terms of their perception limitations. The auditory sense performance to perceive different frequencies is outstanding, with a frequency JND (Just Noticeable Difference) of 0.6% for frequencies around 1000Hz [6]. In contrast the haptic sense has a very poor performance with a frequency JND of 18% [40]. Now, if both modalities are then compared based on their perception range, the frequency hearing range is very wide, with: 0.0032 kHz ~ 16 kHz [22], while the somatosensory sense has a narrow perception range between: 20 Hz ~ 700 Hz[49]. Also, there is evidence that humans are able to tell if a pure tone haptic vibration has the same frequency as a pure tone audio signal. But this has only been proven for low frequencies rates between 50 Hz to 250 Hz [3].
After this rough comparative is evident that the somatosensory sense is unable to detect the frequency changes with the same sharpness and wideness as the auditory sense. Then we consider, that trying to directly map all the audible frequencies or to use audio itself as a haptic signal are inadequate methods to create a synthetic resemblant haptic vibration.

The method employed to translate the MIDI note's frequency and amplitude employs two lineal mappings to establish the frequency and amplitude of the correspondent haptic signal. Therefore, this method is not a straight forward mapping between the auditive perception range and the tactile perception range (see Figure 5.4). Instead, the purposed method takes advantage of the MIDI data structure to narrow the possible frequencies to be mapped. First the lowest (n_{min}) and highest (n_{max}) present in the MIDI file are identified, to define the complete note frequency range (Δn). Then to maximize the frequency range, the number of different playing notes present in the song is identified (n_ϵ). So by using n_ϵ the notes position in the song is narrow by performing the lineal mapping, shown in the Equation 5.1.

$$n_{fh} = n_\epsilon \frac{n_{max} - n}{\Delta n + \frac{\Delta n}{2}} \tag{5.1}$$

Then after all the note's position are mapped inside the unique frequency range in the song, every note will have a mapped position (n_{fl}) inside that mapped range. Then n_{fl} is used for the following mapping, where the note mapped position n_{fl} is then mapped into the human perceivable haptic frequency range. To do so we used another lineal mapping with another set of other parameters to map the audio into a pre-defined frequency mapping range (Δf), that should be in the human frequency perceivable range. This new mapping uses the notes mapped position (n_{fl}), the song's frequency registry range (Δk) and number of unique notes present in the song (n_ϵ). Then the frequency is obtained as described in the Equation 5.2.

$$f_h = f_{max} - \left(\frac{\Delta f}{n_\epsilon}\right) n_{fh} \tag{5.2}$$

We also take advantage of the MIDI command format instructions and use the range of each instrument to map the relative height of each instrument into its respective haptic frequency range. This means that if the instrument plays low frequency notes then it will be mapped on the lower perception frequency range and if it plays high frequency notes then it will be mapped on the high part of the frequency range. To implement this, first the registry of each instrument is identified by comparing if the average of all the notes played by the instrument is lower or higher that average of all the notes played in the song. After identifying each instrument as a "low" or "high" instrument, then a special case of the Equations 5.1 and 5.2 is used to map instruments with a low note registry, as shown in the Equations 5.3 and 5.4.

$$n_{fl} = n_\epsilon \frac{n - n_{min}}{\Delta n + \frac{\Delta n}{2}} \tag{5.3}$$

$$f_l = f_{min} + \left(\frac{\Delta f}{n_\epsilon}\right) n_{fl} \tag{5.4}$$

Then after the user is identified the Equations 5.1 and 5.2 will be used for instruments with a high note registry and the Equations 5.3 and 5.4 will be used for instruments with a low registry. By these means the resulting vibration of low frequency instruments will be lower if compared to instruments with a high pitch.

5.2.2 Amplitude Mapping

Beside the frequency, the amplitude was also mapped using the same method. So notes with a high frequency will have a higher initial amplitude at their attack and low frequency notes will have a lower initial amplitude. So, to map the amplitude of the each note an special version of the equations presented before (see Equations

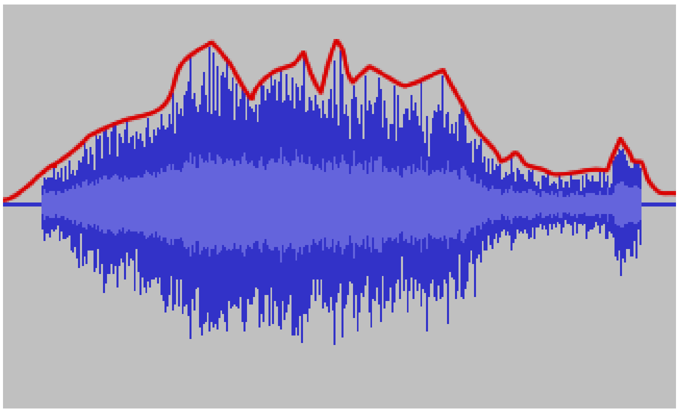


FIGURE 5.5: A simple example of an envelope function (in red) around a sound signal (in blue). The envelope function defines generally defines the amplitude changes of a haptic signal though time.

5.1 and 5.2) are re-defined in order to map the low frequency instruments. Just a few parameters of these equations are modified. So instead using the frequency perception range (Δf) this parameter was changed to the amplitude frequency range value (Δa). Also instead of using the minimum frequency value of the perception range f_{min} this value was substitute by the minimum amplitude of the perception range a_{min}, as mentioned in the Equations 5.5 and 5.6 .

$$n_{ah} = n_{\epsilon} \frac{n_{max} - n}{\Delta n} \tag{5.5}$$

$$a_h = a_{min} + \left(\frac{\Delta a}{n_{\epsilon}} \right) n_{ah} \tag{5.6}$$

5.2.3 Envelope Mapping

The envelope can be mathematically defined as an continuous function that defines the outlining extremes of another function. In simple words it can be defined as a signal that defines the general amplitude changes of another signal though time, as shown in the Figure 5.5. Specifically for synchronized audio-tactile signals this changes may impact the way we perceived both signals, as we already told in Chapter 4, humans are very sensitive to sudden changes in amplitude for both haptics and audio. So, it seem a logical idea that for a highly perception resemblance between a synchronized audio and haptic signal, the envelopes of both signals should have a high resemblance on their respective envelopes; as shown in the Figure 5.3.

5.2.4 ADSR Envelope modelling

The envelope of the haptic signal was defined by considering the envelope properties of the specific MIDI instrument that create the specific audio signal. So, the envelope of the haptic signal will have the same amplitude temporal characteristics as the audio signal, and this was defined by the specific instrument in the MID song. By these means, we tried to define on the haptic signal the amplitude characteristics

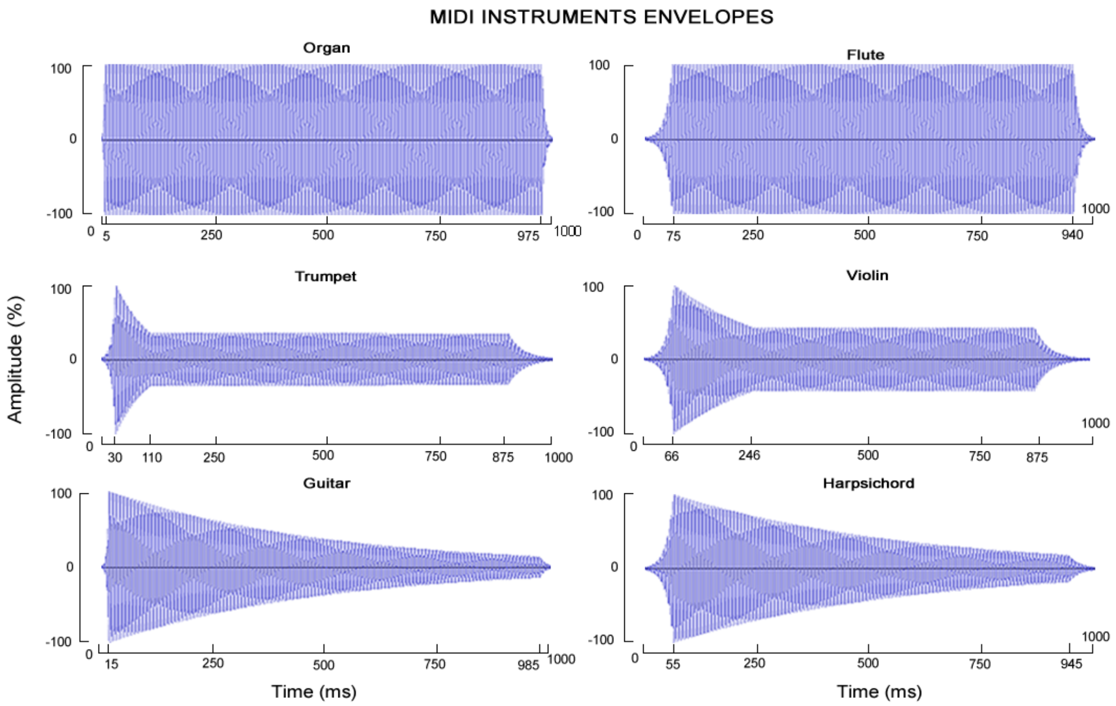


FIGURE 5.6: This figure shows the envelope shapes of some MIDI instruments. From the left down: Organ #1 (MIDI #16), Trumpet (MIDI #57), Acoustic Guitar (MIDI #25), Flute (MIDI #74), Violin (MIDI #40) and Harpsichord (MIDI #7).

of different of different musical instruments (see Figure 5.6).

So the define the envelopes of the haptic signal that represents each music instrument. First we measured and listed the general envelope shapes of different MIDI instruments. To do these measurements we used an oscilloscope and the default Windows MIDI Synthesizer, without adding any extra audio filters to the output audio signal. Then amplitude temporal properties of the most used MIDI instruments was measured. If the envelope of the measured instrument had an ADSR (Attack Decay Sustain Release) shape then the attack, decay, sustain and release phases were measured. If not then only the places with drastic amplitude changes were measured (see Table 5.1).

Subsequently, these measurements were used as parameters for an ADSR envelope to define the haptic signal envelope. So, the haptic signal envelope will have the same overall shape of the MIDI instrument that created the audio signal (see Figure 5.3).

