

RAE

- **TIPO DE DOCUMENTO:** Trabajo de grado para optar por el título de Ingeniera de sonido.
- **TÍTULO:** DESARROLLO DE UN VIDEO JUEGO CON RECONOCIMIENTO DE VOZ PARA LA SELECCIÓN DE COLORES Y FIGURAS GEOMÉTRICAS.
- **AUTOR:** Luisa Fernanda Parra Espinel
- **LUGAR:** Bogotá D.C.
- **FECHA:** Febrero de 2015
- **PALABRAS CLAVE:** Videojuego, reconocimiento de voz, Unity 3D, Praat.
- **DESCRIPCIÓN DEL TRABAJO:** El objetivo principal de éste trabajo es el desarrollo de un videojuego controlado por comandos de voz, haciendo uso de un motor de programación cómo lo es Unity 3D. El videojuego se centra en una temática educativa dónde se pretende lograr un repaso de conocimientos de una manera didáctica, aprovechando la accesibilidad que tienen las personas a una variedad de dispositivos electrónicos los cuales se pueden utilizar como una herramienta pedagógica. Las personas actualmente deben estar preparados para el manejo de recursos tecnológicos así que se debe empezar a proponer una alfabetización digital la cuál le permita a los niños estar familiarizados con éstos dispositivos. Los videojuegos al generar a través de la práctica un constante reto o curiosidad, se convierte en un gran recurso para el refuerzo de conocimientos y habilidades.
- **LINEAS DE INVESTIGACIÓN:** Tecnologías actuales y sociedad -> Análisis y procesamiento de señales -> Acústica y audio
- **FUENTES CONSULTADAS:** C. V. Rojas Rusinque, L. Gómez Sánchez y M. I. Guarnizo Pinto, "los videojuegos como objetos desarrolladores de aprendizaje", referencia publicada en el 2014, Tesis Universidad de San Buenaventura, Facultad de Educación, Bogotá D.C, Colombia, 2014.
- **CONTENIDOS:** Dado que el continuo desarrollo de la tecnología ha obligado al ser humano a estar en constante evolución con ella. Es necesario que desde los primeros años de vida las personas tengan un contacto directo con dispositivos que les permitan desarrollar habilidades para el manejo de recursos tecnológicos. Los videojuegos son muy atractivos para los niños ya que son una fuente de distracción y diversión. Si se logra aprovechar esta motivación que sienten, los videojuegos pueden llegar a ser vistos como objetos desarrolladores de aprendizaje. Por otra parte la interacción que tiene el usuario con el dispositivo también juega una parte importante dentro del interés que puede generar un videojuego, el reconocimiento de voz aún no es tan usado pero es una herramienta que llama la atención por el hecho de que el usuario puede tener una interacción más directa con el dispositivo que esté usando.
- **METODOLOGÍA:** Desarrollar un videojuego manejado a través de comandos de voz por medio del software Unity 3D para sistemas operativos iOS y Windows. El proyecto es una investigación empírico-analítica.
- **CONCLUSIONES:** Se logró un videojuego manejado a través de comandos de voz que cumpliera con los objetivos plantados, se requiere la optimización del reconocimiento generando espacios más adecuados para la utilización del juego.

**DESARROLLO DE UN VIDEO JUEGO CON RECONOCIMIENTO DE VOZ
PARA LA SELECCIÓN DE COLORES Y FIGURAS GEOMÉTRICAS.**

LUISA FERNANDA PARRA ESPINEL

**UNIVERSIDAD DE SAN BUENAVENTRA
FACULTAD DE INGENIERIA
INGENIERIA DE SONIDO
BOGOTÁ
2014**

**DESARROLLO DE UN VIDEO JUEGO CON RECONOCIMIENTO DE VOZ
PARA LA IDENTIFICACIÓN DE COLORES Y FIGURAS GEOMETRICAS.**

LUISA FERNANDA PARRA ESPINEL

Proyecto de grado.

**Tutor:
Ing. Miguel Olivares**

**Asesora:
Ing. Lorena Aldana**

**UNIVERSIDAD DE SAN BUENAVENTRA
FACULTAD DE INGENIERIA
INGENERIA DE SONIDO
BOGOTÁ
2014**

Nota de aceptación

Firma del presidente del jurado.

Firma del jurado.

Firma del jurado.

Dedicado a mis padres por su apoyo incondicional, por creer y darme las herramientas para hacer esto posible.

Agradezco a cada persona que durante este tiempo fue participe del crecimiento y desarrollo del proyecto, a mi tutor Miguel Olivares y especialmente a la ing. Lorena Aldana por su apoyo en la construcción formal del proyecto.

TABLA DE CONTENIDO

INTRODUCCIÓN	14
1. ANTECEDENTES	15
1. PLANTEAMIENTO DEL PROBLEMA	23
1.1. DESCRIPCIÓN DEL PROBLEMA	23
1.2. PREGUNTA PROBLEMA:	23
2. JUSTIFICACIÓN	24
3. OBJETIVOS	25
3.1. Objetivo General:	25
3.2. Objetivos específicos:	25
4. ALCANCES Y LIMITACIONES	25
4.1. Alcance:	25
4.2. Limitaciones:	25
5. MARCO REFERENCIAL	26
5.1. MARCO TEÓRICO	26
5.1.1. PRAAT:	26
5.1.2. RECONOCIMIENTO DE VOZ	32
5.1.3. UNITY 3D:	33
5.1.4. DISEÑO DE VIDEO JUEGOS EDUCATIVOS	33
5.2. MARCO CONCEPTUAL	34
5.2.1. FORMANTES	34
5.2.2. PRINCIPIO Y FIN DE PALABRA	36
5.2.3. FFT	36
5.2.4. FILTRO LPC (LINEAR PREDICTIVE CODING)	37
5.2.5. DENSIDAD ENERGÉTICA	37
5.2.6. TRACTO VOCAL	37

5.2.7.	ADQUISICIÓN DE DATOS	38
5.2.8.	REDES NEURONALES	38
5.2.9.	SENSIBILIDAD EN MICROFONOS	39
5.2.10.	DIAGRAMA POLAR	39
5.2.11.	RANGO DINÁMICO	39
5.2.12.	RESPUESTA EN FRECUENCIA.....	39
5.2.13.	JUGABILIDAD	39
6.	MARCO METODOLÓGICO	40
6.1.	Tipo de investigación	40
6.2.	Línea de investigación	40
6.3.	Hipótesis.....	40
6.4.	Variables.....	41
6.4.1.	Variables Dependientes:.....	41
6.4.2.	Variable independiente:.....	41
6.5.	Metodología del diseño	41
7.	ANÁLISIS INGENIERIL.....	41
7.1.	Diseño del videojuego.....	41
7.1.1.	Atmósfera del Videojuego	47
7.1.2.	Jugabilidad.....	48
7.2.	Parámetros de estudio de las palabras a reconocer	50
7.2.1.	Análisis de formantes de las palabras correspondientes a los Colores:	59
7.2.1.1.	Análisis de la palabra Amarillo.....	60
7.2.1.2.	Análisis de la palabra Azul.....	62
7.2.1.3.	Análisis de la palabra Rojo.	64
7.2.2.	Análisis de formantes de las palabras correspondientes a las Figuras Geométricas:.....	66

7.2.2.1.	Análisis de la palabra Círculo.....	66
7.2.2.2.	Análisis de la palabra Cuadrado.	68
7.2.2.3.	Análisis de la palabra Triángulo.....	70
7.2.3.	Cuadros comparativos entre diferentes capturas.....	71
7.3.	Desarrollo del videojuego	75
7.3.1.	Requerimientos:.....	75
7.3.2.	Funcionamiento del videojuego.....	75
7.3.3.	PROGRAMACION EN UNITY:.....	81
8.	Presentación y análisis de resultados.....	87
9.	CONCLUSIONES.....	90
10.	RECOMENDACIONES.....	92
11.	REFERENCIAS	93
12.	BIBLIOGRAFÍA	96

LISTADO DE TABLAS

Tabla 1. Datos patron polar.....	56
Tabla 2. Datos respuesta en frecuencia.....	57
Tabla 3. Vocales: a,i,o.....	60
Tabla 4. Vocales: a, u.....	62
Tabla 5. Vocales: o.....	64
Tabla 6. Vocales: i, u, o.	66
Tabla 7. Vocales: u, a, o	68
Tabla 8. Vocales: i, a, u, o.....	70
Tabla 9. Tabla de comparación entre colores	73
Tabla 10. Tabla de comparación entre figuras	74
Tabla 11. Requerimientos técnicos videojuego "Pick Up"	75
Tabla 12. Resultados niño 1, 8 años	88
Tabla 13. Resultados niño 2, 7 años	89
Tabla 14. Resultados niño 3, 5 años	89

LISTADO DE FIGURAS

Figura 1. Ventana para la opción de grabar.....	26
Figura 2. Ventana SoundRecorder	26
Figura 3. Ventana para leer un archivo	26
Figura 4. Ventana para crear un sonido desde una fórmula.....	27
Figura 5. Lista de Objetos	27
Figura 6. Opciones de guardar	27
Figura 7. Modo de ingreso a la opción de Edit.....	28
Figura 8. Espectrograma.....	28
Figura 9. Vista de varios parámetro ofrecidos por praat.....	29
Figura 10. Propiedades del espectro	29
Figura 11. Propiedades Frecuencia fundamental	30
Figura 12. Propiedades de intensidad	30
Figura 13. Propiedades de los formantes.....	31
Figura 14. Spectrum	31
Figura 15. LPC.....	32
Figura 16. Diagrama representativo de un sistema de reconocimiento de voz.....	33
Figura 17. Formantes.....	35
Figura 18. FFT	36
Figura 19. Formas del tracto vocal para las vocales i, a y u	38
Figura 20. Metodología del diseño	41
Figura 21 Ejemplo 1 primer diseño.	43
Figura 22. Ejemplo 2 primer diseño	43
Figura 23. Imagen figuras geométricas	44
Figura 24. Personajes del juego.	45
Figura 25. Diseño de pick up	46
Figura 26. Colores.....	47
Figura 27. Figuras	47
Figura 28. Atmósferas del videojuego	48
Figura 29. Colores a identificar.....	50
Figura 30. Figuras a identificar	51

Figura 31. Diagrama flujo de señal.....	51
Figura 32. Grafica de figura 8 y respuesta en frecuencia.....	52
Figura 33. Diagrama de flujo con micrófono Beta 87c	52
Figura 34. Patrón polar cardioide	53
Figura 35. Respuesta en frecuencia micrófono cardioide	53
Figura 36. Diagrama de flujo sin micrófono externo	54
Figura 37. Medición características micrófono Mac BookPro	55
Figura 38. Diagrama polar micrófono Mac BookPro.....	57
Figura 39. Respuesta en frecuencia micrófono Mac BookPro.....	58
Figura 40. Grabación en el Salón de Post- producción. USB con un Micrófono Beta 87c. Niño de 8 años - Amarillo	60
Figura 41. Grabación en el Salón de Post- producción. USB con un micrófono Beta 87c. Niño de 8 años - Amarillo.....	60
Figura 42. Grabación en el Estudio con un micrófono AKG c414. Mujer de 25 años. - Amarillo	60
Figura 43. Espectro- Formantes. Palabra amarillo.....	61
Figura 44. Grabación en el Salón de Post- producción. USB con un Micrófono Beta 87c. Niño de 8 años - Azul.....	62
Figura 45. Grabación en el Estudio con un Micrófono AKG c414. Mujer de 25 años. - Azul.....	62
Figura 46. Grabación en el Salón de Post- producción. USB con un Micrófono Beta 87c. Niño de 8 años - Azul.....	62
Figura 47. Espectro- Formantes. Palabra Azul.....	63
Figura 48. Grabación en el Salón de Post- producción. USB con un Micrófono Beta 87c. Niño de 8 años - Rojo	64
Figura 49. Grabación en el Salón de Post- producción. USB con un Micrófono Beta 87c. Niño de 8 años - Rojo	64
Figura 50. Grabación en el Estudio con un Micrófono AKG c414. Mujer de 25 años. - Rojo	64
Figura 51. Espectro- Formantes. Palabra Rojo.	65
Figura 52. Grabación en el Salón de Post- producción. USB con un Micrófono Beta 87c. Niño de 8 años – Circulo.....	66

Figura 53. Grabación en el Salón de Post- producción. USB con un Micrófono Beta 87c. Niño de 8 años – Circulo.	66
Figura 54. Grabación en el Estudio con un Micrófono AKG c414. Mujer de 25 años. – Circulo.	66
Figura 55. Espectro- Formantes. Palabra Círculo.	67
Figura 56. Grabación en el Salón de Post- producción. USB con un Micrófono Beta 87c. Niño de 8 años – Cuadrado.	68
Figura 57. Grabación en el Salón de Post- producción. USB con un Micrófono Beta 87c. Niño de 8 años – Cuadrado.	68
Figura 58. Grabación en el Estudio con un Micrófono AKG c414. Mujer de 25 años. – Cuadrado.	68
Figura 59. Espectro- Formantes. Palabra Cuadrado.	69
Figura 60. Grabación en el Salón de Post- producción. USB con un Micrófono Beta 87c. Niño de 8 años – Triángulo.	70
Figura 61. Grabación en el Salón de Post- producción. USB con un Micrófono Beta 87c. Niño de 8 años – Triángulo.	70
Figura 62. Grabación en el Estudio con un Micrófono AKG c414. Mujer de 25 años. – Triángulo.	70
Figura 63. Espectro- Formantes. Palabra Triangulo.	71
Figura 64. Imagen del botón para seleccionar un micrófono.	76
Figura 65. Imagen del botón para el ruido de fondo.	76
Figura 66. Reconocimiento de voz.	76
Figura 67. Nivel Figuras Geométricas.	77
Figura 68. Nivel Colores.	78
Figura 69. Nivel 3	79
Figura 70. Funcionamiento del videojuego	80
Figura 71. Niños participantes del proyecto.	88

RESUMEN

En éste proyecto se desarrolla un videojuego educativo manejado a través de comandos de voz, el cuál pretende ayudar al refuerzo de conocimientos por medio de atención, reto, repetición, entre otras características que funcionan como aliadas en el desarrollo del aprendizaje. También generando una manera de interacción no tradicional como lo es el reconocimiento de voz, para generarle al usuario mayor interés hacia el videojuego.

ABSTRACT

In this project an educational video game handled through voice commands, which aims to help strengthen knowledge through attention, challenge, repetition, among other features that work as partners in the development of learning skills. Also generating a nontraditional way of interaction such as the speech recognition to carry potential users greater interest in the game.

INTRODUCCIÓN

La constante necesidad del ser humano por crear nuevas tecnologías conlleva al desarrollo de diferentes sistemas y herramientas para las actividades que realiza a diario. En la interacción entre hombre máquina ya existen modos de comunicación pero actualmente gracias a ese afán de innovar se han generado más estrategias como lo es el reconocimiento de voz, que a partir del procesamiento de señales permite conocer características del interlocutor tanto fisiológicas como del comportamiento del mismo (hábitos lingüísticos, entonación en frases, entre otros) para así lograr tener una interacción por medio del habla con el dispositivo.

Los videojuegos son una forma de interacción muy llamativa ya que les permite a las personas salir de su rutina y entretenerse. Algunos motores de programación para videojuegos permiten la implementación de reconocimiento de voz haciendo de este algo más interactivo y de la experiencia algo más natural para el usuario.

Teniendo en cuenta las múltiples formas de programación de un videojuego se abre un universo de posibilidades en donde la elección está abierta a las necesidades específicas del programador para el diseño y el desarrollo del juego. El diseño de un videojuego comprende varias fases, desde su concepto inicial hasta el producto ya finalizado, para lograr esto se debe llevar a cabo un proceso, el cuál comienza por la idea donde se define la jugabilidad (lo que lo hará divertido y aplicable), seguido de la definición de la historia, el arte conceptual, el sonido, la mecánica y el desarrollo del videojuego donde se describe en donde se implementara, la plataforma en que se va a programar y el lenguaje de programación a usar. Ya para dar por terminado el diseño y desarrollo del video juego se deben hacer unas pruebas con las cuáles se comprueba su jugabilidad y su correcta aplicación.

En este proyecto se pretende desarrollar un videojuego el cuál genere una experiencia más transparente e interactiva entre el usuario y la máquina cuya finalidad será la selección acertada de figuras geométricas y colores por medio de reconocimiento de voz. Dentro del juego se proponen distintos niveles de dificultad para el usuario.

1. ANTECEDENTES



ESTUDIOS HECHOS EN COLOMBIA

- Andrea Lorena Aldana blanco., Julián Ricardo Piñeros Castaño., En el año 2009 para su tesis desarrollaron un algoritmo de reconocimiento de voz con la herramienta MATLAB, la cuál reconoce nueve palabras y a cada una la asocia con una imagen que previamente se predetermino.

En el trabajo dan a conocer los métodos que exploraron dándole una caracterización y entretenimiento al usuario. Finalmente muestran el método con mayor porcentaje de reconocimiento para la aplicación. Universidad de San Buenaventura, Bogotá, Colombia [21].

- Okasa Studios, que son un grupo de profesionales en diseño y arte en el 2012 presentaron una investigación sobre como los videojuegos podrían ser una herramienta educativa y cultural haciéndolos generadores de conocimiento. Presentan un amplio estudio sobre la apropiación cultural que tienen este tipo de juegos en un contexto cultural. Llegan a unir lo divertido con lo educativo, mostrando una discusión sobre los resultados que reunieron por la experiencia a la cual fue expuesto cada usuario. Bogotá, Colombia [24].
- Andrés López Astudillo en el 2010 tomo como ejemplo el videojuego “Rise of the Nations” en su trabajo de tesis, para conocer que usos en el proceso de aprendizaje tenia este videojuego. Llegando a la conclusión que el juego

presenta muchas posibilidades al usarse como herramienta educativa, ya que su aplicación y contexto están bien dirigidas a una población específica. Universidad Icesi, Bogotá, Colombia [25].

- José Orlando Montes de la Barrera., Helman Enrique Hernández Riaño., Jorge Mario López Pereira., Juan Ángel Chica Urzol., escribieron para la revista Educación en Ingeniería en el 2010 un artículo sobre el impacto de los juegos didácticos tienen sobre el aprendizaje de la ingeniería industrial. Para determinar si estas herramientas didácticas cumplían con su función, educar, emplearon un experimento. El estudio da una evidencia empírica relacionada con la actividad de diferentes metodologías de enseñanza y tiene en cuenta las variables que pueden afectar el rendimiento de los estudiantes. Universidad de Córdoba, Montería, Colombia [26].
- Jorge Mario Karam Rozo., En el año 2011 logró una publicación en la revista “Repertorio de Medicina y cirugía” en la ciudad de Bogotá, sobre el paradigma que se tiene por el aprendizaje autónomo y el aporte que tienen los videojuegos a esto. Se basa en que los videojuegos fortalecen el pensamiento creativo, esto lleva a que esta herramienta no sólo sea usada por ocio, sino como estrategia activadora de juicio de meta cognición. Aunque deja el campo abierto para la profundización del tema. Fundación Hospital Infantil de San José, Bogotá, Colombia [29].
- Juan David Dúran Castañeda., Néstor Raúl Cely Garzón., En su tesis presentada para 2011 para la Universidad de la Sabana realizaron un análisis a diferentes juegos con mayor éxito comercial en el país y en el continente sin importar el género. Querían conocer que proyección tenía al usuario, teniendo en cuenta todo el proceso desde la concepción del juego, su diseño, argumento, estilo visual hasta el momento de la interacción con las personas. Dada la importancia que han adquirido los videojuegos ya que se han convertido en la plataforma ideal para la difusión de mensajes positivos o negativos. Facultad de Comunicación, Universidad de la Sabana, Bogotá, Colombia [30].
- Claudia Victoria Rojas Rusinque., Ligda Gómez Sánchez., María Isabel Guarnizo Pinto., en el 2014 Realizaron su tesis en la cual valoraron los usos que pueden llegar a tener los videojuegos y qué apropiación académica pueden llegar a tener. Analizan ésta herramienta como medio de alfabetización y diseñan una ruta para ampliar la visión de los maestros y lograr que los videojuegos sean aplicables en las aulas. Universidad de San Buenaventura, Bogotá, Colombia [37].

ESTUDIOS INTERNACIONALES

- Hamid Sheikhzadeh., Li Deng. Escribieron en 1994 un artículo sobre el comienzo del análisis para el reconocimiento de voz basado en la utilización de filtros gaussianos variables en el tiempo. Con el fin de obtener un algoritmo que

le permita al usuario obtener un resultado normalizado, eliminando ruido y la sensibilidad que puede generar la variación de potencia al hablar. Departamento de ingeniería eléctrica y de computación, Universidad de Waterloo, Waterloo, ON, Canadá. [1].

- Michael M. Cohen., Dominic W. Massaro en 1994 escribieron para la Universidad de California sobre ¿Qué puede decir la síntesis visual del reconocimiento visual del habla? En el documento se discutieron en primer lugar, el cómo el uso de la voz sintética les podía servir para mejorar la percepción de la voz humana, en segundo lugar, el uso de la síntesis para instanciar modelos de producción, y tercero, el uso de estos últimos para guiar el reconocimiento automático del habla. Universidad de California, Santa Cruz [2].
- Enrique Morales., Licenciado en Sociología por la Universidad Pública de Navarra y doctorando investigador de la facultad de ciencias de la información de la Universidad Complutense de Madrid, publicó en 1995 un artículo el cuál trata sobre el uso de los videojuegos como recursos de aprendizaje en la educación primaria. Asegura que los videojuegos conectan a los niños directamente con la informática, sin embargo a nivel educativo no son muy utilizados por sus contenidos. En el trabajo se hacen propuestas para que los educadores dejen de ver a los videojuegos como algo estrictamente lúdico, y vean que también se pueden utilizar como cualquier otra herramienta didáctica para enseñar o repasar conocimientos. Universidad Complutense de Madrid [3].
- Sven Anderson., Diane Kewly-Port., En 1995 escribieron un artículo sobre la importancia que podría tener el uso del reconocimiento de voz en el entretenimiento. Aunque hay unas características de la voz a las que no es tan fácil acceder, ya que se necesitan herramientas muy específicas las cuales no hicieron parte de su investigación. Estudiaron las funciones de dos software de reconocimiento para identificar los pros y contras de cada uno. Universidad de Indiana [4].
- D.G. Jaimieson (1), L.Deng (2), M.Price (1), Vijay Parsa (1), J.Till (3), Hicieron una investigación en 1996 sobre cómo los codificadores de los procesos de reconocimientos de voz generan trastornos sobre la inteligibilidad del habla. Los esquemas de codificación de voz que se estaban desarrollando en ese momento se basaban en cubrir una demanda para telecomunicaciones, los cuáles eran aceptables para la época. Los autores en este artículo se enfocaron en las pérdidas que provocaban estos algoritmos y en cómo modificaban las voces originales. (1) Unidad de investigación de cuidados auditivos. Universidad del Oeste de Ontario, London, Ontario, Canadá, (2) Departamento de Ingeniería Eléctrica y de computación (3) Servicio de Patologías, Centro Medico VA, Long Beach, California [5].

