UNIVERSIDAD DE INVESTIGACIÓN DE TECNOLOGÍA EXPERIMENTAL YACHAY

Escuela de Ciencias Biológicas e Ingeniería

TÍTULO: Design of a Software for Training and Rehabilitation of Speech for Hearing Impaired People

Trabajo de integración curricular presentado como requisito para la obtención del título de Ingeniería Biomédica

Autor:

Pilicita Guanochanga Luis Estalin

Tutor:

Ph.D. Salum Graciela Marisa

Co-tutor:

Ph.D. Zerpa Morloy Levis Ignacio

Urcuquí, Octubre 2020

UNIVERSIDAD YACHAY TECH

Urcuquí, 22 de octubre de 2020

SECRETARÍA GENERAL
(Vicerrectorado Académico/Cancillería)
ESCUELA DE CIENCIAS BIOLÓGICAS E INGENIERÍA
CARRERA DE BIOMEDICINA
ACTA DE DEFENSA No. UITEY-BIO-2020-00030-AD

A los 22 días del mes de octubre de 2020, a las 09:00 horas, de manera virtual mediante videoconferencia, y ante el Tribunal Calificador, integrado por los docentes:

Presidente Tribunal de Defensa	Dr. ALMEIDA GALARRAGA, DIEGO ALFONSO , Ph.D.
Miembro No Tutor	Mgs. VILLALBA MENESES, GANDHI FERNANDO
Tutor	Dra. SALUM , GRACIELA MARISA , Ph.D.

El(la) señor(ita) estudiante **PILICITA GUANOCHANGA, LUIS ESTALIN**, con cédula de identidad No. **1723176820**, de la **ESCUELA DE CIENCIAS BIOLÓGICAS E INGENIERÍA**, de la Carrera de **BIOMEDICINA**, aprobada por el Consejo de Educación Superior (CES), mediante Resolución **RPC-SO-43-No.496-2014**, realiza a través de videoconferencia, la sustentación de su trabajo de titulación denominado: **Design of a Software for Training and Rehabilitation of Speech for Hearing Impaired People**, previa a la obtención del título de **INGENIERO/A BIOMÉDICO/A**.

El citado trabajo de titulación, fue debidamente aprobado por el(los) docente(s):

Tutor	Dra. SALUM , GRACIELA MARISA , Ph.D.

Y recibió las observaciones de los otros miembros del Tribunal Calificador, las mismas que han sido incorporadas por el(la) estudiante.

Previamente cumplidos los requisitos legales y reglamentarios, el trabajo de titulación fue sustentado por el(la) estudiante y examinado por los miembros del Tribunal Calificador. Escuchada la sustentación del trabajo de titulación a través de videoconferencia, que integró la exposición de el(la) estudiante sobre el contenido de la misma y las preguntas formuladas por los miembros del Tribunal, se califica la sustentación del trabajo de titulación con las siguientes calificaciones:

Tipo	Docente	Calificación
Miembro Tribunal De Defensa	Mgs. VILLALBA MENESES, GANDHI FERNANDO	10,0
Tutor	Dra. SALUM , GRACIELA MARISA , Ph.D.	10,0
Presidente Tribunal De Defensa	Dr. ALMEIDA GALARRAGA, DIEGO ALFONSO , Ph.D.	10,0

Lo que da un promedio de: **10 (Diez punto Cero)**, sobre 10 (diez), equivalente a: **APROBADO**

Para constancia de lo actuado, firman los miembros del Tribunal Calificador, el/la estudiante y el/la secretario ad-hoc.

Certifico que *en cumplimiento del Decreto Ejecutivo 1017 de 16 de marzo de 2020, la defensa de trabajo de titulación (o examen de grado modalidad teórico práctica) se realizó vía virtual, por lo que las firmas de los miembros del Tribunal de Defensa de Grado, constan en forma digital.*

Luis Pilicita

PILICITA GUANOCHANGA, LUIS ESTALIN
Estudiante

Firmado electrónicamente por:
DIEGO ALFONSO
ALMEIDA
GALARRAGA

Dr. ALMEIDA GALARRAGA, DIEGO ALFONSO , Ph.D.
Presidente Tribunal de Defensa

GRACIELA MARISA SALUM

Firmado digitalmente por GRACIELA MARISA SALUM
Fecha: 2020.12.03 09:27:07 -05'00'

Dra. SALUM , GRACIELA MARISA , Ph.D.
Tutor

Firmado electrónicamente por:
GANDHI FERNANDO VILLALBA MENESES

Mgs. VILLALBA MENESES, GANDHI FERNANDO
Miembro No Tutor

Firmado electrónicamente por:
KARLA ESTEFANIA ALARCON FELIX

ALARCON FELIX, KARLA ESTEFANIA
Secretario Ad-hoc

AUTORÍA

Yo, **LUIS ESTALIN PILICITA GUANOCHANGA**, con cédula de identidad 1723176820, declaro que las ideas, juicios, valoraciones, interpretaciones, consultas bibliográficas, definiciones y conceptualizaciones expuestas en el presente trabajo; así cómo, los procedimientos y herramientas utilizadas en la investigación, son de absoluta responsabilidad de el/la autora (a) del trabajo de integración curricular. Así mismo, me acojo a los reglamentos internos de la Universidad de Investigación de Tecnología Experimental Yachay.

Urcuquí, octubre 2020

Luis Pilicita

Luis Estalin Pilicita Guanochanga

CI: 1723176820

AUTORIZACIÓN DE PUBLICACIÓN

Yo, **LUIS ESTALIN PILICITA GUANOCHANGA**, con cédula de identidad 1723176820, cedo a la Universidad de Tecnología Experimental Yachay, los derechos de publicación de la presente obra, sin que deba haber un reconocimiento económico por este concepto. Declaro además que el texto del presente trabajo de titulación no podrá ser cedido a ninguna empresa editorial para su publicación u otros fines, sin contar previamente con la autorización escrita de la Universidad.

Asimismo, autorizo a la Universidad que realice la digitalización y publicación de este trabajo de integración curricular en el repositorio virtual, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior

Urcuquí, octubre 2020

Luis Pilicita

Luis Estalin Pilicita Guanochanga

CI: 1723176820

Dedication

I dedicate this work to my parents, sister, friends, and God:

To my mother, María Augusta, for being my adoration and always counting on her help. Her infinite love and unconditional support in the good and bad times have been fundamental in the formation of my personality.

To my father, Luis Alberto, for being an example to follow. For his love, for his good advice and for always sharing his experiences that have allowed me to reach this stage of my life.

To my sister, Evelyn Nathaly, for always being my accomplice and friend. I have always counted on their support and confidence that have allowed me to be who I am today.

To my friends for so many moments and shared experiences.

To God for being present at every step.

Dedico este trabajo a mis padres, hermana, amigos y Dios:

A mi madre, María Augusta, por ser mi adoración y siempre contar con su ayuda. Su infinito amor y apoyo incondicional en los buenos y malos momentos han sido fundamentales en la formación de mi personalidad.

A mi padre, Luis Alberto, por ser un ejemplo a seguir. Por su amor, por sus buenos consejos y por siempre compartir sus experiencias que me han permitido llegar hasta esta etapa de mi vida.

A mi hermana, Evelyn Nathaly, por siempre ser mi cómplice y amiga. Siempre he contado con su apoyo y con su confianza que me han permitido ser lo que hoy en día soy.

A mis amigos por tantos momentos y experiencias compartidas.

A Dios por estar presente en cada paso.

Luis Estalin Pilicita Guanochanga

Acknowledgements

I want to express my gratitude to my advisor, Graciela Salum, for her knowledge, time, and unconditional support during the development of the project and during my academic education.

To my co-tutor Leviz Zerpa for the guidance and advice provided for the development of this project.

To Yachay Tech University and the professors for all the magnificent, shared knowledge.

Luis Estalin Pilicita Guanochanga

Resumen

En todo el mundo, el 5% de las personas sufre de discapacidad auditiva. La discapacidad auditiva genera patologías en la voz y en la comunicación oral porque no tiene la capacidad de desarrollar el lenguaje oral como lo hace una persona sana. La audición tiene la función de generar un control de retroalimentación sobre la producción de la voz. Las personas con discapacidad auditiva no tienen un buen control de retroalimentación por lo que su voz es afectada y esto genera problemas de comunicación. El presente proyecto consiste en el diseño de una aplicación de computadora, enfocada al entrenamiento de la voz en personas con discapacidad auditiva. El programa diseñado este compuesto de 4 aplicaciones y 1 juego, estas aplicaciones permiten el entrenamiento de la intensidad de voz, tono de voz y duración de voz. Las aplicaciones generan distintas respuestas visuales a las señales de voz que son captadas por el micrófono. En el uso de este programa como terapia, se necesita un micrófono de escritorio o un micrófono de computador portátil ubicado a 50 cm de la persona para obtener resultados reales.

Palabras clave: Entrenamiento de la voz, discapacidad auditiva, intensidad de voz, tono de voz, duración de voz, control de retroalimentación.

Abstract

Worldwide, 5% of the person suffers from hearing impairment. Hearing disability generates pathologies in the voice and oral communication because it does not have the ability to develop oral language as a healthy person does. Hearing has the function of generating feedback control over the production of the voice. People with hearing disabilities do not have good feedback control, so their voice is affected, and this generates communication problems. This project consists of the design of a computer application, focused on voice training in hearing impaired people. The program designed is composed of 4 applications and 1 game, these applications allow the training of the intensity of voice, tone of voice and duration of voice. The applications generate different visual responses to the voice signals that are picked up by the microphone. In using this program as therapy, a desktop microphone or laptop microphone located 50 cm from the person is needed to obtain real results.

Keywords: Voice training, hearing impaired, voice intensity, voice tone, voice duration, feedback control.

Index

Tables list

Figures List

1. Introduction

Hearing loss is a decrease in sensitivity and hearing ability in a person. A deaf person is one who cannot hear partially or totally, such as does a person with a normal hearing sense. Hearing loss can be due to genetic causes, birth complications, some infectious diseases, chronic ear infections, use of certain medications, exposure to excessive noise, and aging (Vaughan, 2020). According to the World Health Organization (WHO) (2016) "disabling hearing loss refers to a hearing loss greater than 40 decibels (dB) in the best hearing ear in adults and a hearing loss greater than 30 dB in the best hearing ear in children". Most people with hearing disabilities live in low- and middle-income countries.

There is a close relationship between the sense of hearing and speech production because hearing is important for language learning and for feedback from oral communication. The vocal cords in most deaf people are functional. The problem is that deaf people cannot hear their own voice, and this causes that they cannot modulate their voice and the sounds they produce. The voice is described as the sound resulting from the vibration of the vocal cords, which is amplified by the resonators of the vocal tract (Wolfe et al., 2009). The articulators of the vocal tract modify this sound producing recognizable vowels and consonants, in Spanish language. A pleasant and socially acceptable voice production is highly dependent on emotional, social, and physical conditions; the latter including auditory voice monitoring.

Currently in Ecuador there are private institutions and specialists in speech therapy that help deaf people with voice disorders. These institutions use some pre-established methods and provide their services to deaf people who have hearing aid and cochlear implant technology. On the part of the Ministry of Public Health, there are Type C Health Centers, which offer 12 sessions of speech therapies per day (Ministerio de Salud Publica, 2019) . It should be noted that the 12 daily therapies are distributed among various patients with speech disorders; that is, the 12 daily therapy sessions are not exclusive for people with hearing disabilities who seek to rehabilitate their voice. The therapists' job consists of using graphics, posters, and any visual representation to relate a thing or action to its meaning; they also teach the children lip reading and improve their pronunciation

When searching for information regarding programs for the rehabilitation of the voice in people with hearing loss, two applications were found in the market such as Globus 3 and Speech Viewer III. These applications for Windows offer the possibility of performing voice rehabilitation therapies but have some drawbacks.

Globus 3 is an application aimed at voice training for people with hearing disabilities. It was developed several years ago, the application is optimized for Windows 95 (Lagares, 2005) despite this in 2012 they have managed to make the application compatible with current operating systems of Windows (Lagares, 2020), the application contains several functions that have stopped working. In addition, it has an old graphical interface and not very user-friendly, and as a final comment, the measurement scales are inadequate. In the case of Speech Viewer III, it is an application created by IBM in 1989 and consists of a phonetic viewer. The application has been discontinued by IBM and is not compatible with current Windows operating systems (IBM, 1997).

Conducting an overview of the topic, it was found that there is a need for rehabilitation of the voice in people with hearing loss but the lack of human resources (lack of professionals in the area or the low number of existing rehabilitation sections) and technological (lack of tools functional technologies) make this work impossible.

1.1 Auditory System

1.1.1 Sound

Sound is a pressure wave created by a vibrating object. It is a phenomenon that involves the propagation of longitudinal mechanical waves through a solid, liquid, or gaseous elastic physical medium (generally air). The vibrations generated by a vibrating object cause the particles present in the medium to generate a vibratory movement. In this way, the energy produced by the object is transported through the medium to a receiver (Sueur, 2018). Since the particles move in a direction parallel to the motion of the wave, the sound wave is called a longitudinal wave. Sound is also characterized by being a phenomenon that involves transport of energy, but without transport of matter. In addition, sound does not propagate in a vacuum due to the absence of particles.

In a mathematical way, sound waves can be represented in the temporal domain and in the frequency domain. In the time domain, sound is a sequence of pressure changes (pressure increases or decreases) that occur over time. In the frequency domain, the spectrum (frequency analysis) defines the sound in terms of the tonal components that make up the sound (National Research Council (US), 2004).

1.1.2 The ear

The ear, or vestibule-cochlear organ, is a complex and sensitive organ. Its function is to provide the sensation of hearing and balance to the human body. It acts as a filter in which sound stimuli (mechanical waves) are transformed into information (electrical impulses) that are interpreted by the brain. In general, the ear performs tasks such as detecting, transmitting and converting sounds into electrical impulses (National Research Council (US), 2004). According to Bansal (2013), the auditory system consists of three parts: the outer ear, the middle ear, and the inner ear, as shown in the Figure 1 .

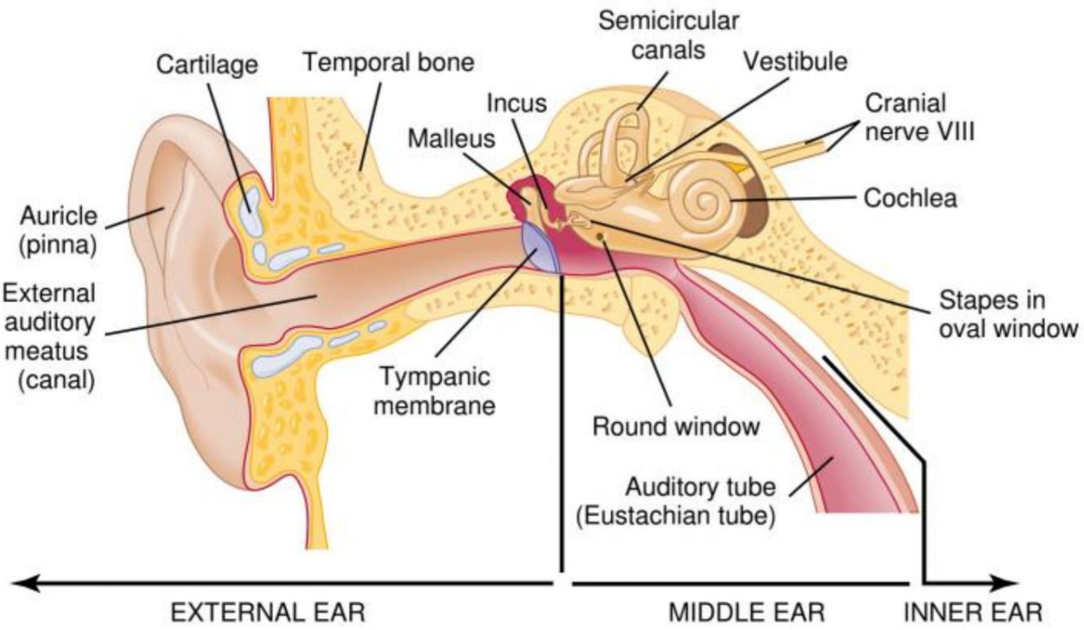


Figure 1: Image of the anatomy of the ear, modified from (Applegate, 2011).

1.1.2.1 Outer ear

The outer ear is made up of the pinna and the ear canal. Its main functions are the protection of the tympanic membrane, the capture of sound waves and the amplification of sound. The ear canal is the protection system of the tympanic membrane, the ear canal is a narrow tube approximately 0.75 cm in diameter and 2.5 cm long. It is covered with hairs and wax-secreting cells. Sound pickup is achieved thanks to the external structure formed by auricular lobule, concha, antihelix, and triangular fossa, scapha and helix (Faddis, 2008). Sound amplification occurs in the ear canal since it functions as a resonator. The ear canal is like a tube that has one end open and the other closed. The fundamental resonant wavelength is equal to four times the length of the tube and in this way the original mechanical wave is amplified.