Instrument	t_1 ms	t_2 ms	t_3 ms	$s = a_h\%$	$r = a_h\%$
Instruments' ADSR Parameters					
Organ	5	0	t_4-25	100	1
Flute	75	0	t_4-60	100	1
Harpsichord	55	0	t_4-55	100	17.39
Guitar	15	0	t_4-15	100	1.36
Trumpet	30	110	t_4-100	32.09	1
Violin	66	376	t_4-225	57.14	1
Cello	40	290	t_4-200	47.45	1
Contrabass	55	355	t_4-225	52.74	1
Simple Envelopes' ADSR Parameters					
Square	0	0	t_4	100	0
Triangular	0	t_4	0	100	0

TABLE 5.1: ADSR envelope parameters, used in the Equation 5.7, in order to generate the haptic envelope of different MIDI instruments. Also the parameters of their simpler envelopes, used in the performed perception experiments (see Chapter 3), are also mentioned in the table.

$$E(x) = \begin{cases} \frac{x*PA}{t_1} * \sin(x*f*2\pi) & x < t_1 \\ \left[\left(\frac{s-a_h}{t_2-t_1} * x\right) - \left(\frac{a_h}{t_2-t_1} * y_1\right) + a_h\right] * \sin(x*f*2\pi) & t_1 \le x < t_2 \\ s * \sin(x*f*2\pi) & t_2 \le x < t_3 \\ \left[s * \exp\left(\frac{\ln(r)*(x-t_3)}{t_4-t_3}\right)\right] * \sin(x*f*2\pi) & t_3 \le x < t_4 \end{cases} \tag{5.7}$$

The final haptic signal is created by using: An ADSR filter defined in Equation 5.7, the mapped haptic frequency. The amplitude is defined with the Equations: 5.2, 5.4 and 5.6 using the MIDI envelope measurements (see Table 5.1). Once the ADSR filter value and the frequency of the initial amplitude are known, then the final haptic signal is computed by Equation 5.7. Finally, by these means the haptic signal could have the same envelope as any MIDI instrument. Also the purposed method will keep the signal frequency constant even if the envelope shape is modified.

5.2.5 Envelope mapping and perception

But we are designing a totally synthetic haptic vibration that resembles audio, so we do not want to directly map the envelope shape of the audio signal into the audio signal. Instead we want to understand in which degree and which characteristics of the signal envelope have more impact on the similarity perception of both signals. By these means we can design a simpler but also resemblant haptic vibration. Then, the proposed strategy can be used to display an audiotactile signal with high similarity, even on limited haptic hardware with narrow bandwidth, weak amplification

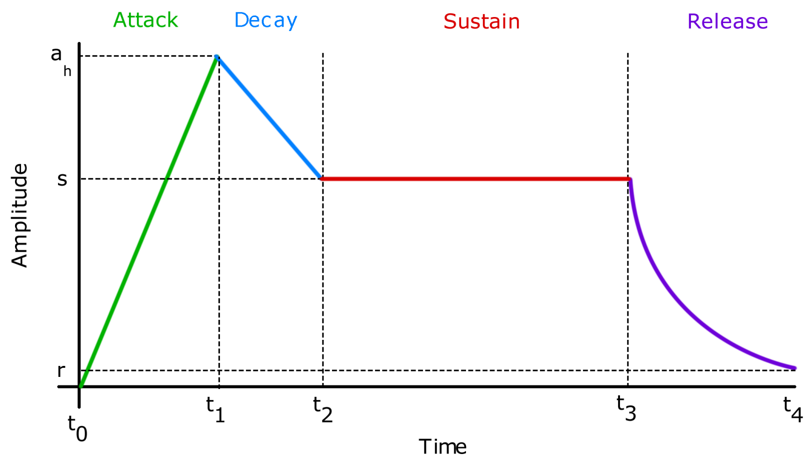


FIGURE 5.7: This figure shows a graphical explanation of the temporal and amplitude parameters used to define the ADSR envelope in the Equation 5.7. The timing parameters are $(t_0, t_1, t_2, t_3, t_4)$ and define the inflexion points between each ADSR phase. While the parameters to control the amplitude are: a_h peak amplitude, s sustain and r final amplitude.

or slow update rate.

To develop the haptic signal we used the results from the perception experiments and directly apply them to the implementation of our envelope mapping technique. The specific details about our perceptual experiments are mentioned in the Chapter 4, but the general findings the perceptual experiments are:

1. To improve the similarity perception, both signals should be synchronized at the attack and their general envelope shape should be similar.

2. The users are not able to distinguish the temporal amplitude characteristics of the signal during the decay or release.

3. The frequency and amplitude variability between notes can mask the tactile envelope correlation perception.

The first point of the findings list was highly considered on the envelope mapping implementation technique. In the first versions of the Haptic Music Player, the synchronization between the haptic signals was not considered at all. So, in the firsts versions there was a 25ms delay between the haptic and the audio signals, that some users reported some not. Doing more research about this subject, we found that Adelstein [1] reported an for audio-tactile asynchrony threshold of 24ms. So the firsts version of the Haptic Music Player had delay between the haptic signals and audio signals very close to the reported threshold value. Because of this implementation negligence, on the firsts public demonstrations of the Haptic Music Player, more or less 10% of the users reported an asynchrony between the audio and haptic signals. So, this let us to informally confirm Adelstein results.

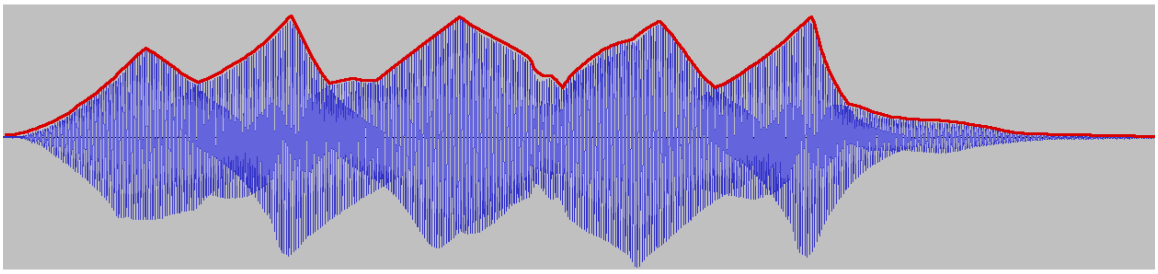


FIGURE 5.8: shows the envelope shape of the MIDI Instrument #50 (String Ensemble #2). The envelope shape of this instrument has to much reverb, making difficult to match all their temporal amplitude characteristics with an ADSR envelope modeling on a haptic signal.

The mentioned asynchrony bug was corrected in successive versions of the Haptic Music Player and the initial attack of both signals was controlled to have a asynchrony of ±5ms. And it was adjusted accordingly to the findings from the performed experiments on the ADSR audiotactile asynchrony perception (4.3.3).

Is also necessary to mention that is technically very difficult to have total control of the asynchrony and reduce it to cero. This problem is principally caused by the latency of the ADC (Analog-to-Digital Converter). When the sound card transforms the digital MIDI signal into an analog audio signal, the audio card itself introduces ±5ms of latency even if real-time drivers are used. Even with specialized real time audio hardware the ADC latency cannot be reduced to cero, it can just be controlled to be around ±2ms in the best case. Therefore, is technically impossible to generate an audio and haptic signals using two audio cards with a full control of the delay between them.

The second point of the findings lists, was also important to properly design the ADSR envelopes of every instrument. As mentioned before, the ADSR envelope shape of the haptic signals were model to have a very similar shape to the envelope of their respective MIDI instrument, as shown in the Figure 5.6. But, there are some specific cases were the proposed ADSR mapping technique were not able to perfectly fit the envelope shape of the MIDI instrument. In one case this can be caused because the velocity of the Note ON MIDI commands was not considered. For example the drums in MIDI are very sensitive to the velocity parameter changes. For this instruments if the MIDI Note ON command specified a very high or low velocity, the envelope of the haptic signal during the decay and release will not perfectly match with its audio counterpart. In another case, there are instruments like the MIDI Orchestra (MIDI instrument #50) that have a audio envelope shape with high reverb, that will be impossible to match using an ADSR envelope; as seen in the Figure 5.8.

Being rigorous, the previously mentioned issues may seem irreconcilable with the ADSR envelope technique, but the the second point of our finding list lets us consider that just by reproducing the general envelope shape of the audio signal on the haptic signal, we can create a highly resemblant vibration.

For the case of the high sensitivity of the MIDI drums to the velocity property, it is not so important if the changes in decay or release do not perfectly match on both signals. What really matters in this case is to match the attack of the audio and haptic signals and model the rest of the haptic envelope with an envelope shape that resembles the general envelope shape after the decay and matches note's length.

In addition, for MIDI instruments that have reverb, it may seem plausible to introduce a sine wave form during the ADSR sustain phase to model the MIDI Audio reverb on haptics, but though some informal experiments we noticed that using an triangular or even an square envelope shape was enough for the haptic signal, mostly because we are not able to perfectly perceive and match those high frequency amplitude variations in both senses.

The third finding from the performed perception experiments, let us consider that is not necessary to perfectly match the temporal amplitude characteristics of both signals, because its perception can be masked by the frequency and amplitude variations between the notes (see Section 4.2.3). In simple words, it seems that the users gave more importance the amplitude variations on the attack of every note confused the users. So this means that the users seem to use the general amplification shape of the haptic and audio signals to find the similarities between them. So, for high pitch instruments or for instruments that play with a high volume a haptic signal with a high amplification is also expected.

5.2.6 Haptic Display Hardware

After the haptic module of the haptic music player computed the haptic signal, the signal should be displayed in a haptic interface so the user will be able to touch it. The specific haptic interface used to display the computed haptic signal was SPIDAR 6G [29]. This haptic interface is composed of 8 motor/sensors mounted on an aluminium frame, all of the motors are symmetrically connected with an string to a plastic ball, so whit the motors pull the plastic ball can levitate around the area inside the aluminium frame. The ball is position and rotation is calculated by taking advantage of the frame geometry and the length measurement of each individual thread, as mentioned in Kim's paper [29].

This haptic interface is ideal to interact with full 3D environments (see Figure 5.9), because provides 6-DOF (Degrees Of Freedom), it has a very low latency. By using the 8 motors it also can display relatively strong forces. And it's update rate of 1kHz allows the display of wide variety of materials.