- Suat Yelderner., Jack H. Rieser., en el 2000 iniciaron una investigación basados en el problema del ruido en la codificación de los algoritmos de voz. Se describe una nueva técnica la cuál esta basada en mantener una forma sinusoidal, con el fin de mejorar la calidad y la inteligibilidad de la voz procesada. En el entorno de la telecomunicación, las señales de voz se degradan dado la cantidad de ruido. Dado este inconveniente, proponen una nueva técnica que genera reducción de ruido basada en la extracción armónica y que se puede integrar a cualquier algoritmo de codificación de voz. Maryland. USA [6].
- Douglass O`Shaughnessy., Realizó un paper en el cuál examina cómo es la comunicación de las personas con las computadoras mediante el reconocimiento de voz. Analiza las teorías de codificación de voz y la simulación espectral que realiza el oído. Emplea en su propio algoritmo tanto la transformada de Fourier como también de forma auditiva la Skale Bark, para la simplificación de la representación espectral usa una decorrelación en coeficientes cepstrales. [7].

El análisis de voz se puede representar de una forma automática dado la forma de onda del discurso, primero hay una entrada, a partir de una base de datos generada por el entrenamiento previo que se le dio al sistema, la palabra luego será almacenada y comparada mediante un proceso de transcripción fonética. En cada palabra de entrenamiento se debe tener especial cuidado con la entonación que esté usando la persona ya que se tendrá que aproximar a la naturalidad del habla humana. INRS-Telecomunicaciones. Montreal, QC, Canadá.

- Carlos Miranda., Reyna Camal., José Cen Magaña., Cinhtia Gonzalez Segura., Sergio Gonzalez Segura., Michael García., Lizzie Narvéez Díaz. En el año 2004 presentaron para la Universidad Autónoma de Yucatán un juego electrónico (Gravedad). Estos estudiantes se preocuparon por una población en específico, los niños con problema de lenguaje dislalia. Presentaron un videojuego que se maneja por comandos de voz. Lo desarrollaron en un lenguaje C y los modelos acústicos los desarrollaron en un software de reconocimiento de voz HTK. Universidad Autónoma de Yucatán, Facultad de matemáticas, unidad de Tizimin, Tizimin, Yucatán, México [8].
- Sanja Grubesa., Tomislav Grubesa., Hrvoje Domitrovic., Escribieron su artículo en el 2005 sobre el reconocimiento del habla combinando FFT, la función de wavelet y redes neuronales. La función de wavelet se utiliza para obtener una aproximación y los detalles promedios del espectro del hablante para así tener las características dadas por las frecuencias. Estos datos son utilizados como entrada para las redes neuronales que es donde se toman decisiones, no sólo se reconoce la identidad de la persona, pero hay que tener en cuenta que ésta decisión puede ser proporcional a muchos factores. Facultad de ingeniería eléctrica y de computación, Universidad de Zagreb, Croacia [9].

- Zhihong Zeng., Jilin Tu., Brian Pianfetti., Ming Liu., Tong Zhang., Zhenqui Zhang., Thomas S. Huang and Stephen Levinson., En el 2005 hicieron un artículo sobre la interacción del hombre con la máquina. El trabajo habla sobre el desarrollo de un algoritmo en donde por medio de sensores de audio y de gestos detecta el estado anímico de la persona que lo esté usando. Llega a detectar hasta 11 estados cognitivos (emociones). Universidad de Illinois, Urbana, Champaign [10].
- José Luis Oropeza Rodríguez., En su artículo “Algoritmos y métodos para el reconocimiento de voz en español mediante sílabas” publicado el 15 de diciembre del 2006, planteó una alternativa diferente a lo que hasta ese momento se venía implementando para el reconocimiento de voz. En su investigación le da prioridad a la sílaba y a su labor en el español. Realizó una segmentación de elementos esenciales:
 1. La función de la energía total en un corto tiempo.
 2. La función de energía de altas frecuencias cepstrales.
 3. Un sistema basado en conocimiento

(Centro de investigación en computación IPN, México D.F.)
[11]

- Cinthia González., Carlos Miranda., Michael García., Sergio González., Analizaron en el año 2007 tres videojuegos con reconocimiento de voz, los cuáles eran usados para la rehabilitación de niños con problemas con lenguaje. Luego desarrollaron su propio software el cuál tenía la filosofía de ser gratuito, haciéndolo así asequible a las familias de bajos recursos. En este trabajo se muestran las pruebas y resultados dando como resultado que la interfaz es usable y motivadora para los niños en sus terapias de rehabilitación. Universidad Autónoma de Yucatán, Facultad de Matemáticas, Unidad Tizimín. Tizimín, Yuc. México. [12].
- Sonia Martín Moreno., Francisco Javier Remesal Escalero., Laura Rivera Rodríguez., Como proyecto en el año 2007 realizaron un videojuego para el aprendizaje de SQL. Aprovecharon la posibilidad de usar los videojuegos para aprender de una forma amena, ese fue uno de los motivos por los cuales realizaron ese proyecto. Universidad Complutense de Madrid Facultad de Informática [13].
- Juan E. Jiménez., Estefanía Rojas., En el 2008 hicieron una investigación sobre los efectos del videojuego *Tradislexia* en los niños disléxicos, para su conciencia fonológica y el reconocimiento de palabra. Analizaron cómo el entretenimiento de multimedia favorecía a los procesos de rehabilitación y que tanto rendimiento ofrece. Facultad de Psicología Universidad de La Laguna, Islas Canarias, España [14].

- Reyna del Rosario Camal Uc., En su proyecto de grado presenta módulos de reconocimiento de voz, empleándolos en la interacción de la persona con un videojuego. El juego fue utilizado en la rehabilitación de niños con problema de lenguaje dislalia. En el trabajo se muestra un análisis comparativo para los tres grupos de hablantes a los que se les realizó la prueba: adultos, niños sin problemas de lenguaje y niños con dislalia. También se describe como es la integración del videojuego con el módulo de reconocimiento de voz. Universidad Autónoma de Yucatán Facultad de Matemáticas, México[15].
- Felix Etxeberria Vlerdi., Escribió un artículo en el 2008 para la revista Electrónica Teoría de la Educación. Educación y Cultura en la Sociedad de la Información. Basando su investigación en el creciente consumo de los videojuegos. La violencia en los videojuegos son los principales temas de la literatura sobre el estudio de los efectos que tienen los videojuegos en la conducta infantil. Pero él presenta un paralelo justificando que estos pueden educar y ser positivos para el crecimiento de los niños. Universidad de Salamanca, España [16].
- Xiaoping Wang., Yufeng Hao., Degang Fu., Chunwei Yuan., Presentaron para el, second International Symposium on Information Technology application, en el año 2008 un proyecto en el cual la interacción audio-visual entre las personas y la maquina es la prioridad. Hicieron varias y nuevas segmentaciones de voz, combinando varios enfoques y teorías. En el artículo muestran los resultados los cuales fueron positivos y demuestran eficiencia del sistema. Universidad del sur-este, Nanjing, Jiangsu, China [17].
- Pawel Zwan., Bozena Kostek., Miembros de la AES, en el 2008 diseñaron un sistema el cual reconoce la calidad de voz dependiendo el tipo de cantante que se le muestre a la máquina. Contaron con una base de datos que contenía más de 2.690 muestras de cantantes entrenados y no entrenados. El programa cuenta con un conjunto de parámetros los cuáles clasifica a la persona. Se aplicaron redes neuronales las cuáles fueron entrenadas y programadas para detectar la calidad de técnica de canto. Departamento de sistemas multimedia, Universidad de Tecnología Gdasnk, Polonia [18].
- Oscar Mayor., Jordi Bonada., Jordi Janer., en el 2009 presentaron un proyecto el cual fue muy llamativo para el público. Su característica más relevante fue la transformación de la voz humana que podía generar, por ejemplo la robotización, cambiar el acento de la persona por la de un extranjero, cambio de sexo de hombre a mujer, entre otras. Fue utilizada en tiempo real lo cuál permitió implementarla en museos y se adaptó para los videojuegos y transformar la voz del usuario o la de cualquiera de los personajes. Universidad Pompeu Fabra, Barcelona, España [19].

- Pablo Caloto Crespo., Manuel Moranchel Edras., Ángel Ruiz Alonso., para su proyecto de grado en el 2009 presentaron una investigación sobre el campo de los modelos acústicos de diferentes idiomas los cuáles integraron con el objetivo de manejar un avatar a través de comandos de voz, descartando el resto de interfaces físicas, como el teclado o el ratón. Universidad Complutense de Madrid [20].
- P.S. Rajakumar., S. Ravi., R.M Suresh., en el 2010 escribieron un artículo en el cuál compararon el rendimiento de dos codificadores de voz, teniendo como indicador la inteligibilidad, que se refiere a que tan comprensible es la voz de salida y la calidad, es decir cómo suenan de “natural” las palabras. Instituto de educación e investigación, Maduravoyal, Chennai, India [22].
- Héctor del Castillo., Sara cortés., Ana Belén García-Varela., Natalia Monjelat., Gloria Nogueiras. En el 2010 realizaron una investigación para su proyecto de grado en el cuál analizan las creencias que tienen los alumnos de educación secundaria con respecto a los videojuegos comerciales y cómo se pueden usar como instrumentos de aprendizaje. Concretamente se centraron en estudiar si se puede aprender o no con esos instrumentos de ocio, en la escuela. Qué tipo de aprendizaje pueden ofrecer los videojuegos comerciales y que características los hacen tan llamativos para los adolescentes. Universidad de Alcalá, España [23].
- Marcela Marín., En el 2011 realizó un trabajo sobre los videojuegos y el impacto educativo que pueden llegar a tener. Se puede evidenciar que los videojuegos son una herramienta de entretenimiento pero con una orientación adecuada pueden ser usados en el ámbito educativo. Universidad Deusto, España [27].
- Rehan Ahmed., Roberto Gil-Pita., David Ayllón., Lorena Álvarez., Presentaron un artículo en el año 2011 en el cuál presentaron un algoritmo que permite separar señales de voz dependiendo de la asignación de fuentes iniciales. La diferenciación de las señales se hace mediante sus frecuencias fundamentales y se compara con las entradas. Universidad de Alcalá, España [28].
- Konstantinos Drossos., Andreas Floros., Kyriakos Agavanakis., Nicoles-Alexander Tatlas., Nikolaos- Grigorios Kanellopoulos en abril del 2012 publicaron una artículo para la AES sobre una aplicación la cuál usaron en una casa “inteligente”, la cuál detectaba el nivel de estrés del usuario. Departamento de Artes Audiovisuales, Universidad Ionian, Islas Jónicas, Grecia [31].
- Francisco Ignacio Revuelta Domínguez., Jorge Guerra Antequera., en el 2012 presentaron un artículo sobre lo que se puede aprender de un videojuego, pero desde la perspectiva netamente del jugador. Investigaron sobre la aplicabilidad didáctica que pueden tener estas herramientas digitales. Basaron su trabajo a resultados estadísticos, aplicando más de 115 encuestas a video jugadores. Universidad de Extremadura, España [32].

- Mark S. Hawley., Stuart P. Cunningham. Phil D. Green., Pam Enderby., Rebecca Palmer., Siddharth Sehgal., Peter O'Neill en el 2013 presentaron un programa el cuál apoya a personas con discapacidad del habla. El software ayuda al usuario a ordenar las palabras de entrada para luego generar una salida la cuál sea más entendible. Las pruebas que se realizaron muestran un alto porcentaje de efectividad y satisfacción de los usuarios. Departamento de nuevas y emergentes aplicaciones de la tecnología para la salud (NEAT), U.K. [33].
- Timo Becker., Yosef Solewicz., Gaëlle Jardine., Stefan Gfrörer., en junio del 2012, hicieron una comparación de voces en condiciones de análisis forense. Dependiendo del sistema en el que se esté ingresando la voz el resultado puede variar ya que también depende de otro tipo de condiciones. Oficina de la policía federal criminal, Alemania [34].
- Petko N. Petkov., Gustav Eje Henter., Bastiaan Kleijin., en el 2013 publicaron para la IEEE un artículo en el cuál habla de su proyecto donde intentaron maximizar el reconocimiento de fonemas a partir de la reducción del ruido en el sistema. Al reducir la distorsión en la entrada el programa logra tener una mejora natural y dar un rendimiento mucho mayor. Universidad Victoria de Wellington, Nueva Zelanda [35].
- Adam Kupryjanow., Andrzej Czyzewski., Presentaron en el 2013 dos algoritmos los cuáles permitían el análisis de las señales de voz en tiempo real. El primer algoritmo es usado para la detección de la voz y el inicio y fin de palabra; y el segundo es el analizador del habla.

Dentro de sus funciones principales están el reconociendo automático del idioma, la emoción o estado de ánimo de la persona. Se evaluaron los algoritmos llegando a la conclusión de que son efectivos y fiables para el usuario. Universidad Gdansk de Tecnología, Polonia [36].

1. PLANTEAMIENTO DEL PROBLEMA

1.1. DESCRIPCIÓN DEL PROBLEMA

El avance tecnológico es una consecuencia de la búsqueda constante del ser humano por tener confort y mayor comodidad en cada área de su vida, como por ejemplo los sistemas o productos que utiliza a diario, el computador u otro dispositivo.

Es muy importante en el diseño de sistemas (hardware y software) que se contemple la efectividad, eficiencia y practicidad ya que son estas las características que generan interés a las personas, sin embargo, el uso exitoso de herramientas, depende directamente de la facilidad en la interacción entre el usuario y la herramienta.

Actualmente se cuenta con métodos de entrada hacia la máquina como lo son el teclado y el mouse, pero el desarrollo e innovación han llevado a otro tipo de formas de controlar estos dispositivos, como lo son las pantallas táctiles y el reconocimiento de voz, los cuáles no son más o menos efectivos, pero si le aportan una interacción más personal o “natural” al usuario.

El reconocimiento de voz es una forma de comunicación entre la persona y el dispositivo que tiene como función que la máquina entienda a la persona, haciendo más transparente la relación entre estos. El hablar con las personas es algo que se hace normalmente todos los días, al pasarlo al ámbito virtual se genera una naturalidad en la interacción. Aunque existe investigación y desarrollo en este tema, la variabilidad presente en el proceso de identificación es alta ya que el usuario no puede repetir de forma exacta la misma frase o palabra con la que esta interactuando con la máquina.

1.2. PREGUNTA PROBLEMA:

¿Cómo por medio del reconocimiento de voz es posible mejorar la interacción hombre máquina en un videojuego que permita la selección de colores y figuras geométricas?

Tener una interacción más eficiente entre hombre – máquina logrando minimizar errores en la experiencia entre estos; incrementar la satisfacción interactiva, hacer más productivas las tareas que rodean a las personas y a las máquinas, creando un videojuego que le permita al usuario a partir de la elección de figuras geométricas y colores por reconocimiento de voz, una forma de entretenimiento.

2. JUSTIFICACIÓN

Los sistemas de reconocimiento de voz ofrecen nuevas posibilidades de satisfacción a necesidades corrientes o prácticas para la relación de las personas con las máquinas, ya que el procesamiento de señales entre otras cosas nos permite conocer características específicas de la voz del usuario, haciendo más personal esta interacción. Depende del uso del sistema el tipo de procesamiento que se le da a la señal ya que las características a estudiar varían según la aplicación que le demos.

El procesamiento de voz consiste en el estudio de la señal de entrada en tiempo real permitiendo extraer características del usuario que permitan identificarlo, se pueden generar códigos para que a partir de la voz se le dé una orden a la máquina y ésta cumpla correctamente con lo que se le está pidiendo, este tipo de avances ha permitido desarrollo de la industria en diferentes campos como lo son: comerciales y comparativas tecnológicas, aprendizajes, laborales, entretenimiento, etc.

Al diseñar un videojuego que por medio de reconocimiento de voz se puedan elegir objetos hace que el usuario se vea atraído no solo por el hecho del entretenimiento que da el juego sino también por la transparencia que genera el “hablarle” a la máquina y que esta responda a sus órdenes.

El desarrollo de este tipo de herramientas a futuro puede mejorar la experiencia de aprendizaje del usuario.

3. OBJETIVOS

3.1. Objetivo General:

Desarrollar un video juego que por medio de reconocimiento de voz permita la selección de figuras geométricas y colores.

3.2. Objetivos específicos:

- Diseñar el videojuego desde la ambientación, tipo de personajes, número de palabras a reconocer y su funcionalidad al responder al comando de voz.
- Determinar todos los parámetros necesarios para el desarrollo del algoritmo de reconocimiento de voz que dará funcionalidad al videojuego.
- Desarrollar el videojuego e implementar el algoritmo de reconocimiento de voz.
- Validar la eficacia del algoritmo de reconocimiento de voz en el videojuego, mediante evaluaciones del número de aciertos del software.

4. ALCANCES Y LIMITACIONES

4.1. Alcance:

Desarrollo de una plataforma que permita de manera didáctica un reforzamiento de los conceptos de tres figuras geométricas y tres colores en dos personas dentro de la primera infancia.

4.2. Limitaciones:

- La herramienta únicamente va a funcionar para 3 colores (Amarillo, azul y rojo) y tres figuras geométricas (Triángulo, cuadrado y círculo).
- El video juego se implementará para dos niños entre los 5 y 8 años, los cuales serán los que grabarán y calibrarán las palabras para el funcionamiento de este.
- Se tocará el tema educativo tangencialmente, ya que el proyecto será propuesto desde el punto de vista ingenieril, es decir el desarrollo de los algoritmos de programación de reconocimiento de voz y de el video juego.
- El video juego estará limitado a un sistema operativo.

5. MARCO REFERENCIAL

5.1. MARCO TEÓRICO

5.1.1. PRAAT:

Es un software libre dedicado al análisis y síntesis del sonido, creado por Paul Boersma y David Weenink.

El programa ofrece diversas maneras de obtener un sonido.

- Grabación de un sonido, la única condición es que el ordenador debe tener un micrófono.

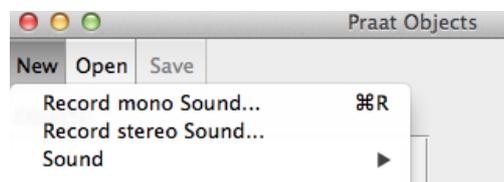


Figura 1. Ventana para la opción de grabar.

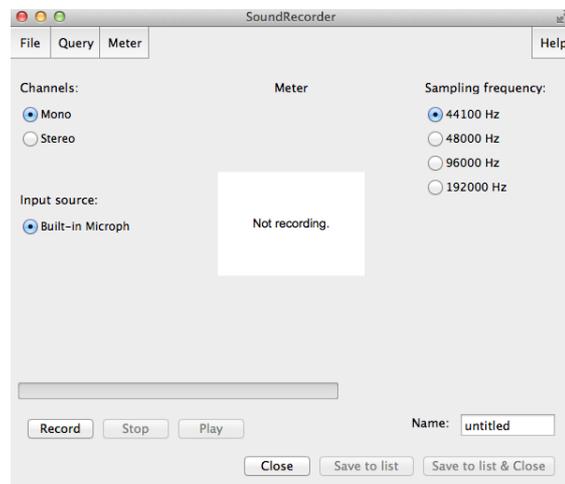


Figura 2. Ventana SoundRecorder

- Leer un sonido desde el disco, se pueden obtener varios tipos de archivos con extensiones estándares de sonido, como, .WAV, entre otras.



Figura 3. Ventana para leer un archivo

- Crear un sonido desde una formula.

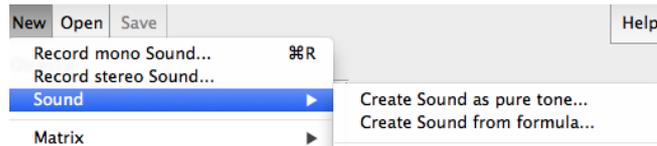


Figura 4. Ventana para crear un sonido desde una fórmula

Luego de tener el sonido en la **lista de objetos** el software permitirá:

- Guardar el sonido, en cualquiera de estos formatos, WAV, AIFC, NeXT/Sun, NIST. Todas las extensiones conservan la calidad y son adecuadas para el correcto funcionamiento de Praat. Ya con el sonido en la **lista de objetos** (List of objects), se podrán guardar en varios formatos en **los menús escritos** (write menu).

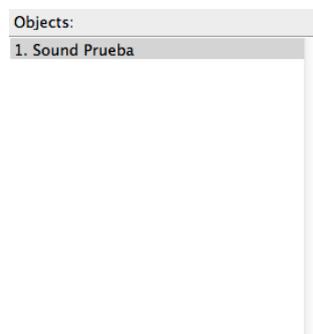


Figura 5. Lista de Objetos

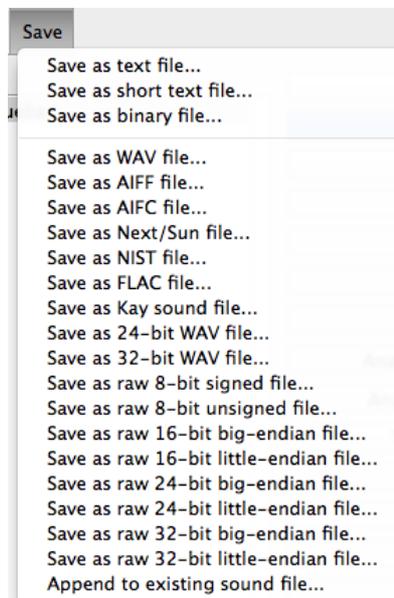


Figura 6. Opciones de guardar

- Desarrollo y análisis de una señal de audio, ya teniendo el sonido en la **lista de objetos** (List of object) se puede seleccionar y entrar a la opción de **editar** (Edit).



