UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias e Ingeniería

Sensory Replacement System For The Creation Of Psychophysical Sensations Of Sound Sources In Space-Fixed Audio Files

.

Johann Isaac Jadán Altamirano

Ingeniería Electrónica y Automatización

Trabajo de fin de carrera presentado como requisito
para la obtención del título de
Ingeniería en Electrónica

Quito, 24 de Diciembre de 2021

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias e Ingenierías

HOJA DE CALIFICACIÓN
DE TRABAJO DE FIN DE CARRERA

Sensory Replacement System For The Creation Of Psychophysical Sensations Of Sound Sources In Space-Fixed Audio Files

Johann Isaac Jadán Altamirano

Nombre del profesor, Título académico **Luis Miguel Prócel, PhD.**

Quito, 24 de Diciembre de 2021

© DERECHOS DE AUTOR

Nombres y apellidos: Johann Isaac Jadán Altamirano

Código: 00138811

Cédula de identidad: 1803548146

Lugar y fecha: Quito, 24 de Diciembre de 2021

ACLARACIÓN PARA PUBLICACIÓN

Nota: El presente trabajo, en su totalidad o cualquiera de sus partes, no debe ser considerado como una publicación, incluso a pesar de estar disponible sin restricciones a través de un repositorio institucional. Esta declaración se alinea con las prácticas y recomendaciones presentadas por el Committee on Publication Ethics COPE descritas por Barbour et al. (2017) Discussion document on best practice for issues around theses publishing, disponible en http://bit.ly/COPETheses.

UNPUBLISHED DOCUMENT

Note: The following capstone project is available through Universidad San Francisco de Quito USFQ institutional repository. Nonetheless, this project – in whole or in part – should not be considered a publication. This statement follows the recommendations presented by the Committee on Publication Ethics COPE described by Barbour et al. (2017) Discussion document on best practice for issues around theses publishing available on http://bit.ly/COPETheses.

RESUMEN

Partiendo de un archivo de audio de un solo canal armónico que originalmente no produce retroalimentación háptica o "ilusiones táctiles" a través de un Vibrotractor de pulsera, se realizarán modificaciones mediante un proceso de modelado sinusoidal. También se realizará una transformación de audio fija en el espacio utilizando HRTF (Función de transferencia relacionada con la cabeza), luego este resultado volverá a sufrir SMS (Síntesis de modelado espectral) en combinación con FX (efectos/transformaciones de audio). Cambiando completamente la respuesta vibratoria del audio reescalando sus frecuencias a un valor inferior, pero manteniendo sus propiedades psicoacústicas. El nuevo espectro se siente a través de los actuadores vibrotáctiles, por lo que la esencia sonora se transforma en esencia vibrotáctil. Generando un sistema de sustitución sensorial basado en las sensaciones psicofísicas producidas sobre la piel. Muy útil para las personas sordas que no pueden oír y sentir la música como las personas no sordas.

Palabras clave: SMS, HRTF, modelamiento sinusoidal, vibrotactor, ilusiones táctiles.

ABSTRACT

Starting from a single harmonic channel audio file that originally produces no haptic feedback or “tactile illusions” through a wristband Vibrotractors, modifications are made through a sinusoidal modeling process. As well performing a Space-Fixed Audio transformation using HRTF (Head-Related Transfer Function), then this result again suffers SMS (Spectral Modeling Synthesis) in combination with FX (audio effects/transformations). Completely changing the vibrational response of the audio by rescaling its frequencies to a lower value, but maintaining its psychoacoustic properties. The new spectrum is felt through the vibrotactile actuators, so the sound essence is transformed into a vibrotactile essence. Generating a sensory replacement system based on the psychophysical sensations produced over the skin. Highly useful for deaf people who cannot hear and feel music like non-deaf people.

Key words: SMS, HRTF, sinusoidal modeling, vibrotactor, tactile illusions.

TABLE OF CONTENTS

INDEX OF TABLES

INDEX OF FIGURES

INTRODUCTION

The simplicity and the amount of information that we can obtain through the skin, has made it a field of research for increasingly immersive experiences in relation to technology. We know that haptics has been used in many things and it is more and more usual to find it as a response of interactions in smartphones or videogame console control. However we can expand those frontiers bringing those sensations to people who do not have one oftheir senses at a disadvantage. It is known that there is research in applying haptics as a sensory expansion in those who have lost a limb and must use a prosthesis [1] to generate stimuli and tactile effects. Nevertheless, this work will be focused on being a complement to sound, so that deaf people can feelmusic in a way they have not experienced before.

When playing a song and convert it to vibration, unlike thevibrations that a solenoid would normally produce, we would only feel the lower frequencies thanks to its bass composition. So instruments with frequency signals in the higher ranges, such as the flute, or violin, may be masked by vibrations from instruments in the lower frequency range, such as bass and drums [2]. Despite the approach is to transmit all frequencies (low, medium and high) through a sinuous modeling and complement it with the spatialization of the sound, generating a new and different type of tactile illusions on the skin.

As well as musical emotion is composed of specific musical characteristics such as rhythm, tempo, meter, timbre and harmony [3], it is desired to convey the same particularities of music experience through vibrations. Therefore, the objective for this work is to develop a sensory substitution tool with vibrotactile feedback, which allows to felt in the skin a sound source from a spatially arranged audio file, using sinusoidal modeling and Head-Related

Transfer Function (HRTF) to create new sensory experiences that supplant sound experiences in hearing impaired people and in turn expand new experiencesin non-deaf people.

STATE OF THE ART

Syntacts: Open-Source Software and Hardware for Audio- Controlled Haptics

Pezent and Cambio [4] begin their research by noting the existence of haptics-related technology, known as vibrotac- tors, which come in different shapes and distributions, such as bracelets, armbands, sleeves, full body suits and chairs. However, the design of signals for the operation of these actuators is complicated, so they decided to create Syntacts, that lowers the technical barrier to synthesizing and rendering vibrations with audio [4], which in addition to being open source multiplatform, allows to manipulate and synthesize a large number of mathematical functions, frequency modulated sequence cues and their spatialization on tactile arrays, each of them useful for haptic design.

In addition, their project is complemented by hardware design based on audio principles, it consists of :

- microcontrollers (e.g., an Arduino) that link the PC to the Integrated Circuits (IC),
- a set of eight individual channels of vibrotractors (voice coil actuators) distributed in a bracelet with a Universal Controller from Engineering Acoustics, Inc. (EAI),
- a digital-to-analog converter (DAC) system responsible for converting digitally represented waveforms, such as music files, into analog signals,
- a differential linear amplifiers specially designed to han-dle eight channels with minimum noise and voltagesources to bias the whole system.

So that together with the specialized software they cancontrol vibrotactile arrays with extremely low latency andare capable of playing complex waveforms. Resulting on acomplete system for Audio-Controlled Haptics.

Hand-to-Hand: An Intermanual Illusion of Movement

The experiment consists in create apparent tactile motion without any object between them using two vibrating handles which they refer to as the Hand-to-Handvibrotactile device, that is a specially built hardware for research which consists oftwo 3D printed handles each containing a voice coil actuator sandwiched between two springs. For the determination of the control space of smooth and consistent motion, they experimented two type of tactile illusions, is about the stimulus-on set asynchrony (SOA) that its modulated for the activation of apparent tactile movement illusion stimuli through two actuators.