We understand that the use of this particular interface seems inappropriate for the kind of interaction that we proposed, because we are not using all the features that SPIDAR G6 can offer. Due to the kind of passive interaction that we propose SPIDAR G6 is only used to display the haptic vibration to the user, so we didn't take advantage of the position, movement or orientation of this haptic interface.

As mentioned before, the main purpose of this research is to find a general method to create a resemblant synthetic haptic vibration from audio signal to improve music listening. It would be trivial just to use and speaker and directly use audio as a haptic signal. So by using SPIDAR, to implement the Haptic Music Player, we proved that is possible to use a haptic device with limited characteristics to display a resemblant audio haptic vibration. Because SPIDAR cannot be directly used as transducer, because it has a low refresh rate (1kHz) compared to the common minimum sampling rate for audio applications (20kHz).

FIGURE 5.9: This figure shows SPIDAR-G6 being used for a 3D virtual reality application. We implemented the Haptic Music Player in SPIDAR-G6 mostly because the performed psychophysical experiments required small latency between the audio and haptic signals. By using SPIDAR-G6 drivers we could achieve a delay time of ±2ms. © [2002] IEEE [29].

2 Mostly all the commercial devices that provide a haptic vibration have very limited capabilities, like: low amplification, slow update rates or a narrow bandwidth. So it is inappropriate to directly use audio on this kind of haptic devices. Therefore, the proposed methodology to create a resemblant synthetic audio-tactile vibration can be applied on a wide variety of haptic devices, like: smartphones, video game controllers, tablets or even in state of the art virtual reality manipulator, like SPIDAR.

5.3 Subjective Evaluation

As we mention at the beginning of this chapter, the principal objective of the Haptic Music Player was to improve the users music listening attention in an specific instrument in the song, by using a synchronized haptic and visual stimuli. But in order to certainly assume that we achieve our objective is to evaluate if the Haptic Music Player can help the user to concentrate their music listening attention.

5.3.1 Listening Focus Evaluation

To evaluate the Haptic Music Player performance, we performed a psychophysical experiment. For this experiment we focused on measuring the amount of time the user takes to focus his listening attention to an specific instrument. And also we focused on measuring the amount of time the users could pay attention to the instrument, without getting confused by the other instruments sound.

It is still and open subject how the mind attention can be accurately and directly measured. So, now a days the mind concentration or self focus can just be measured and studied by indirect methods that are still in debate [14] [12]. So, for this experiment we relied on a self report focus method, were the user hiumself reports if he is paying attention or not. This method to measure attention is fragile because relies on the user self-awareness on his own concentration. On the other hand, if this technique is applied carefully, it can be a simple, straightforward and relatively reliable solution to measure attention.

So for this experiment we relied on a self-awareness technique to measure the users' listening attention. While listening to a song, we shown the users an stimuli related to an specific instrument in the song, then we asked the participants to press a button if they can identify and follow an the instrument with the given stimuli. We also instructed the participants to release the button if they commit a mistake or if they were confused by the sound of the other instruments.

To overcome the disadvantages of the techniques that relied on the users' self-focus-awareness, we decided to minimize the time we asked the user to report their own focus. Then we defined small lapses of time where the users had to focus his attention and report it by pressing a button; we defined this time lapses as "events". For this experiment the events had 15 seconds long. In addition, 3 seconds before every event we used a visual warning to alert the user attention. Then the stimuli (haptic, visual or visual-haptic) was shown to the user for 15 seconds and then the event finished. So, during the event the user were instructed to continuously press the space bar if they can find and follow the instrument related to the given stimuli. So, we measured the users reaction time, by measuring the time from the first note playing to the users first space bar press. And also we measured the total focus time by measuring the total time the user kept the space bar pressed. (see Figure 5.10)

Is important to mention that, the event's timing and instrument were not randomly selected. We selected the start time and instrument of each event carefully, so the instrument would be challenging to find but also were not masked by other instrument playing notes on a different tone but with the same rhythm and timing. So we expected that, the reaction and focus time would be different for each individual event, mostly because the user's reaction and following time would depend on the selected instrument and the particular melody played by the other instruments in the song.

To non verbally specify which instrument the user had to focus on, we create diverse stimuli and synchronize them with an specific instrument in the song. For the experiment we created 3 different of stimuli: haptic, visual and visual-haptic. The haptic and visual stimuli were built by using the techniques previously mentioned in this chapter. And the visual-haptic stimuli was built by using a synchronized combination of a haptic and visual stimulus. Therefore the haptic and visual stimuli represented the notes' pitch, timing and duration.

To compare the haptic and visual modalities in fare terms, we limited the visual information shown to the user. Then we limited the number of notes and instruments shown to the user, so it can be fairly compared to the haptic stimuli. Consequently, the visual stimulus only shown one playing note of one specific instrument in the song. The visual stimuli was limited, but the visual stimulus represented the note's duration and pitch by changing the rectangle's length and height in the screen. So the visual stimuli could provide the same amount of information as the

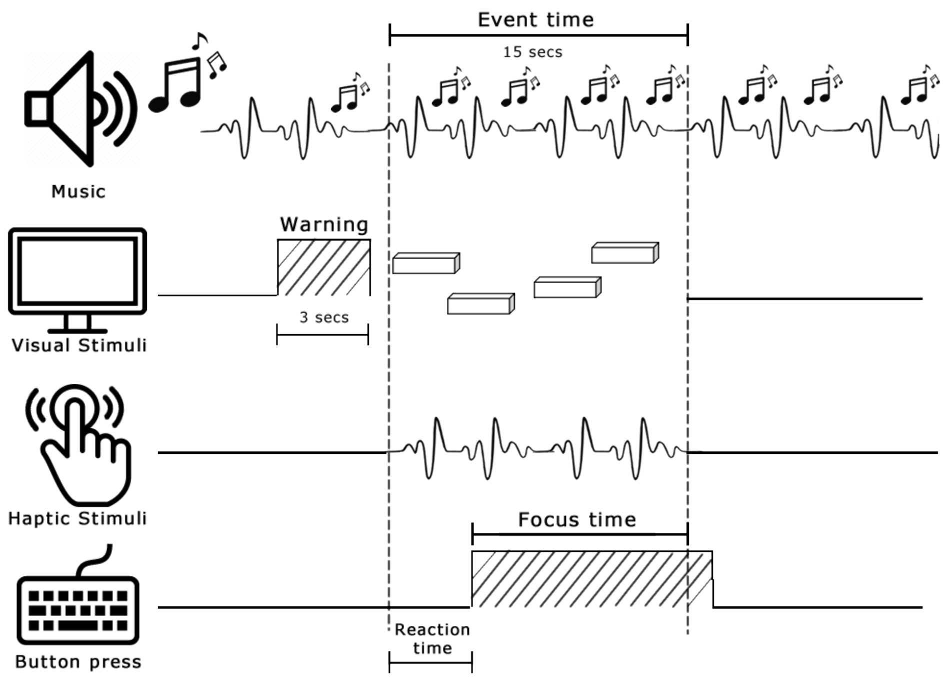


FIGURE 5.10: This diagram shows how the users' listening attention was measured. The listening attention was measured in events that lasts for 15 seconds each. Every song used in the experiment had 9 different events. To alter the user and have all his mental attention a visual warning was displayed 3 seconds before each event. During each event the haptic, visual or visual-haptic stimuli was shown to the subject, for 15 seconds. And the user task was to press the space bar if they were able to find and follow the musical instrument related to the given clue. By these means, we measured the subjects reaction time and the total time they could focus their listening attention on an specific instrument.

haptic stimuli.

Opposite to the previously performed perceptual experiments, described in the section 3, instead of using pure tonal audio signal for this experiment we used music. So for this purpose, we carefully selected 3 different classical songs: Johann Sebastian Bach - Little fugue in G minor (BMV 578), the last movement of the Mozart Adagio and Fugue in C minor (K. 546) and the 4th movement of the Haydn String Quartet Opus 20 No 6. For the experiment we required songs with certain degree of difficulty to find specific instrument in the song, that is why we selected only fugues for this experiment.

Fugues can be defined as music composition technique where a short musical theme is introduced at the beginning of the song and then variations of the original musical theme are gradually introduced on a different voices and tonalities (instrument). Then while a fugue gradually advance the song becomes more and more complex. So when all the voices are playing sometimes is difficult to distinguish an follow individual voice (instrument), because all the other instruments are playing variations of the same short musical theme at the same time. Is also important to mention, that fugues are very complex to compose, because if each individual voice is played individually they have coherent melody by themselves, so trying to match all the variations of each individual voice in a complete song can be really challenging.

Then, we selected fugues because it is challenging for an untrained person to clearly identify and follow an individual voice (instrument) in the song. Consequently, the challenge of the task was not defined in the number of different instruments playing at the same time, instead the challenge of the task was defined on the structural and melodic complexity of the song.

In the experiments we measured the reaction time and the focus time of the participants on each of the 9 events presented in 3 songs for each of the 3 modalities. Due to the timing, duration and instrument of each event was previously selected, we tried to avoid any learning effect between the participants by separating the 3 different modalities for the same song on different days. Then each participant had to complete the experiment on 3 different days, on each day the participant had to perform the experiment for the 3 different songs on 3 different modalities until complete all the combinations. We separated modalities of the same song in different days, and we also wait at least 4 days to perform the experiment again. Then by these means we tried to minimize the learning effect.

In total, we measured the focus and reaction and focus or each participant 9 times per song, using the same event on 3 different modalities (visual, haptic, and visual-haptic) with 3 different songs.

In addition, to the data obtained by measuring the participants reaction and focus time, we also measure the user's subjective perception of the performed task on different modalities. Therefore, after the subjects ended the 3 experimental session we ask them to answer a survey, with 3 visual scale questions. The first question asked the participants if they like or dislike the modality. The second asked if they could easily understand the stimuli. And the third one asked them if they could easily follow the stimuli. (see Figure 5.11)

5.3.2 Experimental Conditions

The experiment was performed by 30 participants, 23 males and 7 females, from 22 to 60 years old any of them reported audition problems. And the experimental software used a laptop (DELL Latitude E5440) with Windows 7 and an Intel Core i5M processor.