Figura 7. Modo de ingreso a la opción de Edit

Aparecerá la ventana de **edición de sonido** (Sound Editor), en donde se podrá acercar y recorrer la imagen como se desee, además de poder seleccionar cualquier parte del sonido.

Como principal herramienta ofrece en análisis espectral, permite generar espectrogramas y segmentos espectrales. El eje horizontal representa el tiempo; el vertical la frecuencia con un intervalo entre 0[Hz] y [5000Hz].

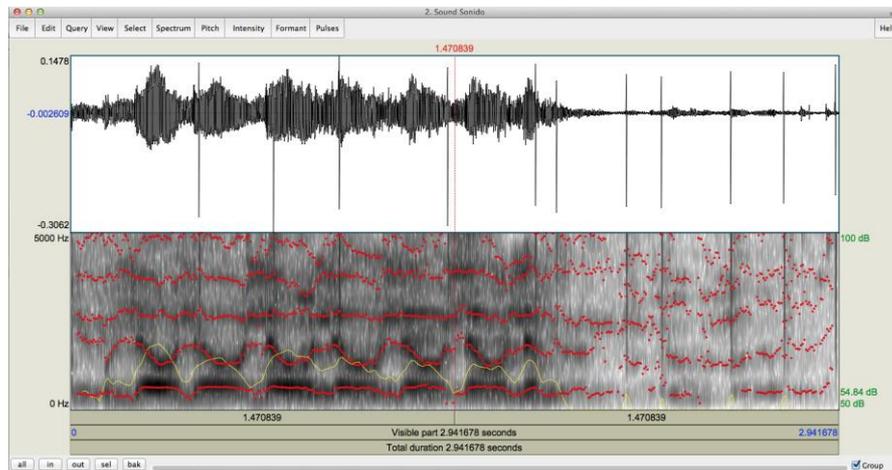


Figura 8. Espectrograma

Entre mayor densidad de energía tenga una parte de la palabra en el espectrograma se verá más oscuro. Por ejemplo si alrededor de los 2[s] a una frecuencia de 4000[Hz] el espectrograma esta con una área oscura, significa que el sonido tiene altos niveles de energía en esa frecuencia para ese periodo de tiempo.

Existe un cursor el cual le permite al usuario verificar el tiempo y la frecuencia en donde se encuentre la mayor cantidad de energía y así encontrar los formantes y los peak espectrales de cada sonido.

Es decir, Praat es un software que permite el análisis de parámetros tales como: la duración, la frecuencia de los formantes, la intensidad o la frecuencia fundamental (pitch).

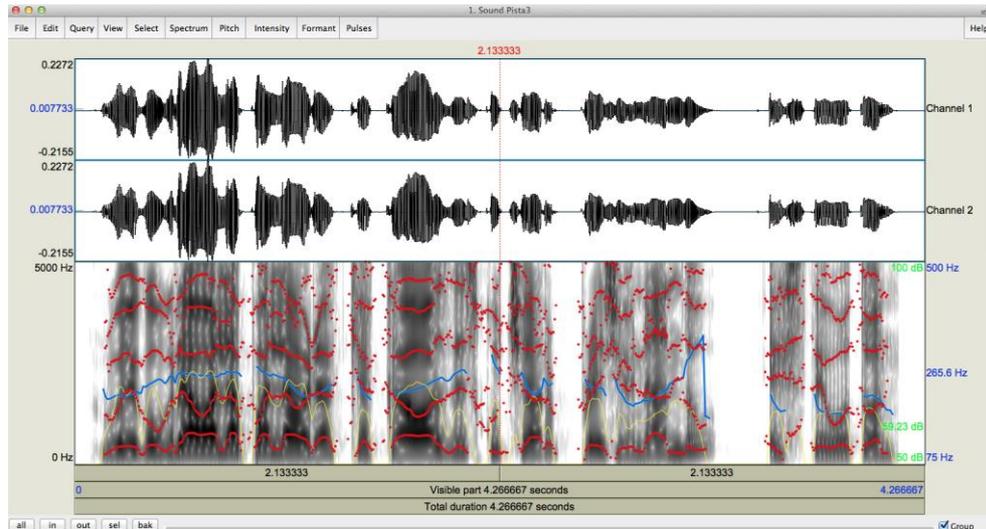


Figura 9. Vista de varios parámetro ofrecidos por praat.

Los principales análisis que se pueden realizar están relacionados con el oscilograma, espectrograma, curva de intensidad, la frecuencia fundamental además de los formantes. Estos se pueden diferenciar en la ventana, ya que cada uno tiene un color que lo representa, el rojo delimita los formantes, el azul la frecuencia fundamental y el verde la intensidad.

Para la obtención de información de cada parámetro se debe tener en cuenta:

- Las propiedades del espectro se pueden ver mejor si en las ventanas solo se visualiza la forma de onda y el espectrograma. En Ajustes del espectrograma (*Spectrogram Settings*) se puede modificar, los intervalos de frecuencia, la amplitud de las ventanas de análisis en donde se pueden ver mejor los armónicos, formantes y el análisis temporal. También es posible controlar la intensidad de los colores, entre más alto sea el valor de estos más fácil será resaltar cada parámetro por separado.

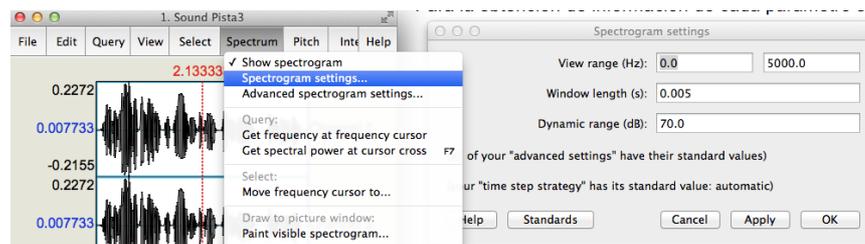


Figura 10. Propiedades del espectro

- Propiedades de la frecuencia fundamental, se debe ver únicamente la curva de entonación para que esta sea mucho más clara, facilitando también el análisis de la sonoridad. Se debe tener en cuenta el tipo de voz que se quiere analizar ya que el rango de frecuencias varía según el caso. Para los niños se espera valores máximos alrededor de 600 [Hz], para las mujeres de 400 [Hz] y, en voces masculinas picos en 300 [Hz]. [38]

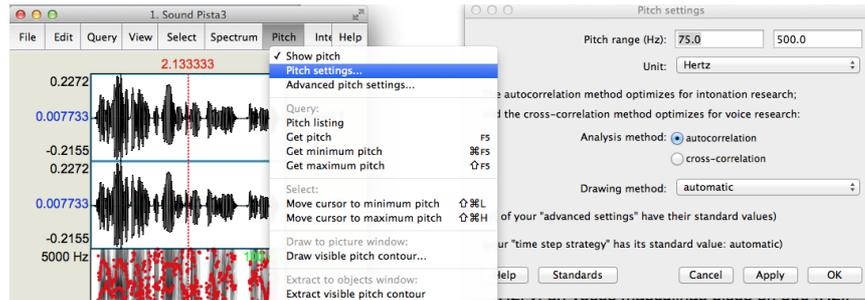


Figura 11. Propiedades Frecuencia fundamental

- c. Las propiedades de la intensidad se pueden analizar únicamente con la curva de intensidad, pero es más claro será verla al tiempo con el espectrograma ya que nos ubicara en la parte de la onda que corresponde al movimiento de la intensidad.

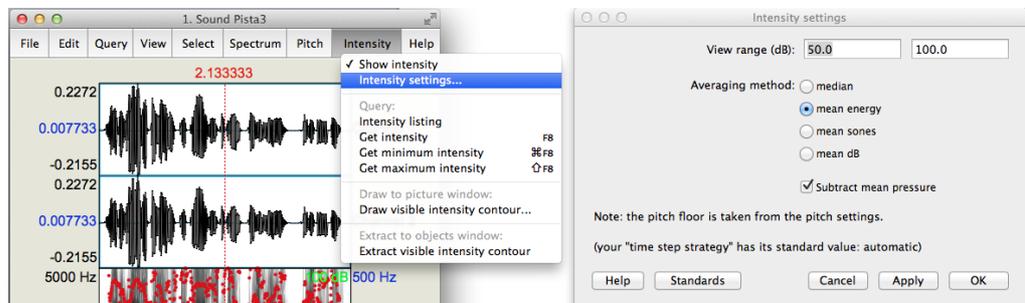


Figura 12. Propiedades de intensidad

El programa por defecto tiene los intervalos de intensidad entre 50 y 100 decibeles pero es recomendable tener el límite inferior en 0 [dB].

Los formantes son la característica más relevante que este software permite analizar, Praat hace un cálculo de estos a partir de LPC (Linear Predictive Coding), lo que permite verlos en toda la trayectoria de la onda. Se pueden ver solos en la ventana, aunque es mucho mejor obtenerlos sobre el espectrograma ya que así los valores serán más fiables.

Para ver cómo se comporta el espectro de algún sonido con respecto al tiempo se debe recurrir algún software que permita procesamiento de señales de voz, donde se puede grabar o importar en varios tipos de archivo de audio logrando obtener el espectrograma y hacer un estudio pertinente. Praat ofrece un análisis que solo funciona para calcular los formantes de las vocales orales, en el caso de las consonantes nasalizadas el programa presenta fallas significativas. Sin embargo este programa es muy útil para encontrar formantes vocálicas así no se encuentren bien definidas. Otra facilidad es que permite modificar la marca de la frecuencia máxima que se desea para encontrar un formante, como se mencionó antes las vocales femeninas son un 10% más altas que las vocales masculinas, el parámetro formante máximo (Maximum formant) cabe calcularlo entre 5500Hz y 5000Hz dependiendo del usuario. El

número máximo de formantes es de 4 (F4), ya que en la ventana de edición estos no se muestran de forma clara o no aparecen.

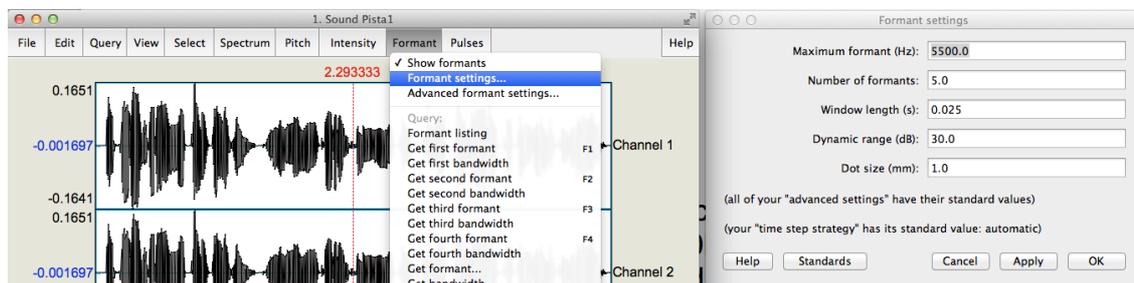


Figura 13. Propiedades de los formantes.

Un analizador de voz para hallar cualquiera de los parámetros anteriormente mencionados, principalmente debe obtener el espectro correspondiente al audio, Praat lo obtiene a partir de dos métodos de análisis: FFT (Fast Fourier Transform), que es la que permite dar a conocer los armónicos del sonido; y LPC (Linear Prediction), que simula un filtro del vocal permitiendo obtener los picos .

Si se desea ver solo la FFT el software lo permite crear desde la opción de edición de sonido (*Sound Editor*). Se ubica el cursor exactamente donde se desea analizar la onda sonora y se selecciona en el menú del espectro (*Spectrum*), la opción vista el fragmento del espectro (*View Spectral Slice*).

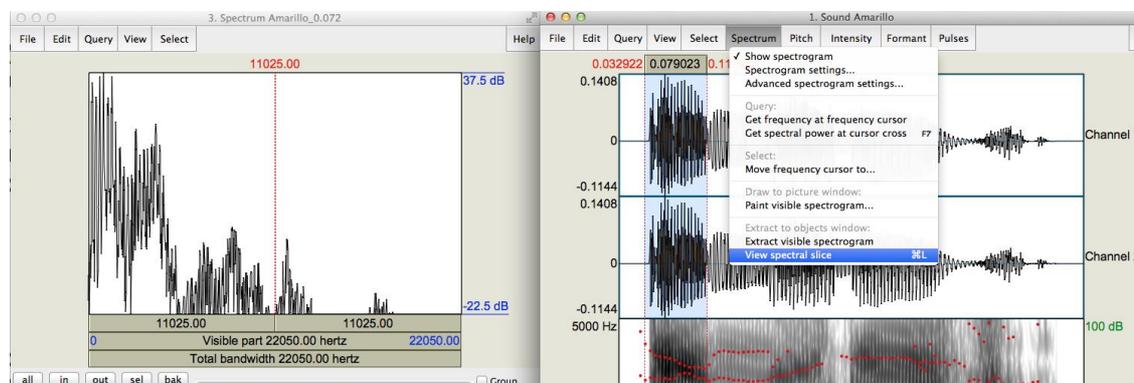


Figura 14. Spectrum

Para darle precisión al análisis el usuario de debe fijar tanto en el oscilograma como en el espectrograma.

El espectro LPC se podrá obtener desde la ventana principal, Objetos de Praat (*Praat Objects*). En la opción de análisis de espectro (*Analyse spectrum*). Se debe escoger el punto exacto en el tiempo donde se desea hallar este parámetro.

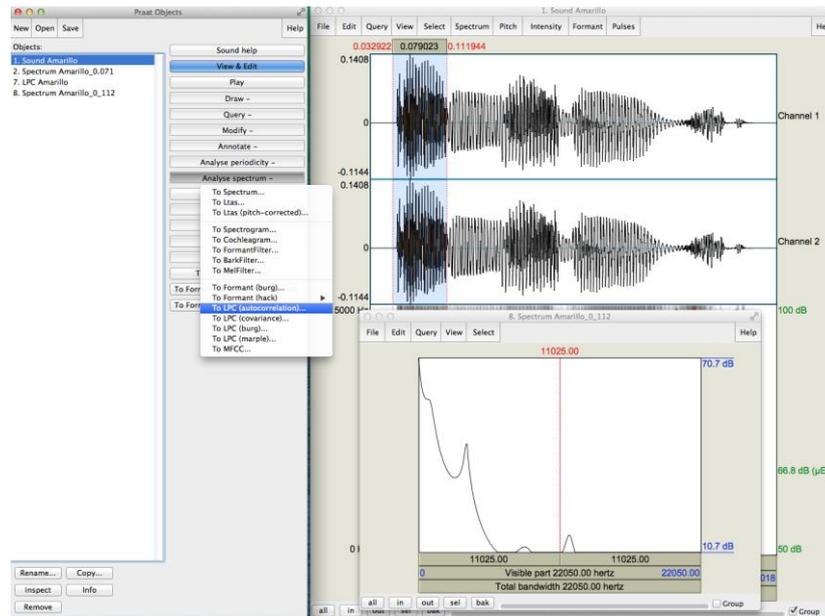


Figura 15. LPC

5.1.2. RECONOCIMIENTO DE VOZ.

El reconocimiento de voz es una herramienta que tiene como objetivo la interacción entre un usuario y un sistema. El inconveniente con este tipo de herramientas es que debe unir diversas informaciones de muchas fuentes, como la acústica, fonética, léxica, sintáctica, semántica y pragmática. Se debe tener en cuenta que habrá errores inevitables y su porcentaje de reconocimiento dependerá del número de usuarios, el número de palabras, la forma del habla de cada persona, entre otros aspectos.

Actualmente la investigación y tecnología relacionadas al reconocimiento de voz viene cada vez más en ascenso, hoy en día es considerable la cantidad de dispositivos y aplicaciones que tienen incorporados opciones para el reconocimiento de voz, así como se ha incrementado el porcentaje de reconocimiento de los mismos, tanto para palabras específicas como para personas.

Un sistema de reconocimiento de voz está formado por varios procesos o modelos, de los cuáles el primero es el modelo acústico que permite detectar el canal de comunicación para así identificar el grado de distorsión presente, seguido a esto se encuentra el modelo lingüístico que se encarga de identificar el idioma, analizando acentos, formas de hablar, etc., luego se procede al modelo semántico que entra a comparar frases en bases de datos para buscar su significado, aunque todavía no llega el momento en que una máquina pueda entender todo lo que una persona le dice, los procesos e investigaciones que se llevan a cabo actualmente pueden hacerlo posible.

A continuación podemos observar un diagrama de los procesos involucrados en un sistema de reconocimiento de voz:

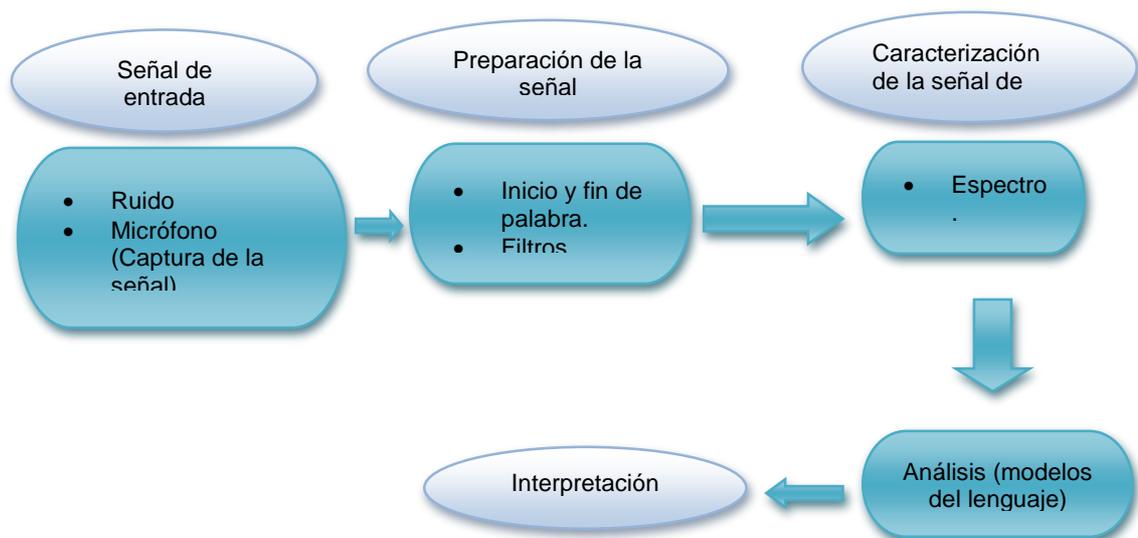


Figura 16. Diagrama representativo de un sistema de reconocimiento de voz

5.1.3. UNITY 3D:

Es un motor de desarrollo para la creación de videojuegos en 2D o 3D. Se pueden desarrollar en cualquiera de estas plataformas, Windows y OS X dando la facilidad de crear juegos para Windows, OS X, Linux, Xbox 360, PlayStation 3, Playstation Vita, Wii, Wii U, iPad, iPhone, Android y Windows Phone. Adicionalmente permite desarrollar juegos para navegador.

Tiene dos versiones, Unity que no tiene ningún costo y Unity pro que por tener características adicionales como textura, cámara oculta, y facilidades de métodos de programación, tiene un costo.

La versión más reciente es Unity 4.2 aunque sus funciones son casi iguales a las versiones anteriores, el cambio principal ha sido en la interfaz gráfica mejorando para el programador el espacio de trabajo.

Unity tiene una página en internet donde se puede descargar las versiones que se deseen, se encuentran demos de juegos, tutoriales, asset store.

Cuando se trabaja en Unity se maneja sobre un proyecto, es decir guarda en una carpeta todas las estructuras, texturas, scripts, assets, efectos de sonido. Para facilitar la organización de todos los detalles del videojuego.

5.1.4. DISEÑO DE VIDEO JUEGOS EDUCATIVOS.

En la actualidad se debe valorar el uso de los videojuegos como medio de alfabetización, éste representa una valiosa herramienta de apoyo pedagógico. Los niños están muy relacionados con consolas de juegos, computadores y diferentes sistemas que les permiten divertirse de manera interactiva, lo que les prepara para el manejo de recursos tecnológicos. Existe afinidad hacia los

videojuegos pues en su mayoría estos llegan a mejorar su agilidad o ayudan en la creación de estrategias para solucionar problemas. Los videojuegos pueden ser herramientas de aprendizaje porque permiten el refuerzo de habilidades a través de la práctica constante (La repetición), el reto y la curiosidad por conocer cada parte del juego además de reducir el espacio que existe en la labor educativa logrando hacerla más próxima a el contexto de los estudiantes. Uno de los objetivos de estas herramientas educativas es lograr el fortalecimiento de la comunicación, la motivación y el uso de estrategias de aprendizaje. La tecnología está al alcance de la mayoría y se puede usar como un foco diferente de interacción llamativa para los estudiantes. Los videojuegos son elementos digitales multimodales, mezclan audio, video, imagen, entre otros. Llevan en ellos un aprendizaje implícito en cuánto a memoria y habilidad de resolver problemas. Por esto son una herramienta perfecta que le permite al usuario no solo distraerse o divertirse con los retos propuestos, si no también aprender. Se puede aprovechar el gusto y la motivación que las nuevas generaciones sienten sobre los videojuegos para aplicarlos a prácticas educativas y ser objetos desarrollares de aprendizajes

5.2. MARCO CONCEPTUAL

5.2.1. FORMANTES

Es el pico de intensidad en el espectro de un sonido (mayor concentración de energía, mayor amplitud de onda en una determinada frecuencia).

El tracto vocal por la configuración de las articulaciones genera un proceso de filtrado gracias a la resonancia por la configuración de los articuladores en este, de ésta manera se determinan los diferentes formantes. [39]

Los formantes permiten distinguir los sonidos del habla, sobre todo las vocales dónde hayamos la mayor concentración de energía, y otros sonidos.