- The first one, “the cutaneous rabbit illusion, involves twovibrotactile actuators are modulated in a timely fashionto create a third illusory perceptual sensation like a rabbit hopping in between the two real stimulators” [5].

- The second one, calls “funneling illusion, involves twoactuators vibrating with different intensity are able tocreate a third in-between point, whose position will bedetermined accordingly to the variation of the intensityof the two vibrations” [5].

Afterwards determinate the optimal parameters for the il-lusion of movement, the model involves posture, speed andthe intensity of the stimuli on the skin of the participants. That concluded with a implementation of a VR interface toachieve perceptual integration of visuo-tactile stimuli. Finally, they conclude that the sensation of illusory movement between two hands is not only restricted to situations in which an object is being held with both hands, or in distributed conFigureurations throughout the body using it as a physical interface.

Modelling Perceptual Elements of Music in a VibrotactileDisplay for Deaf Users: A Field Study

This research based on the Model Human Cochlea (MHC) to build a sensory substitution technique for presenting musicas a equiparable musical emotion similar when non-deaf users listen with audio speakers and makes them to feel the music. The behavior of the human cochlea consists whenit is exposed to different types of frequencies, it vibrates hundreds of hairs according to the stimulus received in the basilar membrane. This generates vibrations that in turn, are reinterpreted as an electrical potential that will be processed by the audio cortex of the brain. Based on this idea, the researchers built a chair equipped with an array of [4×2] voicecoils (VC), which are embedded in the back of a canvas chair.

The emotional complexity of a musical composition involves some characteristics which vary according to the intention and emotion of the song. The challenge is to transmit the same emotions while converting the audio to vibrations, because audible vibrations span frequencies from 20 Hz up to 20 kHz, while our tactile system only detected vibrations ranging in frequency from 10 Hz and 1000 Hz [2]. Therefore, two types of experimentation were carried out.

Track model (TM)

Consists in splitting the audio bytracks, making each instrument or voice vibrate on a separate channel.

Frequency model (FM)

Consists of creating multiplebands based on the separation of an audio in specific frequencies.

Concluded by noting that according to the type of vibration (level of variation and intensity) they can demonstrate emo-tionality, such as the “higher, more varied signals convey joy when interpreting the emotional content of the vibetracks; theslower, less variable, weaker signals produce vibrations thatare difficult to detect, which were either interpreted as sad, or as expressing no emotional information” [2].

Haptic music exploring whole-body vibrations and tactilesound for a multisensory music installation

Mixing an interactive installation with research, resulted in a study on haptic music composition for a multi-sensory installation and how composers create different vibrational sensations based on whole body vibration, this kind of stimulioccurs when a surface shakes and vibrate while affects bodyparts.

One approach was to create an inverse experience inside the installation called Sound Forest, also focused on tactile displays allowing deaf persons to experience music, so that just as there is the head-related transfer function (HRTF) for audio, the same concept can be applied to the human body which resonance different parts of it, which is called the body-related transfer function (BRTF). This was the key tool for creating eight different types of signals, mostly sinusoidal signals with frequency variations. Because gave better results than than modifying the range of amplitude envelopes, in order to provide different haptic sensations.

One of his biggest challenges besides creating the installation was to make the workshops for the composers toexperiment and explore with vibrations. In that way, as Fridand Lindetorp [6] quote “musical representations for hearing-impaired should focus on staying as close to the original aspossible while conveying the physics of the representation via an alternate channel of perception” [7] . Thus they concluded that the installation does have the potential to produce different haptic experiences ranging from unpleasant to pleasant. As long as composers have the possibilities to experiment withinit or have an adaptation of it in

their studio and do create morespecific body illusions and possibilities of the multisensory installation improved through training.

METHOD

To generate tactile stimuli throughout scientific work wehave seen that they have evolved in different ways, such asgiving two similar tactile stimuli at different points on the skin, with control of the temporal relationships, the intensity of thestimuli and the distance between them [8], or as Macaron[9], which uses wave forms with a combined variation offrequency and amplitude that characterizes particular patternsof vibrations such as a heartbeat or the purr of a cat.

Understanding part of the concepts used, we know that the best way to create vibrations is with sinusoids. Therefore, inthe present work we rely on Serra's [10] [11] research, in order to create sinusoidal modeling of an audio file and transportit to a possible tactile effect created by the same musical composition of the instrument. That is, we are looking fora tactile illusion that is not specifically designed to mimic a particular sensation on the skin as we have seen in previous works. Rather it is expected that, by processing the audio and reinterpreting it through a series of algorithms, the same arrangement of the musical composition originally created bythe artist with a certain intention. Including its notes, chords,or even the instrument used for that purpose, will be in chargeof creating tactile illusions. Which in the original audio werenot expected and were not aware that these could be present indirectly.

Therefore, we know that there are certain patterns of frequencies that when felt on the skin we reinterpret them astactile effects, many of them carefully and mathematically designed, such as phantom sensations [12] or funneling [13], however in a musical track in an

almost stochastic way couldor could not present patterns very similar to those that in other investigations have designed [14], evidently these can only be ashort fragment, perhaps rescaled or with different frequenciesand amplitudes, but all of them is not an important problem. The relevant thing to emphasize of this idea is that internally any musical track has the possibility that when transforming it into a haptic file. This contains at least one or a combination of tactile effects, with an own and particular pattern of each input that it has. Likewise that would result that no vibratory pattern would be equal to the other, as long as each musical track is different from each other.

In synthesis, we understand a song as the audible part, themost basic scale of sensorial abstraction, that in addition each one of them possesses more information not necessarily audible. To understand it better we can use the Kantian Doctrine[15] to extrapolate two main ideas. The song would correspond to the "phenomenon" because it is offered to us through our senses and we can experience it through experience, in an easy way whenever we listen to music. On the other hand, the non-perceptible vibratory part would apparently correspond to the"noumena". Because in a certain way it is the thing in its pure existence, many times the unknowable, which is alsocalled the thing in itself, whose particularity is that it can only be reached by an intellectual work. In our case, to bring the metaphor back to engineering, we take that noumena as the object of the generally non-sensory contemplation, and we cross it to the sensory world through two vibrotactile actuators which will make intelligible that essence of the song.

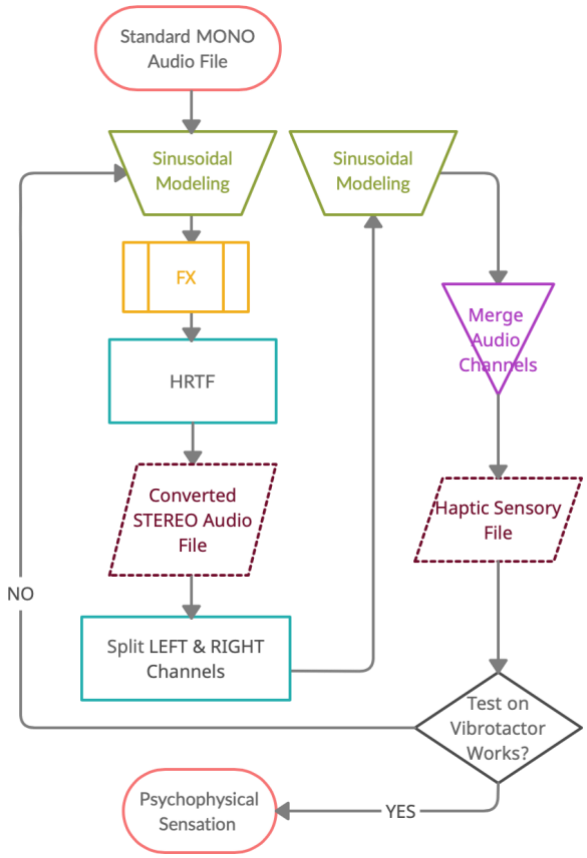

Figure 1. Workflow and methodology to obtain each Haptic File.