The haptic signal was created in real time and then output into a external DAC (Digital to Analog Converter), then the signal was amplified by using a 2 channel digital amplifier (Lepai LP-2020A+). Finally to display the haptic stimulus to the

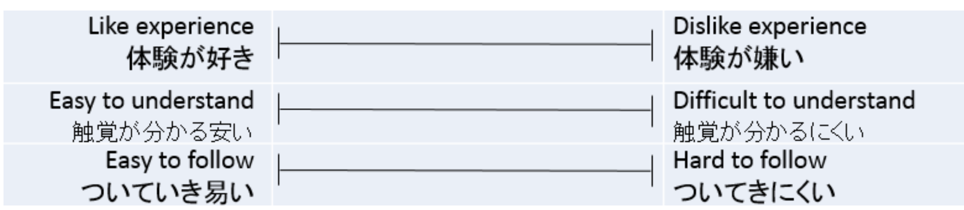


FIGURE 5.11: This figure shown the visual scale questions used in the subjective survey.

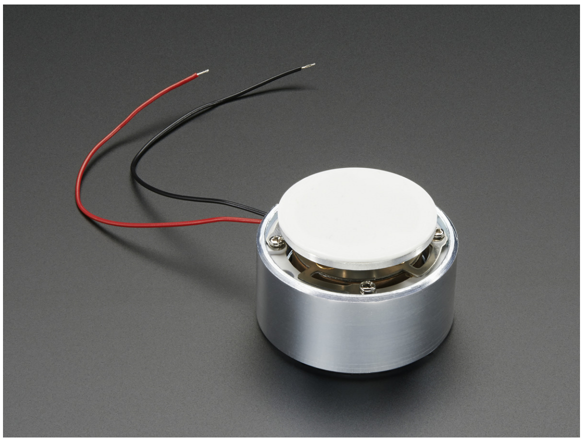


FIGURE 5.12: The figure shows an Adafruit surface transducer, which is the haptic device used to display the haptic signal for the Haptic Music Player evaluation.

participants, we used a surface transducer (Adafruit 4Ohm 5Watt transducer); as shown in the Figure 5.12.

The visual stimuli was limited to have the same characteristics as the haptic signal. Only one note of the selected instrument was shown to the participants, while the note's duration and pitch were represented on the stimulus length and vertical position in the screen. The visual stimulus was shown on the screen of the laptop used to perform the experiment.

To play the music, we used MIDI renditions of the tree previously mentioned songs. We used the default Windows Synthesizer to generate the audio signal and this was displayed with the computer's on-board audio output (Realtek HiDef Audio). The users listened to the song using a Sennheiser MX475 earplugs along with strong industrial full ear earmuffs.

For these specific case we preferred to use a surface transducer instead of SPIDAR, because for this kind of experiment it was not so important synchronize the timing between the audio and haptic signals. We only measured and controlled the attack of the haptic and audio signals to be between $\pm 5ms$. Based on the performed perception experiments findings we can conclude that a strict synchronization control between the haptic and audio signals was not necessary.

5.3.3 Results

To determine whether the different factors had an influence on the participants' responses, a multivariate analysis of variance (MANOVA) was performed. Since all of the songs included voices with a similar pitch register (see Table 5.2), it was possible to average the participants responses with a similar pitch register across different lapses and songs. Hence, the selective listening and detection responses of each measuring lapse were grouped and averaged, across songs, based on their respective highlighted voice pitch register. So for this analysis the lapses were grouped as follows: High (Mozart:Violin-1, Haydn:Violin-1 & Bach:Organ-1), Medium-1 (Mozart:Violin-2, Haydn:Violin-2, Bach:Organ-2), Medium-2 (Mozart:Viola, Haydn:Viola, Bach:Organ-3) and Low (Mozart:Cello, Haydn:Cello, Bach:Organ-4).

Song	Instrument	High	Low	Average
Mozart	Violin-1	89 - G5	55 - G2	75 - Eb4
	Violin-2	84 - C5	55 - G2	67 - G3
	Viola	68 - G#3	48 - C2	60 - C3
	Cello	61- C#3	36 - C1	48 - C2
Haydn	Violin-1	88 - E5	57 - A2	74 - D4
	Violin-2	83 - B4	56 - G#2	69 - A3
	Viola	76 - E4	49 - C#2	63 - Eb3
	Cello	66 - F#3	37 - C#1	52 - E2
Bach	Organ-1	84 - C5	57 - A2	73 - C#4
	Organ-2	80 - G#4	50 - D2	65 - F3
	Organ-3	74 - D4	38 - D1	59 - B2
	Organ-4	50 - D2	24 - C0	38 - D1

TABLE 5.2: The instruments' pitch register and average played note.

Effect	F	df	Error df	P value
Stimulus	11.400	4.000	672.000	**0.000**
Skill	30.816	2.000	335.000	**0.000**
Pitch	5.100	6.000	672.000	**0.000**
Stimulus*Skill	0.192	4.000	672.000	0.943
Stimulus*Pitch	0.873	12.000	672.000	0.574
Skill*Pitch	0.519	6.000	672.000	0.794
Stimulus*Skill*Pitch	0.323	12.000	672.000	0.985

TABLE 5.3: Multivariate ANOVA results table, inclusive of Pillai's trace. All of the main effects (stimulus, skill, & pitch) presented were significant ($p < 0.01$). However there was not a significant interaction between any of the simple main effects, which indicates the lack of any particular significant interaction between the participant's skill level or the voice's pitch with the stimulus condition.

Therefore, a three-factor multivariate ANOVA was performed using the following main effects and levels: the stimulus condition (with 3 levels: visual, haptic, and visual-haptic), the participant's skill (with 2 levels: players and non-players), and the voices' pitch register (with 4 levels: High, Medium-1, Medium-2, and Low); while the participant's detection and listening time were the dependent variables.

The data obtained from the evaluation satisfied almost all of the ANOVA assumptions. The variance and errors between the cells were similar. The measurements were independent from each group because all the participants performed the experiment with a random stimuli and song order. However, as expected for the reaction measurements [43], the detection and the listening time measurements were skewed. This skewness was normalized using a logarithmic transformation; thereafter, the same multivariate ANOVA test was applied to the original and transformed data in order to verify any difference in the statistical analysis between the transformed and untransformed data. Since both tests had the same results and interactions, the original untransformed data were used in this report.

In some cases, the participants were unable to provide an answer or their initial detection was unreflective and immediate (i.e., faster than the average reaction to haptic and visual stimuli < 200ms [13] [44]). In these cases, the generated data were replaced by the lapse's average in that stimulus condition. Only less than the 3% of the data were eccentric or missing, which allowed the use of this specific measure to fix this issue.

The results of the multivariate ANOVA, as shown in Table 5.3, were calculated with SPSS using Pillai's trace. These results show a significant main effect of the stimulus condition ($F(4.0, 672.0) = 11.4, p < 0.01$) and a significant main effect of the participant skill level ($F(2.0, 335.0) = 30.816, p < 0.01$). In addition, the voices' pitch effect was significant ($F(6.0, 672.0) = 5.1, p < 0.01$).

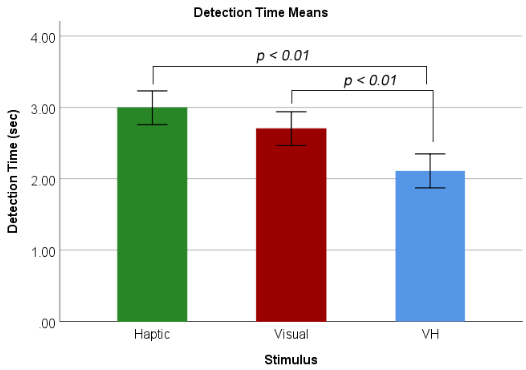
On the other hand, in regard to the simple main effects, a significant effect did not exist between the stimulus type and the participant's skill level ($F(4.0, 672.0) = 0.192, p > 0.01$), suggesting that the performance difference of each stimulus condition was the same, regardless of the participants' abilities. Moreover, a significant effect was not found between the stimulus type and the voices' pitch ($F(12.0, 672.0) = 0.873, p > 0.01$), thus showing that the performance of each stimuli was the same, regardless of the voice's pitch. Further, a significant effect was not present between the participants' skill level and the voices' pitch ($F(6.0, 672.0) = 0.519, p > 0.01$), meaning that players' and non-players' performances were the same regardless of the voice's pitch. Finally, a significant effect involving the combination of the three main effects was not discovered (Stimulus*Skill*Pitch) ($F(12.0, 672.0) = 0.323, p > 0.01$). In summary, none of the simple main effects combinations presented a significant interaction, as shown in the four last rows of Table 5.3.

Since the stimulus main effect was significant, a Tukey HSD pairwise comparison was performed in order to assess which of the feedback stimuli means were significantly different. The results of this analysis, depicted in Figures 5.13a & 5.13b, indicate that the participants' performance with the visual-haptic was significantly better, with faster detection times and longer listening periods, compared to the visual-only and haptic-only conditions. The visual-only and haptic-only means do not show any significant difference in their detection time or listening time.

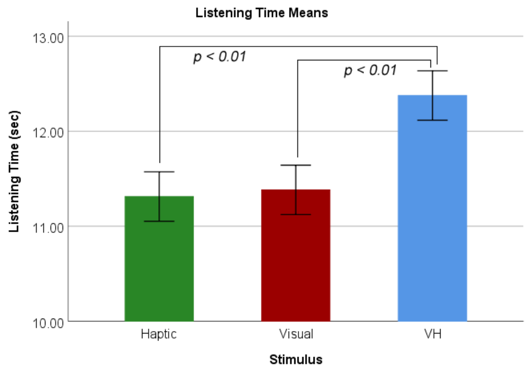
A pairwise comparison between the players and non-players groups was also performed. The results indicate a significant difference between the player and non-player means, for listening time as well as for detection time (see Figures 5.13c & 5.13d). As expected, the participants who had experience playing an instrument obtained significantly faster detections and longer listening periods compared to the participants who did not have experience playing an instrument.

The multivariate ANOVA results indicate that the performance of skilled participants was significantly better compared to that of unskilled participants, although the participants' performance was significantly better in both groups when using the visual-haptic stimuli. As shown in the interaction plots between participants' skill level and the stimulus condition(see Figures 5.14a & 5.14b). No interactions were detected between the participants' skill level and the given stimulus in the detection and listening time measurements. This indicates that the visual-haptic stimuli performance improvement was the same regardless of the participants' music skills.