Sirven para los sistemas de reconociendo de voz (transposiciones de altura del audio digital), se puede hacer un sonido acerca de cada sonido del habla humana puesto que estos poseen una marca característica de formantes entre sí, en otras palabras cada sonido tiene un reparto diferente de la energía sonora entre los diferentes formantes, lo cual permite distinguirlos y clasificarlos.

El oído humano tiene la capacidad de hacer un análisis de formantes de manera inconsciente, esto nos permite identificar los sonidos de manera rápida

En la mayoría de los lenguajes los dos primeros formantes permiten distinguir la mayoría de sonidos vocálicos del habla, que se identifican con la posición de la lengua.

Primer formante: Se identifica por la frecuencia más baja que se genera en la abertura del tracto vocal por la frecuencia de las ondas estacionarias que vibran verticalmente en la cavidad.

Entre más baja esté la lengua, o sea cuanto mayor sea la abertura de la vocal esta frecuencia será más alta.

Segundo formante: Está relacionado con la vibración en dirección horizontal. Permite identificar si la vocal es anterior, central o posterior.

¿Cómo se producen?, por la resonancia del tracto vocal la vibración de las cuerdas vocales se generan ondas sonoras con un espectro de frecuencias bastante destruido, éstas son filtradas por el tracto vocal lo que causa que algunas frecuencia sean reforzadas y otras atenuadas. Las frecuencias reforzadas son precisamente los formantes principales de la emisión sonora.

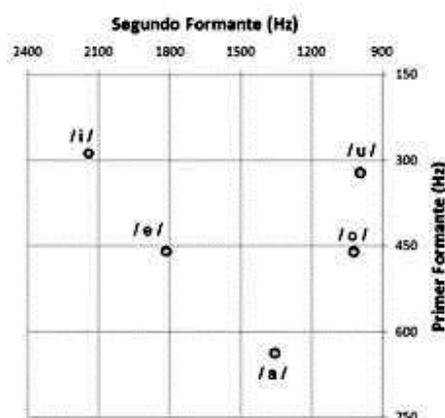


Figura 17. Formantes

Normalmente solo se necesitan los dos primeros formantes para caracterizar una vocal. Sin embargo pueden existir casos en los cuales sea necesario 6 formantes como mínimo para lograr una caracterización adecuada de la vocal. Los formantes posteriores determinan propiedades acústicas como el timbre, entre otras.

Las **consonantes aproximantes**, se caracterizan por variaciones dinámicas de la frecuencia de los formantes principales. En las **oclusivas** y **fricativas**, los formantes son menos claros y esos sonidos se perciben en gran medida gracias al efecto producido en las vocales **adyacentes**, dónde los formantes involucran mayor energía sonora.

La posición de la lengua es la que permite determinar los dos primeros formantes.

F1 tiene una frecuencia más alta entre más baja está la lengua. Es decir entre más grande sea la abertura que se necesite para una vocal, mayor es la frecuencia con la que aparece F1

F2 tiene mayor frecuencia cuanto más hacia adelante esté posicionada la lengua. Entre más anterior es una vocal, mayor es F2.

No todos los sonidos humanos se componen de formantes definidas. Tan solo aparecen en **sonantes**, que incluyen los sonidos pulmonares (vocales, aproximantes y nasales). Las nasales tienen un formante adicional (F3) en torno a los 1500 [Hz]. Las consonantes róticas por su parte presentan pequeñas oclusiones y, en caso de la múltiple ([r] como rio) aparecen formantes vocálicas.

5.2.2. PRINCIPIO Y FIN DE PALABRA

Es un algoritmo el cual permite que es sistema reconozca cuando se empieza y cuando se termina de hablar y así solo dejar la parte de señal que interesa para el procesamiento de señales.

5.2.3. FFT.

La transformada rápida de Fourier o FFT por sus siglas en inglés, es un algoritmo que reduce el tiempo de cálculo de n^2 pasos a $n \cdot \log_2(n)$

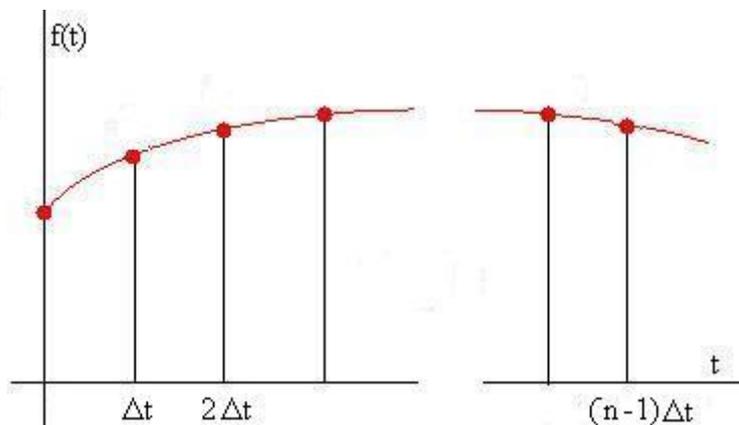


Figura 18. FFT

La función de la transformada de Fourier es convertir un vector o ecuación en función del tiempo t a uno en función de la frecuencia w , de esta manera podemos visualizar el espectro de frecuencias de un vector determinado. [40]

- Espectro de frecuencias, es la medida de distribución de amplitud para frecuencia en un determinado fenómeno ondulatorio.

Una de las aplicaciones es la del tratamiento digital de señales la cual en el caso del reconocimiento de voz es muy útil. Nos permite conocer las características frecuenciales de la cual está compuesta cada señal que entré al sistema y así poder categorizar y comparar.

5.2.4. FILTRO LPC (LINEAR PREDICTIVE CODING)

El modelo del análisis de un filtro LPC se basa en la redundancia presente en las señales del habla, como la periodicidad y variación lenta, esto permite la predicción de una señal muestreada a partir de muestras anteriores, un sistema de predicción lineal dada una señal del habla (palabra), puede definir la función de transferencia del filtro que la ha generado. El resultado del análisis son unos coeficientes que se relacionan con los polos espectrales, por lo que se describe la acción de una función de transferencia en la que únicamente se describen picos.

El número de picos debe ser suficiente para modelar los formantes en la amplitud de la banda de la señal, se trabaja sobre bloques de 20 ms de voz sobre lo que se conoce como modelo corto, se hace esto el suponer que las características de la voz no varían en intervalos pequeños. A partir de estos bloques se determinan los coeficientes de predicción, luego se cuantifican y se envían al receptor junto a otros parámetros. El efecto de la predicción es similar a la correlación entre muestras adyacentes. [41]

5.2.5. DENSIDAD ENERGÉTICA

Concepto utilizado para definir en qué puntos de la señal de habla se encuentran los picos o mayores niveles energéticos, esto nos permitirá identificar los formantes posteriormente.

5.2.6. TRACTO VOCAL

Constituido por la cavidad oral, nasal, la faringe y la laringe, dentro de estas cavidades están los órganos de la articulación los cuales pueden dividirse en activos y pasivos. Los activos son la lengua, mandíbula, velo del paladar y los labios, mientras que los pasivos son los dientes, paladar duro y maxilar superior. Por medio de la modificación y diferentes posiciones que adoptan los órganos articulatorios, el tracto vocal puede tomar varias formas o configuraciones que actuaran las cuales funcionaran como filtros acústicos para el sonido que viene de la laringe. Cada configuración diferente del tracto vocal constituye un filtro diferente y por ello el sonido vocal producido será distinto. Un ejemplo de esto son las vocales, cada vocal tiene una forma distinta en tracto vocal y por lo tanto valores formánticos distintos, lo cual permite diferenciar una de otra.

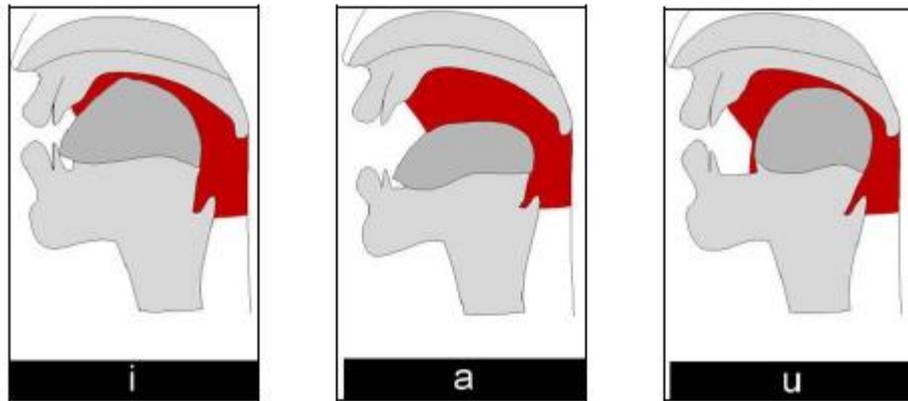


Figura 19. Formas del tracto vocal para las vocales i, a y u

5.2.7. ADQUISICIÓN DE DATOS

La adquisición de datos se inicia con el fenómeno físico o la propiedad física de un objeto (objeto de la investigación) que se desea medir. Esta propiedad física o fenómeno podría ser el cambio de temperatura o la temperatura de una habitación, la intensidad o intensidad del cambio de una fuente de luz, la presión dentro de una cámara, la fuerza aplicada a un objeto, o muchas otras cosas. Un eficaz sistema de adquisición de datos puede medir todas estas diferentes propiedades o fenómenos.

La capacidad de un sistema de adquisición de datos para medir los distintos fenómenos depende de los transductores para convertir las señales de los fenómenos físicos mensurables en la adquisición de datos por hardware. Transductores son sinónimo de sensores en sistemas de DAQ. Hay transductores específicos para diferentes aplicaciones, como la medición de la temperatura, la presión, o flujo de fluidos.

El acondicionamiento de señales suele ser necesario si la señal desde el transductor no es adecuado para la DAQ hardware que se utiliza. La señal puede ser amplificada o des amplificada, o puede requerir de filtrado, o un cierre patronal, en el amplificador se incluye para realizar demodulación.

5.2.8. REDES NEURONALES

Las redes neuronales son sistemas ideados como abstracciones de las estructuras neurobiológicas (cerebros) encontradas en la naturaleza y tienen la característica de ser sistemas desordenados capaces de guardar información.

La forma en que desarrollan su trabajo es esencialmente distinta de la utilizada por las computadoras convencionales. Los procesadores microscópicos del cerebro (neuronas) operan en paralelo y presentan cualitativamente más ruido que los elementos que forman a las computadoras. No ejecutan un programa fijo con base en un conjunto previamente especificado de datos, sino que comunican señales a través de retransmisores que llamamos sinapsis, que

llegan a centros de conjunción llamados los cuerpos de las neuronas y desde los cuales surgen señales eléctricas a través de canales conocidos con el nombre de axones.

La importancia de cada sinapsis en el proceso de retransmisión se actualiza continuamente y lo mismo ocurre con algunas propiedades intrínsecas de las neuronas, proporcionando un sistema de auto programación y adaptación que sustituye a la programación externa de los sistemas de cómputo comunes. Existe así una dinámica de las sinapsis y de las neuronas en el cual los programas y los datos cambian todo el tiempo. [38]

5.2.9. SENSIBILIDAD EN MICROFONOS

La sensibilidad de un micrófono es la eficiencia con que se transforma la presión sonora a tensión eléctrica. Se denomina a la sensibilidad como la relación entre la tensión eléctrica que se encuentra en los bordes del micrófono a circuito abierto y la presión sonora aplicada, utilizando una frecuencia de 1K [Hz].

5.2.10. DIAGRAMA POLAR

El diagrama polar de un micrófono es aquel que muestra la sensibilidad con la que este puede captar el sonido, según el ángulo con que este incida, en el diagrama podemos visualizar 360° es decir una vuelta completa del micrófono. Existen tres tipos básicos de patrones, unidireccionales, omnidireccionales y bidireccionales.

5.2.11. RANGO DINÁMICO

Es el rango de niveles sonoros en los que la señal eléctrica que produce el micrófono es suficientemente alta para ser utilizada, esta característica está relacionada con la amplitud de la onda sonora que llega al micrófono.

5.2.12. RESPUESTA EN FRECUENCIA

Ésta se caracteriza por la intensidad de la señal eléctrica producida por un micrófono, para una amplitud determinada de la presión de la onda sonora, a diferentes frecuencias. Dependiendo de la situación en la que se vaya a usar el micrófono, se hace conveniente que sea plano o que presente refuerzos o atenuaciones para algunas bandas.

5.2.13. JUGABILIDAD

Describe la calidad de algún juego en término del funcionamiento de sus reglas y su diseño como juego. Se puede relacionar con las propiedades que describen la experiencia del jugador al enfrentarse a un videojuego.

6. MARCO METODOLÓGICO

6.1. Tipo de investigación

Tecnologías actuales y sociedad: El video juego controlado por medio de comandos de voz es un producto el cual demuestra el desarrollo de la tecnología y el impacto positivo que puede llegar a tener este tipo de herramientas.

Análisis y procesamiento de señales: La forma que tiene el usuario para interactuar con el videojuego es por medio de su voz. Se hizo un análisis de cada palabra, conociendo así los formantes que las constituyen y cómo es su comportamiento dentro del videojuego. Esto quiere decir que se necesitó de una plataforma que permitiera el procesamiento en tiempo real de señales, en éste caso para el procesamiento de la voz del usuario.

Acústica y audio: Para entender el comportamiento de cada palabra que en el juego se va a reconocer se hicieron capturas previas de la voz. Se hicieron tres muestras diferentes para tener un punto de comparación y poder así explicar el óptimo funcionamiento del videojuego. Una de las muestras se hizo dentro de un recinto considerando los conceptos de grabación y acústica para tener la mayor calidad posible en la captura.

6.2. Línea de investigación

La Facultad de Ingeniería de Sonido de la Universidad de San Buenaventura, sede Bogotá plantea unas líneas de investigación, este proyecto aplica a:

Tecnologías actuales y sociedad -> Análisis y procesamiento de señales -> Acústica y audio

6.3. Hipótesis

A través de la elección de figuras geométricas y colores será posible que los niños de primera infancia afiancen conocimientos de una manera didáctica a partir de un video juego con reconocimiento de voz.

6.4. Variables

6.4.1. Variables Dependientes:

- Grabaciones de las voces
- Algoritmo de reconocimiento de voz
- El desarrollo del video juego

6.4.2. Variable independiente:

- Software de desarrollo
- Tiempo de procesamiento del software.
- Resultados de calibración del sistema

6.5. Metodología del diseño

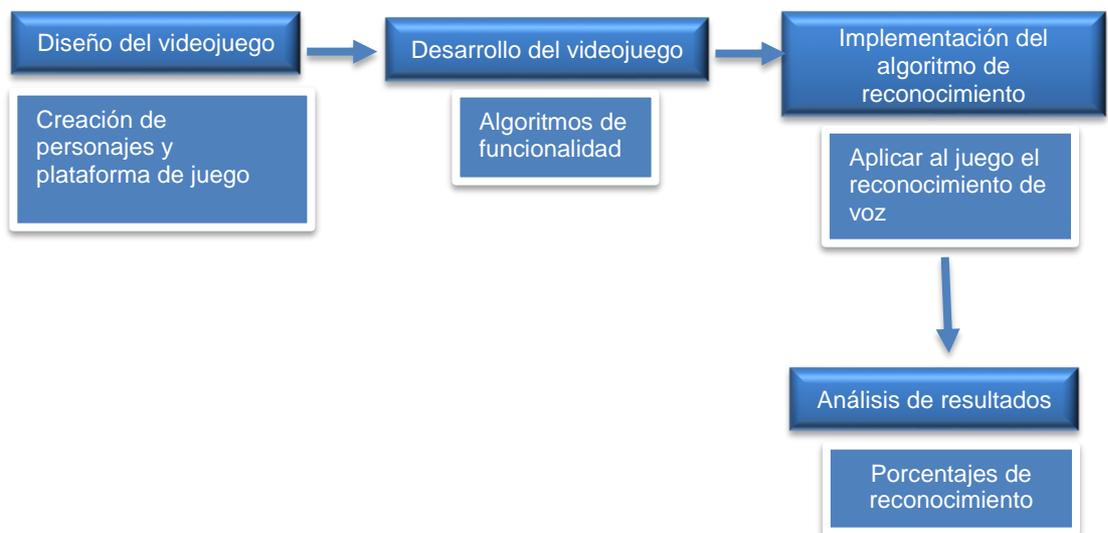


Figura 20. Metodología del diseño

7. ANÁLISIS INGENIERIL.

7.1. Diseño del videojuego.

Teniendo en cuenta que el principal fin de un videojuego es proporcionar al usuario entretenimiento a partir de los múltiples retos planteados, se pensó en

que esto se podría aprovechar para que el usuario aprendiera algún tema de una manera interactiva. Se propuso un videojuego educativo, el cuál le permitiera al usuario el refuerzo de conceptos básicos sobre tres colores y tres figuras geométricas.

Se diseñó todo el ambiente y la lógica del juego para este propósito.

DISEÑO DEL VIDEO JUEGO:

Desde el principio de la propuesta se dejaron claros cuáles iban a ser los personajes del videojuego, con los cuáles el niño iba a tener que interactuar debido a los conceptos propuestos para generar un fortalecimiento en el aprendizaje.

El diseño partió desde la necesidad de generar una experiencia de repaso de tres colores (amarillo, azul y rojo) y tres figuras geométricas (círculo, cuadrado y triángulo).

Según la investigación previa, la teoría en la cual el videojuego se basó fue en la del constructivismo de Piaget [37], la cuál habla sobre la construcción del conocimiento a partir de la experiencia. El niño iba a tener una guía sobre la cuál poder proponer una solución a cada reto que le sería propuesto.

La idea original fue enfocar el videojuego sólo en la memoria de él usuario, se diseñó la ambientación y los personajes para este fin.

El primer nivel sólo tendría 3 imágenes con las figuras geométricas de colores correspondientes a lo propuesto desde el inicio del desarrollo del proyecto. Aparecería en pantalla un tablero con un camino el cuál el niño tendría que memorizar, para luego repetir en un tablero vacío.

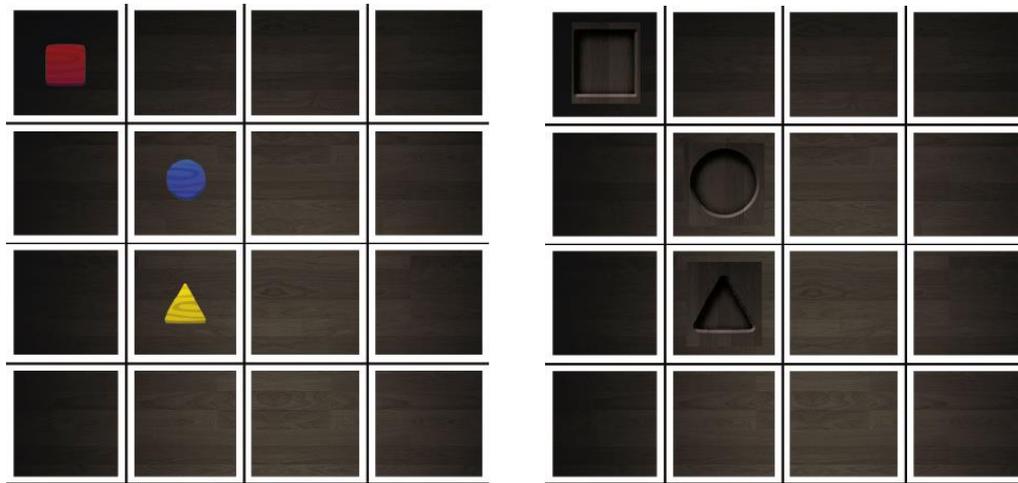


Figura 21 Ejemplo 1 primer diseño.

A medida que avanzará de nivel el camino de figuras sería más largo, y así crecería la curva de aprendizaje de forma coherente.



Figura 22. Ejemplo 2 primer diseño

Las figuras geométricas y los colores propuestos en el proyecto estarían combinados. Se diseñó un tablero con un color madera oscuro que le permitiera al usuario distinguir de forma sencilla los colores y la forma de cada imagen, y los conceptos propuestos estarían siendo correctamente usados en el videojuego.

Se probó con niños de 5 a 8 años los cuáles mostraron interés en la propuesta del videojuego. El problema estuvo en la aplicación del reconocimiento de voz, ya que el software tendría que reconocer casi simultáneamente dos palabras. Por ejemplo para las imágenes anteriores la primera casilla se llenaría si el niño dice "circulo azul", lo cuál genero muchos problemas y se decidió cambiar la forma del juego.

Teniendo en cuenta el problema del diseño anterior se decidió que los conceptos a repasar tenían que ser personajes independientes.

En este caso el reto estaría en que cuando el niño viera las figuras geométricas se centrara en su forma y no se distrajera con el color lo cuál podría confundirlo para lograr la finalidad del videojuego. Y de la misma forma con los personajes de los colores, pero en este caso darle más importancia al color y no a la forma.

El juego en el que se basó el siguiente diseño se llama “fruit ninja” el cuál consta de varios personajes en forma de frutas que están cayendo aleatoriamente durante cierto tiempo propuesto en los diferentes niveles que lo componen, el objetivo del usuario es “destruir” las frutas con sus dedos antes de que caigan al fondo de la pantalla, gracias a la pantalla touch el juego simula el contacto con cada personaje.

La propuesta fue hacer lo mismo pero en éste caso con los colores y figuras geométricas propuestas anteriormente y que la interacción no fuera con las manos del usuario si no con su voz. En éste caso destruirá la imagen hasta que diga la palabra correspondiente al color o a la figura geométrica que se encuentre en ese momento en pantalla.

Las figuras geométricas fueron diseñadas con colores secundarios (naranja, verde y fucsia), pero sólo en el borde de la imagen para así lograr el objetivo de darle la prioridad en este caso a la forma y no al color.



Figura 23. Imagen figuras geométricas

Los personajes para los colores primero fueron una especie de mancha de pintura, pero al momento de mostrarlo a un niño, él se sentía confundido por la forma en que caían los colores.