In order to obtain that information the workflow "Figure. 1" was proposed, it starts with a single channel audio, then this is converted to sinusoidal modeling in which some audio effect could be applied. Those ones will be explained later and will be relevant for this experiment. Consequently that result will suffer the HRTF process whose objective is to enhance the tactile illusions that can be created for sound sources in Space-Fixed Audio Files with spatial arrangement, whose output returns two audio channels. To which another sinusoidal modeling will be performed respectively. The workflow is concluded by merging the audios of each channel into a single stereo file that aims to produce psychophysical sensations. In the following we will break down the workflow into four mainareas, which are important to deepen in order to understandthe operation of the proposed topic.

The Audio File

As previously mentioned, some audio files converted to vibration will just not vibrate and will not produce any effect because of their own nature as an instrument and the frequency range in which they operate. For that same reason it will be used an instrument that normally should not naturally produce any haptic vibration directly in the actuators. Therefore the audio to use corresponds to a few seconds ".wav" file with a jazz melody composed only by a saxophone, a musical instrument belonging to the woodwind family.

In addition, it is relevant to mention why a saxophone and not another instrument to corroborate the hypothesis. In thefirst place, the saxophone is not a polyphonic instrument, that is to say that it can only play one note at a time as the great majority of wind instruments. This is importantin order to facilitate the spectral analysis, because the notes will not overlap each other, or other three issues: feature variations caused by sound mixtures, the pitch dependency of timbres, and the use of musical context [16]. Secondly, speaking of frequencies, we know that the saxophone is an harmonic instrument because of its approximately conical body. That means that the saxophone when producing a sound, each note of this will be composed of several simultaneous waves of different frequencies and amplitudes, of which the first one will be determined by a vibration whose fundamental frequency is constant. Besides, each musical octave would correspond to duplicate that frequency, this is called timbre and is the characterization of the sound according to the amount of harmonics present in it, as well as their frequencies and volume that reaches each one of them respectively.

Hence, a given musical note will be determined by a main vibratory motion, which is the fundamental harmonic, incomplement to several secondary vibrations (harmonic series) associated to the same instant in time [17]. This can be synthesized in "(1)", where L is the length of the cone and T is the temperature of the external medium. This is closely related

to the elements of the Fourier series and precisely for the following steps of the workflow some of these transformations will be performed, such as the Short Time Fourier Transform (STFT).

$$f_n = \frac{n}{L} \cdot 5.01 \cdot \sqrt{T} \qquad (1)$$

Sinusoidal Modeling + FX

First, it is imperative to define what the following algorithmis and how it works. The research starts from the fact that the sounds produced by musical instruments, sound sources or any physical system itself, can be modeled as the sum of a set of sinusoids plus a residual noise. The relevance of using these algorithms is that we understand that the sinusoidal (deterministic component) normally corresponds to the main vibration modes of a system. So they are useful to create haptic effects and fulfill the proposed objectives. Therefore, this algorithm is able to detect the magnitude, frequency and phase of the partial trajectories present in the original sound. Which is called the deterministic component [10].

That is, through spectral signal processing, we can representthe signals of a sound with different shapes and different designs or transformations. These alternative representationsin the frequency domain can be transformed and inverted to produce new sounds. In our case for this part of the workflow we required that the input response sound is the same at the output. With the difference that the waves that compose the sound have changed to a sinusoidal spectrum. Such sinusoidal spectra are obtained using the Short-Time Fourier Transform (STFT) and the prominent spectral peaks that are detected and incorporated into the existing partial trajectories using a peak continuation algorithm [10]. Though that process of using the STFT may promise more flexible representations, but it may also compromise both sound fidelity and computational time.

We must clarify that within the whole sinusoidal processing there are many ramifications. However the model to be used is based on algorithms based on the sinusoidal plus residual model, which allows to process the information and model the spectrum as sinusoids and another part as the residual. About the noise, residual or stochastic component we can say that this results from first obtaining the deterministic signal with additive synthesis, and then subtracting it from the original waveform in the time domain. Phase information is especially important for subtracting the deterministic component and finding the residual. Therefore, in the spectral modeling of musical sounds, sinusoids and noise are represented as two separate components [18].

Serra [10] states that this analysis methodology is quite successful, because the sinusoidal plus residual representation is very flexible while maintaining good sound fidelity, and the representation is quite efficient. At the same time, this sinusoidal representation algorithm is focused to model only the stable partials of a sound. To continue with the explanation of the algorithm we have listed all its subprocesses, all the software is implemented in MATLAB, whose framework is based on Spectral Modeling Synthesis (SMS). Which we can understand as a set of spectrum-based techniques and related implementations for the analysis, transformation and synthesis of an audio signal [11].

STFT

The first step of analysis corresponds to the calculation of the magnitude and phase spectra of the current frame. It is on these spectra that the sinusoids are tracked and it is decided whether a part of the signal is considered deterministic or noise. The calculation of the spectra is performed usingthe STFT [19]. This transform is linked to several parameters,such as window size, window type, FFT size and frame rate. These must be carefully selected

according to the audio to be processed. I mean each sound source will have its particular parameters. For the audio of a saxophone, we have used theparameters on Table 1.

Table 1. SMS algorithm parameters

Spectral Modeling Synthesis Variables Settings		
Variable Name	*Value*	*Description*
SR	22050	Sampling rate
w1Length	2048	Analysis window size
n1	256	Analysis window hop size
nPeaks	100	Number of peaks detected
nSines	50	Number of sinuosoids to track and synthetise
minSpacePeaks	2	Minimum space (bins) between two picked peaks
zp	2	Zero-padding coefficient
rgain	1.	Gain for the residual component
MaxFreq	11000	Maximum frequency [Hz] for plottings
MinMag	-100	Minimum magnitude [dB] for plottings

Hence a good resolution of the spectrum is needed, since the process that tracks the partials has to be able to identify the peaks that correspond to the deterministic component.

Windowing

In order to find a good resolution ofthe spectrum, it is important to find a precise window. As this corresponds to the consequent step in the process of converting a signal in the time domain to its representationin the frequency domain. This operation consists of selecting a number of samples of the sound signal and multiplying its value by a window function [11]. Regarding this process it is important to make two important points. Because the results will change according to the type of window and the size of it.

For example, in case of taking very long windows we will have loss of temporal resolution but good frequency resolution; and inversely in case of shorter windows. That phenomenon is called the time-frequency trade-off. There must be a balance between both to have a good frequency resolution and, therefore, a good measurement of the frequencies of the partials.

However, most sounds cannot afford these adjustments and achieve perfect synthesis, so one must be weighted against the other. The same goes for noise, it turns out that to model the stochastic part of sounds, such as percussion instruments, bow noise in string instruments, breath noise in wind instruments or even breath noise in voices. Then filter them correctly from the other base, a good temporal resolution is needed, so frequency resolution can be left a bit aside [10].