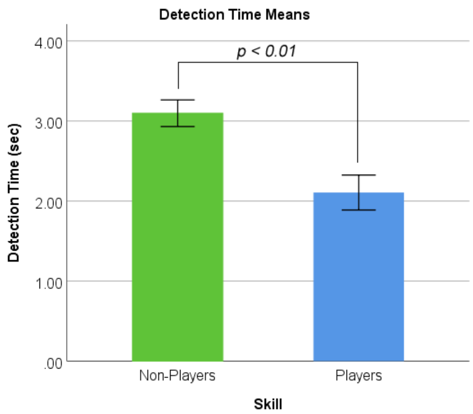
Furthermore, as shown in the interaction plots featuring the voices' pitch and the stimuli (see Figures 5.14c & 5.14d), the participants performed better when using the visual-haptic stimulus across voices with a different pitch. This indicates that, regardless of the voices' pitch, the participants had longer listening periods and faster detections with the visual-haptic stimulus, compared with the single modality stimuli. The same plots showed an interaction between visual and haptic conditions



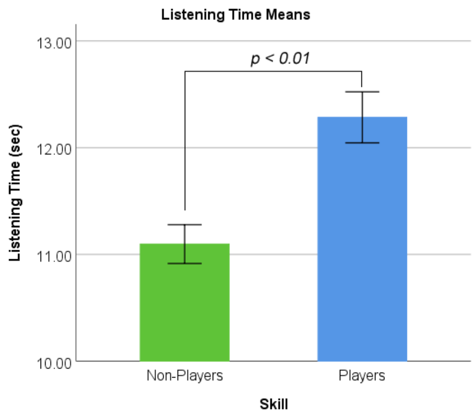
(A) Detection mean time between stimulus conditions



(B) Listening mean time between stimulus conditions



(C) Detection mean time between players and non-players



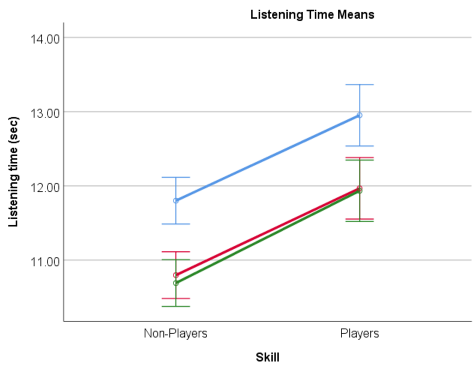
(D) Listening mean time between players and non-players

FIGURE 5.13: Bar plots that compare the detection and listening time means between stimuli conditions and the participants' skill level, where error bars indicate a 95% confidence interval. Plots (a) and (b) show a significant better performance of the players versus the non-players, with faster detection times and longer selective listening periods, while (c) and (d) show the significant performance increase of the visual-haptic stimulus (VH) compared with the single modality stimuli, with faster detection times and longer listening periods.

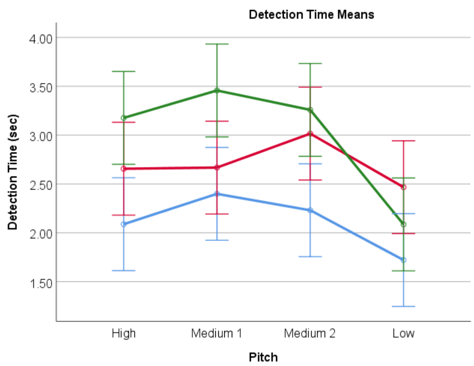
only for voices with a low pitch register, where the haptic stimuli obtained better results. Although, the lack of interaction between the voices' pitch and the stimuli conditions (see Table 5.3) do not justify a post-hoc comparison.

In addition to the psychometric evaluation, a subjective perception questionnaire was used to compare the participants' perception of the performed task in different stimulus conditions. A graphic rating scale with three questions was used: "*Did you like the experience?*", "*Was the stimulus easy to understand?*" and "*Was the stimulus easy to follow?*". To compare the results, a single factor ANOVA, between the conditions, was performed. The results, as shown in Figure 5.15, indicate that the participants found the visual-haptic condition more understandable and easier to follow compared to the visual-only and haptic-only conditions. Also, the participants showed a high likeness for the haptic-stimuli; it was almost as high as the visual-haptic likeness results.

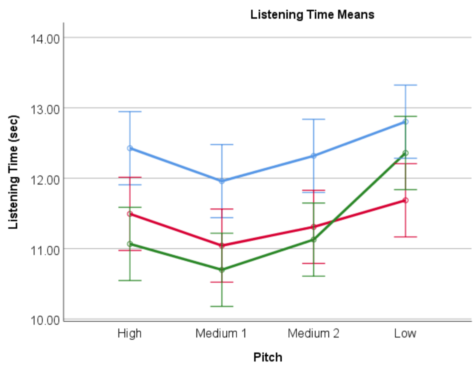
(A) Detection time between stimulus conditions and the participants' skill.



(B) Listening time between stimulus conditions and the participants' skill.



(C) Detection time between the voices' register and the stimulus conditions.



(D) Listening time between the voices' register and the stimulus conditions.

FIGURE 5.14: Interaction plots between skill*stimulus and voice*stimulus, where error bars indicate a 95% confidence interval. Plots (a) and (b) suggest that players and non-players had a better performance with the visual-haptic stimulus compared to single modality stimuli. The performance difference between stimuli was the same, regardless of the participants' music skills. Plots (c) and (d) also show a performance improvement with the visual-haptic stimuli across voices with different pitches. Therefore, the improvement caused by the visual-haptic stimuli was not related to the participants' skill level or the voice's pitch.

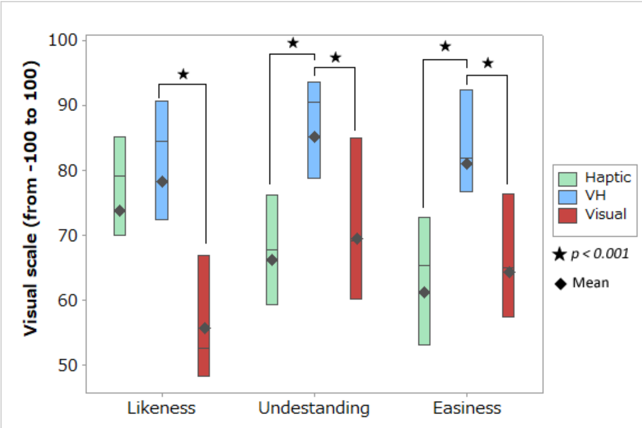


FIGURE 5.15: Box plot of the visual scale questionnaire results, which shows the participants' likeness, understanding and the easiness of the task across modalities. The boxes demonstrate the median confidence area, the diamonds show the measurements' mean and the stars indicate the mean statistical difference between stimulus conditions.

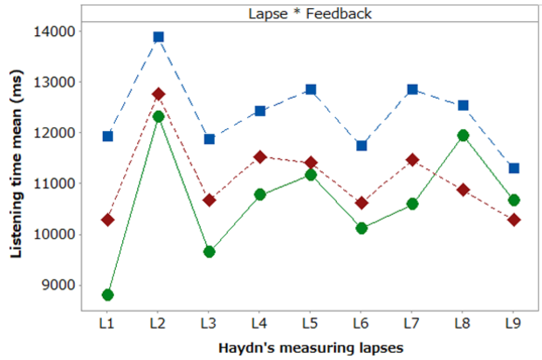
5.3.4 Detailed analysis

In the previous section, we demonstrated that the participants' performance improved with visual-haptic stimuli, regardless of the participants' skill level and the voices' pitch registry. However, the variance differences between the lapses measurements strongly indicate that the unique musical structure of each measuring lapse was a factor in the experiment (see Figure 5.16). Yet, since an objective or numerical strategy to measure the complexity of a polyphonic melody does not exist, the melody complexity of each measuring lapse was not viewed as a factor in the previous multivariate ANOVA.

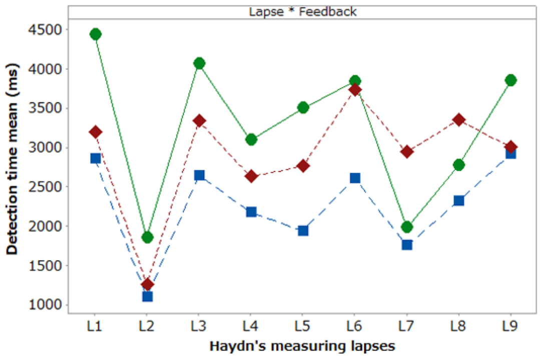
Instead, we performed an individual multivariate ANOVA analysis of each song. Then, considering the measuring lapses and the stimuli conditions as the main effects, a two-factor multivariate ANOVA was performed using the following main levels: the stimuli condition (with 3 levels: visual, haptic, and visual-haptic) and the measuring lapses (with 9 levels, one for each measuring lapse). The participants' detection and listening times were considered as dependent variables. By these means, a significant interaction between the measuring lapses and the stimulus conditions was identifiable.

The results of the multiple multivariate ANOVA are shown in Table 5.4. The table of results shows the significant main effect of the stimulus condition across the three songs (Mozart $F(16, 1566) = .352, p < .001$, Haydn $F(16, 1566) = .097, p < .001$ and Mozart $F(16, 1566) = .216, p < .001$). Also, there is significant main effect of the measuring lapses, although this interaction was expected due to the melodic difference of each measuring lapse.