Aprovechando el concepto del videojuego, en donde las imágenes iban a estar cayendo aleatoriamente se decidió hacer una forma de gotas para los colores, ya que este concepto para el niño ya es familiar por la lluvia y sería un diseño sencillo el cual permitía al niño no confundirse con la forma y así reconocer de manera rápida el color.



Figura 24. Imagen colores

Ya que se tenía en el anterior diseño un fondo oscuro para resaltar las imágenes, se insistió en este concepto pero con forma de nubes para darle un contexto al jugador que conoce y lo puede relacionar bien con algo con lo que esta familiarizado (lluvia), aprovechando la forma de gotas de los colores y que van a estar cayendo.



Figura 24. Personajes del juego.

Dado el diseño del juego se propusieron varios nombres los cuáles tuvieron relación con el objetivo de cada nivel, el cuál es destruir imágenes pero teniendo en cuenta que en este caso la interacción es con la voz.

Al escoger el tipo de letra nos basamos en una forma que fuera amigable con el niño. Después de probar algunas se decidió en una letra despegada para mayor entendimiento y con un tono de tiza, cómo se veía en un tablero antiguo.



Figura 25. Diseño de pick up

Como es un juego educativo la primera palabra que se puso en la lista para el nombre fue Learn. Luego se tuvo en cuenta el objetivo que era destruir las imágenes y se añadió la palabra crush, pero en medio de la búsqueda apareció la palabra pick up la cuál hace referencia en español a aprender, captar, adquirir, lo cual es la característica más relevante del videojuego y da la impresión al niño que debe “atrapar algo” esto dicho por los jugadores con los cuales se probó el juego y si hizo una votación para el nombre final.

Título: Pick Up
Logo:



Pick Up es un videojuego educativo, el cuál ésta enfocado en tres figuras geométricas (círculo, cuadrado, triángulo) y tres colores (amarillo, azul, rojo).

Actualmente los niños de primera infancia tienen la oportunidad de tener dispositivos electrónicos cerca y en la mayoría de los casos a su total disposición, ya que la globalización se ha encargado de que en cada hogar sea casi una necesidad acceder a este tipo de herramientas.

Aprovechando esta etapa de desarrollo de los niños dónde están aprendiendo a reconocer las diferentes palabras, objetos y sensaciones que en su diario los rodean; se puede llegar a generar un vínculo con él dispositivo electrónico que le ayude al niño a reforzar didácticamente algunos conocimientos, ya que el videojuego le permite explorar su funcionamiento jugando e interactuando, a la vez que se va fomentando un pensamiento y una solución para cada reto propuesto. Partiendo del construccionismo (filosofía educativa desarrollada por

Seymour Papert), que se basa en la construcción activa de conocimiento a partir de la experiencia. (Constructivismo de Piaget). [37]

A continuación se hace una breve descripción de las herramientas utilizadas en el juego:

COLORES:



Figura 26. Colores

FIGURAS GEOMETRICAS:

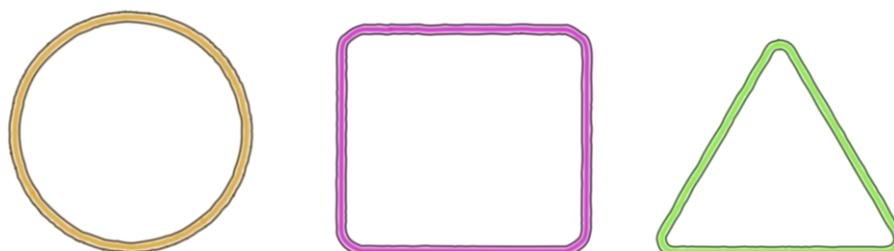


Figura 27. Figuras

Pertenciente a un contexto donde el niño ya conoce las figuras geométricas y los colores, el objetivo del juego será afianzar conocimientos basándonos en la repetición y en la interacción didáctica con el dispositivo que será por medio de la voz.

7.1.1. Atmósfera del Videojuego

Ya que es un juego educativo que se centra específicamente en tres colores (amarillo, azul, rojo) y en tres figuras geométricas (círculo, cuadrado, triángulo)

se pensó en un diseño sencillo que consta de un fondo oscuro, que permite a cada objeto ser reconocido de forma inmediata en el momento que aparezca en pantalla.

Las imágenes de los colores y las figuras geométricas estarán cayendo aleatoriamente dependiendo del nivel en el que él usuario se encuentre, se optó por darle forma de gotas a los colores como si estuviera lloviendo y formas de nubes al fondo para darle un contexto al jugador.



Figura 28. Atmósferas del videojuego

7.1.2. Jugabilidad

Pick Up es un juego en primera persona, sólo permite tener un jugador al tiempo, éste debe pronunciar la palabra correspondiente al color o a la figura geométrica que en ése momento esté en pantalla para así destruirla y lograr obtener puntos, la idea principal es obtener una puntuación de 10 en menos del tiempo propuesto en cada nivel.

Es importante aclarar que al igual que con otros juegos de este género se hace uso de controladores que permitan una interacción más atractiva para el jugador con el videojuego, en éste caso específicamente se usará el reconocimiento de la palabra dándole así facilidad al usuario para jugar, ya que en cada nivel necesitará sólo de su voz.

Se debe tener en cuenta que el videojuego cuenta con una fase de repaso en la que el jugador podrá ver y escuchar las instrucciones para conocer el funcionamiento de cada botón correspondiente a la previa grabación de cada palabra, para así poder empezar cada nivel.

Para diseñar la curva de aprendizaje del usuario se hicieron 3 niveles, la dificultad del nivel colores y del nivel figuras geométricas es la misma ya que en estos dos niveles, él usuario sólo estará reconociendo colores (amarillo, azul, rojo), o figuras geométricas (cuadrado, círculo, triángulo). La dificultad aumenta en el nivel figuras geométricas y colores, ya que en este se combinan las seis imágenes y el usuario tendrá que identificar cada una combinando los conceptos anteriormente repasados en los anteriores niveles, esto se describe de la siguiente manera:

- **NIVEL DE REPASO:**

Acá el jugador podrá ver y escuchar como es el procedimiento que debe llevar a cabo para que el juego funcione correctamente. Primero la selección del micrófono, luego que botón es el indicado para grabar el ruido de fondo, posteriormente cuál botón oprimir mientras graba cada comando y finalmente donde está la opción de “jugar” para iniciar el nivel. Tendrá la opción de escuchar cada palabra antes de grabarla para así tener un ejemplo de cómo debe ser pronunciada, ya que de esto depende la correcta jugabilidad en cada nivel.

- **PRIMER NIVEL:**

Nivel Figuras geométricas:

Las figuras geométricas se diseñaron pensando en que no fueran de los mismos colores en los que el juego se enfoca, por esto se optó por colores secundarios los cuales con el fondo oscuro resaltan y le permite al jugador concentrarse en la figura y no en el color.

Las figuras geométricas (cuadrado, círculo y triángulo) caerán aleatoriamente durante dos minutos, el usuario tendrá que decir la palabra correspondiente a la imagen que esté en pantalla y así sumar puntaje hasta lograr llegar a 10.

- **SEGUNDO NIVEL:**

Nivel Colores:

Se diseñaron imágenes con forma de gotas para que el usuario logre concentrarse en el color que está representando dentro estas y no confundirse con otros objetos que puedan llamar más su atención.

Las gotas de colores van a ir cayendo aleatoriamente con intervalos de tiempos diferentes, el usuario tendrá que decir la palabra del color correspondiente que este en ese momento en pantalla para así destruirla y sumar puntaje. Para completar el objetivo tendrá 2 minutos de juego.

- **TERCER NIVEL:**

Nivel Figuras geométricas y Colores:

En este nivel el jugador tendrá que estar más concentrado ya que caerán aleatoriamente colores y figuras geométricas y se pondrá a prueba los conocimientos repasados en los anteriores niveles.

Para ver más sobre el diseño del videojuego, Anexo A

7.2. Parámetros de estudio de las palabras a reconocer

Para que un sistema logre hacer un reconocimiento de voz se debe llevar a cabo un análisis de todas las características posibles de la señal de entrada y así tener suficiente información para poder comparar cualquier otra señal y luego poder hacer una correlación que permita el reconocimiento.

El software PRAAT permite hacer un completo análisis de cualquier señal de entrada, y así dar pautas para usar sus características de la forma que le interese al usuario.

Los Formantes al ser picos de intensidad en el espectro de una señal o en otras palabras, una concentración de energía en donde se puede ver la mayor amplitud de la onda en determinada frecuencia, es uno de los parámetros que le permite al usuario diferenciar las palabras, ya que cada sonido del habla humana tiene una distribución energética diferente entre los diferentes formantes, lo cual permite clasificar y categorizar. El análisis de las palabras en este trabajo se centró en el análisis de los formantes de cada una.

Palabras a reconocer:

Colores: Amarillo, azul y rojo.



Figura 29. Colores a identificar

Figuras Geométricas: Círculo, cuadrado y triángulo.



Figura 30. Figuras a identificar

Los formantes varían según el emisor y las condiciones acústicas del recinto dónde se realice la captura de las palabras, vale aclarar que en éste caso para un previo análisis de los formantes contenidos de las palabras a trabajar se realizaron tres grabaciones con diferente condiciones.

La primera, con la voz de una mujer de 25 años en el estudio digital de la universidad de San Buenaventura con un micrófono AKG c414.

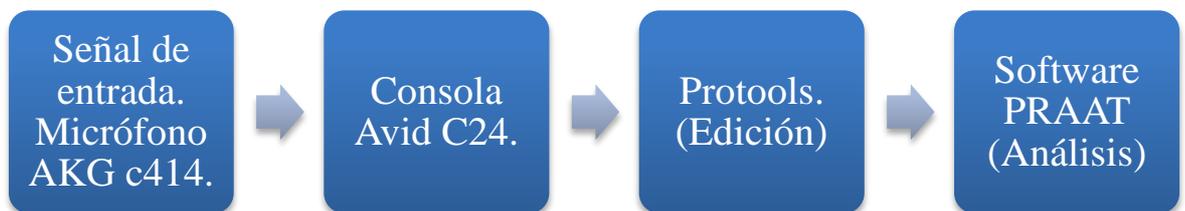


Figura 31. Diagrama flujo de señal

El propósito de esta grabación fue tener una señal lo más fiel posible, evitando en gran porcentaje que fuera alterada por condiciones externas.

El micrófono condensador AKG c 414 cuenta con una sensibilidad de 23 mV/Pa y ofrece nueve patrones polares para permitir una captura fiel a cada aplicación que se le necesite dar. Se escogió para éste caso tener el micrófono en figura 8, ya que al revisar sus características este patrón polar es donde el micrófono muestra una respuesta en frecuencia más plana que el resto.

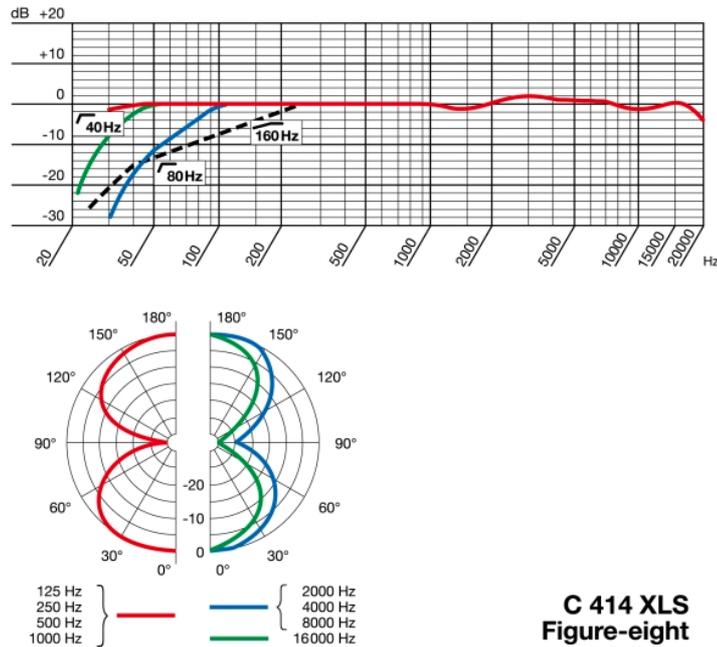


Figura 32. Grafica de figura 8 y respuesta en frecuencia.

La segunda grabación fue hecha por un niño de 8 años para hacer el análisis de formantes más cercano al usuario real que va a tener el software y a las frecuencias que va a tener que responder adecuadamente. Se realizó en el salón de Post- producción de la universidad de San Buenaventura, sede Bogotá con un micrófono de condensador Beta 87c que cuenta con una sensibilidad de 2mV, fue escogido ya que su uso normal está dirigido a la voz, es decir cubre el rango de frecuencias necesario para una respuesta buena en el caso de esta aplicación.

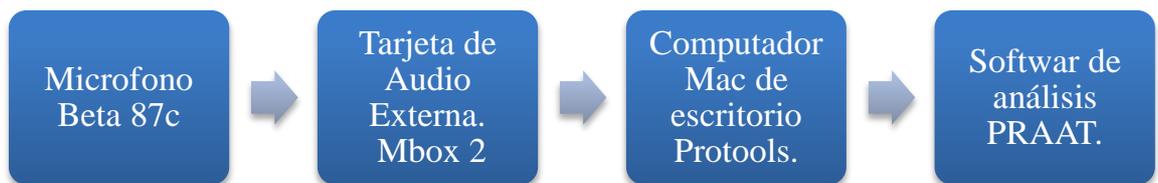


Figura 33. Diagrama de flujo con micrófono Beta 87c

Consta de un patrón polar cardiode el cual da una buena direccionalidad al usuario en este caso.

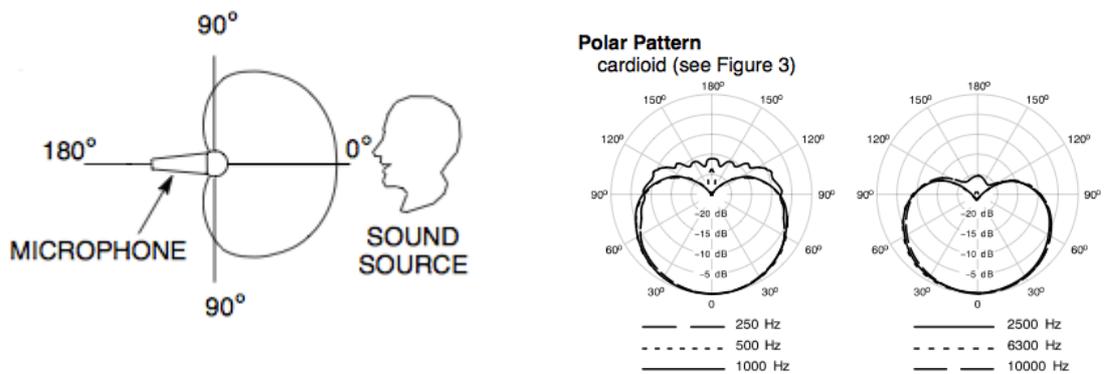


Figura 34. Patrón polar cardioide

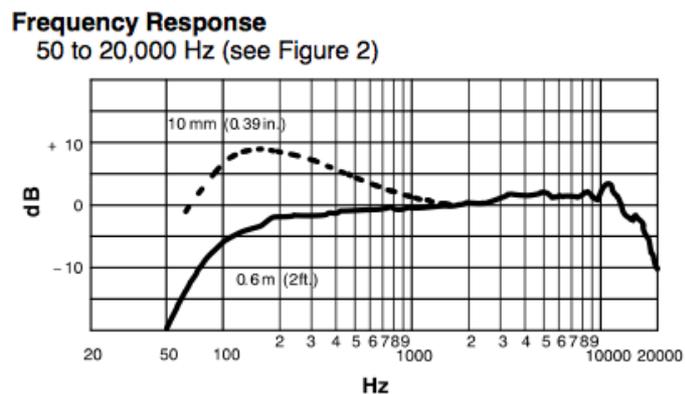


Figura 35. Respuesta en frecuencia micrófono cardioide

Sus características ofrecen una respuesta adecuada para frecuencias medias, las cuáles son las que se utilizarán en el caso de este videojuego ya que el usuario lo maneja por medio de su voz, por esto fue escogido para las pruebas de formantes que se realizaron.

La grabación tres fue realizada por el micrófono integrado de un Mac book pro, ya que será el sistema de real interacción con el usuario.



Figura 36. Diagrama de flujo sin micrófono externo

Las únicas especificaciones técnicas que da el fabricante sobre el audio del computador son:

Audio

- Altavoces estéreo con refuerzo de graves
- Micrófono omnidireccional
- Puerto para auriculares
 - Compatible con los auriculares con mando y micro de Apple para el iPhone
 - Compatible con salida de audio (digital/analógica)

[43]

Las cuales no son suficientes para saber si es un sistema de interacción adecuado para el reconocimiento de voz aplicado al desarrollo del videojuego.

Para hacer la caracterización de este micrófono se realizó una medición en el estudio 5.1 de la Universidad de San Buenaventura, Sede Bogotá. La cuál pretendió dar características cómo, patrón polar y respuesta en frecuencia.

Las mediciones se intentaron hacer en un lugar lo más aislado posible del ruido. Por esto fueron realizadas en el estudio 5.1 de la Universidad de san Buenaventura, sede Bogotá.

Aunque el ruido de fondo es bajo (45.6 dB) según la normativa tendría que ser superado en 35 dB la emisión de la fuente. Pero al no conocer exactamente el valor de sensibilidad del micrófono integrado del Macbook Pro se decide no superar los 80dB para no saturarlo.

En la gráfica de respuesta en frecuencia se puede ver que en bajas y altas frecuencias el micrófono no tiene una respuesta buena, pero como en este proyecto importa es el rango medio, donde se comprende la voz, aunque no es plana su respuesta si es lo suficientemente buena para usarlo como sistema de interacción para el videojuego (mientras se controle en gran medida el ruido de fondo y las características del hablante, claridad y pronunciación)



Figura 37. Medición características micrófono Mac BookPro

Ruido de fondo	45,6dB		
Fuente (1KHz)	80dB		
Frecuencia	100Hz	1KHz	10KHz
0° - 360°	-30	-24	-51
15°	-42	-20	-60
30°	-42	-33	-60
45°	-40	-30	-61
60°	-43	-30	-64
75°	-53	-33	-60
90°	-52	-42	-65
105°	-50	-45	-69
120°	-47	-47	-69
135°	-55	-42	-75
150°	-54	-40	-75
165°	-50	-47	-78
180°	-46	-51	-73
195°	-45	-49	-75
210°	-43	-40	-70
225°	-42	-43	-72
240°	-53	-51	-69
255°	-47	-49	-69
270°	-52	-42	-65
285°	-53	-33	-62
300°	-43	-30	-64
315°	-40	-30	-62
330°	-42	-33	-60
345°	-43	-20	-60

Tabla 1. Datos patron polar

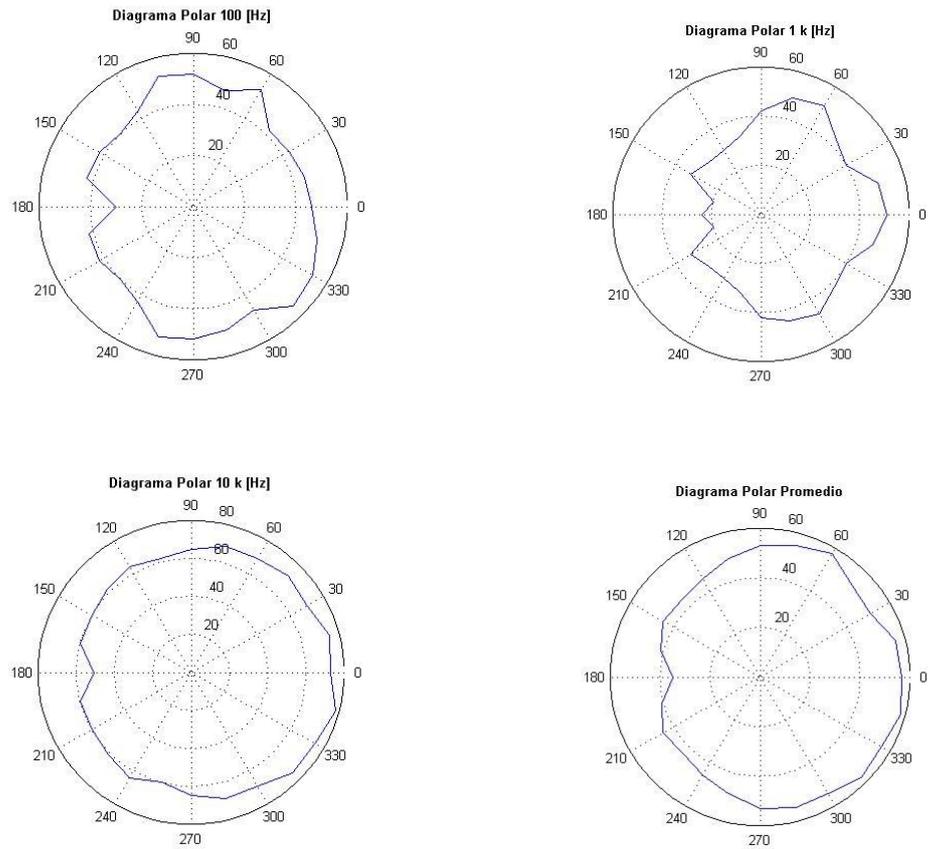


Figura 38. Diagrama polar micrófono Mac BookPro

RESPUESTA EN FRECUENCIA	
	1m
20Hz	-68
31.5Hz	-64
65Hz	-78
125Hz	-42
250Hz	-33
500Hz	-15
1KHz	-22
2KHz	-24
4KHz	-30
8KHz	-51
16KHz	No aplica