The tympanic membrane separates the outer ear from the middle ear and has a thickness of approximately 0.1 mm, both it and the walls of the ear canal are elastic. Its function is to receive sound waves and transfer them to the middle ear (Newman, 2008).

1.1.2.2 Middle ear

The middle hatred or tympanic cavity is formed by the ossicles and the Eustachian tube. The function of the middle ear is to transform mechanical pressure waves into mechanical energy. In addition to this, it is responsible for transmitting and amplifying the sound that travels from the tympanic membrane to the oval window. The transformation to mechanical energy occurs by the movement of the ossicles, which are a chain of bones made up of the malleus, incus, and stapes. The malleus is attached to the eardrum and the stapes to the round or oval window. When the mechanical waves collide with the tympanic membrane, they produce a vibration that generates a movement of the ossicle chain. This movement of the ossicles generates movement in the oval window, so that the malleus, incus, and stapes transform the vibrations of the tympanic membrane into hydraulic waves in the inner ear. The Eustachian tube helps balance the pressure in the middle ear. A balanced pressure is required to obtain adequate transfer of sound waves (Hawkins, 2018).

The amplification in the middle ear occurs in two ways: 1) the transmission of mechanical force from the hammer to the stirrup generates a lever action, which provides a 30% increase in force. 2) the area of the oval window is smaller than the area of the tympanic membrane, as it has the same force, but in areas of different sizes. The pressure is greater in the membrane with a smaller area (Newman, 2008).

1.1.2.3 Inner ear

The inner ear is made up of the cochlea, vestibule, and semicircular canals. The functions of the inner ear are to maintain the balance of the person and transform mechanical energy into electrical signals. Both the vestibule and the semicircular canals are responsible for balance. The vestibule is made up of the utricle that detects vertical acceleration and the saccule that detects horizontal acceleration. The semicircular canals are the horizontal canal, the posterior canal and the superior canal; each of these canals detect the angular acceleration in a different orthogonal axis (Newman, 2008). The cochlea is the most important part of the auditory system as it is responsible for converting the mechanical energy of sound into electrical impulses; electrical impulses are sent through the auditory nerve to the auditory centers of the brain for interpretation.

1.1.2.4 Hearing loss and deafness

The human ear perceives frequencies from 20 Hz (lowest pitch) to 20 kHz (highest pitch). A person with normal hearing picks up sound intensity levels ranging from 0dB (hearing threshold) to 120 dB (pain threshold). People with hearing disabilities have different hearing ranges than those mentioned above.

The World Organization has classified deafness in 4 different levels of severity. The first is "slight / mild", the hearing loss is in a range between 26-40 dB, the second is the level of severity "moderate" which is between 41-60 dB, the third is the level of severity "severe" with a range of 61-80 dB and lastly the severity level "profound" that corresponds to a hearing loss of 81 dB or greater. Disabling hearing loss is considered to occur when in adults the hearing loss exceeds 40 dB in children the hearing loss is greater than 30 dB (WHO, 2016).

1.2 The Voice

The voice is defined as the sound produced by the vibration of the vocal cords, which is amplified by the resonators of the vocal tract. The voice is generated when air from the lungs flows through the vocal cords. This flow of air creates pressure to the point where the vocal cords vibrate (Zhang, 2016). The voice is as unique and different in each person as is the fingerprint. It helps to define the personality, mood, and health. The most important characteristics for speech intelligibility are pitch, intensity, and duration.

1.2.1 Pitch

Pitch is sound whose amplitude changes as a sinusoidal function over time. The pitch allows to classify the sound into bass sounds, medium sounds, and high sounds; pitch is related to frequency. At higher frequencies, the pitch is higher. For a pure sound, the pitch is determined mainly by the frequency, although it can also change with pressure and envelope.

1.2.1.1 Frequency

Frequency is a quantity that measures the number of repetitions per unit of time of any periodic phenomenon or event such as sound waves. Frequency is represented by the letter (f) and its unit is Hertz (Hz).

1.2.1.2 Voice frequency

The range of conversational frequencies of the human voice is between 250 and 3,000 Hz, which is why they have traditionally been considered the most important for understanding the word. However, some phonemes are between 4,000 and 8,000 Hz (Rodríguez, 2015).

1.2.2 Voice intensity

Intensity of sound is the property that makes sound is captured as loud, normal, or soft. It is related to the intensity of the acoustic wave; that is the amount of energy that is flowing through the medium because of the propagation of the sound wave. In other words, the intensity of the voice is the sound force that reaches the listener. An excessively loud voice is irritating and is related to aggression and authoritarianism. On the contrary, too weak a voice makes understanding difficult and projects an image of insecurity and hesitation.

1.2.2.1 Decibels

Sound intensity is defined in a wide range, where the lowest sound intensity that the human ear can hear is taken as $10^{-12}\,\frac{W}{m^2}$ while the most intense sound that the human ear can hear without causing permanent damage is taken as $1\ \frac{W}{m^2}$. To facilitate compression and work with intensity, we work on the decibel scale (Newman, 2008).

The formula for calculating decibels for sound intensity is:

$$I_{dB} = 10\log_{10}\left[\frac{I}{I_o}\right] \tag{1}$$

where "I_o" is the reference intensity or the hearing threshold $10^{-12}\,\frac{W}{m^2}$ and "I" is the intensity in $\frac{W}{m^2}$ that wants to be converted to decibels. Table 1 shows the intensity levels of some common situations-environments.

Table 1

Sound Intensity Levels and Intensities

Sound intensity level (dB)	Intensity	Example/effect
0	1×10^{-12}	Threshold of hearing at 1000 Hz
10	1×10^{-11}	Rustle of leaves
20	1×10^{-10}	Whisper at 1-m distance
30	1×10^{-9}	Quiet home
40	1×10^{-8}	Average home
50	1×10^{-7}	Average office, soft music
60	1×10^{-6}	Normal conversation
70	1×10^{-5}	Noisy office, busy traffic
80	1×10^{-4}	Loud radio, classroom lecture
90	1×10^{-3}	Inside a heavy truck; damage from prolonged exposure
100	1×10^{-2}	Noisy factory, siren at 30 m; damage from 8 h per day exposure
110	1×10^{-1}	Damage from 30 min per day exposure
120	1	Loud rock concert; pneumatic chipper at 2 m; threshold of pain
140	1×10^{2}	Jet airplane at 30 m; severe pain, damage in seconds
160	1×10^{4}	Bursting of eardrums

Note. Adapted from Sound Intensity, by *Ling et al., 2020*

1.2.2.2 Voice intensity

According to table 1, the standard voice intensity level of a normal conversation is 60dB. Additional information indicates that the intensity level in a normal conversation 0.9 meters away varies from 60 dB to 65 dB (Srivastava, 2014), while a normal conversation performed 1.5 meters away varies between 60 dB and 70 dB (Chasin, 2007).

1.2.3 Duration

The quality of sound that refers to a sound's length is duration in seconds. This quality of sound allows us to identify between long sounds and short sounds. It is related to persistence; it is a scientific term that says how long the sound wave exists. Determining the duration of a sound is complicated since the duration of the pronunciation of a word is strictly linked to the size of the same word, a comparison would be the word "bear" versus the word "incomprehensibility".

1.3 Noise

Noise is all unwanted acoustic information that is present in different environments. All information not required becomes noise. It is an unavoidable component in normal environments and is often heard at work or at home, coming from ventilation, heating, cooling, traffic systems, etc. As can be seen in table 1, the average sound intensity in a home is 30 dB, while the average intensity in an office is 40 dB. When studying human voice processing, all sound sources foreign to the voice are considered noise.

1.4 Voice Masking

Masking is the phenomenon of one sound interfering with the perception of another sound (Hansen, 2015). It is produced by unwanted sounds that inevitably interfere with the speech signal. The noises that mask the human voice can be ambient noise in an office, noise from appliances, traffic, among others. When a soft sound and a loud sound are heard at the same time, the soft sound is not perceived by the human ear. The soft sound is masked by the loud sound. Loud sound has a greater masking effect if soft sound remains within the same frequency range. Masking also occurs when the soft sound is outside the frequency range of the loud sound. Masking depends on the intensity of the "masking" and "masked" signals, as well as the frequencies (Andrade et al., 2015).

1.5 Relationship Between Hearing Loss and Speech

Hearing loss causes impairments in oral communication, so hearing loss is directly related to speech intelligibility. The degree of impairment depends on the degree of hearing impairment. In other words, the greater the degree of hearing impairment, the greater the degree of impairment in speech. While people with mild hearing loss have resonance problems, people with severe impairments may lack intensity and frequency control, among other disturbances (Coelho et al., 2015). The main voice disorders in people with hearing disabilities are breathing, articulation and phonation, the latter being the most important (Ubrig et al., 2011).

1.6 Sound Level Meter

It is an instrument that has a built-in microphone and an electrical circuit (responsible for audio processing) that allows measuring the sound pressure level, expressed in decibels. It is designed to respond to sleep sound like how a human ear does and provides objective and reproducible measurements of sound pressure level (Andrade et al., 2015).

Next, some tools used in the present work will be defined and explained.

1.7 Tools for Software Programming

1.7.1 Microsoft Visual Studio

Microsoft Visual Studio is an Integrated Development Environment (IDE) that is available for most important operative systems such as Windows, MacOs, and Linux. An IDE is an application that provides comprehensive services for software development. Microsoft Visual Studio is compatible with multiple programming languages such as C ++, C #, Visual Basic .NET, F #, Java among others (Microsoft, 2019).

1.7.2 C Sharp language

C Sharp ("C#") language, is a programming language designed by the well-known Microsoft company that has been available since 2002. C# language is an object-oriented programming language that is a branch of computer science and uses as its own name indicates objects and their interactions to design applications and computer programs. It should be noted that an object in programming is an entity that combines state (it is the object's data), behavior or method (which defines what operations the object can do) and identity (it is the differentiating factor from the other objects). The main advantages of C

are its compatibility, its relevance in the computer field and the support that is provided by Microsoft (Microsoft, 2017).

1.7.3 Windows Forms

Windows Forms ("WinForms" for short) is a graphical user interface (GUI) class library. It is a sophisticated object-oriented wrapper around the Win32 API, which enables the development of Windows desktop and mobile applications. WinForms is primarily event driven. An application consists of several forms (displayed as windows on the screen), which contain controls (labels, buttons, text boxes, lists, etc.) with which the user interacts directly. In response to user interaction, these controls generate events that the program can handle to perform tasks. In simpler words WinForms is a tool for developing dynamic and graphical interfaces.

1.7.4 Library

In computing, a library is a set of packages and functions that are available for common use in a specific programming language. A typical library might contain compilers, utility programs, packages for mathematical operations, etc. (Oxford University, 2020). They are tools or functions already coded that are useful for the development of new applications. Unlike an executable program, a library is not used autonomously, but its purpose is to be used by other programs, independently and simultaneously. Applications are generally developed with a set of libraries, which are detailed in Table 2.

Table 2

List of libraries used in software development

Library	Description	Reference
Naudio	NAudio is an open source audio application programming interface (API) for .NET written in C # by Mark Heath, with contributions from other developers. It is intended to provide a complete set of tool classes that can be used to develop an audio application. It provides a set of classes that facilitate the development of applications that record, play back, and manipulate audio files. Some of the most important functions for recording and playing audio files, creating loops, and manipulating audio signals in real time. The Naudio library is available on NuGet (Microsoft's management pack).	(Heath, 2019)
ScottPlot	ScottPlot is a free and open source graphics library for .NET that makes it easy to view data in a variety of formats with just a few lines of code. ScottPlot is an open source library that was created by Scott Harden and collaborations with various developers. User controls are available for WinForms and WPF to allow interactive visualization of data. The ScottPlot library can be installed using NuGet.	(Harden, 2019)
FftSharp	FftSharp is a library that is available in NuGet, this library has a collection of Fast Fourier Transformation (FFT) tools that can be used in digital signal processing. FftSharp is provided under the permissive MIT license, making it suitable for use in commercial applications. FftSharp targets .NET Standard and has no dependencies, so it can be easily used in cross-platform .NET Framework and .NET Core applications.	(Harden, 2018)

1.8 Data acquisition and information processing

1.8.1 Data acquisition

The acquisition of the voice signal consists of taking samples from the real world (analog system) to generate data that can be manipulated by a computer or other electronic devices (digital system). One of the most widely used techniques for converting an analog signal to digital is pulse code modulation (PCM).

PCM is a technique for converting an analog signal to digital (bits). It is the standard form of digital audio in computers, compact discs, digital telephony, and other similar applications. Figure 2 shows the process of digitizing an analog signal, this process consists of three stages: sampling, quantization and encoding. The two important factors in the PCM of sound signals are the sample rate and the bit depth.

The sampling frequency is the number of samples of a signal that are taken in 1 second, it can also be interpreted as the discretization of the signal in time. The Bit Depth in sound is the number of bits of information recorded by each sample. The Depth bit represents the capture resolution of an audio signal in relation to amplitude or height (volume). Bit depth be the discretization of the signal in amplitude. Bit Depth determines the dynamic range of an audio signal, that is, it determines the maximum and minimum decibels that a signal can have when recorded.

During the sampling process, the analog signal continuous in amplitude and time is converted into a signal that is continuous in amplitude, but discrete in time. In this process, the sampling frequency is considered, which corresponds to the number of samples that are taken in 1 second of the signal. The more samples are taken, the less information is lost from the original signal (Forouzan, 2013). Music is generally sampled at 44,100 Hz. The quantization process corresponds to the amplitude discretization of the signal that is continuous in amplitude and discrete in time, resulting in a signal that is discrete in amplitude and in time. 16 bits is the quantity commonly used in digitizing audio signals. 16 bits has a resolution of 65,536 possible values, this in turn represents that noise signals of up to 96 decibels can be stored. Finally, the coding corresponds to the interpretation of the information in parameters that the computer can interpret (0 and 1).

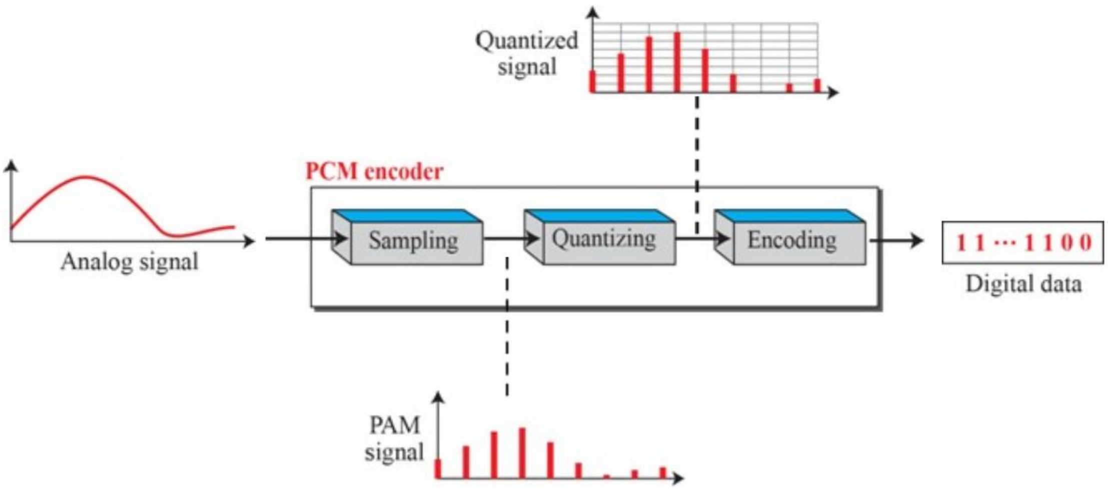


Figure 2: PCM encoder and steps to convert an analog audio signal to digital. Source: (Forouzan, 2013)

1.8.2 Waveform Audio Format (".wav")

Waveform audio format or WAV is the format for storing sound in files developed jointly by Microsoft and IBM. Support for WAV files was built into Windows 95, which made it the de facto standard for sound files on PCs. They contain sounds such as effects, music, and voice recordings. WAV sound files end with the ".wav" extension and can be played by almost all Windows applications that support sound.

WAV files are an application of the Resource Interchange File Format data storage method. The data is stored in chunks, which contain a four-character label and the number of bytes in each chunk, offering the ability to easily extend the format in the future. The equivalent file type for Mac is the AIFF format. WAV files can include 8-bit to 16-bit samples with frequencies between 11,025 Hz and 44,100 Hz. Highest possible quality, 16-bit at 44,100 Hz. Bitstream encoding typically uses Linear Pulse Modulation (LPCM) (IBM, 1991).