Due to the fact we are working with the sound of a harmonic saxophone, according to the literature for such harmonic sounds, the actual size of the window will change as the pitch changes. In order to ensure a constant time-frequency compensation for the whole sound. So the window decision is very relevant. Particularly as our base audio has no low frequencies, percussion, vocals or stochastic arrangements, we can skip limiting time-resolution and focus on frequency resolution. Therefore, considering all these trade-offs and after long trial and error selections, a 92 dB Blackmann-Harriswindow has been chosen because its main lobe includes mostof the energy. Moreover, this transformation window should have the smallest possible number of significant bins, sincethis will be the number of points to be generated per sinusoid [11].

Zero-padding

In the time domain increases the number of spectral samples per Hz and thus increases the accuracy of the simple peak detection [11]. So for this first zero padding the objective is to avoid the time-frequency mismatch. Whereby zeros are added to the windowed signals to have alonger FFT and thus increase the frequency resolution. This is also called frequency domain interpolation, which results in more accurate detection of simple peak detection.

Pitch detection

In the world there are a large number of sounds, many of which are naturally not perfectly periodic, nor do they have well-spaced peaks. Those peaksare defended as a local maximum in the magnitude spectrum. The algorithm will search for them and plot a group of peak trajectories through the instantaneous frame as it progresses in time [11]. So the algorithm advances frame by frame (depending on the width of the window) and one by one overlapping each other, until completing the total time of the audio. For each one of them it looks for its peaks, which also explains the processing time it takes.

The best way to search for them is to have restrictions for the frequency range and the magnitude threshold. That search for the set of peak trajectories, through each frame also implies a possible finding of a fundamental frequency. That frequency can be defined as the common divisor of the harmonic series that best explains the spectral peaks found in the current frame, namely the periodicity. If it does exist, the algorithm will have more information and simplify the tracking of the partials.

Spectral interpolation

The scheme consists of zero-pading, which means filling the vector with zeros so that the total number of samples is equal to the next power of two. Thus we understand it as a quadratic spectral interpolation. For the case of the algorithm this is fundamental since it helps us to compensate the trade-off between timeand frequency. On the other hand, we use only the samples immediately surrounding the sample of maximum magnitude, that is enough to refine the estimation with an accuracy of 0.1% [11].

Pitch estimation

This corresponds to an optional and not necessarily mandatory step, generally used when the input sound is known to be monophonic and pseudo-harmonic. We define the pitch as the

common divisor of the harmonic series that best explains the spectral peaks found in a given frame. In other words the mismatches between the harmonics generated and the maximum frequencies measured are averaged over a fixed subset of the available peaks. A recursive process is necessary for its finding, which also depends on the choice of a suitable window for the Fourier analysis. This will result in obtaining the fundamental frequency, which in turn, will facilitate the following steps of the analysis.

Peak continuation

This algorithm consists of organizing the peaks into time-varying trajectories, starting from a sound that is synthesized using additive synthesis. While the sinusoidal model assumes that each of these peaks is part of a frequency trajectory, the peak continuation algorithm assigns each peak to a given track. In the process, it returns the estimated magnitude, frequency and phase of the prominent peaks in a given frame, sorted by frequency. To obtain a good partial-residual decomposition, the peak continuation process must be well performed, so that the stable partials of the sound will be identified. The less restrictive the peak detection stage is, more faithful will be the reconstruction of the original sound after synthesis. Finally, this algorithm is constantly updated and varies according to the type of audio it processes, which is called “guides”. Because in the case of harmonic sounds, these guides are initialized according to the harmonic series of the detected fundamental frequency. While for inharmonic sounds, each guide is created dynamically[11]. Then, according to their frequency and magnitude, these guides form trajectories from the peaks. Such a spectral trajectory models a sinusoidal representation for the whole time-varying sound.

Residual analysis

Once the stable partials of asound have been identified, they are ready to be subtracted rom the original signal to obtain the residual component. As mentioned above, the

modeling of the stochastic part of the sounds depends on all the previous steps. The final step is the transformation and synthesis of the sound. Which are performed in the frequency domain; generating sinusoids, noise or arbitrary spectral components, and summing them all in a spectral frame. For each track mentioned above, a single Inverse Fast Fourier transform (IFFT) is performed which returns the audio in time in a very efficient way.

It is important to mention that for each of the steps, most of the techniques presented are designed to work correctly with monophonic sounds, while some depend on the pseudo-harmonicity of the input signal.

Within the steps mentioned above, an additional one can be added, that is, the effects. In reason of the inherent nature ofthe signals we are processing, the result of the spectral analysisis very suitable for modifications. It turns out that while the file is in frequency it is easier to transform the spectrum and find new sounds and more immediate representations of the sound signal. For the present work we will focus only on two: the "Harmonizer" (W7) and the "Vibrato and Tremolo"(W12). Thus were selected mainly because these tracks beforethe HRTF were the ones that produced more sensations in the vibrotactile band. The following is a brief description of eachof them:

- Harmonizer: To create the effect of a harmonizing vocal chorus, offset versions of the original voice (with the same timbre) are added and forced to be in tune with the original melody [11].
- Vibrato and Tremolo: This effect are common effects used on different types of acoustic instruments, including the human voice. Both are low frequency modulations: vibrato is applied to the frequency and tremolo to the amplitude of the partials [11].

HRTF

In addition to the perception of loudness, pitch and timbre, human hearing also includes spatial perception, the subjective perception of the spatial properties of sound. From the sound of a sound source, the auditory system can determine the spatial position of the sound source in terms of directionand distance. That is useful for aiding visual attention in searching for objects and alerting humans (or animals) to avoid possible dangers. Naturally humans are primed to be able to perceive sound waves that when radiated simultaneously, the combined pressure in both ears generates information in the brain. That we reinterpret as spatial information from multiple sound sources and in turn form spatial auditory events.

From that mechanism arises HRTF and in turn the basis ofstereo and multichannel surround sound reproduction. Therefore, it is feasible to recreate that situation through certain conditions of relative level and arrival time of each sound perceived by the ear. In this way a virtual sound source is created in a spatial position where no real source exists [21]. Studies on the summed localization of two sound sources (two-channel stereo localization) had already been carried out decades ago. Where the famous stereophonic sine law was synthesized. Which shows that the spatial position of the virtual sound source (summing sound source) is completely determined by the amplitude ratio between: the two speaker signals and the separation angle between the two speakers with respect to the listener. But this law being irrelevant to the frequency and head radius [22].

However, HRTF generates a new approach. As frequency increases, the normalized HRTF magnitudes vary with frequency and azimuth in complex ways. This complexity is attributed to the global filtering effects of the head, pinna, torso and ear canal. That is, the radiated sound reaches its destination after interacting with anatomical structures, such as being diffracted and reflected by the head, torso and pinnae. This is what they called binaural sound pressures, which are something similar to an auditory illusion containing

localization information. Thus, auditory localization consists of determiningthe apparent or perceived spatial position of a sound source in terms of its direction and distance relative to the listener. In the literature following psychoacoustic studies, they have determined that there are two key types of conditions. Theseare the interaural time difference (ITD), and interaural level difference (ILD).