Tabla 2. Datos respuesta en frecuencia

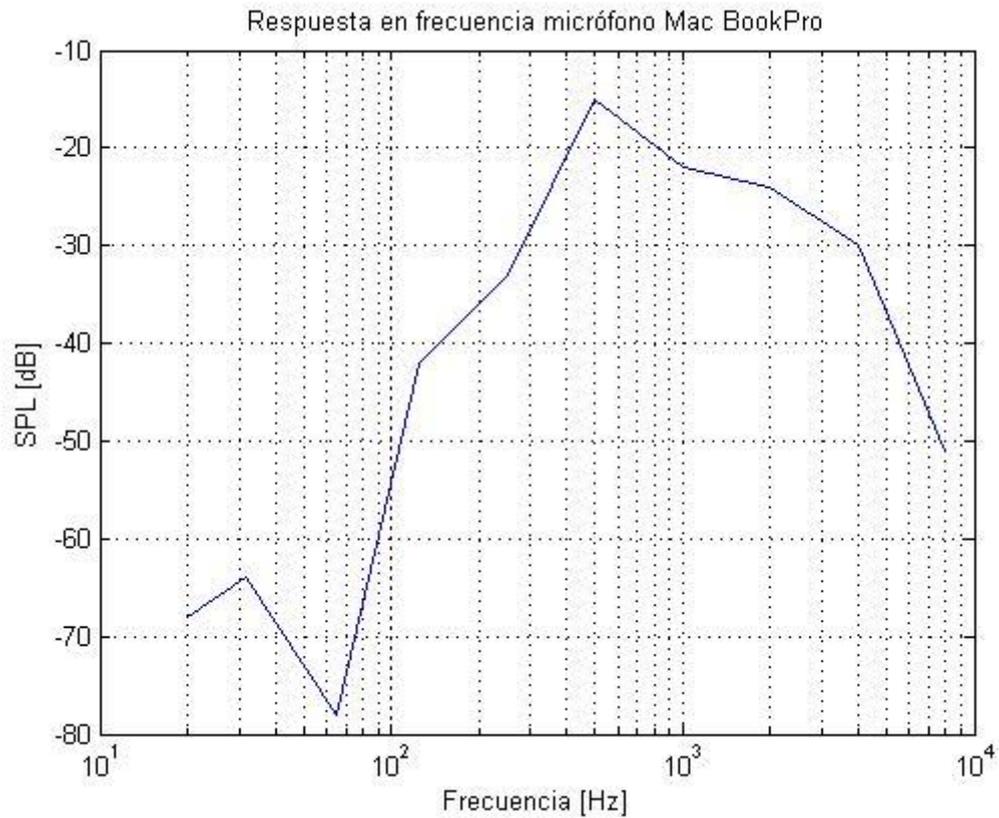


Figura 39. Respuesta en frecuencia micrófono Mac BookPro

El siguiente es el código utilizado en Matlab para graficar las figuras anteriores.

```

microfonointegrado(25,:) = microfonointegrado(1,:);
t100 = microfonointegrado(:,1);
t100 = t100';
t1k = microfonointegrado(:,2);
t1k = t1k';
t10k = microfonointegrado(:,3);
t10k = t10k';
teta = linspace(0,2*pi,25);

figure (1)
polar(teta,t100);

hold on

    title('Diagrama Polar 100 [Hz]','FontWeight','bold');

    grid on

    figure (2)
polar(teta,t1k);

hold on

    title('Diagrama Polar 1 k [Hz]','FontWeight','bold');

```

```

        grid on
        figure (3)
polar(teta,t10k);

hold on

        title('Diagrama Polar 10 k [Hz]', 'FontWeight', 'bold');

        grid on

for i=1:25
    a=[t100(1,i),t1k(1,i),t10k(1,i)];
    prom(1,i)=mean(a);
end

figure (4)
polar(teta,prom);

hold on

        title('Diagrama Polar Promedio', 'FontWeight', 'bold');

        grid on

        figure (5)
frec=[20 31.5 65 125 250 500 1000 2000 4000 8000];
respfrec=[-68 -64 -78 -42 -33 -15 -22 -24 -30 -51];
semilogx(frec,respfrec), title('Respuesta en frecuencia micrófono
Mac BookPro')
xlabel('Frecuencia [Hz]')
ylabel('SPL [dB]')
grid on

```

7.2.1. Análisis de formantes de las palabras correspondientes a los Colores:

Las frecuencias presentadas en las siguientes tablas son una referencia de los formantes correspondientes a cada vocal, pero su valor puede variar según la edad, género o condiciones acústicas del recinto en el que se captura la señal. Sin embargo, son una buena referencia pues los valores estandarizados oscilan cerca del valor mencionado.

Se realizaron tres distintas grabaciones, con condiciones, transductores y usuarios diferentes para lograr verificar los Formantes que componen cada palabra que va a reconocer el videojuego. Otro objetivo de la siguiente comparación es hacer visible como afectan las condiciones externas y cómo depende de estas el correcto análisis de una señal de entrada.

7.2.1.1. Análisis de la palabra Amarillo.

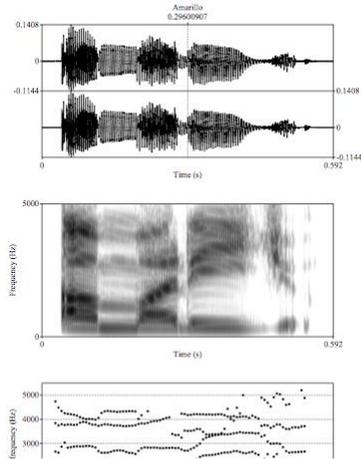


Figura 42. Grabación en el Estudio con un micrófono AKG c414. Mujer de 25 años. - Amarillo

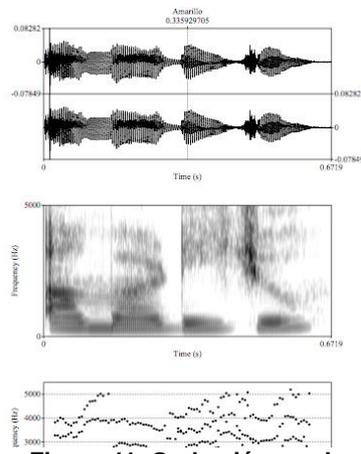


Figura 41. Grabación en el Salón de Post- producción. USB con un micrófono Beta 87c. Niño de 8 años - Amarillo

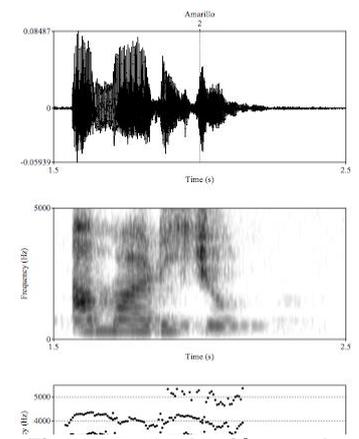


Figura 40. Grabación en el Salón de Post- producción. USB con un Micrófono Beta 87c. Niño de 8 años - Amarillo

VOCAL	Primer Formante	Segundo Formante
/a/	600Hz	1200Hz
/i/	200Hz	2200Hz
/o/	400Hz	900Hz

Tabla 3. Vocales: a,i,o.

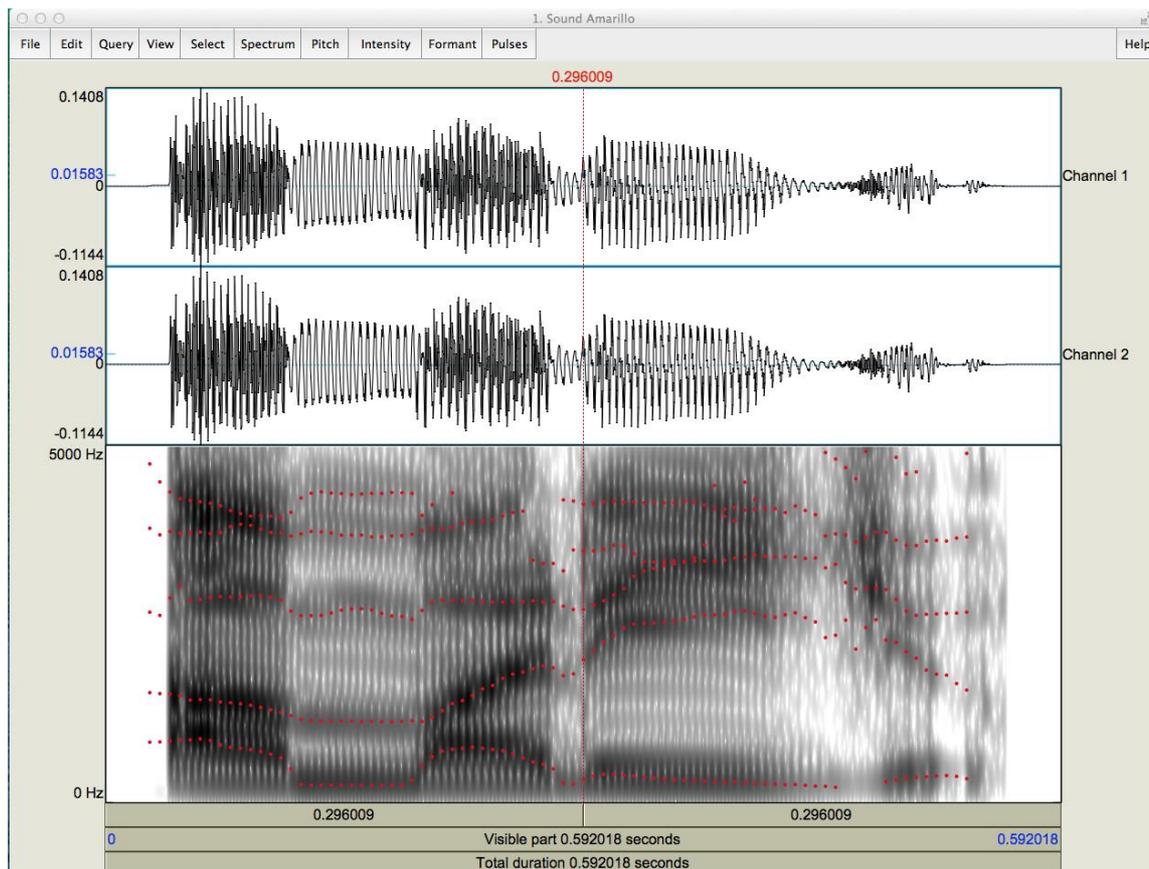


Figura 43. Espectro- Formantes. Palabra amarillo

En la primera parte de la imagen se puede observar la forma de onda de la señal, la cual tiene partes en las cuales es más notorio la concentración de energía. En la parte de abajo de la imagen se encuentran los fonemas que componen la palabra Amarillo.

Se puede observar que en el primer intervalo de tiempo está la vocal /a/ repetida dos veces, la cual está compuesta por el primer formante más o menos en 600 [Hz] y el segundo formante en una frecuencia mucho más aguda con un valor mayor a 1200 [Hz]. En el segundo intervalo de tiempo se vuelve a notar una suma energética considerable, distinguiendo la vocal /i/, la cual está compuesta por el primer formante en 200 [Hz] y el segundo en 2200 [Hz] aproximadamente. Por ultimo aunque menos notorio encontramos la vocal /o/ formada por el primer formante en 400 [Hz] y el segundo formante cerca de 900 [Hz].

7.2.1.2. Análisis de la palabra Azul.

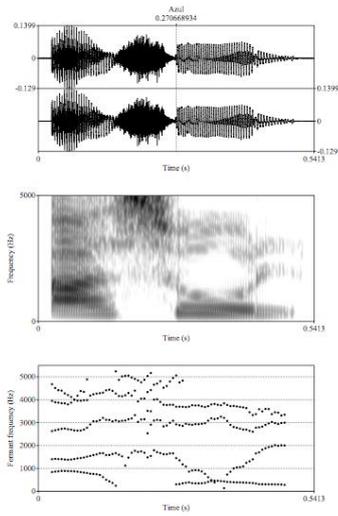


Figura 45. Grabación en el Estudio con un Micrófono AKG c414. Mujer de 25 años. - Azul

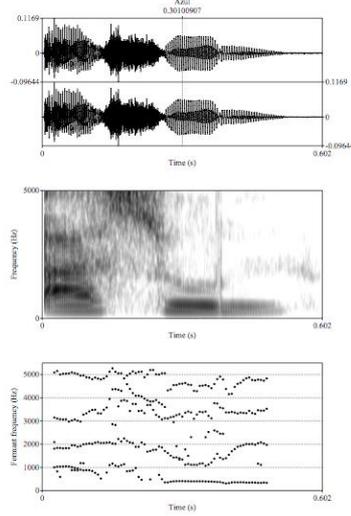


Figura 46. Grabación en el Salón de Post-producción. USB con un Micrófono Beta 87c. Niño de 8 años - Azul

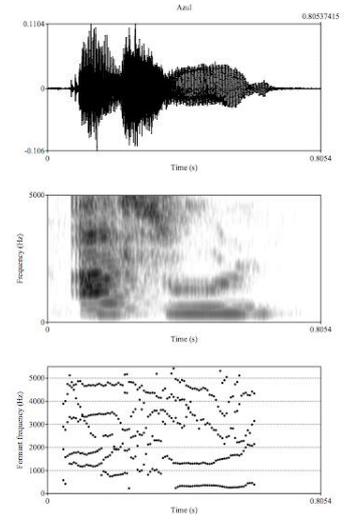


Figura 44. Grabación en el Salón de Post-producción. USB con un Micrófono Beta 87c. Niño de 8 años - Azul

VOCAL	Primer Formante	Segundo Formante
/a/	600Hz	1200Hz
/u/	300Hz	900Hz

Tabla 4. Vocales: a, u.

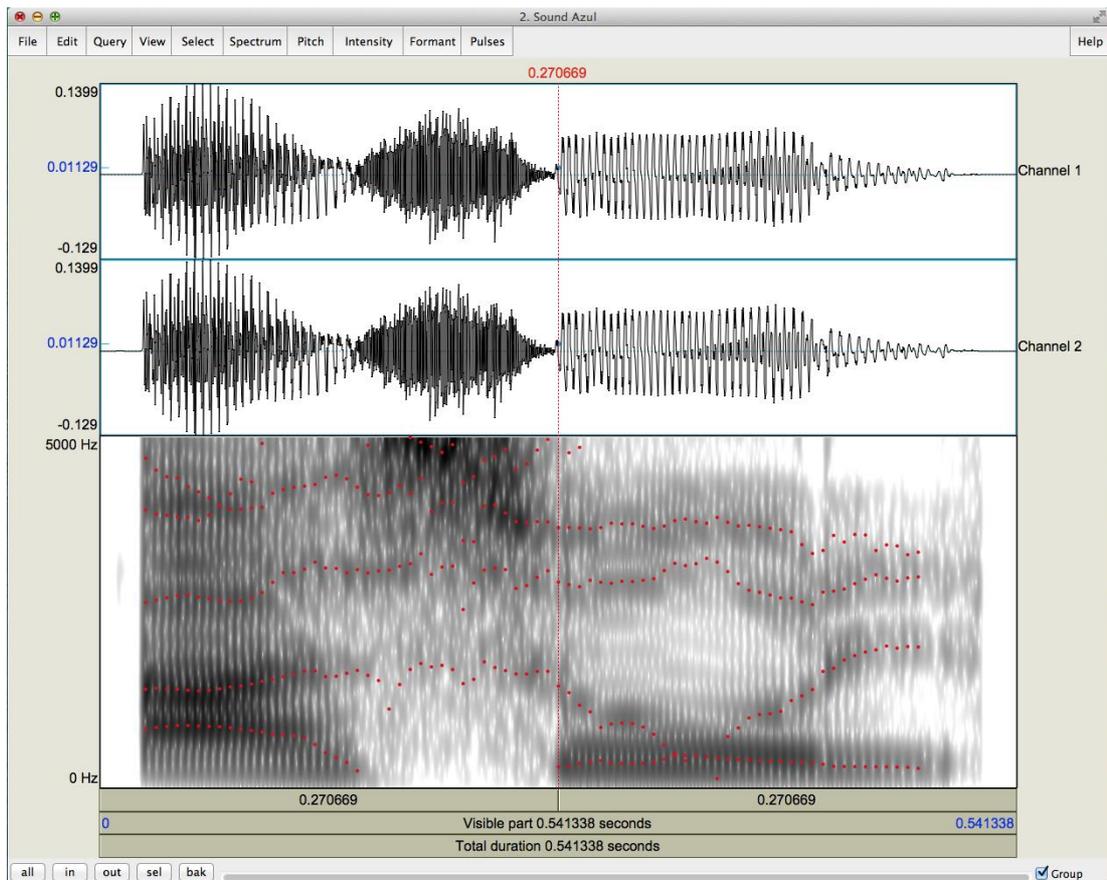


Figura 47. Espectro- Formantes. Palabra Azul.

La palabra Azul está compuesta por dos vocales, /a/ la cual en este caso se nota más su primer formante, obteniendo concentración energética en 600 [Hz] y /u/ que tiene su primer formante en 300 [Hz] y el segundo en 900 [Hz].

7.2.1.3. Análisis de la palabra Rojo.

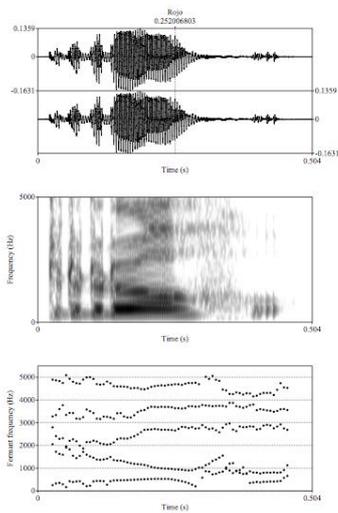


Figura 50. Grabación en el Estudio con un Micrófono AKG c414. Mujer de 25 años. - Rojo

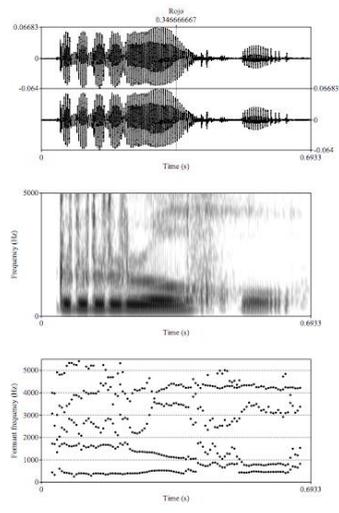


Figura 49. Grabación en el Salón de Post- producción. USB con un Micrófono Beta 87c. Niño de 8 años - Rojo

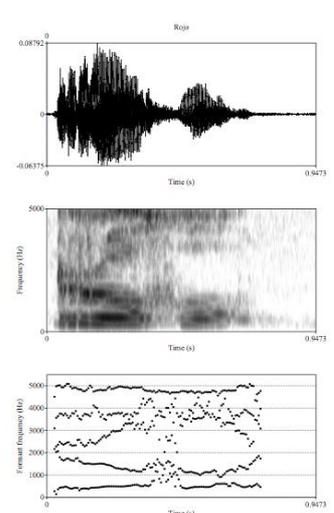


Figura 48. Grabación en el Salón de Post- producción. USB con un Micrófono Beta 87c. Niño de 8 años - Rojo

VOCAL	Primer Formante	Segundo Formante
/o/	400Hz	900Hz

Tabla 5. Vocales: o.

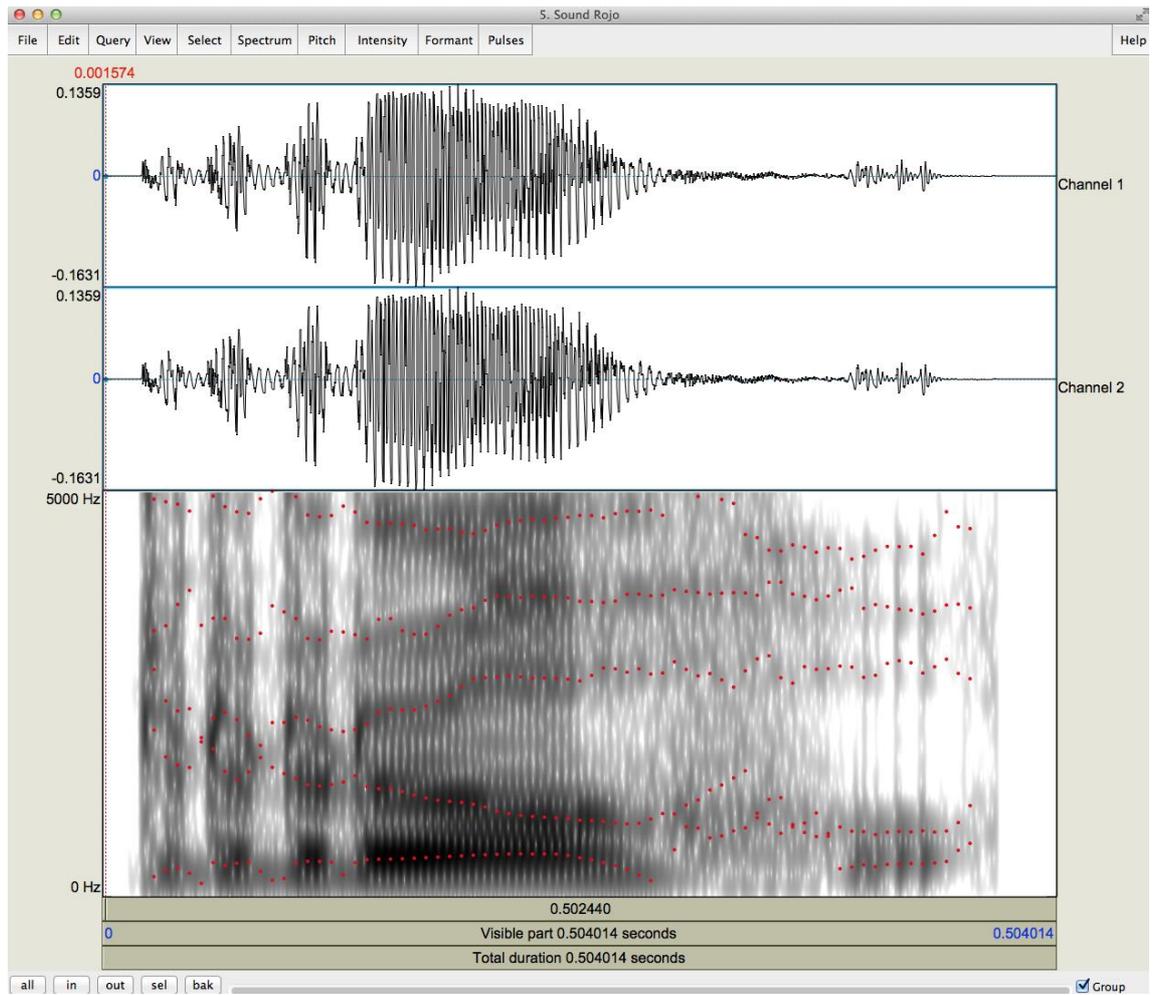


Figura 51. Espectro- Formantes. Palabra Rojo.