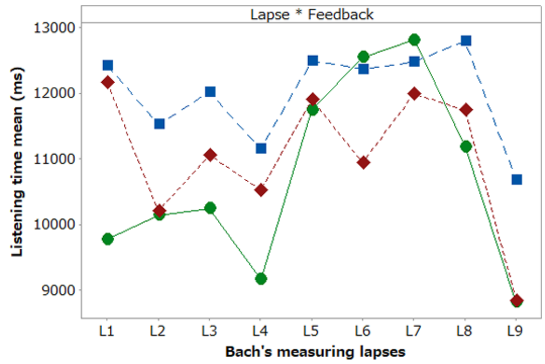
On the other hand, as in the general analysis, there was no significant effect between stimuli and the measuring lapses in any of the three songs (Mozart $F(32, 1566) = .051, p > .05$, Haydn $F(32, 1566) = .038, p > .05$ and Mozart $F(32, 1566) = .052, p > .05$). The lack of any significant interaction between the measuring lapses and the stimulus condition indicates that the performance difference between the stimuli was the same across measuring lapses and songs. The interaction plots also reflected the same results, with parallel lines across most of the measuring lapses (see Figure



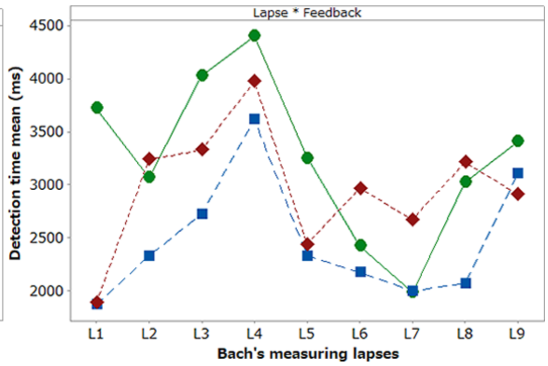
(A) Haydn's listening time interaction plot



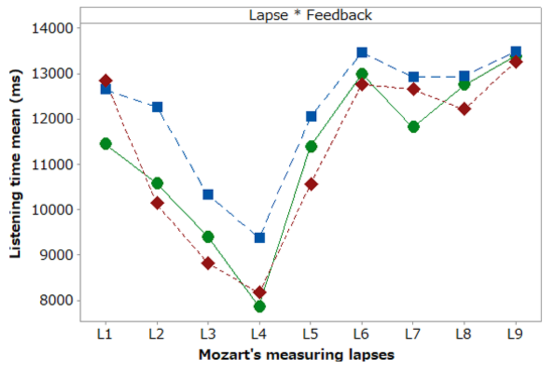
(B) Haydn's detection time intereaction plot



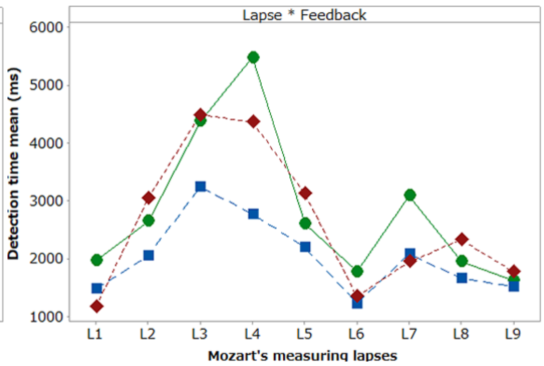
(C) Bach's listening time interaction plot



(D) Bach's detection time interaction plot



(E) Mozart's listening time interaction plot



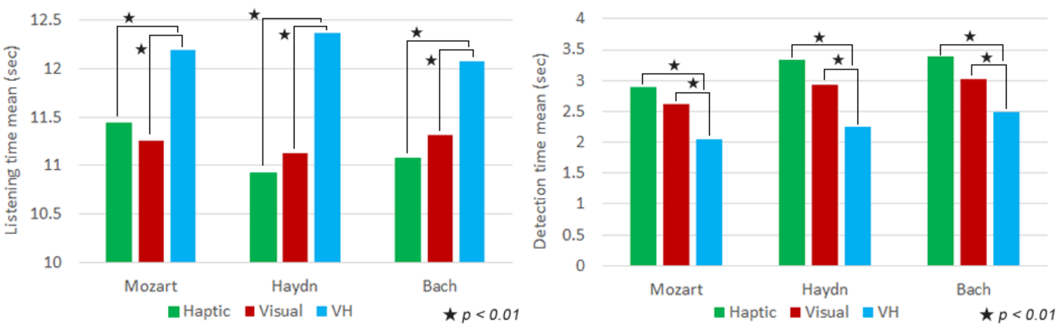
(F) Mozart detection time interaction plot

FIGURE 5.16: The detection and selective listening time interaction plots of each song. In general, the visual-haptic condition presented faster detection times and longer selective listening periods. Additionally, in most cases, the interaction lines between the stimuli ran parallel with each other, which demonstrated the lack of a significant performance difference caused by an specific combination of the given stimuli and measuring lapse. Therefore, the stimuli condition performance difference did not significantly change across measuring lapses.

Song	Effect	F	df	Error df	P value
Mozart	Lapse	.352	16.000	1566.000	**.000**
	Stimulus	.048	4.000	1566.000	**.000**
	Lapse*Stimulus	.051	32.000	1566.000	.144
Haydn	Lapse	.097	16.000	1566.000	**.000**
	Stimulus	.068	4.000	1566.000	**.000**
	Lapse*Stimulus	.038	32.000	1566.000	.563
Bach	Lapse	.216	16.000	1566.000	**.000**
	Stimulus	.034	4.000	1566.000	**.000**
	Lapse*Stimulus	.052	32.000	1566.000	.114

TABLE 5.4: The individual multivariate ANOVA results of each song were performed with Pillai's trace. The lapse and the stimulus conditions had a significant effect ($p < 0.05$) in all songs. However, there is no significant interaction between the lapses and the stimuli in any of the songs. This finding indicates that the participants' performance was not affected by any particular combination of the stimulus condition and the melody.

5.16) showing no significant individual interaction across the stimuli and the measuring lapses.



(A) (B)

FIGURE 5.17: Results of the pairwise comparison between the stimulus' modalities. The results, obtained with Tukey HSD, indicate a significant difference between the visual-haptic stimulus and the single modality conditions in all of the songs. Also, the results associated with the songs do not show any statistical differences between the visual-only and haptic-only conditions.

Since the stimulus main effect was significant, three Tukey HSD pairwise comparisons were performed, for each song, in order to find which feedback stimuli means were significantly different. The results, as shown in Figures 5.17a & 5.17b, indicate a significantly better participant performance when using the visual-haptic stimulus compared to the single modality stimuli. The results also reveal the lack of a significant difference between visual-only and haptic-only conditions, as they had the exact same results across the three songs.

In conclusion, the multivariate ANOVA results and the pairwise comparison results indicate that regardless of the song or the measuring lapse, the participants' performance was significantly better with the visual-stimuli, while there was no significant difference between the single modality conditions. And the lack of conclusive evidence of the interaction between the stimulus condition and the measuring lapses do not allow to precisely indicate a relation lapses melodic structure with an specific stimulus condition.

Overall, in the general and detailed multivariate analyses, none of the simple main effect combinations presented a significant interaction (see Table 5.3 & 5.4)As such, the lack of clear statistical evidence does not allow for a post-hoc analysis of any of the simple main effects interactions. Therefore, consistent evidence of the interaction between and specific stimulus with an specific kind of melody does not exist. However, we suspect that, due their own nature, the visual stimuli should more effective to display the melodic structure of the song, while the haptic stimuli should more effectively highlight the pitch of a specific instrument. Yet, in order to demonstrate this hypothesis, it may be necessary to perform a different study with specific music stimuli.

Chapter 6

Conclusion

In this dissertation we proposed a novel multimodal listening environment, called the Haptic Music Player. We also performed several perception experiments to understand the best strategy to build a synthethic haptic signal that resembles an audio signal, for this specific multimodal listening environment.

In specific, the Haptic Music Player introduced a novel way to map the notes' pitch, duration and the particular instrument waveform shape into a resemblant haptic vibration. To our knowledge, the previous attempts to do so, didn't considered the capabilities and limitations of the auditive and touch senses in order to propose their strategy. We claim this, because we were not able to find a similar proposal were specific multimodal perceptual experiments were performed to define the best strategy to modify or create the haptic signal. Mostly all the similar research previously presented, mentioned or partially considered the auditive and touch sense limitations. In addition the effectiveness of the system, by using redundant visual-tactile stimuli to enhance the listeners selective listening, was also evaluated.

So contrary to other proposal we focus this research on understanding how humans perceive a multimodal audiotactile stimuli to propose a method to create a resemblant synthetic haptic vibration. The results of these experiments let us generally understand how do we perceive multi-modal audiotactile stimuli. And each specific experiment was designed to understand and identify which properties of the audiotactile signal had more impact on the similarity perception. For example: the experiments #1 and #2 were focused on understanding if the users could properly identify the envelope correlation of audiotactile stimuli when the frequency and amplitude of the signals changed though time. The experiment #3 let us understand that the users are able to identify the envelope correlation between the audio and tactile with a diverse envelope shapes. The experiment #4 showed us that the users cannot perceive asynchrony between the signal at the decay, sustain or release phases. And finally the experiment #5 summaries the findings of the previous experiments showing us that is possible to synthetically create haptic signals with a higher subjective similarity even if the audio and haptic waveforms of the signals are not similar.

All the previously mentioned findings, from the performed perceptual experiments, can be synthesized on three general guidelines.

1. To improve the similarity perception, both signals should be synchronized at the attack and their general envelope shape should be similar.

2. The users are not able to distinguish the temporal amplitude characteristics of the signal during the decay or release.
3. The frequency and amplitude variability can mask the envelope correlation perception.

In specific we consider that these guidelines could be useful to create resemblant audiotactile stimuli for diverse applications and devices. By following these guidelines it is possible to design simpler haptic signals that could be displayed in low end haptic devices. The following of these guidelines, seems ideal for entertainment systems that use low end haptic devices, that are not able to directly display an audio signal as a haptic one.

For example, the Occulus VR controllers have a low end haptic actuator, with a refresh rate of 320Hz and a frequency range between 0~160Hz. Due to its limited characteristics it is incapable of directly display an audio signal. So if this specific device is used to display a audiotactile signal, due the haptic actuator limitations, it may be necessary modify the audio signal or create a synthetic haptic signal. And for these specific case these particular guidelines may result useful to simplify the haptic signal and optimize its perceived resemblance to the audio signal.

Also the Haptic Music Player implementation shows that is possible efficiently use a haptic device (Spidar-G) [29], which is not specifically designed to display audio signals, to display a highly resemblant synthetic audiotactile signal. Therefore this specific guidelines can be directly applied on diverse kinds of applications like: video games, virtual reality controllers, cellphone and tablet applications and even in for cinema theaters.

The performed psychophysical evaluation of the systems probed beyond any doubt that the Haptic Music Player enhances the listeners selective listening attention into an specific instrument or melody. Even if the melody is hard to follow as in the classical fugues used for the evaluation. Therefore the system could be easily used or adapted into for this purpose.

On the other hand, we are aware that the data analysis methods and the methodology of the proposed psychophysical experiments did not fully resemble and follow the classical psychophysics methods to measure specific psychophysical metrics. Therefore, the precision and generality of our results remain unclear and need future discussion, however this doesn't invalidate our conclusions. So, based on our results, we purpose some practical design guidelines for audio-tactile stimuli envelope in order to increase the cross-modal similarity perception.