1.8.3 Active Noise Cancellation

Active noise cancellation is a method of reducing and eliminating unwanted sounds or noise. Active noise cancellation includes an electroacoustic or electromechanical system that cancels the primary source of noise (unwanted), based on the principle of superposition; specifically, a secondary noise of equal amplitude and opposite phase is generated and combined with the primary noise, thus the result is the cancellation of both (Andrade et al., 2015).

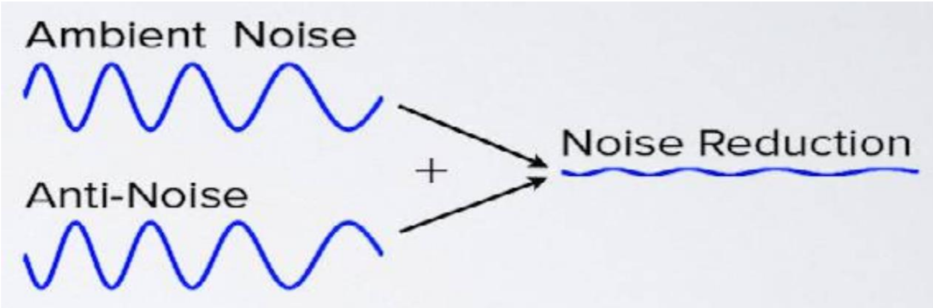


Figure 3: How active noise cancellation works modified from (Muthukumaran, 2016)

1.8.4 Root Mean Square (RMS)

Root Mean Square (RMS) is an engineering measure, created to measure the amplitude of a signal. RMS is used to characterize the "average" of continuously variable signals such as audio, electrical signals, sound, etc. The RMS measurement resembles the way a human ear perceives the audio signal. When the amplitude of the signal is measured from the PCM, it is measuring the peaks of the signal (Burg et al., 2014). Human hearing is like RMS in that the human ear does not hear peaks, but instead hears an average volume.

The formula to calculate the RMS is:

$$V_{RMS} = \sqrt[2]{\frac{\sum_{i=1}^{n}(S_i)^2}{n}} \qquad (2)$$

where n is the number of samples taken and S_i is the i^{th} sample. (Burg et al., 2014)

When working with audio signals in real time the RMS is calculated from the values of the range determined by the bit depth. The bits store the amount of voltage that is picked up by the microphone, therefore the unit of RMS is volts. The RMS unit of measurement depends on the unit of measurement of the samples, it is even possible to work with dimensionless quantities

1.8.5 Fourier Transform and Fast Fourier Transform

The Fourier Transform (TF) is the mathematical tool that allows us to go from a representation in the temporal domain to a representation in the frequency domain, as represented in Figure 3. In general, the Fourier Transform is a mathematical procedure that maps any stationary analog signal to an infinite series of sinusoids, each with a certain amplitude and phase. The Fast Fourier Transform (FFT) is an algorithm optimized for the implementation of the "Discrete Fourier Transform" (DFT).

The Fourier transform formula for discrete signals is:

$$\mathbb{F}[f(t)] = X(\omega_k) = \sum_{n=0}^{N-1} x(t_n)e^{-j\omega_k t_n}, k = 0, 1, 2, \ldots, N-1 \tag{3}$$

where $x(t_n)$ is the input signal at the moment t_n, t_n is the n^{th} number of sample, and $e^{-j\omega_k t_n}$ is the discrete polling phasor (Kernel Function) (Petale, 2019). Further $\omega_k = \frac{2\pi k}{N}$. The Fourier transform when converting the signal from the time domain to the frequency domain, its unit of temporal measurement in seconds, is replaced by the Hertz (unit of frequency).

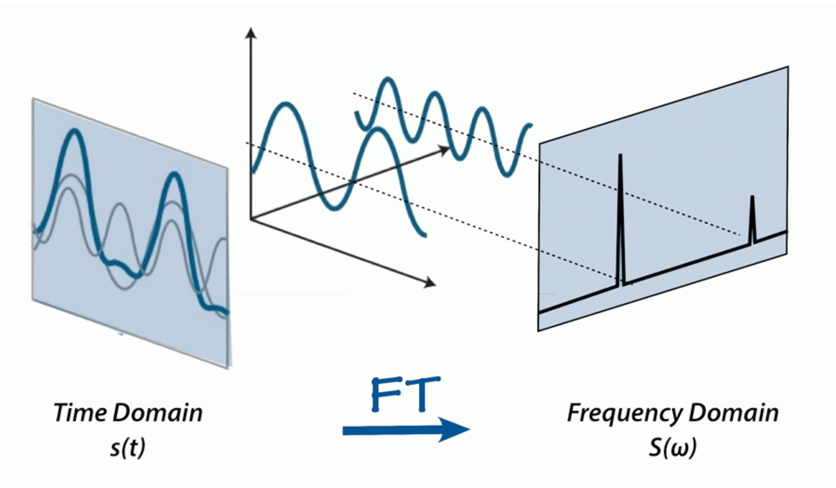


Figure 4: Graphical representation of the Fourier transform, modified of (Petale, 2019).

1.9 Cross correlation

In signal processing, cross-correlation is a measure of the similarity between two signals or waveforms, cross-correlation allows to quantify the degree of similarity between two signals, frequently used to find relevant characteristics in an unknown signal through comparison. with another that is known. The correlation of a signal with itself is called "autocorrelation". The correlation can be defined in terms of the convolution as:

$$\phi_{ab}[n] = a[n] * b[-n] = \sum_{k=-\infty}^{\infty} a[n]b[n+k] \quad (4)$$

where a [n] and b [n] are the signals to compare and "*" represents convolution.

Convolution is a mathematical operation on two functions in frequency domain (in this case a[n] and b[-n]) that produces a third function expressing how the shape of one is modified by the other. The third function is also in the frequency domain.

The amplitude of each sample in the signal generated by the cross-correlation (developed by convolution) shows how much the input signal resembles the target signal at that point. This means that a peak will occur in the cross-correlation signal for each target signal that is present in the received signal. In other words, the value of the cross-correlation is maximized when the target signal is aligned with the same characteristics in the received signal (Smith, 1999).

2 Problem statement

Worldwide, there are 466 million people who suffer from hearing loss, 34 million of whom are children. It is estimated that by the year 2050, more than 900 million people will suffer from hearing impairment, this represents that one in 10 people will suffer a disabling hearing loss (Vaughan, 2020). In 2020, in Ecuador a total of 67,621 people with hearing disabilities registered with the Ministry of Public Health are reported, of which 54.70% are men and 45.29% are women. 65.18% of all people have a degree of disability less than 50%, on the other hand, 34.93% of people with hearing impairment have a degree of disability greater than 50% (CONADIS, 2020). According to statistics, there is a significant number of people with hearing impairment. However, a low awareness-knowledge of the severity of the problems caused by hearing loss is observed. People have the perspective that hearing loss is a nuisance and do not consider hearing loss a major health problem.

Hearing loss, particularly severe and profound hearing loss, is frequently associated with voice disorders, especially if it is not treated or if it is treated in an inappropriate way. But also, if treated properly and in a timely manner, hearing loss can be accompanied by subtle abnormalities in voice control and intonation. Voice disorders in hearing loss are caused by distorted feedback control of the voice itself and can be characterized by reduced control of the pitch of the voice with too high or unstable fundamental frequency, a wobbling, sometimes pressed or creaky voice sound, rhinophonia aperta or a dysprosodic voice (Neumann et al., 2016). For this reason, hearing impairment interferes with the intelligibility of speech, causing a negative impact on the listener and crucially compromising the social integration of the individual.

The treatment or rehabilitation of the voice in people with hearing disabilities is an area of little study and it is also an area that does not receive adequate attention. The treatment and rehabilitation of people with hearing impairment is prioritized in hearing skills and, therefore, the voice is not the focus of therapy in people with hearing loss (Coelho et al., 2015). As mentioned in the introduction in Ecuador there are speech therapists, but these people are not prepared to work with people with hearing loss. The Ecuadorian health system consists of type C health centers, which offer 12 language rehabilitation sessions a day. These rehabilitation sessions are so distributed between people with hearing loss with speech problems and people who only have speech pathologies. Worldwide, technological tools have been sought to help solve the treatment of this area of health and 2 main applications have been found. These applications are Globus 3 and Speech View, these tools were created a long time ago and are no longer functional and compatible with current technology.

The treatment and rehabilitation of the voice in people with hearing loss is a great gap that has not been covered yet, due to the lack of human resources (speech therapists specialized in handling people with hearing disabilities) and technology (applications for voice training focused on people with hearing disabilities). Therefore, the development of a technological means focused on the rehabilitation of the voice in people with hearing loss is of vital importance.

3 Objectives

3.1.1 General Objectives

To development of a computer program for voice training and rehabilitation for people with hearing loss to improve and modulate verbal communication in these people.

3.1.2 Specific Objectives

- To design applications that contain audio processing algorithms that follow the parameters of a normal conversation, that capture an audio signal with the computer microphone and that generates a response, so that the user can use the application to train the pitch, intensity and duration of the voice.

- To design a graphical user interface so that the responses generated with each application generate visual alerts to make the applications friendly to people with hearing disabilities.
- To design a game that is entertaining and useful for voice training and that its operation is governed by the parameters that are considered normal in a conversation.

4 Methodology

In the development of the application for voice training in people with hearing disabilities, various calculation and conversion methods were used to obtain answers that are useful and of interest. The development of the software was carried out on a computer and with the help of mathematical and computational resources and a series of useful libraries for the processing of audio signals. The tools and methods used will be explained below.

The application was developed on a computer with an Intel (R) Core (TM) i7-8750H processor with a CPU @ 2.20GHz, with 16 Gb of RAM memory. The computer has an Intel Cannon Lake-H/S sound card - cAVS (Audio, Voice, Speech). The sound card has an integrated microphone with a capacity to record up to 16 bits at 48 kHz.

The application was created in the integrated development environment (IDE) Microsoft Visual Studio 2019 version for Windows 10. The Windows Forms application (.NET Framework 4.7.2) was used for the development of the interface and within this tool the C # programming language for algorithm development.

4.1 Data acquisition and information processing

4.1.1 Data acquisition

Voice signals produced in real time are used for data acquisition. As a first step, the analog signal generated by the human voice is captured and it is converted into a digital signal. For this, the PCM encoder is used in which, the digital signal is captured with a sampling frequency of 44,100 Hz at 16 bits.

4.1.2 Optimization

To obtain an algorithm that was optimized an efficient, it was decided to work with half of the samples taken. So, what is done is to initialize a variable "i = 0" and we define this variable to be equal to the samples to be taken, starting from the initial number, the samples "i + 2" are taken. This is done to avoid over-computation and decrease computer processing.

4.1.3 Active noise cancellation

Active noise cancellation is used to eliminate ambient noise and avoid voice masking. This process is done in two steps. As a first step, a secondary signal is generated that is equal to the original signal but with opposite amplitude, it means that the aptitude of the secondary signal has a different sign from the original signal. As a second step, the generated signal is superimposed on the original signal, thus eliminating noise. A parameter has been defined in which active noise cancellation will be active. When the signal values in PCM that are within the range that goes from -94 to 94 the noise cancellation algorithm will start working. Canceling values that are in the range -94 to 94 is equivalent to canceling all signals that have a sound intensity equal to or less than 40 dB. 40 dB is used since this sound intensity corresponds to the existing ambient noise in an office or in a house as shown in table 1.

4.1.4 Amplitude normalization

The normalization of the amplitude of the signal is an important step to know and simplify the amplitude with which it is working. This step allows to carry out the RMS calculation. For the normalization of the signal, the following formula has been used:

$$AN_i = \left|\frac{A_i}{32768}\right| , \qquad i = i + 2 \tag{5}$$

where A_i is the amplitude signal in bits is the amplitude in volts of the signal at i^{th} instant, 32,768 is the maximum number (maximum amplitude in volts) that 16 bits can store. 16 bits can store 2^{16} or 65536 different values (between positive and negative values), these values represent the amplitude of the sound intensity. The range of signed integers that can be stored in 16 bits is –32,768 to 32,767. The quantity obtained is applied absolute value to work with positive values. Amplitude in this case is expressed as a number between 0 and 1, where 1 represents the maximum amplitude in the sound file and 0 is the minimum amplitude that the sound signal can reach. Due to A_i and 32768 are voltage measurements, the value of AN_i is a dimensionless quantity.

4.1.5 Root Mean Square (RMS) calculation

The RMS calculation is done from the normalized amplitude. The RMS calculation formula was modified from formula (2), due to the optimization that was carried out and so that it gives us the desired result. The formula is as follows:

$$V_{RMS} = \sqrt[2]{\frac{\sum_i^n (AN_i)^2}{\frac{n}{2}}} \tag{6}$$

where i = 0 and increases by i = i + 2, A is the i^{th} sample size and $\frac{n}{2}$ is the number of samples taken divided by 2 because we work with half of the total samples. AN_i is a dimensionless value for this V_{RMS} also is a dimensionless quantity.

4.1.6 Calculation and Calibration of Decibels

The calculation of decibels is carried out in two stages. In the first stage, decibels are obtained, but the value obtained lacks a calibration; that is, it is not a value that adjusts to a real measurement. Since the signal we are working with is measured in volts, the formula for calculating decibels is used from the signals in volts. When remaking a previous RMS of the signal, the formula must be modified and we have the following form:

$$dB_p = 20 * \log 10(\text{RMS}) \tag{7}$$

where RMS replaces $\frac{V}{V_{ref}}$. dB_p is a weighted decibel measurement that lacks calibration and measurement accuracy. The formula was modified from Burg et al. (2014).

For the second stage, a calibration was performed so that the value dB_p was adjusted to a real decibel measurement. Then formula (7) is modified and it is as follows:

$$dB = 92{,}8 + 20 * \log 10(\text{RMS}) \tag{8}$$

4.1.7 Fast Fourier transform

For some of the programs, it is required to have the obtained signals but in the frequency domain, for which the signal found in the time domain must be transformed into a signal in the frequency domain. To obtain the signal in the frequency domain, the algorithm of the fast Fourier transform is used. For this, we used a function *fftpower* and *fftfreq* from the *FftSharp* library. The *fftpower* function calculates the power spectral density of the signal expressed in decibels. The *fftfreq* function calculates the different frequencies of the signal expressed in Hertz. After this, the power spectral density data and the frequency signal data are aligned.

4.1.8 Cross correlation and Signal Similarity Percentage.

Two audio files in ".wav" format are required for signal correlation, so it is not possible to work in real time. As a first instance for the correlation of signals, it is required that the signals be in the frequency domain to carry out the comparison. As a next step, the correlation between the signals is carried out, for which the signals are convolutional and generate a third signal. To generate the third signal, a matrix "c []" was created that was completed with the values obtained from the convolution. The formula (9) was modified from formula (4). This formula is used to perform the convolution.

$$c[i] = \sum_{i=-m}^{m} (a[n] \cdot b[n+i]) \tag{9}$$

where m is the total number of samples, a[n] is the voice signal 1 and b[n + i] in the voice signal 2 moving.

From this signal "c", the percentage of similarity between the signals is calculated. For this, an autocorrelation of signal 1 (matrix "d") is carried out and the maximum value is taken from the generated signal. In the same way, the maximum value of matrix "c" is taken and a division is performed. The maximum value is taken because this value represents when the signals generate the greatest similarity.

$$\%\ of\ similarity = \frac{maximum\ value\ of\ matrix\ c}{maximum\ value\ of\ matrix\ d} * 100 \tag{10}$$

4.2 Inclusive Graphical User Interface with Hearing-Impaired People

In Microsoft Visual Studio IDE, programming in Windows Forms is linked to development of the graphical user interface. Therefore, the programming and the graphic interface are developed together. For the development of the graphical interface, Windows Forms own tools such as button, label, richTectBox, comboBox, timer, pictureBox, progressBar, checkBox were used, and the scottPlotUC tool from the ScottPlot library was also used.

- Button: It is used for specific functions such as open, start, select, return to the main menu, and close.
- Label: It is used to display texts and numerical results.
- RichTextBox: Used to display dynamic texts.
- ComboBox: It is used to add different lists of options among which the user can select.
- Timer: Used to create functions and algorithms that depend on time.
- PictureBox: Used to create add images and characters.
- ProgressBar: Used to display results that are seen as progress or load charts.
- CheckBox: It is used to activate or deactivate options manually, the decision to activate or deactivate is made by the user.
- ScottPlotUC: is the tool used to graph the signals to be projected.

4.2.1 Inclusive hearing-impaired people

To make the application easy to use for the hearing impaired, visual alerts have been incorporated that indicate variations in response to voice. The color red has been used to indicate an excess of intensity of the voice, the color green to indicate normality and the color purple to indicate a deficiency in the intensity of the voice. A series of graphs has also been used to draw the behavior of the voice in both time and frequency and also to serve as a visual guide for training. Additionally, text notifications were used to indicate the performance of the person during the use of the application.