En el caso de la palabra Rojo solo se tiene una vocal, la cual está presente a partir del segundo intervalo de tiempo en la gráfica, como la consonante *j* no tiene mayor energía en comparación a las vocales, se ve como las dos /o/ parecen una sola, con el primer formante en 400 [Hz] y el segundo formante en 900 [Hz]

7.2.2. Análisis de formantes de las palabras correspondientes a las Figuras Geométricas:

7.2.2.1. Análisis de la palabra Círculo.

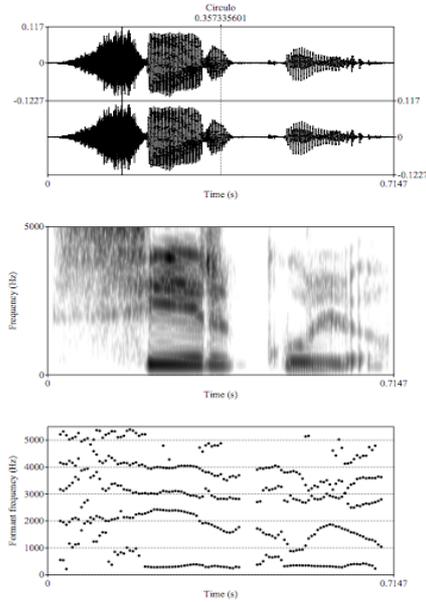


Figura 54. Grabación en el Estudio con un Micrófono AKG c414. Mujer de 25 años. – Círculo.

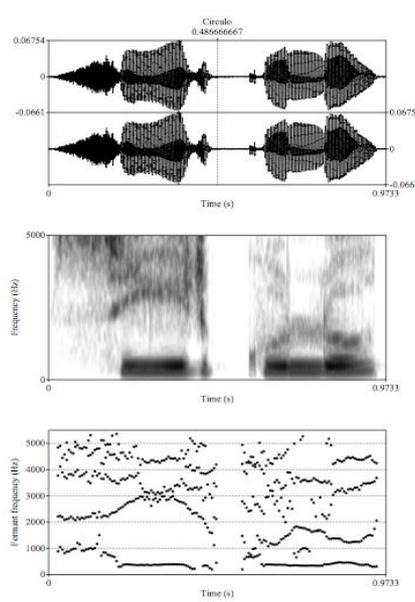


Figura 53. Grabación en el Salón de Post- producción. USB con un Micrófono Beta 87c. Niño de 8 años – Círculo.

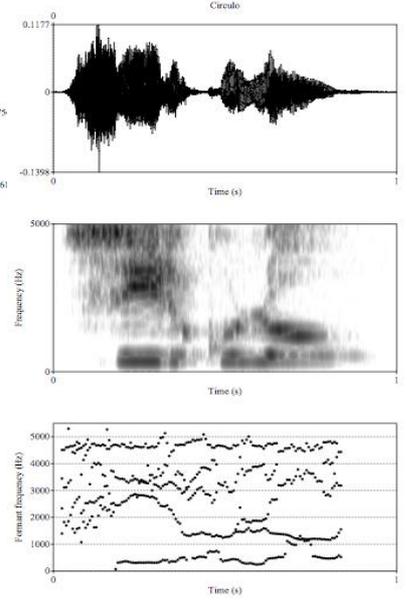


Figura 52. Grabación en el Salón de Post- producción. USB con un Micrófono Beta 87c. Niño de 8 años – Círculo.

VOCAL	Primer Formante	Segundo Formante
/i/	200Hz	2200Hz
/u/	300Hz	900Hz
/o/	400Hz	900Hz

Tabla 6. Vocales: i, u, o.

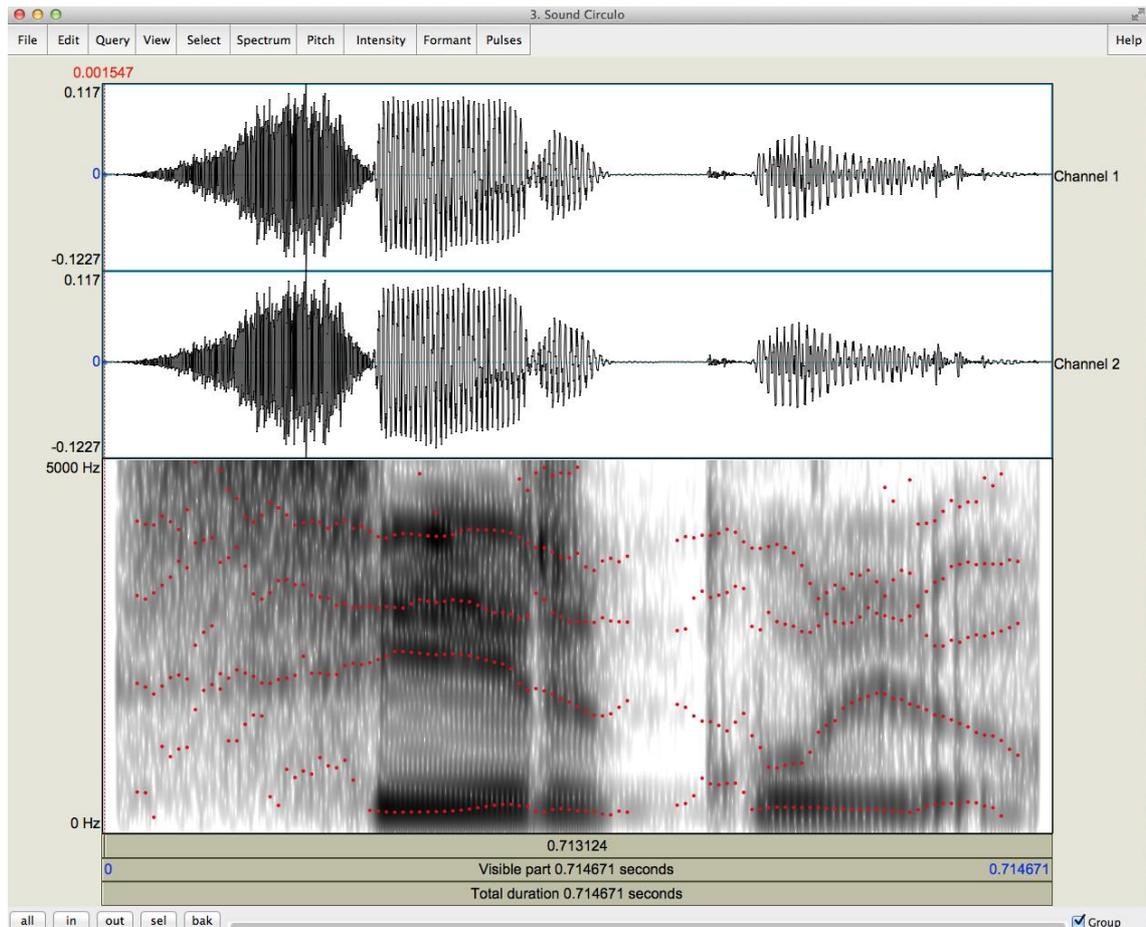


Figura 55. Espectro- Formantes. Palabra Círculo.

En esta imagen es evidente como las vocales contienen mucha más energía que las consonantes. La primera vocal que se encuentra en la palabra círculo es /i/ la cual tiene su primer formante en 200 [Hz] y el segundo aproximadamente en 2200 [Hz]. Inmediatamente se nota el cambio de vocal, no tanto en el primer formante si no en el segundo el cual si varia bastante. La vocal /u/ contiene su primer formante en 300 [Hz] y su segundo formante en 900Hz y por último la vocal /o/ la cual es muy parecida en sus formante a la /u/ pero en este caso se ve con menos energía.

7.2.2.2. Análisis de la palabra Cuadrado.

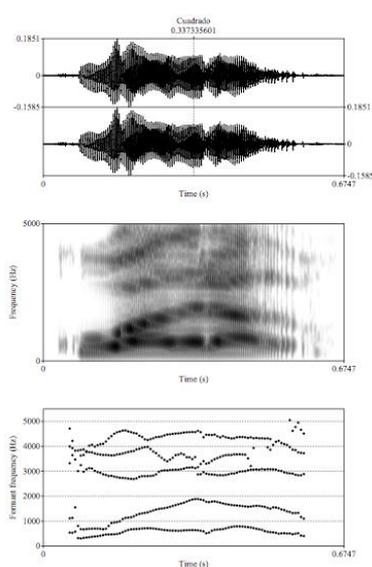


Figura 58. Grabación en el Estudio con un Micrófono AKG c414. Mujer de 25 años. – Cuadrado.

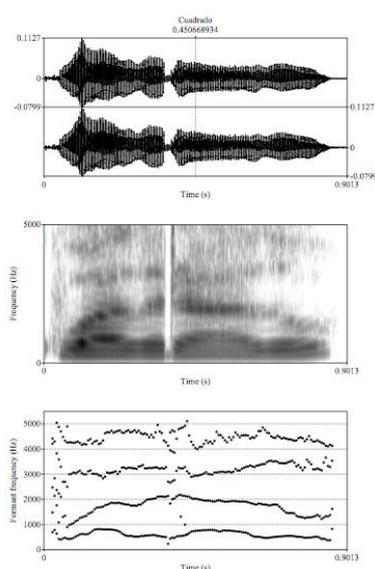


Figura 57. Grabación en el Salón de Post- producción. USB con un Micrófono Beta 87c. Niño de 8 años – Cuadrado.

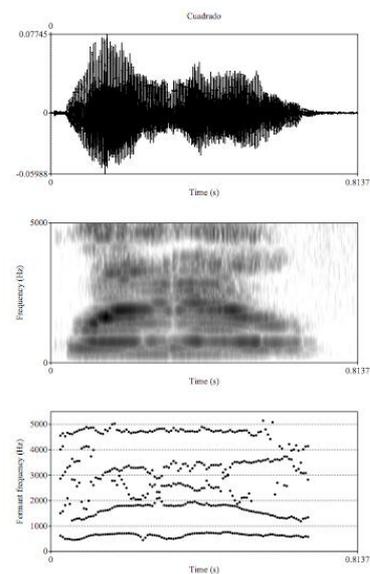


Figura 56. Grabación en el Salón de Post- producción. USB con un Micrófono Beta 87c. Niño de 8 años – Cuadrado.

VOCAL	Primer Formante	Segundo Formante
/u/	300Hz	900Hz
/a/	600Hz	1200Hz
/o/	400Hz	900Hz

Tabla 7. Vocales: u, a, o

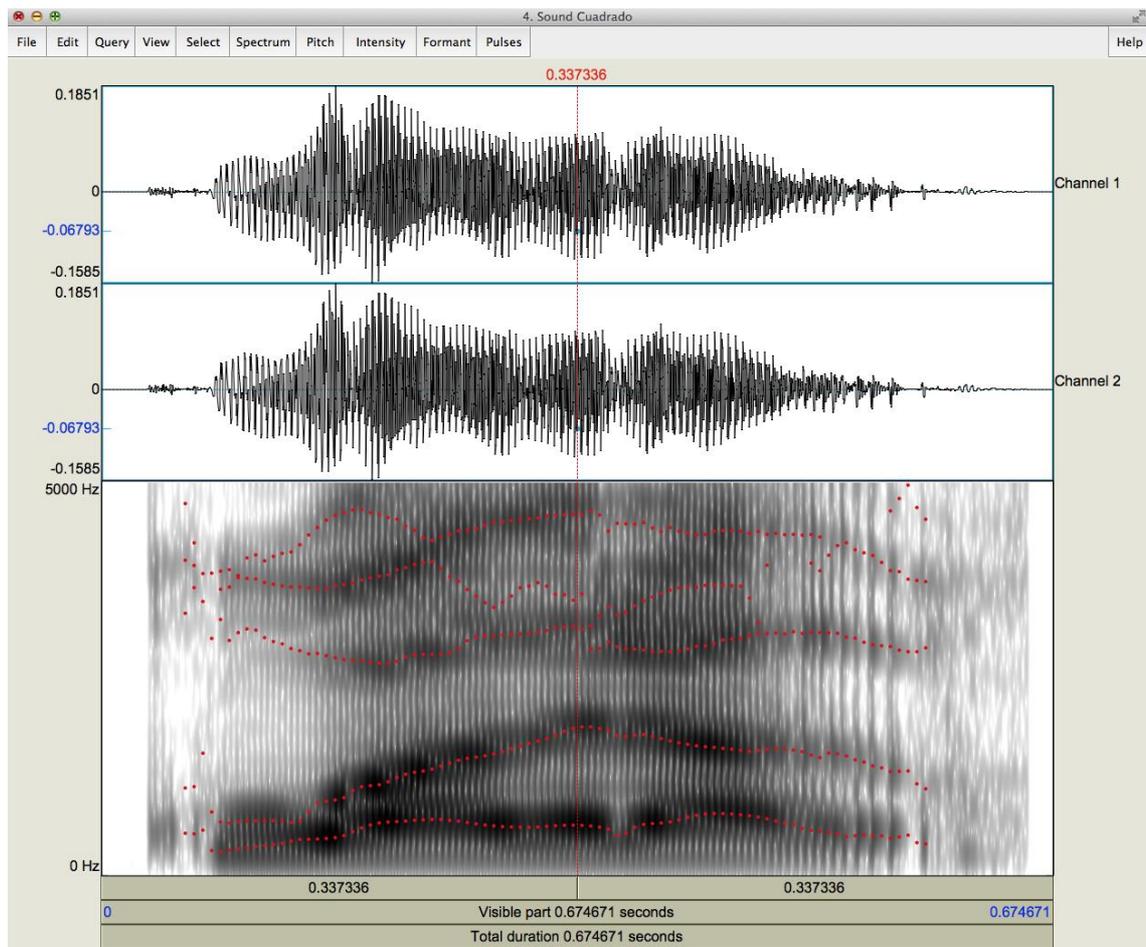


Figura 59. Espectro- Formantes. Palabra Cuadrado.

La palabra cuadrado al tener la vocales tan cerca la una a la otra al momento de pronunciarlas se ve en todo el espectro una cantidad energética considerable.

La /u/ se puede diferenciar ya que su primer formante es el más agudo de todos, en una frecuencia aproximada de 300 [Hz], luego se encuentra la vocal /a/ con el primer formante en 600 [Hz] y por último la /o/ la cual tiene su primer formante en medio de las dos anteriores (400 [Hz]).

7.2.2.3. Análisis de la palabra Triángulo.

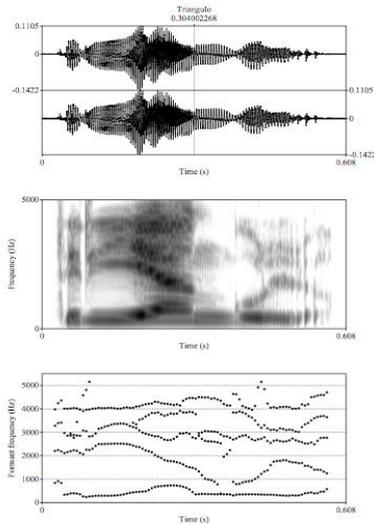


Figura 62. Grabación en el Estudio con un Micrófono AKG c414. Mujer de 25 años. – Triángulo.

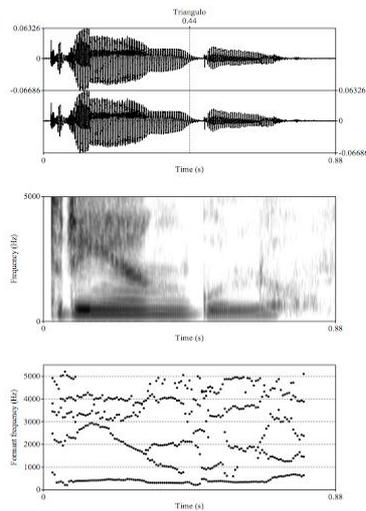


Figura 61. Grabación en el Salón de Post-producción. USB con un Micrófono Beta 87c. Niño de 8 años – Triángulo.

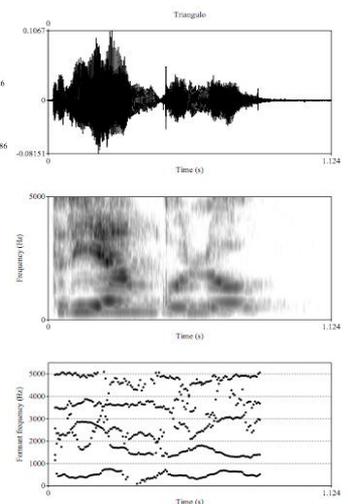


Figura 60. Grabación en el Salón de Post-producción. USB con un Micrófono Beta 87c. Niño de 8 años – Triángulo.

VOCAL	Primer Formante	Segundo Formante
/i/	200Hz	2200Hz
/a/	600Hz	1200Hz
/u/	300Hz	900Hz
/o/	400Hz	900Hz

Tabla 8. Vocales: i, a, u, o.

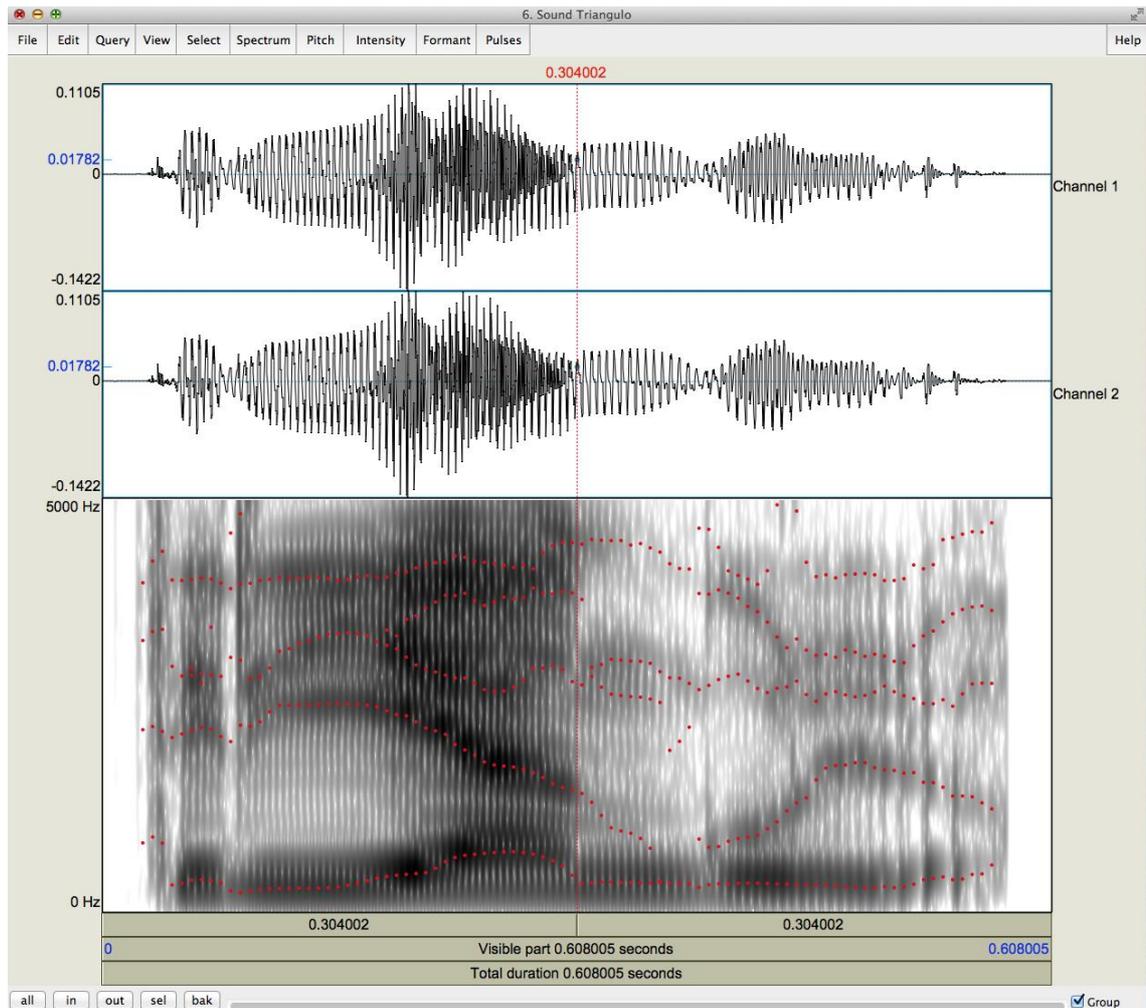


Figura 63. Espectro- Formantes. Palabra Triangulo.

De las palabras trabajadas en este proyecto triángulo es la que mayor cantidad de vocales contiene. En el primer intervalo de tiempo se puede ver la vocal /i/ por su primer formante tan bajo, en 200 [Hz], luego más o menos en la mitad de la gráfica se puede distinguir la vocal /a/, la cual tiene su formante en 600 [Hz] y en el último tramo de la señal se ven las vocales /u/ y /o/ muy parecidas, dado a sus formantes tan cercanas.

7.2.3. Cuadros comparativos entre diferentes capturas.

El siguiente cuadro comparativo permite ver como los formantes pueden variar en su intensidad, claridad o en sus valores dependiendo de la calidad de la captura y de la persona que haya grabado la señal de entrada.

Cómo se mostró anteriormente se realizaron tres grabaciones distintas (lugar, características acústicas, transductor, personas) para tener un punto de comparación y llegar a consolidar los formantes de los cuáles están compuestos cada palabra que se usará en el videojuego.