They have also determined that one of the best ways to perceive this effect without being immersed in a room created with loud speakers in all possible directions is through the use of headphones. Which can successfully replicate a hearing experience equivalent to the condition of real free-field sources. This means that the parameters to be considered are based on anthropometry. Thus the acoustic transmission processes, from an earphone to the listener's eardrum, are characterized by the response of the earphone, as well as by the acoustic coupling between the earphone and the external ear [23]. In our case, the earphone is replaced by vibrotactile actuators, which in theory should satisfy the same conditions.

Considering that we are talking about a 360° virtual immersive space in all directions, we understand that there can be infinite possibilities of combinations between them. Therefore, in order to make the measurements easier, we have focused only on the azimuth. This is defined as the inter-aural time difference(ITD) which represents the difference in the arrival times of a sound wavefront at the left and right ears. Similarly, there can also be an inter-aural difference (IID) which refers o the difference in amplitude generated between the right and left ears by a sound in the free field. This is known as duplex theory [24].

So the perception of sound in the "left-right" or "right-left" direction corresponds to the perception of the location of a sound in the azimuthal plane. The way it works is very simple, a sound is perceived closer to the ear to which the first wavefront arrives, which translates into a greater lateral displacement. While the other ear would receive

the out-of-phase wave. It is important to note that the lateral displacementmust correspond approximately to the diameter of the head (no more than 1500 Hz), otherwise aliasing would occur and the effect would be indistinguishable.

To implement it we used Reaper, which is a free software and a complete workstation of digital audio production application. The DearVR micro plug-in from DearReality was installed inside Reaper. It also has a very friendly GUI as shown in "Figure. 2". This software is in charge of managing the 3D binural audio technology that simulates human spatial hearing through headphones [25]. Which as already said breaks the usual barrier of sound sources inside the headphones and externalizes them. Hence the sound would seem to emanate from a specific point, anywhere within a three-dimensional sphere of 360 °.

Figure 2. Graphical interface of DearVR plug-in belonging to DearReality for Reaper Application.

Hardware Implementation

The physical implementation of the circuit illustrated in"Figure. 3" is mainly composed of an Arduino DUE, as it has two analog output pins. Those digital to audio DAC1 and

DAC2 are pins that provide real analog outputs with 12-bit resolution (4096 levels) with "theanalogWrite()" function. These pins canbe used to create an audio analog output using the Audiolibrary [20]. That output is connected to an audio amplifier (XH-M543) which in turn its amplified output is routed to the wristband Vibrotractors. In addition to the Arduino, it isconnected to an SD card module (HR0104 TF), which allows to manage the audio files in an easier, faster and simpler way.

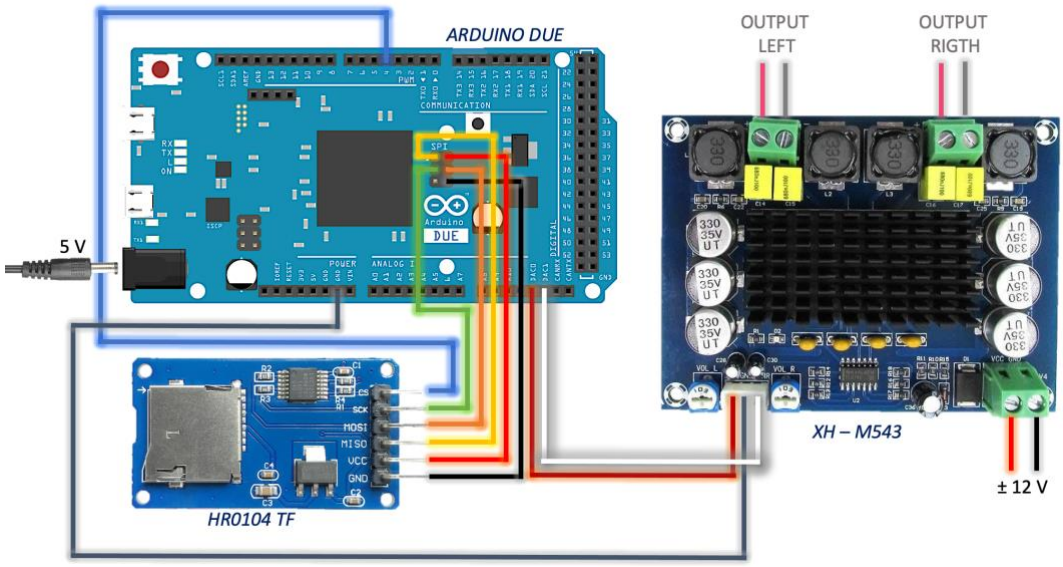

Figure 3. *Implemented physical circuit without wristband Vibrotractors.*

Finally, wristband Vibrotractors consist of a voice coiling contained in a 3D printed mold which makes it easy to place it on the skin somewhere on the body, in this case on the forearms. Each of the two actuators as input has a 3.5 mm audio jack which receives the signal from the amplifier.

EXPERIMENT

For the experimentation, a sample space of 10 people was taken into account, 5 women and 5 men between 18 and 23 years old (MD = 20; SD = 1.69), each one of them wore each of the two actuators respectively on their right and left arms. Their arms rested on a flat surface

and the distance between them was the distance corresponding to the distance between each person's shoulders. To ensure that the haptic sensations would predominate and not interfere with external means, an attempt was made to neutralize the other senses. They were blind folded and isolated from all external environmental noise and odors, in a room prepared for this experimentation. For the experimentation, three files were presented.

- W0 - haptic file without effects
- W7 - haptic file with Harmonizer effect
- W12 - haptic file with Vibrato and Tremolo effect

Before playing each haptic file, the original audio file was presented without any kind of modification (without SMS or HRTF). Which would serve as a point of comparison and contrast for the internalization of their own answers. In addition to this, both the experiment and each question was carefully explained without interfering or inducing answers, seeking the most realistic results possible. These were divided into three sections, one for each audio, each containing the same four questions. These were:

- *"Rate the sensation of the vibration."* A Lickert scale between 0 and 7 that aimed to rate how well the vibration felt, if you felt any tactile effect and above all to compare it with the original audio.
- *"Describe in your own words what you felt, try to be very descriptive."* A qualitative question that sought to obtain more diverse answers, similes and metaphors of what they might have felt.
- *"Which of these most closely resembles what youfelt."* It consisted of presenting a list of common tactile illusions, users could select more than one.
- *"In case you felt a spatial movement on your skin, what was it like?"* This question is related to the HRTF process that the audio suffered and if anyperson could detect it on their skin.

RESULTS

To understand the results we must first analyze the resulting output after the SMS algorithm, as we can see in the workflow the output corresponding to the peak detection is obtained twice. The first spectral analysis corresponds to the transformation of the original audio into a sinusoidal model through SMS at “Figure. 4a”. The second response is obtainedone for each audio channel (Left - Right) from the HRTF response and the spectral modeling accompanied by its effect, as can be seen in “Figure. 4b” and “Figure. 4c” respectively.

To interpret this graph we understand that on the X-axis is the time, while on the Y-axis is the frequency, each of the lines (guides) contained represent the spectral reinterpretation of the audio. Serra [10] defines the output of the deterministic analysis as a set of amplitude and frequency functions with a break point for each frame. From these functions, a series of sinusoids can be synthesized to reproduce the deterministic part of the sound. This model in turn consists of peaks associated to a local maximum, which are part of a frequency trajectory and the peak continuation algorithm is responsible for assigning each peak to a particular track [11]. Hence the image contains many associated frequencies for the same time, each one called partial.