4.2.2 Auditory Parameters

In the programming and execution of the graphical user interface, parameters and limits that function as indicators were defined; for example, visual alerts are generated if the person who is training speaks with an inappropriate voice intensity

The parameters used are those that are considered normal in a conversation between two people 1.5 meters apart. The range goes from 50 dB (minimum) to 70 dB (maximum).

4.3 Voice graph database

A graphic database has been created that contains a graphic record of the correct pronunciation of the vowels and the most important consonants in Spanish language found in Annex 1. The graphic record shows a two-dimensional plane in which the frequencies of the voice signal (measured in Hertz unit) are found on the x-axis and the voice intensity (measured in decibels) on the y-axis. The database was created by a 23-

year-old person with a normal or neutral voice register; that is, a voice that is neither high nor low. The database is in Annex 1.

4.4 Letter Sound Database

A sound database has been created. The database contains ".wav" files that are approximately 2 seconds long. Each sound file contains the pronunciation of the vowels in Spanish language (a, e, i, o, and u) found in Annex 2. The database was created by a 23-year-old person with a normal or neutral voice register, the voice was recorded with a sampling frequency of 44,100 Hz at 16 bits. The voice recording was carried out in an environment with a noise intensity equivalent to the noise in an average home (~40 dB). The database is in Annex 2.

5 Results and discussion

A computer program was developed that is useful for voice training in people with hearing disabilities. The program is composed of 4 applications and 1 game, Figure 5 shows the flow chart of the developed program. In this flowchart you can see the different activities that the program carries out to generate a feedback.

A main menu has been designed. It contains the different applications, and their menu works as an application launcher. The applications have been programmed to generate visual feedback in response to the different components of the voice signal, such as, voice level intensity, pitch, and duration.

For the correct operation of the applications, an optimization process has been carried out so that the signal processing is carried out in real time. The optimization and operation results of the applications are detailed below.

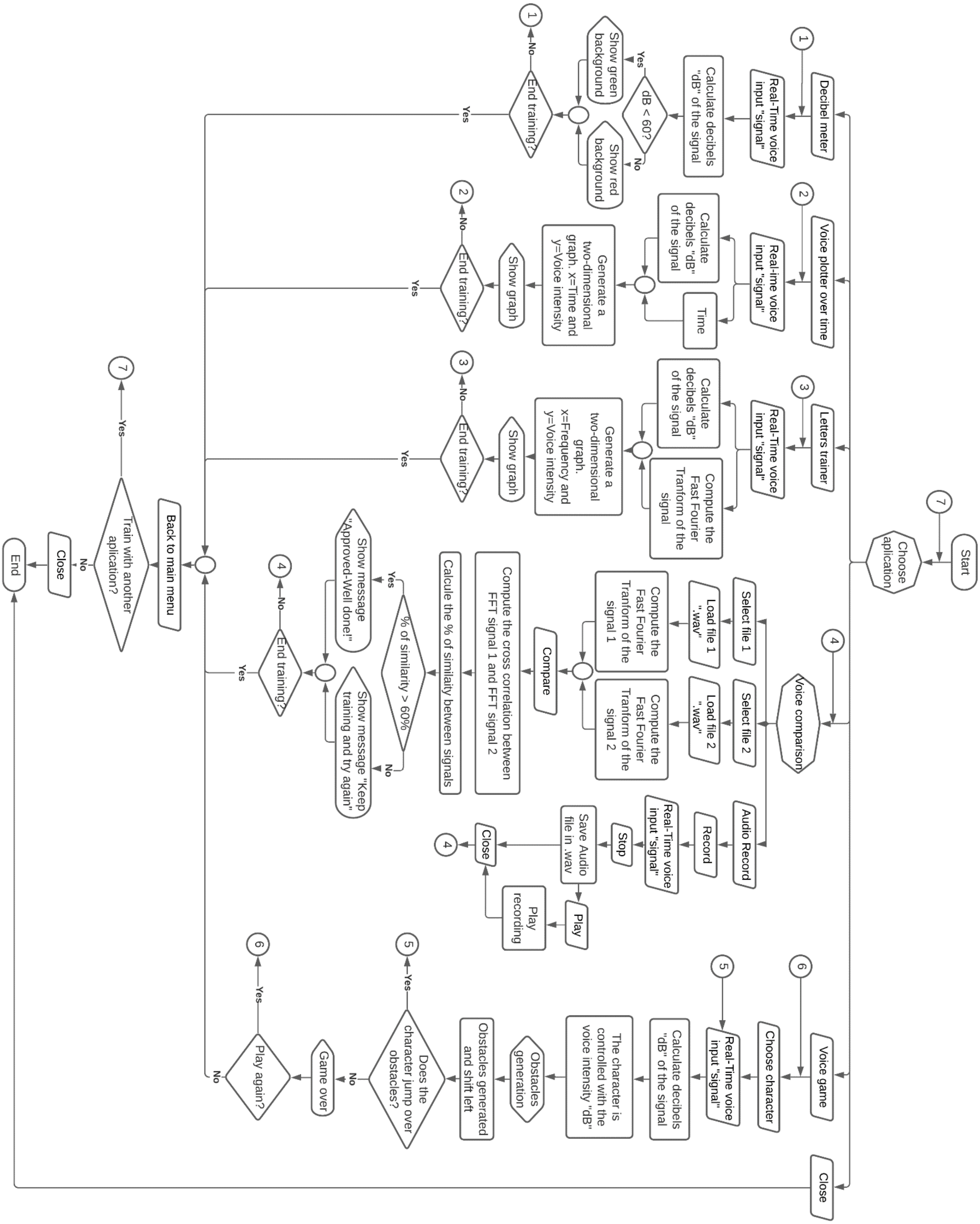


Figure 5: Voice trainer program flowchart

5.1 Optimization

The optimization process carried out helps the real-time signal processing to run smoothly and without delays as the number of samples that need to be analyzed is reduced. The optimization process carried out also helps to reduce the amount of processing that the computer needs to run the program. A program developed by Andrade et al. (2015) used the noise cancellation algorithm to process signals in real time. They worked with all the sampled values and as results found that "Due to the time it takes to capture and process audio, the software generates a delay that prevents the generated anti-noise signal from being reproduced in real time". When the optimization process is carried out, it is working with 22,050 samples per second. This number continues to be large, so the loss of samples is irrelevant and there is a correct processing of the signal in real-time.

5.2 Main menu

A main menu was designed to function as a launcher for the different applications for voice training. Figure 6 shows the block diagram of the structure and operation of the start menu. The application for voice training in hearing impaired people is composed of 5 useful functions for training, which are "Decibel meter", "Voice plotter over time", "Letter coach", "Voice comparison", and "Voice game". Figure 7 shows the graphical user interface of the main menu that contains the different launch buttons for the 5 functions mentioned above and in addition to this a button to close the program.

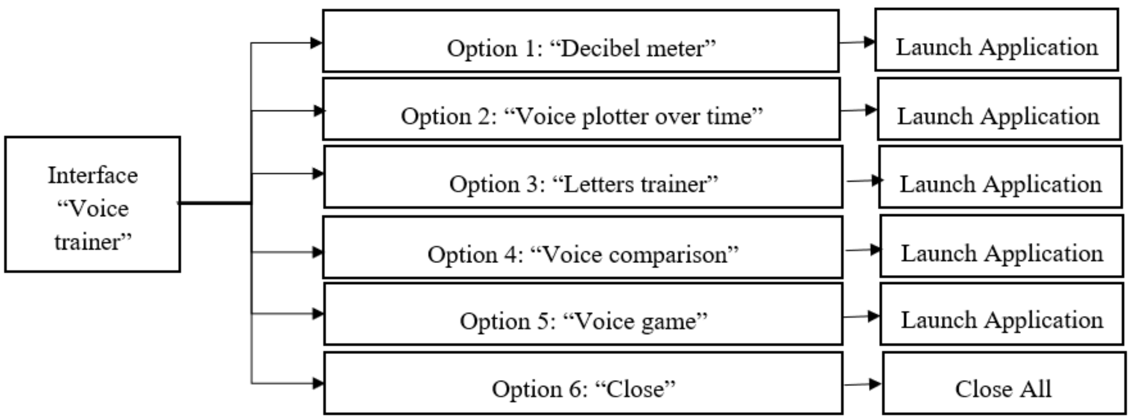


Figure 6: Main menu block diagram

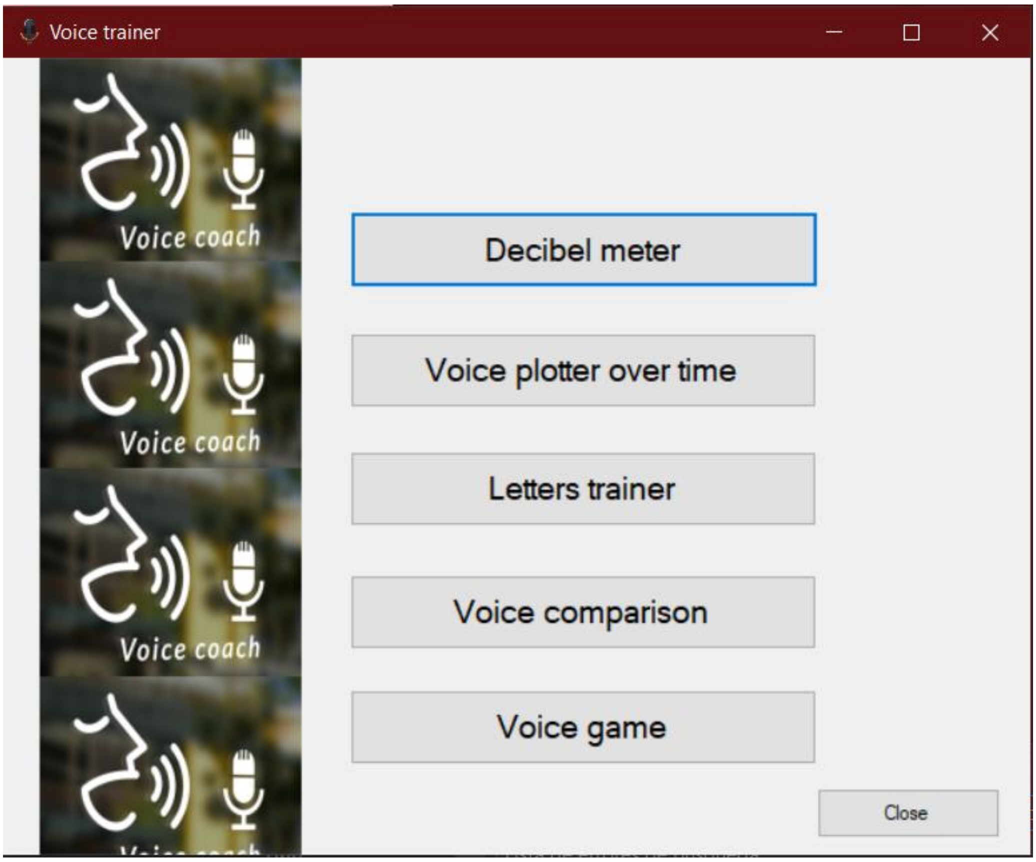


Figure 7: Main menu graphical user interface

5.3 Decibel Meter

The “Decibel meter” is a similar application to a sound level meter, but with some variations such as: the sound level meter measures the peak sound pressure level of the signal, while the developed application measures the RMS of the signal. In addition to this, the program generates a visual response at the voice level. Figure 8 shows the block diagram of the decibel meter.

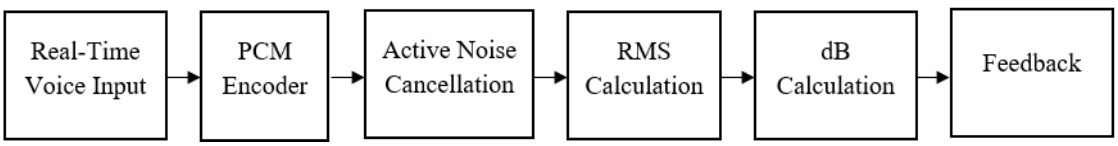


Figure 8: "Decibel meter" block diagram

In the block diagram of Figure 8 we can see that a voice signal is taken in real time and finally a visual response is generated. In block 1, the voice signal is taken in real time. In the second block, the signal is digitized to PCM with a Depth bit of 16 bits and a sampling frequency of 44,100 Hz. Here, the optimization is also performed so that of the 44,100 samples per second, 22,050 samples are used. In block 3, the noise cancellation algorithm is used and the audio signals that are in the range from -94 V to 94 V are eliminated. The 16-bit architecture in the sound section can store values ranging from –32,768 V to 32,767 V, these values represent the amplitude of the sound. The -94 V and 94 V are values of sound amplitude assigned by the 16-bit architecture, the value -94 V or 94 V is equivalent to 40 dB, so the range of values from -94 V to 94 V contains the different sound intensities lower than 40 dB. In the fourth block, the noiseless audio signal is normalized by dividing the maximum value of the bits (32,768 V) and the RMS value is calculated. In the fifth block the conversion to decibels is carried out.

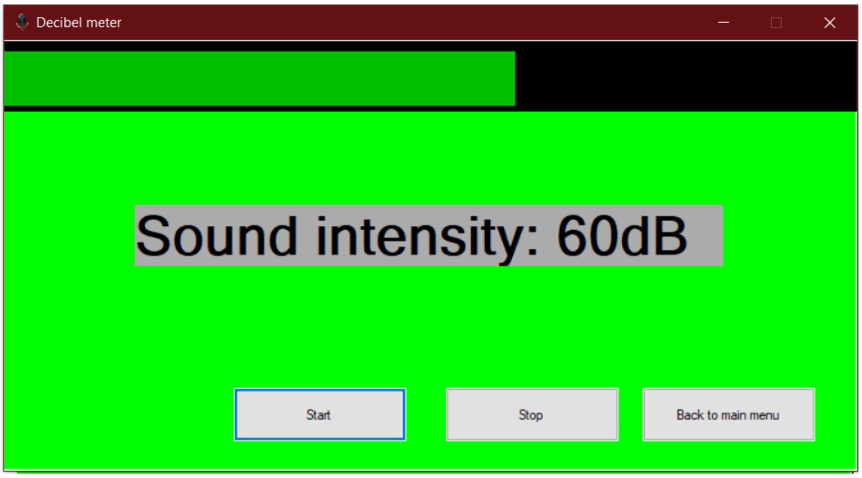


Figure 9: "Decibel meter" graphical user interface that show the feedback to a voice under the limit.

Finally, with the obtained value, a parameter is defined to generate a visual response; the parameter was defined at 70 dB. So, when a person uses a voice level lower than 70 dB, the program projects a green background, as can be seen in Figure 9.

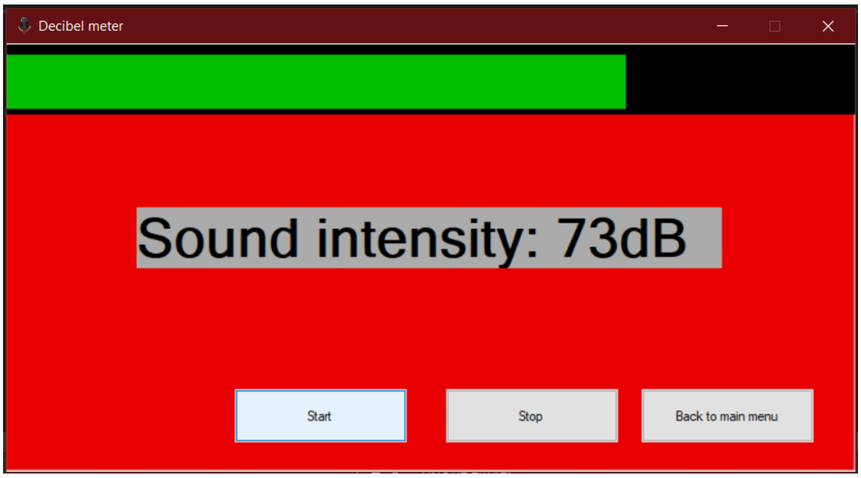


Figure 10: "Decibel meter" graphical user interface that show the feedback to a voice over the limit.

On the contrary, when the person uses a voice level higher than 70 dB the application changes the background of the application and displays a red background, as shown in Figure 10.

The application helps to control the level of intensity of the voice that is being used, this generates that the person who is using the application can modulate their voice intensity until reaching a suitable voice intensity. Controlling the level of intensity of the voice helps the individual to function in society. Some people speak too loudly. At the other extreme are those who can barely be heard. The volume of the voice should be appropriate in strength and intensity and should be varied to add emphasis and express emotions (Toastmasters International, 2011).