Even if the user's experience of sound and vibration are subjective, we measured them robustly and in a quantitative manner. Then, we consider that the purposed design guidelines on the relationship between the physical stimulation and human perception can be directly used on different cross-modal applications to avoid unnecessary prototyping.

6.1 Future Work

6.1.1 Hapbeat

One of the latest and more outstanding efforts is a commercial product publicly available called HapBeat [55] [56]. HapBeat is a wearable haptic interface, some versions of this product can be used in the neck other versions of it can be used as a belt around the chest. This haptic interface directly uses the an audio signal as a haptic signal. The signal can be input, trough a cellphone or any other audio source, then the signal is amplified and finally the amplified is directly used to drive one or two motors to pull a little string. The string, which can be around the neck or the chest of the user is then pulled accordingly to the audio signal which produces a shock or vibration sensation where the user is wearing it (see Figure 6.1).

(A) (B)

FIGURE 6.1: The current and last versions of the Hapbeat wearable haptic interface, which may be an ideal haptic device for the next revision of the Haptic Music Player [©HapBeat].

The principal purpose of this interface is to enhance the haptic sensation of diverse multimedia environments like: cinema, interactive cinema, video games and also music listening. This device is very versatile and it is even portable. It seems to be an ideal haptic interface to implement a synthetic haptic vibration using the findings on the performed psychophysical experiments. So, it may be ideal to implement a mobile version of the Haptic Music Player which uses HapBeat as a haptic interface. This Haptic Music Player version may take advantage of the psychophysical experiments results and the HapBeat characteristics to crate an isolated and clear haptic signal of each instrument in a song.

6.1.2 Music Teaching and Audiation

During the time working in this project and after many different demonstrations in several conferences [5] [20], we found that the Haptic Music Player was entertaining and interesting to almost all the demonstration participants. But on the other hand this virtual environment received hard critics from different haptic experts. The most recurrent question asked by them, beyond the system entertainment capabilities, was: "What kind of application the system may have?". Therefore, in this section were are going to mention some specific applications that the proposed system have beyond just entertainment.

During the demonstrations where the Haptic Music Player was shown, we noticed that young children from 12 years old and bellow, were naturally attracted to the system and usually they come back to tried 2 or 3 more times. Also some of them invited or brought their relatives or friends to try the system.

In addition, due to the language barrier sometimes it was difficult to explain the demonstration to the children. But in most of the cases it seemed that the children didn't need any explanation to understand the system. So, for small children we

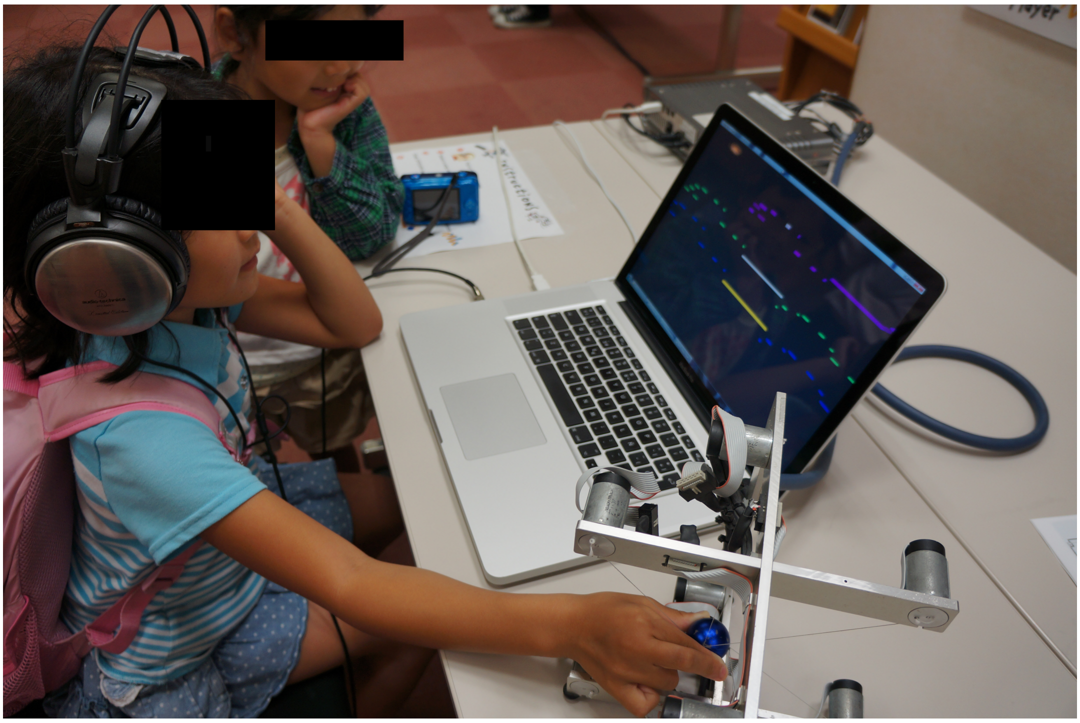


FIGURE 6.2: The Haptic Music Player used by two kids in a congress demonstration.

just let them use the system freely without any other explanation. Even without any instruction or explanation, the children were able to concentrate on the music for a long periods of time. For the demonstration we used a 5 minute MIDI rendition of a Mozart classic music concert, usually young children cannot pay attention to this specific kind of music for long periods of time, but when they were using the Haptic Music Player most of the children could actively listen to the complete 5 minute song without to many distractions. So these informal observations let us believe that the system could be used for music education (see Figure 6.2).

From these informal observations during the demonstrations, we started investigate about music teaching for young children. One of the pioneers of children music education research was Edwin Gordon. He made major contributions to the study of music aptitudes, explore the music development in infants the study of music aptitudes [16]. But he is also know to key the term, *Audiation*. He defined Auditation as the musical equivalent of thinking in language. Or in precise words, audiation is the human capacity of imagine or comprehend music in the mind. It is almost the same as thinking words when speaking in a foreigner language. Edwin Gordon proposed that audiation is developed during infancy on the first 9 years [52], after this term the subjects audiation do not develop to much further. Dr. Edwin also proposed several methods to study and measure audiation, like the: PMMA (Primary Measures of Music Audiation) designed for children younger than 9 years old and IMMA (Intermediate Measures of Music Audiation) designed for children older than 9 years old [17].

Now days Gordon's theories and findings are applied by diverse music educators all over the world. Gordon recurrently mentioned that is extremely important to develop the audiation of young children, mostly because their audition development will be directly proportional to their future musical abilities. That is why many music education experts focused their attention on developing new methods and games to develop children's audiation.

Due to the positive impression that children had during the Haptic Music Player demonstration, and after some investigation of the games and techniques used by profesionals to teach audiation to children. We consider that it could be a good idea to explore if a resemblant audiotactile signal could be used to teach children how to audiate. Currently, music teachers use games to help the children to image specific music pattern or complete them. We believe that the explanation reveal naturally and the game will be more fun and interesting to children if a virtual system similar to the Haptic Music Player is used.

Bibliography

[1] Bernard D. Adelstein et al. "Sensitivity to Haptic-audio Asynchrony". In: *Proceedings of the 5th International Conference on Multimodal Interfaces*. ICMI '03. Vancouver, British Columbia, Canada: ACM, 2003, pp. 73–76. ISBN: 1-58113-621-8. DOI: 10.1145/958432.958448.

[2] S. Allin, Y. Matsuoka, and R. Klatzky. "Measuring just noticeable differences for haptic force feedback: implications for rehabilitation". In: *Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2002. HAPTICS 2002. Proceedings. 10th Symposium on*. 2002, pp. 299–302. DOI: 10.1109/HAPTIC.2002.998972.

[3] M. Ercan Altinsoy and Sebastian Merchel. "Haptic and Audio Interaction Design: 5th International Workshop, HAID 2010, Copenhagen, Denmark, September 16-17, 2010. Proceedings". In: Springer Berlin Heidelberg, 2010. Chap. Cross-Modal Frequency Matching: Sound and Whole-Body Vibration, pp. 37–45. ISBN: 978-3-642-15841-4. DOI: 10.1007/978-3-642-15841-4_5.

[4] ANSI. *American National Standard Acoustical Terminology*. 1994.

[5] Malika Auvray and Christian Duriez. *Haptics: Neuroscience, Devices, Modeling, and Applications: 9th International Conference, EuroHaptics 2014, Versailles, France, June 24-26, 2014, Proceedings*. Vol. 8618. Springer, 2014.

[6] Jacob Benesty, M. Mohan Sondhi, and Yiteng (Arden) Huang. *Springer Handbook of Speech Processing*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2007, p. 65. ISBN: 3540491252.

[7] Albert S Bregman. "Auditory streaming is cumulative." In: *Journal of Experimental Psychology: Human Perception and Performance* 4.3 (1978), p. 380.

[8] Anne Caclin et al. "Tactile "capture" of audition". In: *Attention, Perception, & Psychophysics* 64.4 (2002), pp. 616–630.

[9] Angela Chang and Conor O'Sullivan. "Audio-haptic feedback in mobile phones". In: *CHI'05 extended abstracts on Human factors in computing systems*. ACM. 2005, pp. 1264–1267.

[10] Marsha G. Clarkson, Rhonda L. Martin, and Sheridan G. Miciek. "Infants' perception of pitch: Number of harmonics". In: *Infant Behavior and Development* 19.2 (1996), pp. 191 –197. ISSN: 0163-6383. DOI: http://dx.doi.org/10.1016/S0163-6383(96)90018-1. URL: http://www.sciencedirect.com/science/article/pii/S0163638396900181.

[11] ALPS Alpine Company. *HAPTIC™ Reactor Vibrational Feedback Device*. https://www.alps.com/prod/info/E/HTML/Haptic/. [Online; accessed 02-February-2018]. 2018.

[12] Patricia J. Flowers. "Patterns of Attention in Music Listening". In: *Bulletin of the Council for Research in Music Education* 148 (2001), pp. 48–59. ISSN: 00109894. URL: http://www.jstor.org/stable/40319077.

[13] Christopher D. Frith and D. John Done. "Routes to action in reaction time tasks". In: *Psychological Research* 48.3 (1986), pp. 169–177. ISSN: 1430-2772. DOI: 10.1007/BF00309165. URL: https://doi.org/10.1007/BF00309165.