At the same time, the behavior of each partial varies according to the signal. It means that they have a variable behavior in time and therefore each partial will be different from each other. This also indicates that they change accordingto the type of input processed. In our case we are processing harmonic signals, as indicated in the previous section. This implies that each of the guides created during the analysis are decomposed into their frequencies, the line with the lowest frequency represents the fundamental frequency, the others corresponding to the same instant of time correspond to their harmonics. Finally, it is

worth noting that each section of the spectrum differentiated by a valley in the peaks could correspond to a different frequency and therefore can be reinterpreted as a musical note.

For “Figure. 4a” we notice that the audio has a lot of information, the spectrum contains an array of musical notes and a lot of high frequencies around 11000 [Hz], among its top values. As we know that those frequencies are high frequencies and those sounds when reproduced through the vibrotactile actuator, simply do not vibrate as we have explained before. In process to make that change, according to the workflow first a HRTF transformation was performed which potentiates the possible intrinsic tactile effects and in turn returns a stereo sound. After that the second SMS is processed one for each channel, including the effects for spectral transformation. Theresults are very clear, the peak measurements are very sensitive to the transformations because as soon as modifications are applied to the analysis data, parts of the sound that were not audible in the original can be made audible [10].

In our study, parts of the sound that were not felt in the original file are made sensitive on the skin. We note in both “Figure. 4b” and “Figure. 4c” that the frequency component has been reduced by more than half, obtaining for example in (b), we have among its highest frequencies values between 2000 [Hz] and 4500 [Hz]. Without taking into account a couple of atypical values, but still these peaks do not exceed 6000 [Hz].

We can synthesize the results as the separation of our file into two channels, which complement each other to create a stereo sound and a greater vibrotacile sensation. Then we observe that the whole spectrum is compressed to lower frequencies, keeping the same patterns, notes and particularities of the original spectrum. This implies that we maintain in essence basic characteristics of the original piece. That perhaps may be related to the intention of the interpreter or musical artist, a psychoacoustic background or any related

emotion. With the difference that now all the new spectrum is now feelable through the vibrotactile actuators. Therefore the sound essence is transformed to a vibrotactile essence.

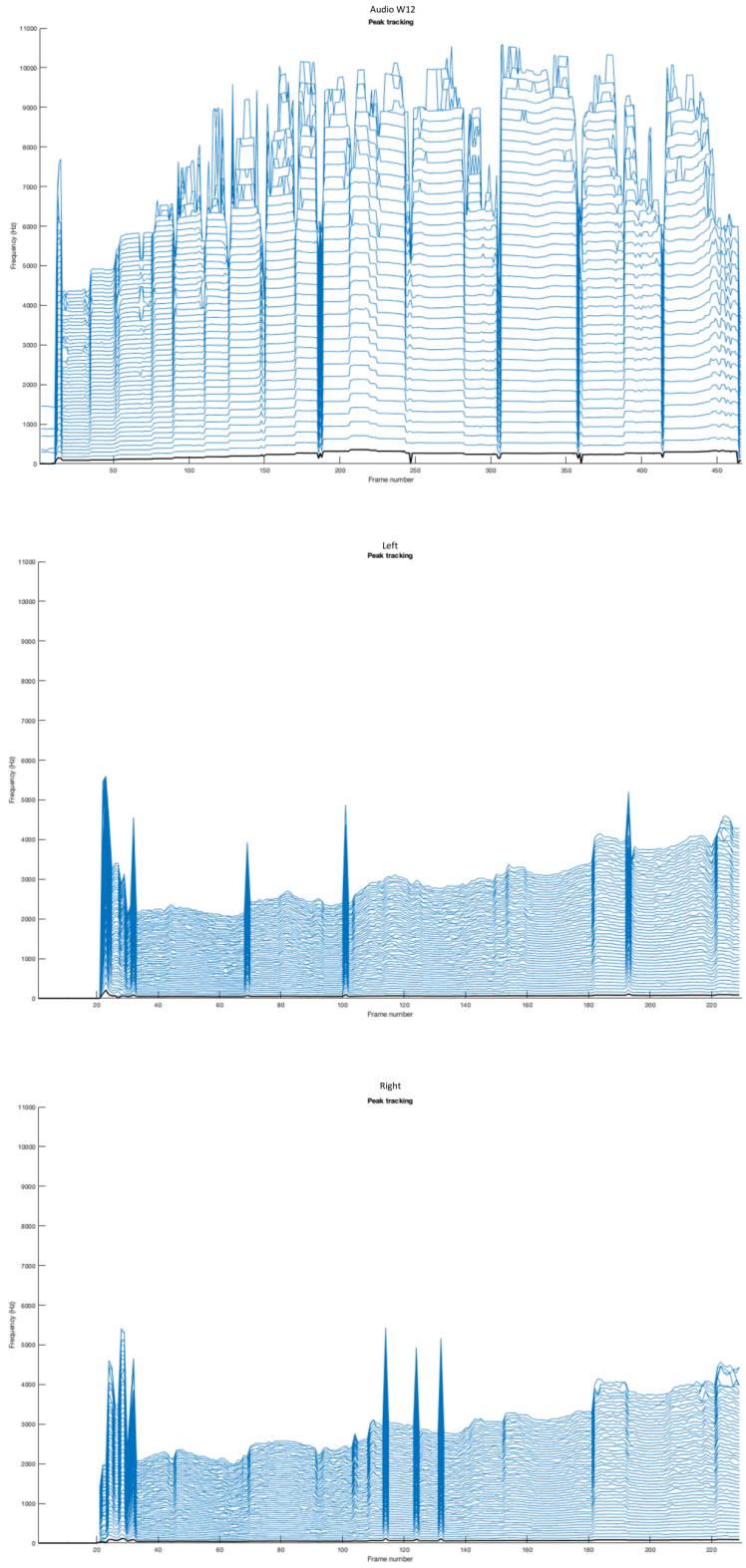

Figure 4. S*pectral analysis of a. Audio W12 with Vibrato and Tremolo effect b. Audio W12 left channel without effect after HRTF c. Audio W12 right channel without effect after HRTF*.

Once these statements are made, we can interpret the results of the audios tested, mainly to compare and contrast them. Whose fundamental objective was to increase tactile effects to a given audio. For this purpose and after the Lickert scale we can perform some statistics to determine which is the best version of them. As we observe in “Figure. 5”, we have thecomparison of three files, W0 (No effect), W7 (Harmonizer)and W12 (Vibrato and Tremolo), associated with a number ofthe mentioned scale.