As mentioned in the introduction, Speech Viewer III is a program for voice training, this program contains a functionality called "intensity range", this function aims to generate a visual response to the intensity of the voice that the user is using. This application contains objects such as balloon, rocket and saxophone that vary their size or position in response to the range of voice intensity that the user emits. In this application the maximum and minimum limits of voice intensity are defined qualitatively. The limits must be calibrated manually with a voice signal; calibration is usually done by a speech therapist. In other words, the maximum and minimum limits depend on the discretion of each speech therapist and not on well-established parameters (Ortí, 2014). On the other hand, I developed the program "Decibel meter", the aim is that from a real-time voice signal a visual response to the level of voice strength is generated. The program displays the decibels of the voice (calculated in a similar way to how the ear hears sound “RMS”) and generates a visual response based on parameters or limits allowed for the intensity of the voice. The decibel limits of the voice have been defined from a standard of a normal conversation developed at 1.5 meters. So, the limits have been defined quantitatively and do not require calibration.

5.4 Voice Plotter Over Time

The program "Voice plotter over time" has a similar processing to the program "Decibel meter" as we can see in Figure 11, with the difference that it adds the characteristic of time.

This feature allows you to create a record of the behavior of the intensity of the voice (measured in decibels) versus time.

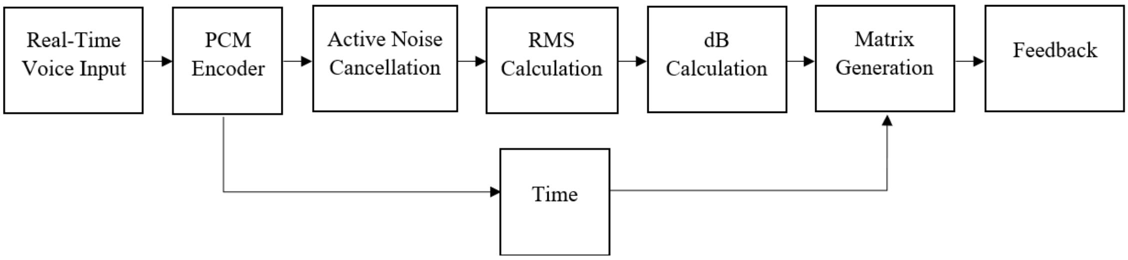


Figure 11: "Voice plotter over time" block diagram.

In the first block, the signal is captured in real time. Next, in the second block, the signal is digitized to PCM with a Depth bit of 16 bits at 44,100 Hz, due to the optimization of these 44,100 Hz and 22,050 Hz are used.

From block two, two paths are generated, in block of the time, a normalization is made, the total number of samples is used and they are divided into 22,050 and in this way the time is generated, where 1 second represents 22,050 samples, 2 seconds is 44,100 samples, 3 seconds are 66,150 and so on until reaching 10 seconds.

The other way is block 3 in which the active noise cancellation algorithm is used, which eliminates signals with a sound intensity lower than 40 dB.

In the fourth block the signal is normalized by dividing the signal amplitude for 32,768 V and the RMS calculation is performed.

In the fifth block, the decibel calculation is performed from RMS. In the sixth block, a matrix is generated in which it is filled in with decibels and time. Finally, a two-dimensional graph is generated, in the "x" axis the time is placed while the "y" axis is placed the sound intensity in decibels, in this way a graph is created in which the change of voice intensity over time.

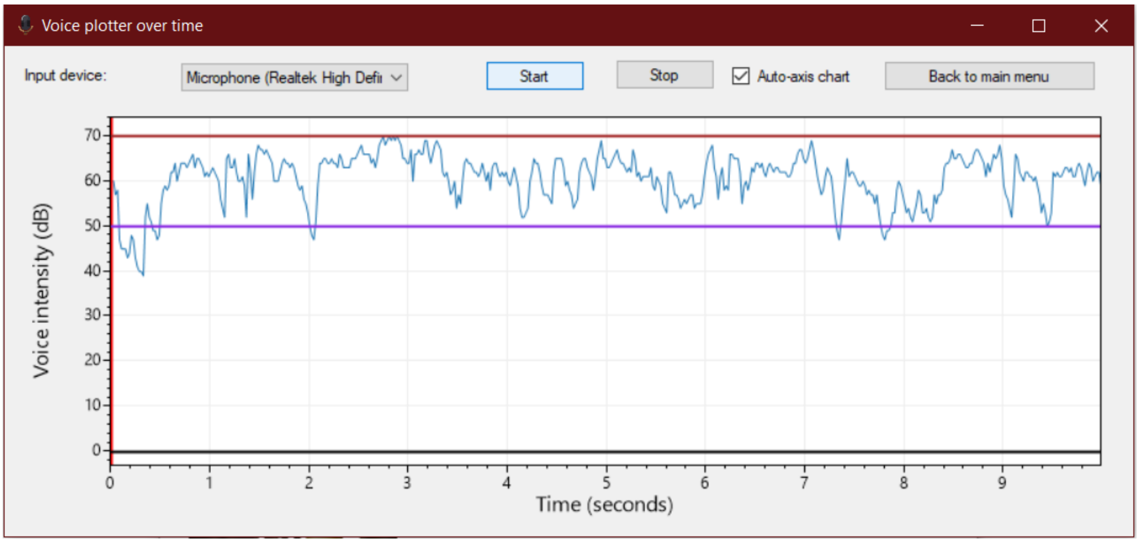


Figure 12: “Voice plotter over time" graphical user interface that show the feedback of a voice performed with normal intensity.

As can be seen in Figure 12, the parameters within which a normal conversation is found have been defined, at least there is a purple line located at 50 dB and a maximum red line located at 70 dB.

The application has two training purposes. The first purpose is that the person who is training with this system can see the behavior of their voice and that they modulate the intensity of voice so that it is within the parameters of a normal conversation. The second purpose is for people to observe the duration of the word they are speaking and see if they are speaking too fast or too slow.

Then, the application captures the voice in real time and generates the record of the voice behavior in amplitude of sound intensity and in duration. According to Sundberg et al. (2011), the intensity and duration of the words play an important role in the comprehensibility of speech. Feelings and emotions have acoustic profiles. For example, anger and happiness are usually linked to a higher intensity of voice and a faster speed of speech, while boredom and pain are characterized by a low intensity of voice and a slow speed of speech.

An application for Android devices called “Sound meter pro” has been found, this application has been developed by Smart Tools co. This application measures the level of sound intensity in the environment or place where the measurement is made. This application has the functionality of generating a report of the behavior of the sound intensity versus time; In other words, it generates a two-dimensional graph in which on the x axis it places the time measured in seconds, while on the y axis it places the sound intensity measured in decibels. The application measures the peaks of sound intensity just like a traditional sound level meter does (Smart Tools Co., 2015). Comparing Sound meter pro with the developed application "Voice plotter over time", firstly, the developed application measures decibels from RMS (model that resembles human hearing). Another difference is that the application developed thanks to active noise cancellation suppresses sound signals with intensity lower than 40 decibels. And as a final point, it is applied during the creation of the record of sound intensity versus time, the graph has 1 upper limit and 1 lower limit that allows to identify the range of voice intensity that should be used.

5.5 Letters Trainer

The “Letters trainer” application is an implementation that has been developed from the calculation of the Fourier transform as can be seen in Figure 13. The application takes voice signals in real time and generates a graph of the voice in the frequency domain versus the amplitude of the intensity of the voice.

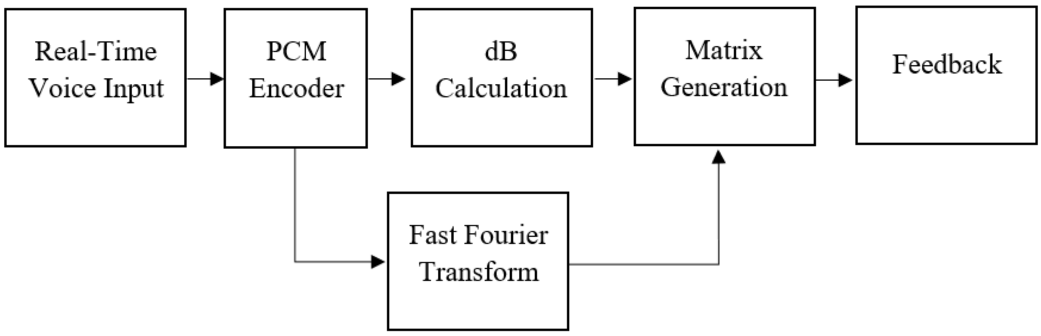


Figure 13: "Letters trainer" block diagram.

As we can see in Figure 13, in this system in block 1, the signal is captured in real time. In the second block the signal is digitized at 16 bits and 44,100 Hz. Here, an optimization is carried out in which we work with the samples that are in even positions. From this block two signals emerge. From the block of the Fast Fourier Transform, the different frequencies of the voice signal are obtained. On the other hand, in block 3, the PCM is normalized by dividing the PCM values for 32,768 V (maximum value that 16 bits architecture can store), with the value obtained the decibels are calculated. In block 4 a matrix is generated that is filled in with frequencies and decibels. In the last block, a two-dimensional graph is generated, on the x-axis the frequencies (in Hertz) of the voice signal are placed while on the y-axis the voice intensity (in decibels) are placed.

As we can see in Figure 14, the graphical user interface is made up of two parts. The first, or upper part, is the area in which an image of the data base of the register of the normal pronunciation of different letters is projected. The second part, or the lower part, has the system that graphs the voice in the frequency domain against the intensity of the voice in decibels; this is done in real time.

As already mentioned, the application has a database that contains the graphic record of the correct pronunciation of the vowels and the most important consonants in Spanish language.

This application focuses on the training and modulation of the tone of voice and sound intensity. In this application, the person with hearing disabilities has to select which letter they want to train, after that the person has to pronounce the letter they selected and try to make the graph of their voice look as similar as possible to the graph of the database projected.

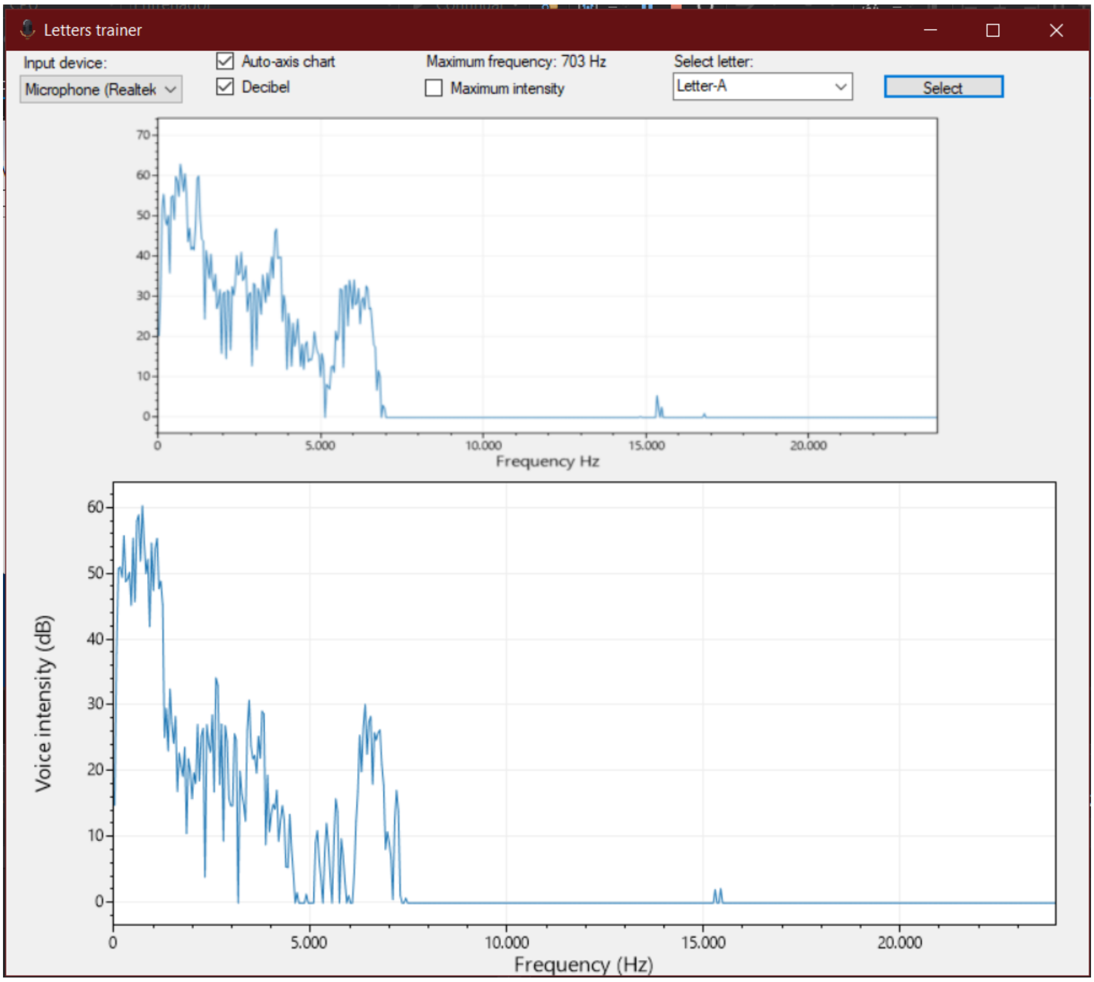


Figure 14: "Letters trainer" graphical user interface that shows the performance of the application during a person trains the pronunciation of the letter “A”.

According to Coelho et al. (2015), people with hearing impairment due to their hearing impairment are not able to have a correct feedback control. This is an important component for the correction of pitch and the intensity of voice. Oxenham (2008) carried out a study that found that people with hearing disabilities have a more imitated perception than people with cochlear implants, which makes speech difficult to understand. The program proposes an alternative to auditory feedback control that is normally used to modulate the pitch. The program provides visual feedback control to understand and modulate the tone and intensity of voice in people with hearing disabilities.

The Department of Electronic Engineering of the Polytechnic University of Madrid has developed a digital signal processing card focused on the rehabilitation of speech in people with hearing disabilities. The VISHA system has the ISOTON program, this program allows you to work with intensity, pitch, and sound. It offers a visualization of a role model that serves as a reference for the person who debuts. A speech specialist is responsible for preparing the models that the patient must imitate with his or her voice or that of another person (Perez, 2007). On the other hand, we have the system that I have developed "Letters trainer", this application does not need external components for its operation. The only component that the application requires is the sound card that is installed by default in desktop or laptop computers and a microphone. In addition to this, the Letters trainer application has a large database of the correct pronunciation of letters in Spanish language.

5.6 Voice Comparison

The “Voice comparison” allows to compare two audio streams in “.wav” format and return a percentage of similarity between the two signals. As can be seen in Figure 15, in block 1 and 2 two voice signals are loaded in “.wav” format.

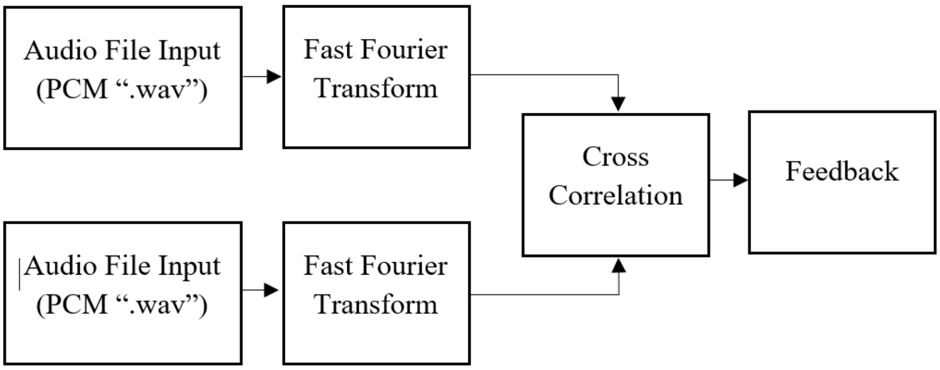


Figure 15: "Voice comparison" block diagram.

In blocks 3 and 4, the fast Fourier transform algorithm is used to change time domain signals to frequency domain signals. In the fifth block, a cross correlation between two signals is made. This generates a new signal. From this, the percentage of similarity between the two signals is calculated.

Finally, a feedback is generated from the percentage of similarity. When the percentage of similarity is higher than 60% as shown in Figure 16, a message appears saying "Approved - Well Done!".