[14] John M. Geringer, Clifford K. Madsen, and Dianne Gregory. "A Fifteen-Year History of the Continuous Response Digital Interface: Issues Relating to Validity and Reliability". In: *Bulletin of the Council for Research in Music Education* 160 (2004), pp. 1–15. ISSN: 00109894. URL: http://www.jstor.org/stable/40319214.

[15] George A. Gescheider et al. "Vibrotactile intensity discrimination measured by three methods". In: *The Journal of the Acoustical Society of America* 87.1 (1990), pp. 330–338. DOI: 10.1121/1.399300. eprint: http://dx.doi.org/10.1121/1.399300. URL: http://dx.doi.org/10.1121/1.399300.

[16] Edwin Gordon. *Learning sequences in music: Skill, content, and patterns: A music learning theory*. GIA Publications, 2003.

[17] Edwin E Gordon. "A factor analysis of the musical aptitude profile, the primary measures of music audiation, and the intermediate measures of music audiation". In: *Bulletin of the Council for Research in Music Education* (1986), pp. 17–25.

[18] Andrew H. Gregory. "Listening to Polyphonic Music". In: *Psychology of Music* 18.2 (1990), pp. 163–170. DOI: 10.1177/0305735690182005. URL: https://doi.org/10.1177/0305735690182005.

[19] Eric Gunther, Glorianna Davenport, and Sile O'Modhrain. "Cutaneous Grooves: Composing for the Sense of Touch". In: *Proceedings of the 2002 Conference on New Interfaces for Musical Expression*. NIME '02. Dublin, Ireland: National University of Singapore, 2002, pp. 1–6. ISBN: 1-87465365-8.

[20] "Haptical Music Player". In: Entertainment Computing 2013 (2013), pp. 210–213. URL: https://ipsj.ixsq.nii.ac.jp/ej/index.php?active_action=repository_view_main_item_detail&page_id=13&block_id=8&item_id=96002&item_no=1.

[21] David Hecht and Miriam Reiner. "Sensory dominance in combinations of audio, visual and haptic stimuli". In: *Experimental Brain Research* 193.2 (2009), pp. 307–314. ISSN: 1432-1106. DOI: 10.1007/s00221-008-1626-z. URL: https://doi.org/10.1007/s00221-008-1626-z.

[22] Henry E Heffner and Rickye S Heffner. "Hearing Ranges of Laboratory Animals". In: *Journal of the American Association for Laboratory Animal Science* 46.1 (2007), pp. 20–22.

[23] A. J. M. Houtsma and J. Smurzynski. "Pitch identification and discrimination for complex tones with many harmonics". In: *The Journal of the Acoustical Society of America* 87.1 (1990), pp. 304–310. DOI: http://dx.doi.org/10.1121/1.399297. URL: http://scitation.aip.org/content/asa/journal/jasa/87/1/10.1121/1.399297.

[24] David Huron. "Voice Denumerability in Polyphonic Music of Homogeneous Timbres". In: *Music Perception: An Interdisciplinary Journal* 6.4 (1989), pp. 361–382. ISSN: 0730-7829. DOI: 10.2307/40285438. eprint: http://mp.ucpress.edu/content/6/4/361.full.pdf. URL: http://mp.ucpress.edu/content/6/4/361.

[25] Inwook Hwang, Hyeseon Lee, and Seungmoon Choi. "Real-time dual-band haptic music player for mobile devices". In: *IEEE transactions on haptics* 6.3 (2013), pp. 340–351.

[26] ISO ISO. "226: 2003: Acoustics–Normal equal-loudness-level contours". In: *International Organization for Standardization* (2003).

[27] M. Karam, F. A. Russo, and D. I. Fels. "Designing the Model Human Cochlea: An Ambient Crossmodal Audio-Tactile Display". In: *IEEE Transactions on Haptics* 2.3 (2009), pp. 160–169. ISSN: 1939-1412. DOI: 10.1109/TOH.2009.32.

[28] Kevin Kelly. *The Music Animation Machine*. http://www.musanim.com/. [Online; accessed 15-January-2016]. 1985.

[29] S. Kim et al. "Tension based 7-DOF force feedback device: SPIDAR-G". In: *Proceedings IEEE Virtual Reality 2002*. 2002, pp. 283–284. DOI: 10.1109/VR.2002.996540.

[30] Kuzu. *Open MIDI Project*. 2002. URL: http://openmidiproject.osdn.jp/.

[31] Project Petrucci LLC. *International Music Score Library Project - IMPSLP*. http://imslp.org/wiki/Musikalisches_Opfer,_BWV_1079_(Bach,_Johann_Sebastian). [Online; accessed 15-January-2016].

[32] J. Lofvenberg and R.S. Johansson. "Regional differences and interindividual variability in sensitivity to vibration in the glabrous skin of the human hand". In: *Brain Research* 301.1 (1984), pp. 65 –72. ISSN: 0006-8993. DOI: http://dx.doi.org/10.1016/0006-8993(84)90403-7. URL: http://www.sciencedirect.com/science/article/pii/0006899384904037.

[33] Karon Maclean and Mario Enriquez. "Perceptual design of haptic icons". In: *In Proceedings of Eurohaptics*. 2003, pp. 351–363.

[34] Clifford K. Madsen and John M. Geringer. "Differential Patterns of Music Listening: Focus of Attention of Musicians versus Nonmusicians". In: *Bulletin of the Council for Research in Music Education* 105 (1990), pp. 45–57. ISSN: 00109894. URL: http://www.jstor.org/stable/40318390.

[35] Suranga Nanayakkara et al. "An enhanced musical experience for the deaf: design and evaluation of a music display and a haptic chair". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM. 2009, pp. 337–346.

[36] Oarih. *Wikipedia*. [Creative Commons - Share Alike Licence v3.0]. 2016.

[37] Daichi Ogawa et al. "HapTONE : Haptic Instrument for Enriched Musical Play". In: (), pp. 2–3.

[38] Earl R Parker. "OpenStax Anatomoy and Physiology". In: (2016). [Rice University, Creative Commons Licence v4.0].

[39] Lorenzo Picinali et al. "Spectral Discrimination Thresholds Comparing Audio and Haptics for Complex Stimuli". English. In: *Haptic and Audio Interaction Design*. Ed. by Charlotte Magnusson, Delphine Szymczak, and Stephen Brewster. Vol. 7468. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2012, pp. 131–140. ISBN: 978-3-642-32795-7. DOI: 10.1007/978-3-642-32796-4_14. URL: http://dx.doi.org/10.1007/978-3-642-32796-4_14.

[40] H. Pongrac. "Vibrotactile Perception: Differential Effects of Frequency, Amplitude, and Acceleration". In: *Haptic Audio Visual Environments and their Applications, 2006. HAVE 2006. IEEE International Workshop on*. 2006, pp. 54–59. DOI: 10.1109/HAVE.2006.283803.

[41] Marie Ragert, Merle T. Fairhurst, and Peter E. Keller. "Segregation and Integration of Auditory Streams when Listening to Multi-Part Music". In: *PLOS ONE* 9.1 (Jan. 2014), pp. 1–9. DOI: 10.1371/journal.pone.0084085. URL: https://doi.org/10.1371/journal.pone.0084085.

[42] R. R. Riesz. "Differential Intensity Sensitivity of the Ear for Pure Tones". In: *Phys. Rev.* 31 (5 1928), pp. 867–875. DOI: 10.1103/PhysRev.31.867. URL: http://link.aps.org/doi/10.1103/PhysRev.31.867.

[43] Maxwell Roberts and Riccardo Russo. *A student's guide to analysis of variance*. Routledge, 2014.

[44] Jose Shelton and Gideon Praveen Kumar. "Comparison between auditory and visual simple reaction times". In: *Neuroscience & Medicine* 1.1 (2010), pp. 30–32.

[45] Raimond Spekking. *Wikipedia*. [Creative Commons - Share Alike Licence v4.0]. 2018.

[46] Charles Spence. "Explaining the Colavita visual dominance effect". In: *Progress in brain research* 176 (2009), pp. 245–258.

[47] Ville Tikka and Pauli Laitinen. "Designing Haptic Feedback for Touch Display: Experimental Study of Perceived Intensity and Integration of Haptic and Audio". In: *Haptic and Audio Interaction Design* 4129 (2006), pp. 36–44. ISSN: 03029743. DOI: 10.1007/11821731_4.

[48] Michael Tippach. *ASIO4ALL - Universal audio driver for WDM Audio*. Mar. 2003. URL: http://www.asio4all.com/.

[49] Ronald T. Verrillo. "Effect of Contactor Area on the Vibrotactile Threshold". In: *The Journal of the Acoustical Society of America* 35.12 (1963), pp. 1962–1966. DOI: http://dx.doi.org/10.1121/1.1918868.

[50] Ronald T Verrillo. "Subjective magnitude functions for vibrotaction". In: *Man-Machine Systems, IEEE Transactions on* 11.1 (1970), pp. 19–24.

[51] RonaldT. Verrillo and GeorgeA. Gescheider. "Enhancement and summation in the perception of two successive vibrotactile stimuli". English. In: *Perception and Psychophysics* 18.2 (1975), pp. 128–136. ISSN: 0031-5117. DOI: 10.3758/BF03204100. URL: http://dx.doi.org/10.3758/BF03204100.

[52] Darrel L Walters. "Edwin Gordon's music aptitude work". In: *The Quarterly Journal of Music Teaching and Learning* 2.1-2 (1991), pp. 64–72.

[53] Jeremy M Wolfe et al. *Sensation & perception*. Sinauer Sunderland, MA, 2006.

[54] James K. Wright and Albert S. Bregman. "Auditory stream segregation and the control of dissonance in polyphonic music". In: *Contemporary Music Review* 2.1 (1987), pp. 63–92. DOI: 10.1080/07494468708567054.

[55] Yusuke Yamazaki, Hironori Mitake, and Shoichi Hasegawa. "Tension-based wearable vibroacoustic device for music appreciation". In: *International Conference on Human Haptic Sensing and Touch Enabled Computer Applications*. Springer. 2016, pp. 273–283.

[56] Yusuke Yamazaki et al. "Hapbeat: single DOF wide range wearable haptic display". In: *ACM SIGGRAPH 2017 Emerging Technologies*. ACM. 2017, p. 12.