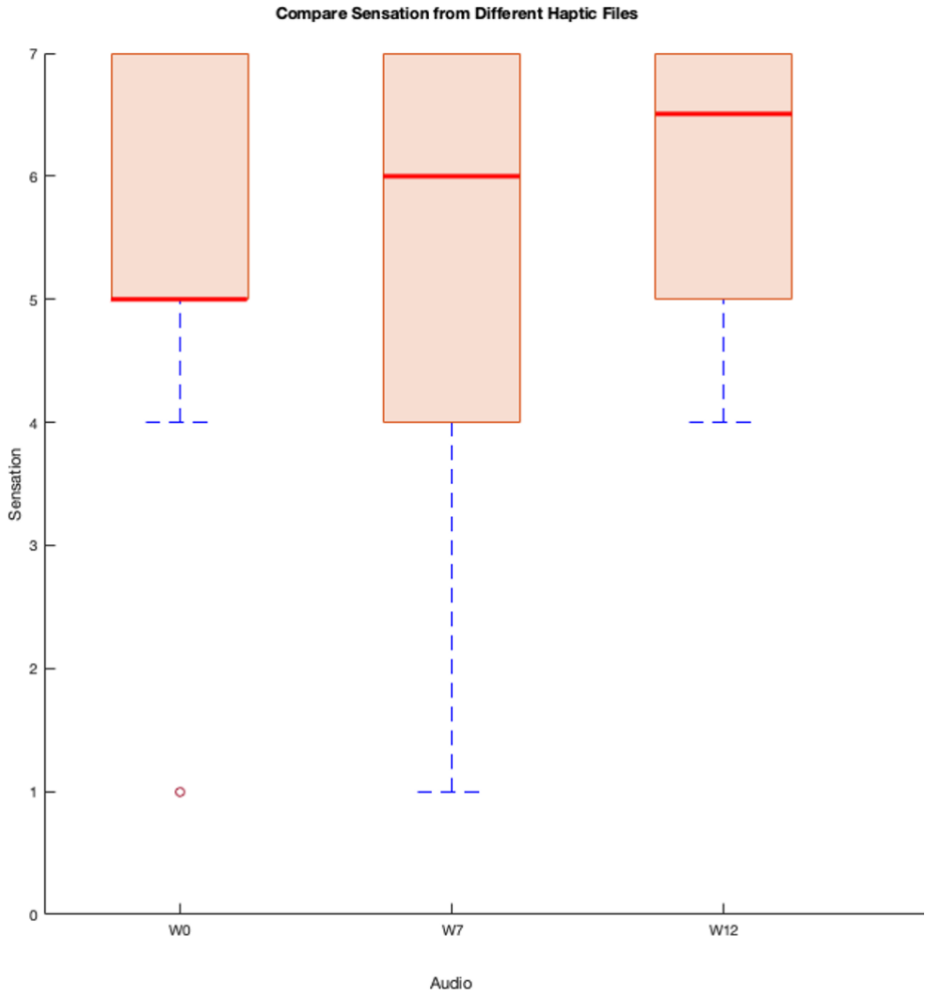

Figure 5. *Box plot comparing the Psychophysical Sensation of each file*.

We are interested in the fact that this result does not have outliers, that its values are concentrated in a certain section of responses, that it has a balanced mean at the center of its data and finally a low standard deviation that ensures that most of the participants in the

experiment have felt the same way. In the first place we can discard W7, because in spite of having a more or less centered mean of its data (MD = 5.3). It has a very high standard deviation (SD = 2.0575), which in this case is a lot of data dispersion with respect to the mean.

Next, we can compare the versions between W0 and W12.We notice that the mean of W0 is much lower than expected (MD = 5.2) being at the bottom of the box, even though it had an acceptable standard deviation (SD = 1.8135). This leaves as the only and best option the W12 file. That it has a very acceptable mean (MD = 6) which is equivalent to greater tactile and satisfactory sensations, remembering that it is very close to 7 which is the maximum value that would ideally be expected. At the same time, the standard deviation of W12 (SD = 1.1547) is the smallest of the three files, indicating concordance in the responses and therefore reliability in the data obtained.

In addition to the statistical data, other results consideredqualitative were obtained. These explore the sensations described by people, for example: “I felt as if something wasspinning on top of my arm as it moved.”, “More frequent and stronger vibrations, it felt like a scream”, “Felt very strong and felt pins and needles in some parts of my right arm like a small pin prick and stronger touches”, “Felt touches on the arm passing from the upper part of the arm”. On the other hand, with respect to the other questions, we can observe “Figure. 6”.

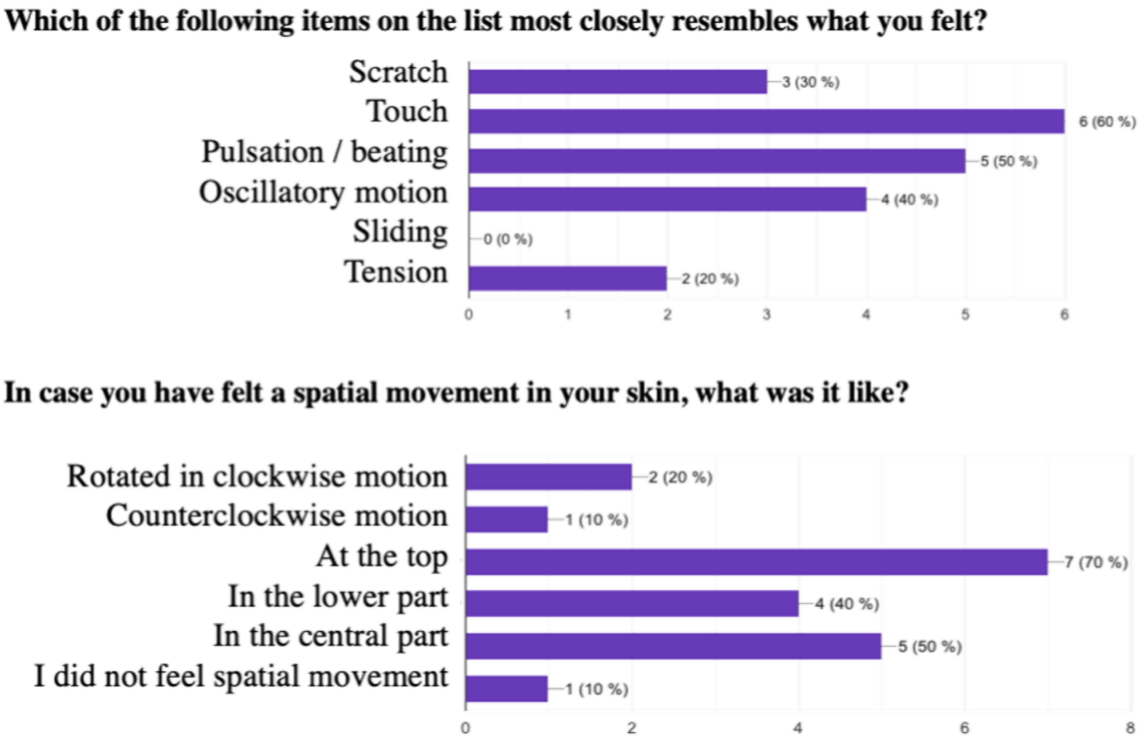

Figure 6. *Qualitative results of 10 individuals.*

Corroborating that indeed psychophysical sensations have been created in various spectra, their responses resulting in anamalgam of the particular sensations of individuals. Which, depending on the case, share common characteristics. After all, perception and sensations are subjective and unique to each individual.

DISCUSSION

Originally it was expected that the proposed workflow would be able to transmit tactile sensations. Similar to those that have been studied in the literature and at the same time that these could be spatially localized on the skin, understanding the arm as a three-dimensional solid in which this virtual sound sphere would be imaginary contained, which is very common to spatialize sound in environments related to HRTF audios. Replacing the headphones by actuators. In brief they could have been the same thing with the difference

that one contains a certain membrane that moves the air and producessound. While the other has no membrane but moves directly on the skin and produces vibration.

However, the results obtained were partially different, but no less interesting. The tactile effects were expressed in the same way and in many different versions directly related to the particular spectrum of the input. What was not expected is that HRTF, does not directly convey spatial information since only a couple of people could feel it. However, HRTF served as a catalyst between tactile illusions to boost each of their effects.