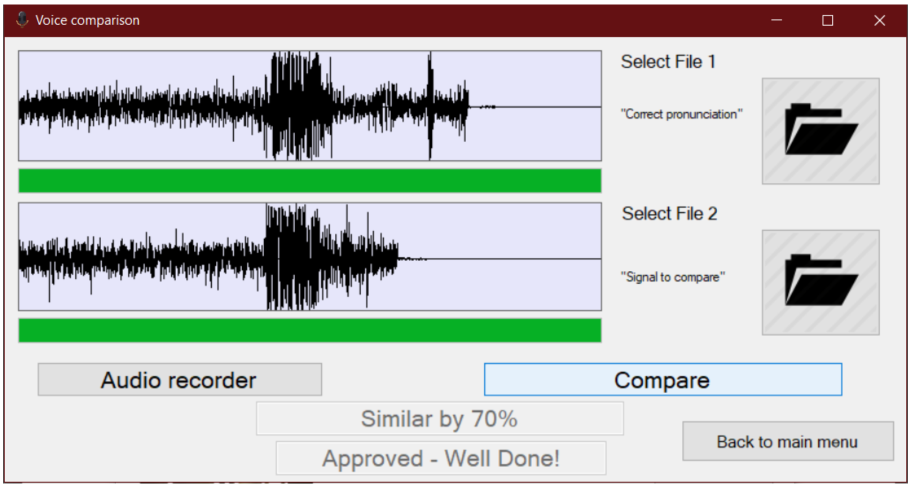


Figure 16: "Voice comparison" graphical user interface that shows the comparison between 2 audio files that has a similarity greater than 60%.

When the percentage of similarity between the audio sequences is less than 60%, as shown in Figure 17, a message is generated indicating that the message "Keep training and try again". Using cross-correlation to compare two signals is the best way to obtain a percentage of similarity between them; thus, it is based on differences in shape and frequency and not on differences in amplitude (Palacios, 2015).

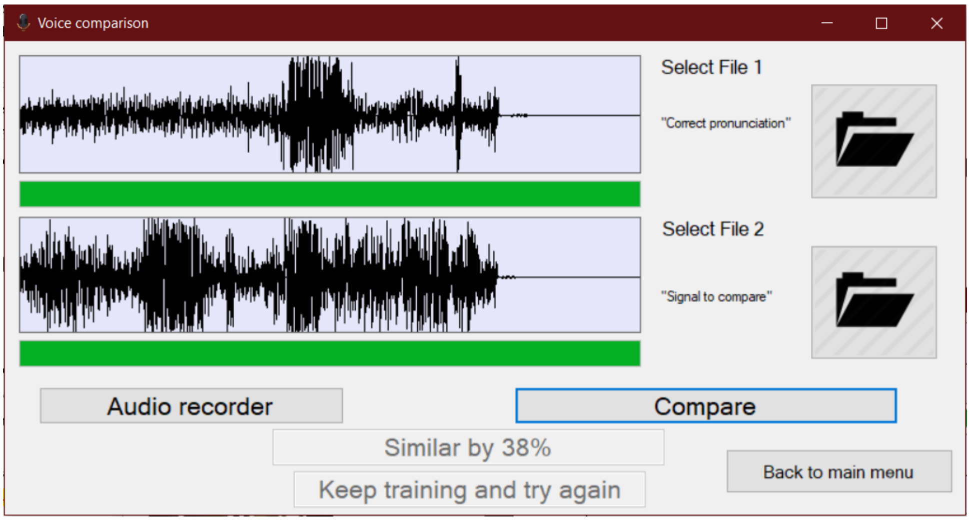


Figure 17:" Voice comparison" graphical user interface that shows the comparison between 2 audio files that has a similarity less than 60%.

In the graphical user interface, as we can see in figure 16 and 17, the necessary tools have been implemented to record new audio sequences, load the audio files, and compare the audio files.

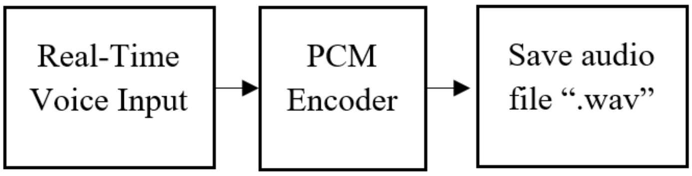


Figure 18: Block diagram of audio recording application.

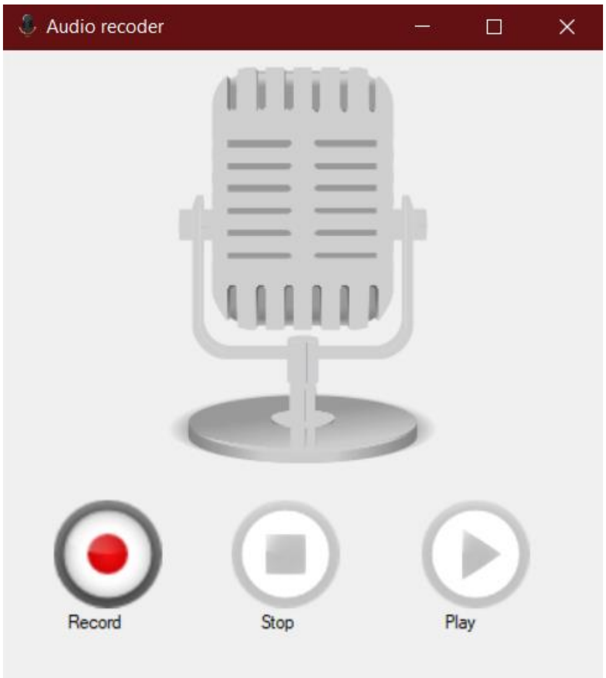


Figure 19: Graphical user interface of the audio recording application.

Figure 18 shows the block diagram of the sound recording system. The recording system is composed of three fundamental blocks. In the first block, the voice signal is captured. In the second block, the analog signal is digitized at 16 bits and at 44,100 Hz. In the third block, the audio sequence is saved in ".wav" format.

Figure 19 shows the graphical user interface of the sound recorder, which is composed of three basic buttons that are record, stop and play the recorded file.

For this application a database of correctly pronounced vowels in Spanish language (a, e, i, o, and u) has been generated. The different vowels of the database are in Annex 2.

The use of the application is aimed at evaluating the tone of the voice of the person who is training. What the application does is compare a correctly pronounced audio sequence with an audio sequence recorded by the person who is trained and then, the application generates a percentage of similarity between the sequences. The voice comparison application even can compare long speech sequences that contain words and even sentences.

It has been found that Palacios, (2015) carried out a research project called "Therapy station to improve verbal communication of people suffering from deafness and linguistic disorders. In this project, a useful application for voice training was developed". The program consists of visually and mathematically comparing a voice signal of the correct pronunciation of the letters a, e, i, o, u, r, l, m and n versus a signal produced by the person who trains. The percentage of similarity is obtained from the use of cracked correlation. As a conclusion of the project, they found that the use of cross correlation is the best way to obtain a percentage of similarity between two signals. On the other hand, the program that I have developed "Voice comparison" compares two audio sequences and generates a percentage of similarity. One of the main differences between the programs is that the developed program can compare different audio sequences, that is, people or speech therapists can generate more voice records. This increases the range of letters or words to be trained and in the same way improves the rehabilitation process since each of the records created must be personalized for each patient. One of the coincidences between the two applications is that the similarity calculation is calculated from developing a cross correlation.

5.7 Voice Game

The "Sound Game" is an entertaining application for training of the intensity and duration of the voice. In Figure 20, it can be seen the application that is made up of two main information chains.

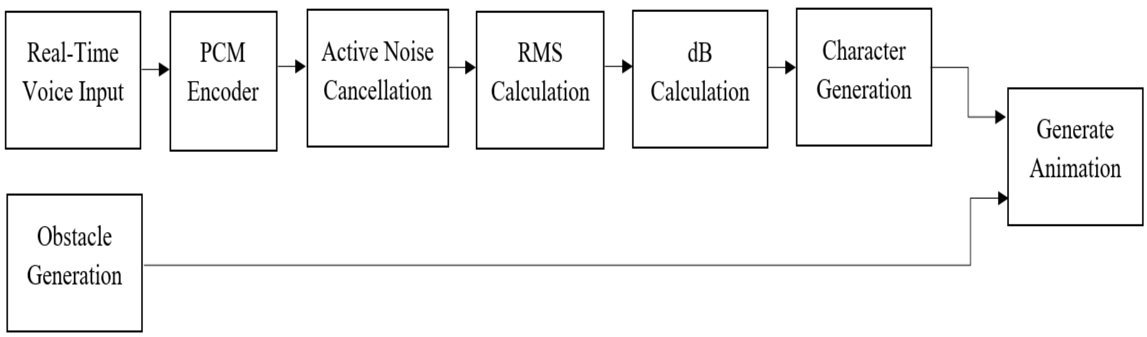


Figure 20: "Voice game" block diagram.

In the first chain there is a sequence of 6 blocks. This sequence of blocks is responsible for controlling the movement of the character. In block 1 a voice signal is taken in real time. In block 2, this signal is digitized at 16 bits at 44,100 Hz. The samples that are in even positions are used for optimization. In block 3 the noise is eliminated with the noise cancellation algorithm, thus eliminating sound intensities less than 40 dB. In the fourth block, the normalization of the divided PCM values for 32,768 V is performed and the RMS calculation is applied. The fifth block calculates the sound intensity of the signal in decibels. In the sixth block, from the calculated decibels, a character is generated that moves according to the increase or decrease in sound intensity (up and down).

The second chain of information corresponds to the generation of obstacles. This action gradually generates obstacles that move from right to left. The obstacles have been calibrated so that the sound intensity level required for the character to overcome the obstacles is within the range of 50 dB to 70 dB and so that the voice signal must be maintained for 1 second.

Once the two information chains have been generated, they are joined to create a game in which the objective is for the character (controlled with the intensity of the voice) to fly over the obstacles that move from right to left. What the program does is that the person who is training modulates her intensity of voice and maintains that intensity for a period. In the graphical interface, the implementation of various animated characters can be observed, as can be seen in the drop-down menu in Figure 21. The different characters are in Annex 3.

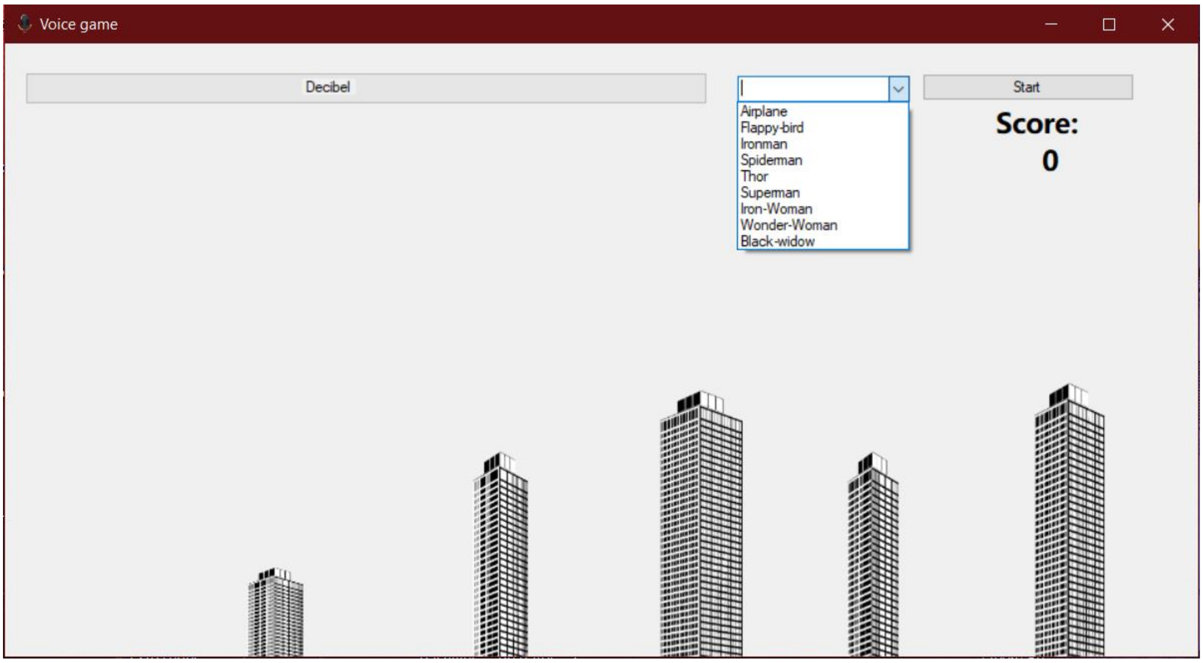
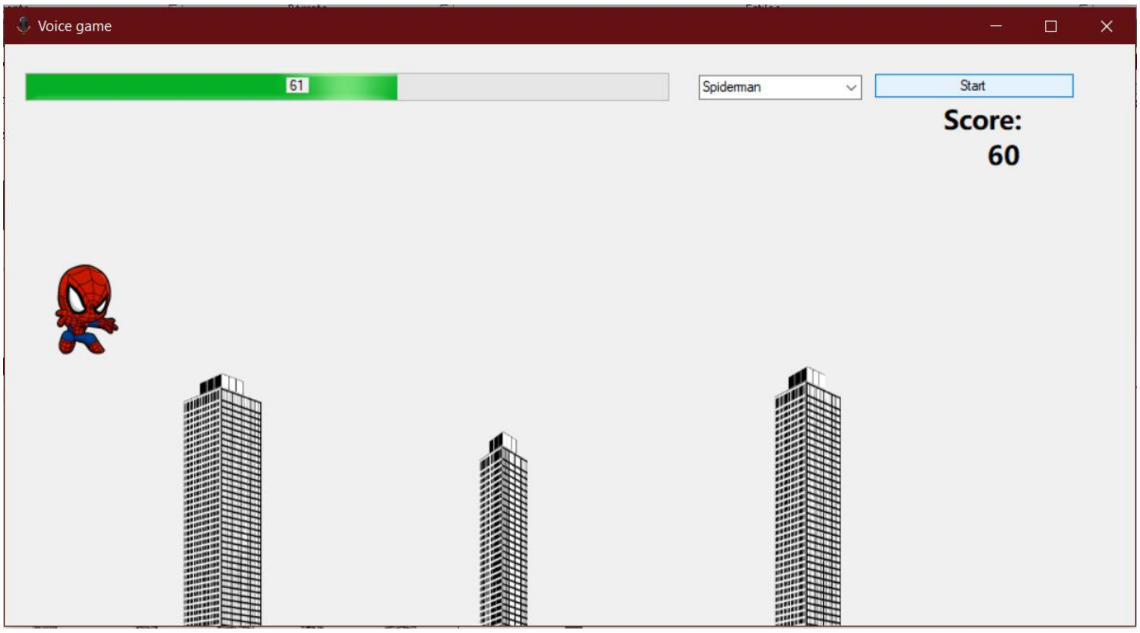


Figure 21: " Voice game" graphical user interface with character menu displayed.

In the graphical interface, there is also a counter that indicates the distance traveled during the game. Figure 22 shows the game in operation.

Figure 22: "Voice game" with Spiderman character during game performance.

The tools commonly used by speech therapists such as posters and drawings are effective when rehabilitating the voice of a child with hearing impairment. Despite this, these tools are boring and outdated, making rehabilitation even more difficult. It is difficult to capture the attention of children, but new technologies have the potential to create new entertaining rehabilitation methods.



Video games with a rehabilitation focus are a good alternative for the rehabilitation of infants. Alves (2015) conducted a study of the impact that games have on rehabilitation sessions in children with disorders of the upper and lower extremities, he found that both patients, parents and therapists increased their taste for attending rehabilitation sessions. It was found that play capitalized on the child's motivation to be successful at play and that children did not perceive therapy as a nuisance but rather as entertaining.

As mentioned in the introduction, Globus 3 is one of the most recognized programs for voice training in hearing impaired people. This program contains different modalities, one of them is "Balloon Trip"; This application is a game that allows voice intensity training. The game consists of a hot air balloon that is controlled with the intensity of the voice. As a first step in the application, the obstacles that the balloon must overcome must be created; For this, a healthy person or a speech therapist must define the position in which the obstacles are placed. The obstacles that are generated are of constant height; that is, the height of the obstacles does not vary depending on the sound intensity used.

In the same way, the balloon rises to a constant height when the sound intensity is higher than that predefined by the program (Lagares, 2005). On the other hand, the developed game "Voice game" consists of a wide range of fictional characters, which are controlled with the intensity of the voice used. The decibels of the voice are responsible for controlling the flight height of the character. Additionally, obstacles are generated automatically and these obstacles must always be overcome with sound intensities greater than 50 dB.

6 Conclusions and Recommendations

6.1 Conclusions

Statistical data indicate that there is a significant number of people with hearing impairment locally and globally. People with hearing disabilities are unaware of the negative impact of hearing loss on voice and speech production. Associated with this, there is a lack of human and technological resources focused on the treatment of the voice in people with hearing disabilities because usually the treatments in these people are focused on solving the lack of hearing.

In this thesis, a functional computer program has been developed aimed at the training and rehabilitation of the voice in people with hearing disabilities. The program has the necessary tools to improve verbal communication in these people and to make the monitoring of the evolution of treatment. The program is made up of four applications and one game.

The four applications: “Decibel meter”, “Voice plotter over time”, “Letter coach”, “Voice comparison”, and “Voice game”, are useful for voice training in hearing impaired people. The applications focus on training the intensity, pitch, and duration of the voice. The different applications have been calibrated to work under the parameters in which a normal conversation takes place. Also, they have a graphical user interface that is friendly to people with hearing disabilities since they are designed to generate different visual alerts in response to the voice signal of the person who is training.

A game called "Voice game" has been created that is useful for training the intensity and duration of the voice. This game is aimed at boys and girls with hearing disabilities since it has a series of animated characters. This game has been designed so that children with hearing disabilities do not see rehabilitation as boring and frustrating, what this game seeks is that the training is fun and satisfying.

The application is useful for voice training in people with hearing disabilities with or without the help of a speech therapist. The application can be used as a teaching tool during a rehabilitation session with a therapist. It can also be used independently without the help of a professional since the application has the necessary tools to be used independently.

6.2 Recommendations

It is recommended that the place where the speech rehabilitation program is to be used for people with hearing disabilities be quiet or that the noise does not exceed 40 dB. This is because excessive noise can generate wrong values and responses. An advantage of working in a quiet place is that people who are trained have a greater concentration.

Since the database of the correct pronunciation of the letters has been made by an adult and because children and adults have different tones of voice, it is recommended to generate new audio sequences with tones of voice similar to that of the person. Also, the therapist can introduce the sounds in different languages. People will train with the help of the sound recorder for the correct operation of the “Voice comparison” application.

To use the program, it is recommended that the computer microphone be placed in front of the person who is going to train with 50 centimeters of separation between the microphone and the person. This to obtain good results and because in the programming it was done in this way.

7 References

Alves, G. G. (2015). *Move With Me : Serious Game for Rehabilitation of Children with Disorders of the Upper and Lower Limbs* [Instituto Superior Técnico, Lisboa-Portugal]. https://fenix.tecnico.ulisboa.pt/downloadFile/563345090414239/55906_GabrielAlves_ExtendedAbstract.pdf

Andrade, E., Azamar, F., & Rodríguez, M. (2015). *Development of software for processing audio signals and multimedia enhancement* [Instituto Politécnico Nacional]. https://tesis.ipn.mx/jspui/bitstream/123456789/21925/1/Tesis_v2.6.pdf

Applegate, E. (2011). Auditory Sense. In *The anatomy and physiology learning system* (Vol. 86, Issue 7, p. 386). Saunders/Elsevier. https://books.google.com.ec/books?id=c8nsAwAAQBAJ&pg=PA210&dq=anatomy+ear+system&hl=es-419&sa=X&ved=2ahUKEwjrwIe49YvrAhWIl-AKHUmYAR4Q6AEwAnoECAkQAg#v=onepage&q&f=false

Bansal, M. (2013). Anatomy and Physiology of Ear. In *Diseases of Ear, Nose and Throat* (pp. 3–3). Jaypee Brothers Medical Publishers (P) Ltd. https://doi.org/10.5005/jp/books/11788_1

Burg, J., Romney, J., & Schwartz, E. (2014). Decibels. In *Digital Sound & Music Concepts, Applications, and Science*. http://digitalsoundandmusic.com/chapters/ch4/

Chasin, M. (2007). Decibel (Loudness) Comparison. *Centre for Human Performance & Health, Ontario, Canada.*, *l*, 95–97.

Coelho, A. C., Medved, D. M., & Brasolotto, A. G. (2015). Hearing Loss and the Voice. In *Update On Hearing Loss*. InTech. https://doi.org/10.5772/61217

CONADIS. (2020). *National Agenda for Equality of Disabilities 2017-2021*. Consejo Nacional Para La Igualdad de Discapacidades. https://www.consejodiscapacidades.gob.ec/estadisticas-de-discapacidad/

Faddis, B. (2008). Structural and Functional Anatomy of the Outer and Middle Ear. In W. Clark & K. Ohlemiller (Eds.), *Anatomy and Physiology of Hearing for Audiologists* (pp. 93–102). Thomson Delmar Learning.

Forouzan, B. (2013). *Data Communications and Networking* (5th ed.). Mc Graw Hill.

https://www.mheducation.com/highered/product/data-communications-networking-forouzan/M9780073376226.html

Hansen, C. (2015). *Fundamentals of Acoustics*. University of Adelaide. https://www.who.int/occupational_health/publications/noise1.pdf

Harden, S. (2018). *FftSharp: A Collection of FFT Tools for .NET*. GitHub. https://github.com/swharden/FftSharp

Harden, S. (2019). *ScottPlot: Interactive Plotting Library for .NET*. GitHub. https://github.com/swharden/ScottPlot

Hawkins, J. E. (2018). Human Ear - Cochlear Nerve and Central Auditory Pathways. *Encyclopedia Britannica*. https://www.britannica.com/science/ear/Cochlear-nerve-and-central-auditory-pathways

Heath, M. (2019). *NAudio: Audio and MIDI library for .NET*. GitHub. https://github.com/naudio/NAudio

IBM. (1991). Waveform Audio File Format (WAVE). In *Multimedia Programming Interface and Data Specifications 1.0*. Microsoft Corp. http://www-mmsp.ece.mcgill.ca/Documents/AudioFormats/WAVE/Docs/riffmci.pdf

IBM. (1997). *SpeechViewer III Support Information* (pp. 1–9). ftp://service.boulder.ibm.com/sns/spv3/spv3supt.htm

Lagares, J. (2005). Globus 3. In *Projecte Fressa*. https://www.uv.es/bellochc/logopedia/globus3.pdf

Lagares, J. (2020). *Educational Resources of Jordi Lagares Roset*. Toti PM. http://www.xtec.cat/~jlagares/

Ling, S., Sanny, J., & Moebs, B. (2020, March 26). *Sound Intensity*. Physics LibreTexts. https://phys.libretexts.org/Bookshelves/University_Physics/Book%3A_University_Physics_(OpenStax)/Map%3A_University_Physics_I_-_Mechanics%2C_Sound%2C_Oscillations%2C_and_Waves_(OpenStax)/17%3A_Sound/17.04%3A_Sound_Intensity

Microsoft. (2017). *C# Language Specification*. Microsoft Docs. https://docs.microsoft.com/es-es/dotnet/csharp/language-reference/language-specification/introduction

Microsoft. (2019). *Visual Studio IDE-Code Editor*. Visual Studio. https://visualstudio.microsoft.com/es/

Ministerio de Salud Publica. (2019). *Presidency of the Republic of Ecuador »Transparency*. https://www.presidencia.gob.ec/los-centros-de-salud-pueden-resolver-un-80-de-la-patologias/

Muthukumaran, B. (2016). *Active Noise Control System Design and Implementation*. https://www.semanticscholar.org/paper/Online-Active-Noise-Control-System-Design-and-Muthukumaran-Jayakandhan/1f19a9fd97840b5bb7f98889c4b98159feb9c6e0#extracted

National Research Council (US). (2004). Basics of Sound, the Ear, and Hearing. *Hearing Loss: Determining Eligibility for Social Security Benefits*, *Dc*, 42–69. https://www.ncbi.nlm.nih.gov/books/NBK207834/

Neumann, K., Mürbe, D., & Wouters, J. (2016). *Voice Disorders in Case of Hearing Loss*. http://www.pevoc.org/pevoc12/abstracts/RT04.pdf

Newman, J. (2008). Sound. In *Physics of the Life Sciences* (pp. 281–284). Springer New York. http://link.springer.com/10.1007/978-0-387-77259-2_11

Oxenham, A. J. (2008). Pitch Perception and Auditory Stream Segregation: Implications for Hearing Loss and Cochlear Implants. *Trends in Amplification*, *12*(4), 316–331. https://doi.org/10.1177/1084713808325881

Oxford University. (2020). *Program library*. Encyclopedia.Com. https://www.encyclopedia.com/computing/dictionaries-thesauruses-pictures-and-press-releases/program-library

Palacios, S. (2015). *Therapy Station to Improve Verbal Communication of People Suffering from Deafness and Linguistic Disorders*. https://pdfs.semanticscholar.org/cec0/10ee3c0fd67d28c79d6a0f4bbb599f994a36.pdf

Perez, T. (2007). Technical Resources for Students with Hearing Disabilities. In *NNTT Special Education*.

Petale, M. (2019). *Fourier Transform: (Theory & Solved Examples)*. BLURB. https://books.google.com.ec/books?id=gGtuDwAAQBAJ&printsec=frontcover&dq=fourier+transform&hl=es&sa=X&ved=2ahUKEwjbyfOqq5TrAhXRxVkKHdx6

C9UQ6AEwAXoECAYQAg#v=onepage&q=fourier transform&f=false

Rodríguez, A. (2015). *Determination of Hearing Thresholds in the Spanish Population* [Universidad Autonoma de Madrid]. https://repositorio.uam.es/bitstream/handle/10486/667533/rodriguez_valiente_antonio.pdf?sequence=1

Smart Tools Co. (2015). *Sound Meter (v1.7) manual*.

Smith, S. (1999). Correlation. In *Digital Signal Processing* (Second Edi, pp. 136–139). California Technical Publishing. www.DSPguide.com

Srivastava, A. K. (2014). Noise: A Nauseating Nuisance for Human Life Abhishek Kumar Srivastava. In *ZENITH International Journal of Multidisciplinary Research* (Vol. 4, Issue 2). FEBRUARY. http://www.indianjournals.com/ijor.aspx?target=ijor:zijmr&volume=4&issue=2&article=024

Sueur, J. (2018). *What Is Sound?* (pp. 7–36). Springer, Cham. https://doi.org/10.1007/978-3-319-77647-7_2

Sundberg, J., Patel, S., Björkner, E., & Scherer, K. R. (2011). Interdependencies Among Voice Source Parameters in Emotional Speech. *IEEE Transactions on Affective Computing*, *2*(3), 162–174. https://doi.org/10.1109/T-AFFC.2011.14

Toastmasters International. (2011). *Your Speaking Voice*. https://www.toastmasters.org/~/media/B7D5C3F93FC3439589BCBF5DBF521132.ashx

Ubrig, M. T., Goffi-Gomez, M. V. S., Weber, R., Menezes, M. H. M., Nemr, N. K., Tsuji, D. H., & Tsuji, R. K. (2011). Voice Analysis of Postlingually Deaf Adults Pre- and Postcochlear Implantation. *Journal of Voice*, *25*(6), 692–699. https://doi.org/10.1016/j.jvoice.2010.07.001

Vaughan, G. (2020). *Deafness and Hearing Loss*. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss

WHO. (2016). *Grades of Hearing Impairment*. World Health Organization; World Health Organization. http://www.who.int/deafness/hearing_impairment_grades/en/

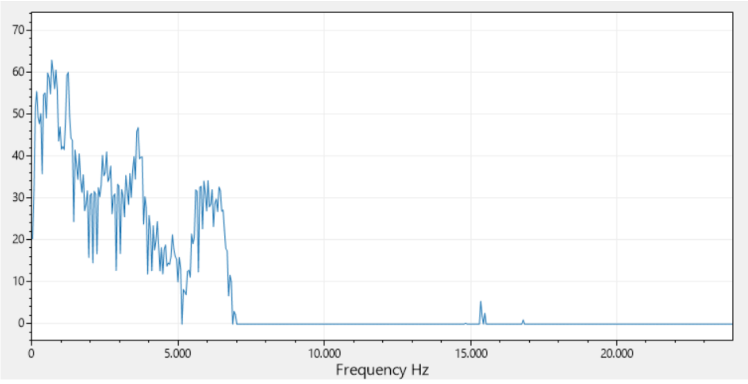
Wolfe, J., Garnier, M., & Smith, J. (2009). Vocal Tract Resonances in Speech, Singing, and Playing Musical Instruments. *HFSP Journal*, *3*(1), 6–23. https://doi.org/10.2976/1.2998482

Zhang, Z. (2016). Mechanics of Human Voice Production and Control. *The Journal of the Acoustical Society of America*, *140*(4), 2614–2635. https://doi.org/10.1121/1.4964509