This study also served to deepen in more areas of knowledge, starting from the creation of the musical piece which is much more than amplitude and frequency. It could become happiness or sadness when heard, or in turn produce a memory or be associated with something that goes beyond the physical and enters the psychological. That information is as relevant as the generation of the functions themselves, unfortunately all that concept in the current status quo deaf people miss it. Because the information could not be transformed to a non-audible medium correctly. Simply were vibrations similar to a noise or stochastic signals. So far in this research, where it is intended that most of the wave data is transmitted reliably and in turn produce through vibrations the same effect that would produce a sound in the psyche/brain of people.

CONCLUSION

Finally, after all the experimentation carried out, we have obtained certain files that effectively transmit tactile illusions through a simple specialized hardware. Which have also been tested and statistically corroborated with experimentation on humans. The psychophysical sensations can indeed be perceptible. In general, the best results were obtained among the case studies, and the one that produced the most tactile illusions

corresponds to the Vibrato and Tremolo effect. Which also underwent an HRTF transformation prior to that effect. That Space-Fixed Audio transformation only varied on the azimuth by 360° in a clockwise direction. As a result, the processed audio changes spectrally in a notorious way, rescaling and/or compressing each of its frequencies to much lower equivalents. If we talk about fundamental frequencies we could say that the process is similar to reinterpreting the audio some octaves lower, producing sounds with less frequency and therefore more bass sound. With the particularity that it maintain the original shape of the seed file. Concluding effectively in a Sensory Replacement System, in which what before was sound, now is a vibrotactile illusion. Safeguarding properties of the essence of the audible track. It is presentedas a very good alternative for deaf and non-deaf people. So they can feel the music (literally and metaphorically) in a no conventional way. It is more realistic and closer to that would correspond to the audible part.

REFERENCES

[1] D. S. Alles, ”Information Transmission by Phantom Sensations,” in IEEE Transactions on Man-Machine Systems, vol. 11, no. 1, pp. 85-91, March1970, DOI: 10.1109/TMMS.1970.299967.

[2] Karam, Maria Nespoli, Gabriel Russo, Frank Fels, Deborah. (2009). Modelling Perceptual Elements of Music in a Vibrotactile Display for Deaf Users: A Field Study. Proceedings of the 2nd International Conferences on Advances in Computer-Human Interactions, ACHI 2009.249-254. 10.1109/ACHI.2009.64.

[3] D. P. Ellis. Extracting information from music audio. Commun. ACM,49(8):32–37, 2006.

[4] E. Pezent, B. Cambio and M. K. O’Malley, “Syntacts: Open-Source Software and Hardware for Audio-Controlled Haptics,” in IEEE Transactions on Haptics, vol. 14, no. 1, pp. 225-233, 1 Jan.-March 2021, DOI:10.1109/TOH.2020.3002696.

[5] Dario Pittera, Marianna Obrist, and Ali Israr. 2017. Hand-to-hand: an intermanual illusion of movement. In Proceedings of the 19th ACM International Conference on Multimodal Interaction (ICMI ’17). Association for Computing Machinery, New York, NY, USA, 73–81.DOI:10.1145/3136755.3136777

[6] Frid, E. Haptic music-exploring whole-body vibrations and tactile sound for a multisensory music installation.

[7] S. C. Nanayakkara, L. Wyse, S. H. Ong, and E. A. Taylor, “Enhancing Musical Experience for the Hearing Impaired Using Visual and Haptic Displays,” Human Computer Interaction, vol. 28, no. 2, pp. 115–160,2013.

[8] Burtt, H. E. (1917). Tactual illusions of movement. Journal of Experimental Psychology, 2(5), 371–385.

[9] Schneider, O. S., MacLean, K. E. (2016). Studying design process and example use with Macaron, a web-based vibrotactile effect editor. 2016 IEEE Haptics Symposium (HAPTICS). DOI:10.1109/haptics.2016.7463155

[10] Serra, Xavier. (1997). Musical Sound Modeling with Sinusoids plus Noise.

[11] Zolzer, U. (2011). DAFX: Digital Audio Effects, Second Edition. InJ. Bonada, X. Serra, X. Amatriain, A. Loscos,U. Z ̈olzer (Ed.), Chapter 10: Spectral Processing (pp. 393-445). John Wiley Sons, Ltd.DOI:10.1002/9781119991298.ch10

[12] Ali Israr and Ivan Poupyrev. 2011. Tactile brush: drawing on skinwith a tactile grid display. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’11). Association for Computing Machinery, New York, NY, USA, 2019–2028.DOI:10.1145/1978942.1979235

[13] J. Lee, Y. Kim and G. Jounghyun Kim, “Rich Pinch: Perceptionof Object Movement with Tactile Illusion,” in IEEE Transactionson Haptics, vol. 9, no. 1, pp. 80-89, 1 Jan.-March 2016, DOI:10.1109/TOH.2015.2475271.

[14] Jongman Seo and Seungmoon Choi, “Initial study for creating linearly moving vibrotactile sensation on mobile device,” 2010 IEEE Haptics Symposium, 2010, pp. 67-70, DOI: 10.1109/HAPTIC.2010.5444677.

[15] Beade, Ileana Paola; La doctrina kantiana de los dos mundos y surelevancia para la interpretación epistémica de la distinción fenómeno /cosa en sí; Universidad de Tarapacá. Escuela de Psicología y Filosofía; Límite; 8; 27; 12-2013; 19-37

[16] Kitahara, Tetsuro Goto, Masataka Komatani, Kazunori Ogata, Tet-suya Okuno, Hiroshi. (2005). Instrument Identification in Polyphonic Music: Feature Weighting with Mixed Sounds, Pitch-Dependent TimbreModeling, and Use of Musical Context.. 558-563.

[17] Azuara de Pablo, G. (2016). Estudio paramétrico de la acústica de la boquilla del saxofón y fabricación según características.

[18] Serra, X. and J. Smith. 1990. Spectral Modeling Synthesis: A Sound Analysis/Synthesis System based on a Deterministic plus Stochastic Decomposition. Computer Music Journal 14(4):12–24.

[19] Allen, J.B. 1977. Short Term Spectral Analysis, Synthesis, and Modification by Discrete Fourier Transform. IEEE Transactions on Acoustics, Speech and Signal Processing, 25(3):235–238.

[20] Arduino. (2021). Arduino Store. Retrieved from Arduino Due:https://store-usa.arduino.cc/products/arduino-due

[21] Xie, B. (2013). Head-related transfer function and virtual auditory display (Second Edition). J Ross Publishing.

[22] Boer K. de (1940). Stereophonic sound reproduction, Philips Tech. Rev.1940(5), 107-114.

[23] Wightman F.L., and Kistler D.J. (1989). Headphone simulation of free-field listening, II: psychophysical validation, J. Acoust. Soc. Am. 85(2),868-878.

[24] Cheng, C.I. Wakefield, G.H.. (2001). Introduction to head-related transfer functions (HRTFs): Representations of HRTFs in time, frequency, and space. AES: Journal of the Audio Engineering Society. 49. 231-249.

[25] Dear Reality GmbH. (2021). Binural Mixing Enter The World Of Spatial Audio. Retrieved from DearVR Micro User Manual:https://lm.dearvr.com/files/dearVRMICROManual.pdf