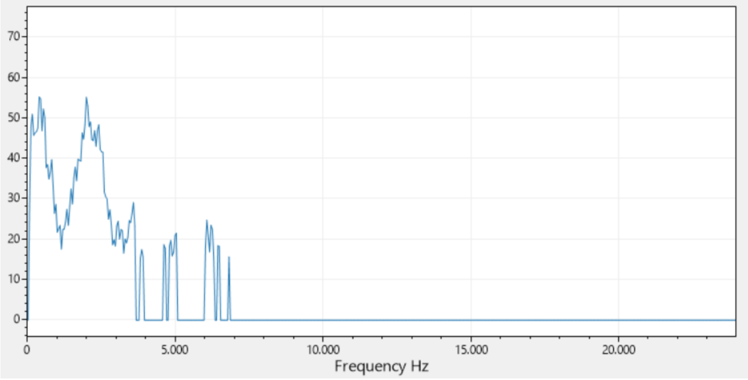
8 Annexes

Annex 1. Voice graph database

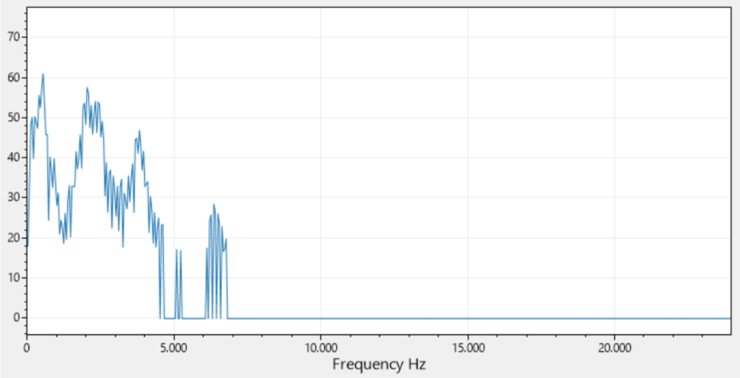
Graphic letter A



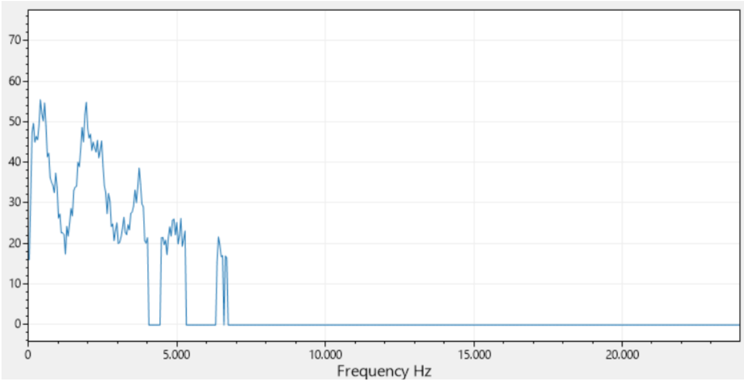
Graphic letter B



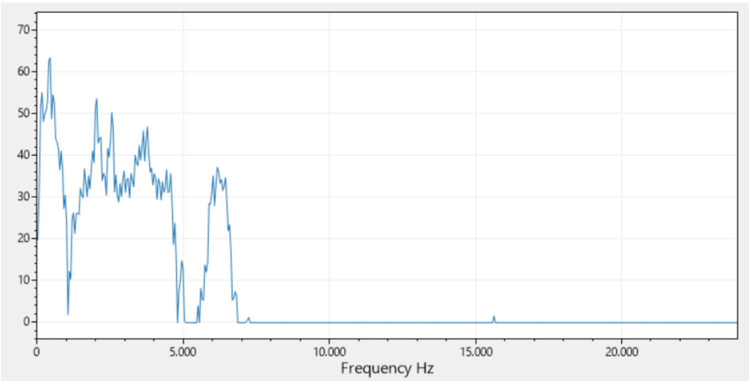
Graphic letter C



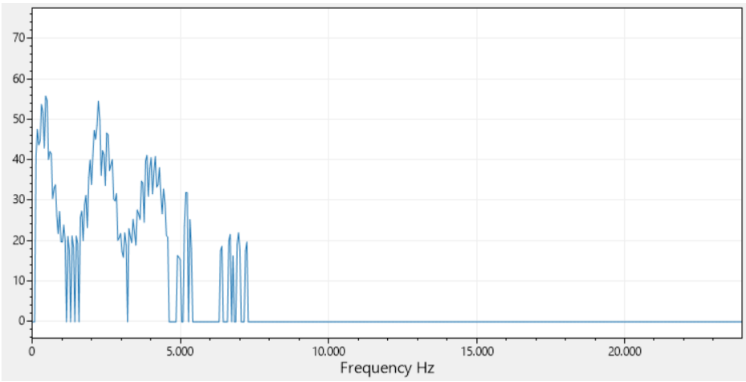
Graphic letter D



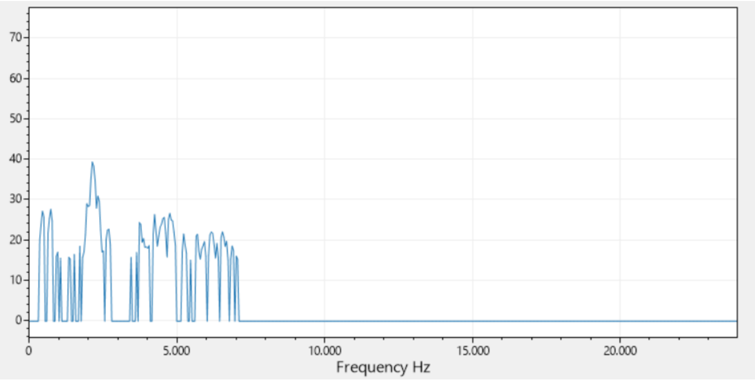
Graphic letter E



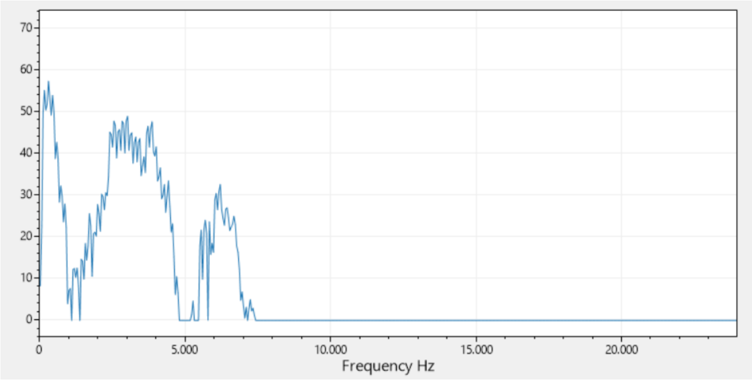
Graphic letter F



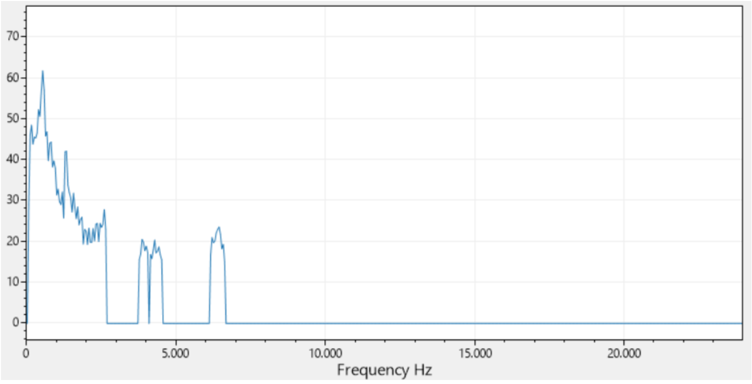
Graphic letter G



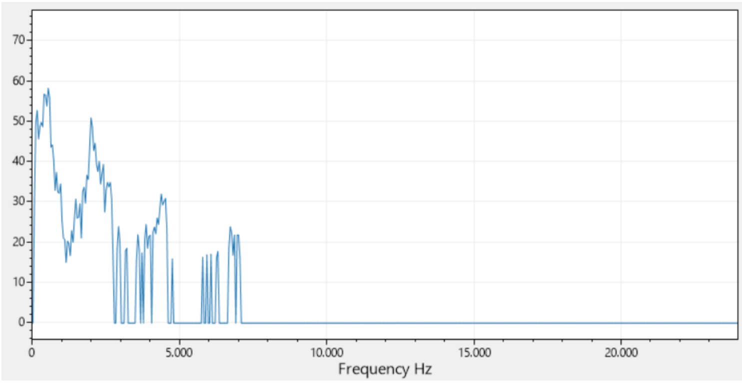
Graphic letter I



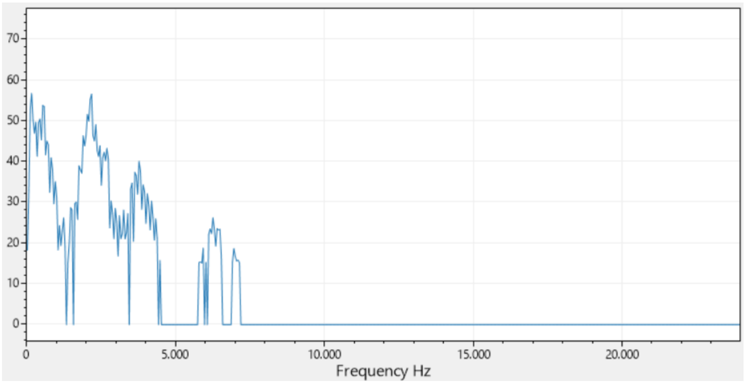
Graphic letter K



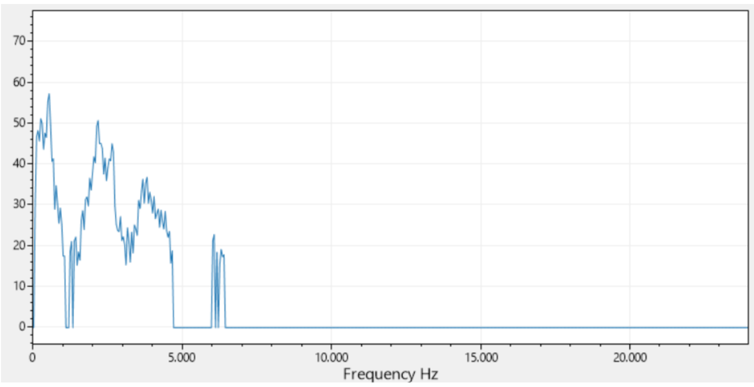
Graphic letter L



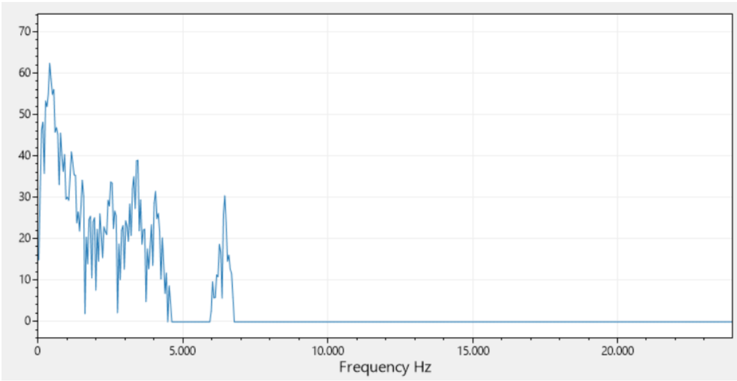
Graphic letter M



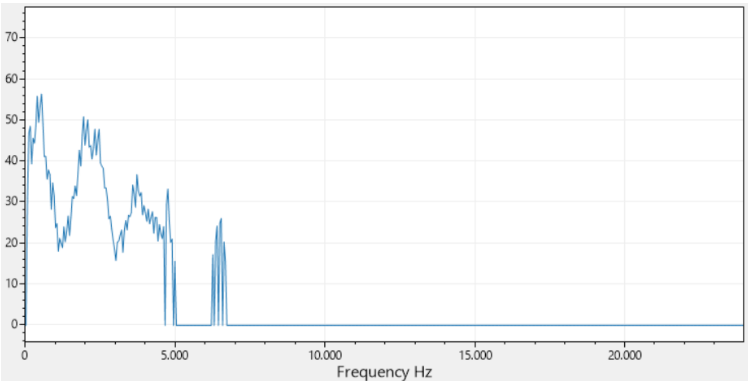
Graphic letter N



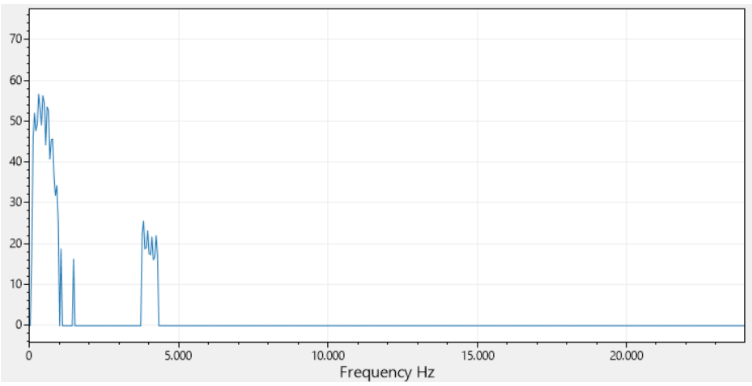
Graphic letter O



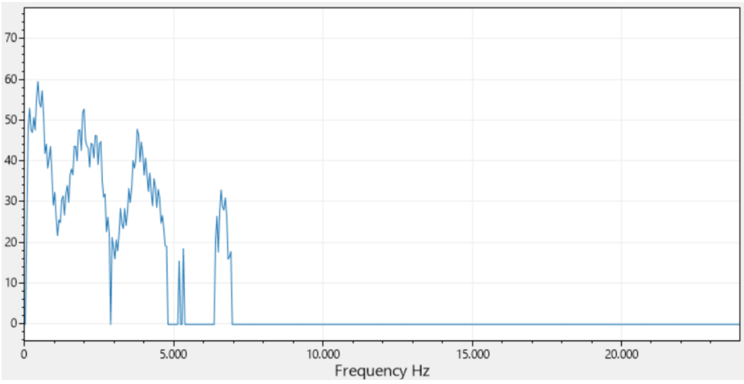
Graphic letter P



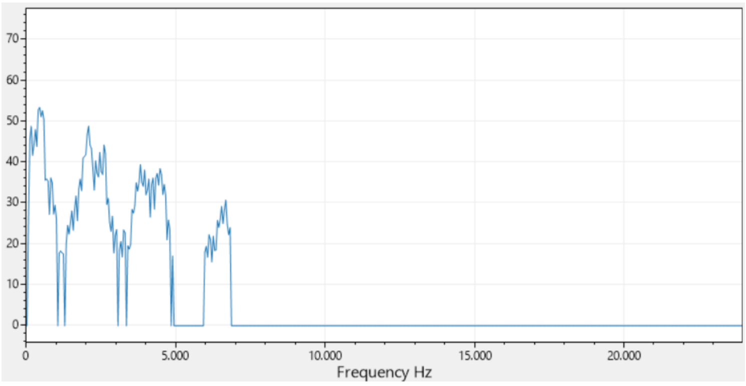
Graphic letter Q



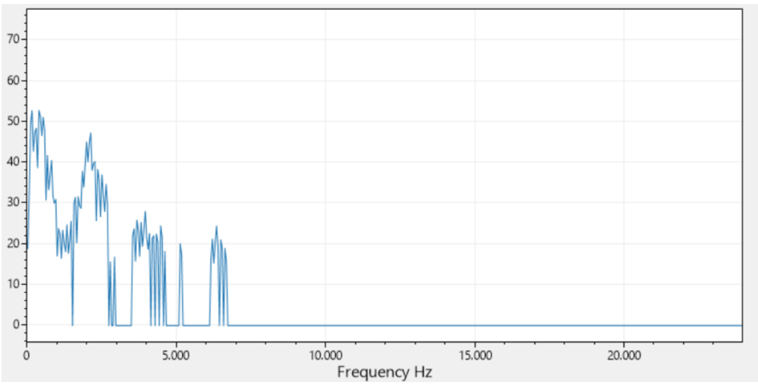
Graphic letter R



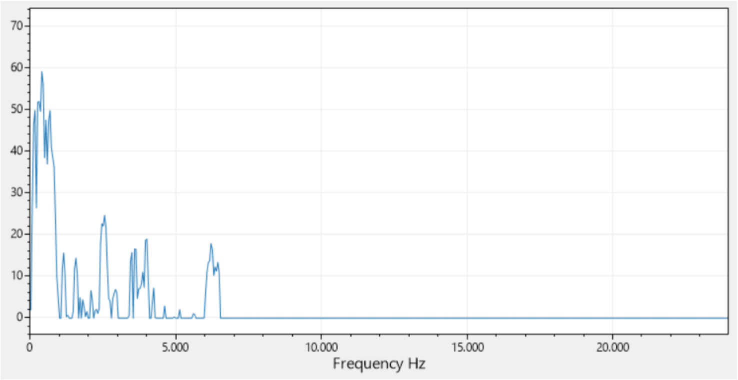
Graphic letter S



Graphic letter T

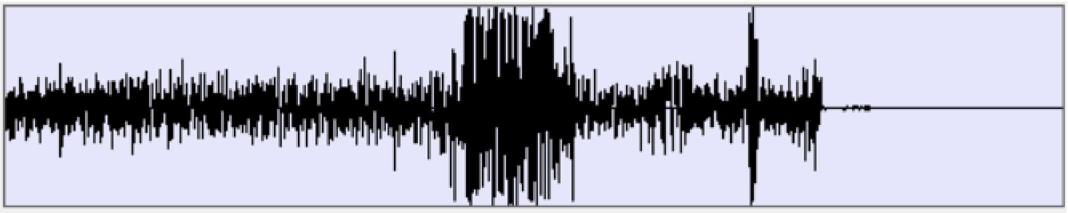


Graphic letter U

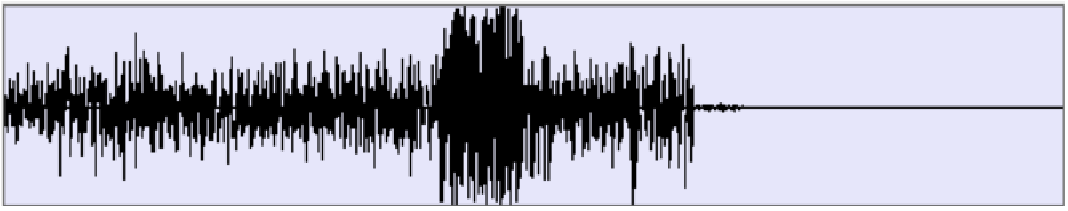


Annex 2. Letter Sound Database

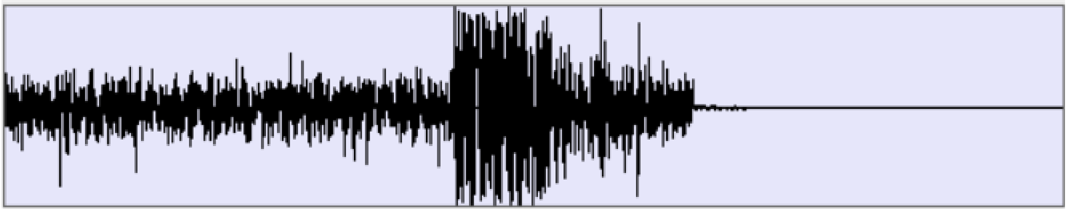
Graphic representation of the pronunciation of the letter A



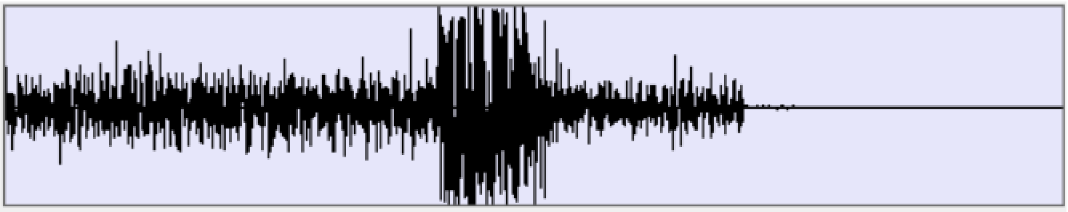
Graphic representation of the pronunciation of the letter E



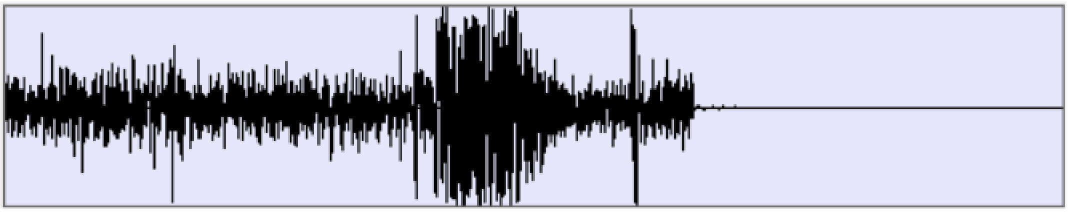
Graphic representation of the pronunciation of the letter I



Graphic representation of the pronunciation of the letter O



Graphic representation of the pronunciation of the letter U



Annex 3. Characters available in the game