Voz \ Color	AMARILLO	AZUL	ROJO
<p style="text-align: center;">Voz femenina estudio</p>	<ul style="list-style-type: none"> • Al estar en un lugar con mejores características acústicas se ven con mayor facilidad las partes de mayor concentración energética. • Las dos vocales /a/ que contiene esta palabra se ven claramente sus formantes y se pueden diferenciar del resto de vocales de la palabra, • La vocal /o/ tiene concentración energética en frecuencias bajas, y por esto se diferencia del resto. 	<ul style="list-style-type: none"> • Aunque las vocales generalmente es en donde se concentra la mayor cantidad de energía en una palabra, no siempre se puede diferenciar sólo por las vocales. En este caso la consonante /z/ tiene una gran cantidad de energía en frecuencias altas que nos permite reconocer fácilmente esta palabra. • También se ven claros los formantes que componen la vocal /a/. • En la vocal /u/ lo que más resalta es su formante más bajo en 300Hz y por esto se puede diferenciar del resto. 	<ul style="list-style-type: none"> • Al ser una palabra compuesta sólo por una vocal se puede notar con mayor facilidad como es ésta la que tiene mayor energía durante la grabación. • La vocal /o/ está presente en casi toda la grabación, lo cual es notorio en el espectro que muestran las gráficas. • Los formantes están ubicados en frecuencias bajas.
<p style="text-align: center;">Voz niño micrófono Beta 86</p>	<ul style="list-style-type: none"> • Aunque en el espectrograma se ve con menor nitidez (por las características de la grabación) en la gráfica de los formantes podemos observar que al ser una voz de 	<ul style="list-style-type: none"> • Existe mayor ruido en esta grabación se puede observar ya que en la primera parte se ve una concentración energética diferente, pero aun así se 	<ul style="list-style-type: none"> • Se ve claramente el primer formante de la vocal /o/ que por ser la única vocal de esta palabra la permite reconocer con facilidad. • Por las características del micrófono y que

	<p>niño, habrá mayor concentración energética en frecuencias altas en comparación a la gráfica de la mujer.</p> <ul style="list-style-type: none"> • A pesar de las diferencias de las voces, los Formantes los cuáles conforman la palabra siguen siendo los mismos. 	<p>pueden ver los primeros formantes de la vocal /a/ teniendo en cuenta la referencia de la grabación anterior.</p> <ul style="list-style-type: none"> • El formante más bajo de la vocal /u/ sí se ve con claridad en esta grabación. 	<p>se usó una interfaz de audio de buena calidad la señal de entrada es adecuada para el análisis.</p>
Voz niño micrófono integrado iMac	<ul style="list-style-type: none"> • Esta grabación aunque es la que contiene mayor ruido no permite corroborar los valores de cada formante. • Da una buena respuesta en frecuencia lo que nos permite confirmar que éste micrófono es una buena herramienta de interacción para el videojuego. 	<ul style="list-style-type: none"> • Los formantes se ven de forma clara. Se logra diferenciar la vocal /a/ y la /u/. • Sigue habiendo una característica especial en esta palabra y es la gran concentración de energía en la consonante /z/. 	<ul style="list-style-type: none"> • Aunque en la señal el ruido genere concentración energética en frecuencia altas (5KHz). Se ve claramente que la mayor cantidad de energía esta en la parte de inferior de las gráficas, en donde están ubicados los formantes de la palabra Rojo.

Tabla 9. Tabla de comparación entre colores

Voz \ Figura	CIRCULO	CUADRADO	TRIANGULO
Voz femenina estudio	<ul style="list-style-type: none"> • En los primeros segundos de la grabación aunque se ve que existe alguna señal no se marca ninguna frecuencia con claridad. • La vocal /i/ tiene concentración energética en gran parte del espectro, lo cual permite diferenciarla 	<ul style="list-style-type: none"> • Las frecuencias de cada formante se ven con claridad en cada gráfica. Esto nos permite asegurar que el programa tiene una buena caracterización de la señal de entrada. • La concentración energética se ve en la mitad de la grabación ya que es allí donde se 	<ul style="list-style-type: none"> • Lo formantes de la vocal /i/ se marcan casi al principio de la señal, pegado empiezan los formantes de la vocal /a/, se parecen en su primer formante pero se logran diferenciar gracias al segundo. • Las dos últimas vocales de esta palabra también

	<p>rápidamente,</p> <ul style="list-style-type: none"> Las vocales /o/ y /u/ al ser vocales cerradas tienen su energía en frecuencias bajas, pero se logran diferencias gracias al segundo formante de cada una. 	<p>encuentran las vocales en esta palabra.</p>	<p>tienen un comportamiento parecido en su primer formante.</p> <ul style="list-style-type: none"> Esta palabra por tener la mayor cantidad de vocales es la más difícil de diferenciar para un reconocimiento de palabra
<p>Voz niño micrófono Beta 86</p>	<ul style="list-style-type: none"> Se ve con claridad el primer formante de la vocal /i/ en 200 Hz. Aunque todas las vocales de esta palabra tienen su primer formante con valores parecidos, se logran diferenciar gracias a sus segundos formantes y a la diferente distribución energética que existen entre ellos. 	<ul style="list-style-type: none"> El niño hizo más pausada la grabación, lo que generó un espacio en blanco en la señal de entrada, por consecuencia en el espectro que vamos a analizar. Se ven las frecuencias de los formantes que componen cada vocal de la palabra. La concentración energética se encuentra en un rango de frecuencias medio – bajas. 	<ul style="list-style-type: none"> En éste caso se marcan con menor claridad los segundos formantes. Comparado con la gráfica de la grabación anterior se puede diferenciar las concentraciones energéticas en cada parte de la señal. Hay gran presencia de frecuencias medio - bajas
<p>Voz niño micrófono integrado iMac</p>	<ul style="list-style-type: none"> Se nota la cantidad de frecuencias altas que genera el ruido del micrófono. De igual manera se logran diferenciar los formantes para cada vocal. 	<ul style="list-style-type: none"> Nuevamente presencia de frecuencias altas las cuales no son propias de la palabra. Se diferencian con claridad los primeros dos formantes de cada vocal que compone la palabra. 	<ul style="list-style-type: none"> Aunque los formantes no se ven de forma exacta, si se pueden diferenciar cada uno en las partes específicas donde se encuentran las vocales en la palabra.

Tabla 10. Tabla de comparación entre figuras

7.3. Desarrollo del videojuego

Se desarrolló el videojuego en la plataforma de programación Unity 3D la cual permite la creación de juegos en dos o tres dimensiones.

Para este caso se usó una vista en dos dimensiones dado el propósito del videojuego.

7.3.1. Requerimientos:

Requerimientos técnicos videojuego "Pick Up"		
	Hardware	Software
PC o Mac	Procesador Intel core 2 duo o compatible, 2 Gb RAM mínimo, 200mb en disco duro para instalación, conexión USB 2.0 libre, con periféricos (mouse, teclado).	Windows xp, 7 o Mac OsX 10.6 en adelante, flash player 10.2. ADOBE Air (dependiendo del sistema operativo los requerimientos de versión del AIR pueden variar).
Sonido	Sistema de sonido estereofónico de mediana potencia.	
Controlador	Reconocimiento de voz. Micrófono integrado o externo.	

Tabla 11. Requerimientos técnicos videojuego "Pick Up"

7.3.2. Funcionamiento del videojuego

Al abrir el videojuego, el usuario tendrá contacto visual con el menú de inicio, en donde podrá tomar la decisión de escoger entre dos opciones: **Jugar** y **salir**, que se ejecutarán al tener contacto con el botón de la interfaz gráfica que estará en ese momento en la pantalla.

En la escena después de seleccionar la opción "**Jugar**" se cargará la animación del Nivel de repaso, en el cual se le asesorará al niño primero como

seleccionar el micrófono que será la manera de interacción con el juego, luego se indicara cual es el botón para grabar el ruido de fondo, para así seguir con cómo debe grabar cada palabra para todos los niveles. Por último se mostrara el botón para iniciar el juego.

Cuando el usuario ya está familiarizado con los botones que tendrá que usar para grabar cada palabra, estará en el Nivel de Figuras geométricas, donde ya podrá aplicar lo sugerido.

En cada nivel se le da la opción al usuario de escuchar la palabra grabada, igual que de grabar las veces que sea necesario hasta que la palabra este correctamente pronunciada, y así iniciar cada nivel de la manera más óptima.

- Primero debe seleccionar el botón que se le muestra para acceder a un micrófono.

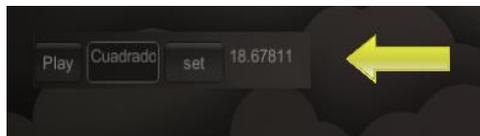


Figura 64. Imagen del botón para seleccionar un micrófono.

- Luego aparecerá el botón para grabar el ruido de fondo.



Figura 65. Imagen del botón para el ruido de fondo.

- Una vez grabado el ruido de fondo se podrán grabar las palabras correspondientes a los comandos las cuales serán usadas para la jugabilidad del videojuego, es decir para destruir la imagen que esté en pantalla.



Figura 66. Reconocimiento de voz.

Fuente propia

Nivel Figuras Geométricas.



Figura 67. Nivel Figuras Geométricas.

Fuente propia.

En el Nivel Figuras Geométricas se presentan los personajes dada por cada figura (Círculo, Cuadrado y Triángulo).

Cada vez que en el nivel el usuario diga la palabra correspondiente a la figura que esté en la pantalla esta será destruida y así se sumarán puntos debido a que el sistema reconocerá la palabra en el momento que el usuario la diga. Para este Nivel existe un total de 10 puntos posibles que se tendrán que conseguir en menos de dos minutos.

Si el niño no consigue completar el nivel aparecerá en pantalla la escena de **Perdiste**, donde le dará las dos opciones al usuario, **intentarlo de nuevo** o **salir**.

Cuando el nivel sea superado aparecerá la escena de **Ganaste**, donde tiene dos opciones, continuar con el **siguiente** nivel o **salir**.

Nivel Colores



Figura 68. Nivel Colores.

Fuente propia

Empieza el nivel Colores, en este nivel se presentan tres personajes, las gotas de colores (Amarillo, azul y rojo).

Ya que cada palabra fue previamente grabada el usuario tendrá que identificar y decir la palabra del color que esté en la pantalla en el intervalo de tiempo que la gota esté cayendo desde la parte superior de la pantalla hasta el momento justo antes de caer completamente, para así sumar puntos y poder seguir con los siguientes niveles.

Para este Nivel existe un total de 10 puntos posibles que se tendrán que conseguir en menos de dos minutos.

Si el niño no consigue completar el nivel aparecerá en pantalla la escena de **Perdiste**, donde el usuario podrá **intentarlo de nuevo** o **salir**.

Cuando el nivel sea superado aparecerá la escena de **Ganaste**, podrá continuar con el siguiente nivel o salir del juego.

Si la opción que se elige es continuar, llegara al último nivel del videojuego.

Nivel Colores y Figuras Geométricas



Figura 69. Nivel 3

Fuente propia

En el Nivel Colores y Figuras Geométricas se presentan los seis personajes, es decir todas las figuras geométricas y colores que el videojuego da para que el usuario pueda repasar conceptos.

En este nivel caerán aleatoriamente los colores y las figuras geométricas generándole al usuario un nivel mayor de concentración.

Para este Nivel existe un total de 10 puntos posibles que se tendrán que conseguir en menos de tres minutos.

Si el niño no consigue completar el nivel aparecerá en pantalla la escena de **Perdiste**, donde le dará dos opciones al usuario, **intentarlo de nuevo** o **salir**.

Cuando el nivel sea superado aparecerá la escena de **Ganaste**, el usuario tiene dos opciones, **repetir el nivel** o **salir**.

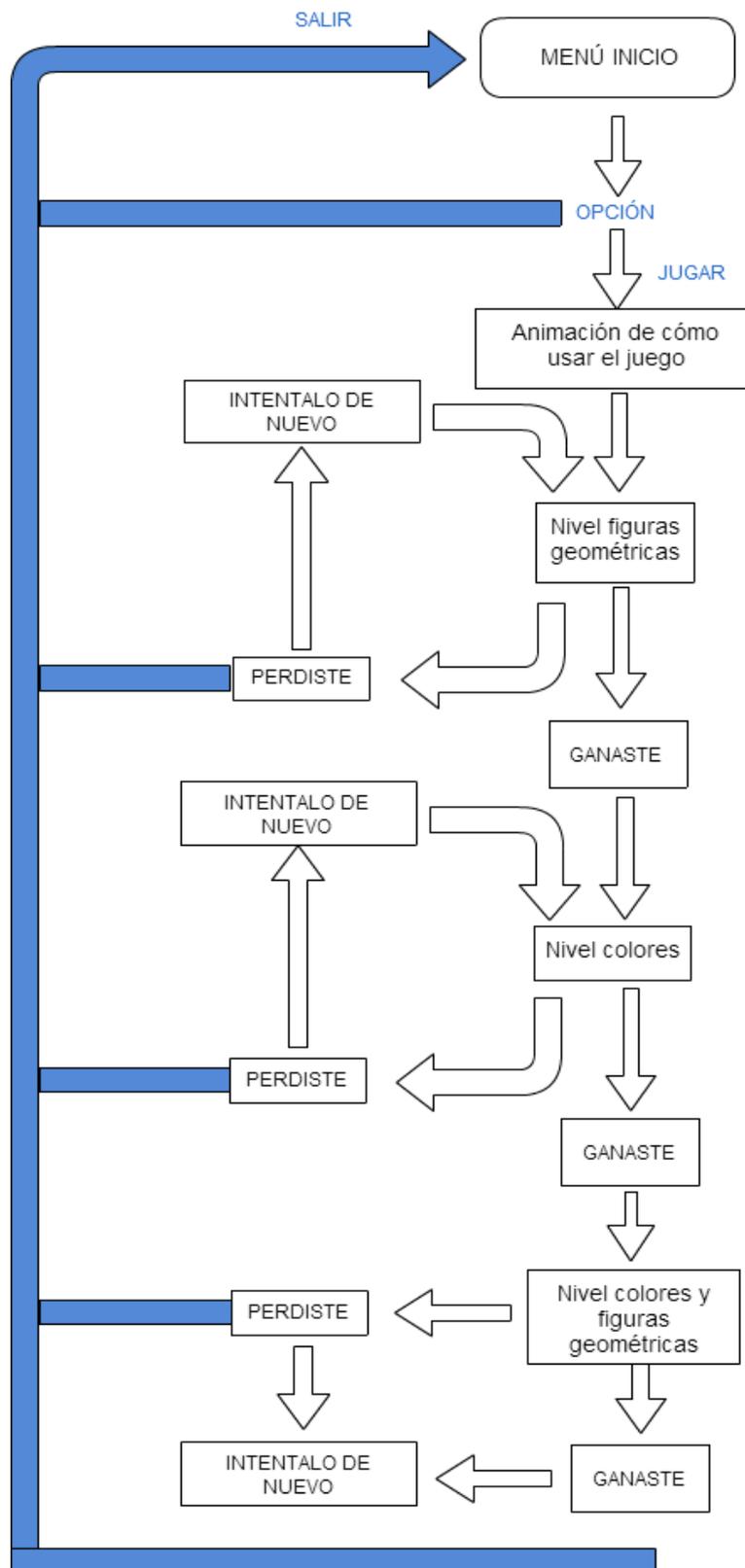


Figura 70. Funcionamiento del videojuego

7.3.3. PROGRAMACION EN UNITY:

La programación se hizo en un lenguaje C# la cuál es una programación orientada a objetos.

Se realizó todo el desarrollo del videojuego mediante varias escenas con sus respectivos scripts y se planteó jugar mediante la navegación entre éstas.

- **Scene de Inicio:** Aquí se establece el inicio de la música, la función para los dos botones (pasar a otra escena y salir) y el plano en donde están ubicados el fondo y las imágenes para los botones.

El único script que tiene esta escena es el de Jugar, que está programado para pasar entre escenas cuando el botón sea presionado dependiendo de la escena donde se encuentre el usuario.

```
using UnityEngine;
using System.Collections;

public class Jugar : MonoBehaviour {

    float x;

    float y;

    public FigurasGotas fiG;
    public figuras fi;
    public Colores co;
    public AudioClip boton;

    void Start () {
        x = transform.localScale.x;
        y = transform.localScale.y;
    }

    // Update is called once per frame
    void Update () {

    }

    void OnMouseEnter(){
        transform.localScale = new Vector3 (x + 0.05f, y + 0.05f, 1);
    }
    void OnMouseExit(){
        transform.localScale = new Vector3 (x - 0.05f, y - 0.05f, 1);
    }
    void OnMouseDown(){
        if (Application.loadedLevelName.Equals ("SceneInicio")) {
            if (this.gameObject.name.Equals ("BotonSalir")) {
```

```

        Application.Quit();
        reproducirSonido (boton);
    }
    if (this.gameObject.name.Equals ("Bjugar")) {
        Application.LoadLevel("Repaso1");
        reproducirSonido (boton);
    }
}
if (Application.loadedLevelName.Equals ("SceneGanaste1")) {
    if(this.gameObject.name.Equals("Bsiguiente")){
        Application.LoadLevel("SceneInicioColores");
        reproducirSonido (boton);
    }
    if(this.gameObject.name.Equals("botonTerminar")){
        Application.LoadLevel("SceneInicio");
        reproducirSonido (boton);
    }
}
if (Application.loadedLevelName.Equals ("SceneGanaste2")) {
    if(this.gameObject.name.Equals("Bsiguiente")){
        Application.LoadLevel("SceneInicioFigurasColores");
        reproducirSonido (boton);
    }
    if(this.gameObject.name.Equals("botonTerminar")){
        Application.LoadLevel("SceneInicio");
        reproducirSonido (boton);
    }
}
if (Application.loadedLevelName.Equals ("SceneGanaste3")) {
    if(this.gameObject.name.Equals("botonTerminar")){
        Application.LoadLevel("SceneInicio");
        reproducirSonido (boton);
    }
    if(this.gameObject.name.Equals("BotonIntentalo")){
        Application.LoadLevel("SceneInicioFigurasColores");
        reproducirSonido (boton);
    }
}
if (Application.loadedLevelName.Equals ("ScenePerdiste")) {
    if(this.gameObject.name.Equals("botonTerminar")){
        Application.LoadLevel("SceneInicio");
        reproducirSonido (boton);
    }
    if(this.gameObject.name.Equals("BotonIntentalo")){
        Application.LoadLevel("SceneInicioFiguras");
        reproducirSonido (boton);
    }
}
}
}

```

```

if (Application.loadedLevelName.Equals ("ScenePerdiste2")) {
    if(this.gameObject.name.Equals("botonTerminar")){
        Application.LoadLevel("SceneInicio");
        reproducirSonido (boton);
    }
    if(this.gameObject.name.Equals("BotonIntentalo")){
        Application.LoadLevel("SceneInicioColores");
        reproducirSonido (boton);;
    }
}
if (Application.loadedLevelName.Equals ("ScenePerdiste3")) {
    if(this.gameObject.name.Equals("botonTerminar")){
        Application.LoadLevel("SceneInicio");
        reproducirSonido (boton);
    }
    if(this.gameObject.name.Equals("BotonIntentalo")){
        Application.LoadLevel("SceneInicioFigurasColores");
        reproducirSonido (boton);
    }
}
if (Application.loadedLevelName.Equals ("SceneFigurasColores")) {
    if (this.gameObject.name.Equals ("Bjugar")) {
        fiG.comenzar = 1;
        this.gameObject.renderer.enabled = false;
        reproducirSonido (boton);
    }
}
if (Application.loadedLevelName.Equals ("SceneFiguras")) {
    if (this.gameObject.name.Equals ("Bjugar")) {
        fi.comenzar = 1;
        this.gameObject.renderer.enabled = false;
        reproducirSonido (boton);
    }
}
if (Application.loadedLevelName.Equals ("sceneColores")) {
    if (this.gameObject.name.Equals ("Bjugar")) {
        co.comenzar = 1;
        this.gameObject.renderer.enabled = false;
        reproducirSonido (boton);
    }
}
}
void reproducirSonido(AudioClip clip){
    float volumen = 1.0f;
    AudioSource.PlayClipAtPoint (clip, transform.position,volumen);
}
}

```

- Para la escena de animación del nivel de repaso se tuvieron que crear 4 scene ya que el plug in de reconocimiento de voz no permitía un libre manejo del mismo.

Para generar una solución se planteó en obtener las imágenes adecuadas para la explicación del proceso que el usuario debía seguir e intentar hacerlo lo más claro posible dentro de una animación

Este es ejemplo de un script para que aparezca la animación con una imagen previamente escogida, con una animación y en una posición adecuada y los más parecida posible al que verá el jugador en cada nivel.

```
using UnityEngine;
using System.Collections;
public class PasarNR2 : MonoBehaviour {

    int mov = 0;

    void Start () {
    }

    void Update(){
        mov++;
        if (this.gameObject.name.Equals("arrow-A") & mov <= 50) {
            this.gameObject.transform.position += new Vector3(-0.005f,0,0);
        }
        if (this.gameObject.name.Equals ("arrow-
A") & mov > 50 & mov <= 100) {
            this.gameObject.transform.position += new Vector3 (0.005f, 0, 0);
        }
        if (this.gameObject.name.Equals("arrow-
A") & mov > 100 & mov <= 150) {
            this.gameObject.transform.position += new Vector3(-0.005f,0,0);
        }
        if (this.gameObject.name.Equals("arrow-
A") & mov > 150 & mov <= 200) {
            this.gameObject.transform.position += new Vector3(0.005f,0,0);
        }

        if (Application.loadedLevelName.Equals("Repaso1") & Time.timeSinceLe
velLoad >= 5) {
            Application.LoadLevel("Repaso2");
        }
        if (Application.loadedLevelName.Equals("Repaso2") & Time.timeSinceLe
velLoad >= 5) {
            Application.LoadLevel("Repaso3");
        }
    }
}
```

```

        if (Application.loadedLevelName.Equals("Repaso3") & Time.timeSinceLevelLoad >= 5) {
            Application.LoadLevel("Repaso4");
        }
        if (Application.loadedLevelName.Equals("Repaso4") & Time.timeSinceLevelLoad >= 5) {
            Application.LoadLevel("SceneInicioFiguras");
        }
    }
}

```

- Ya en los tres niveles se implementó el paquete de reconocimiento de voz, el cuál se adquirió directamente con la empresa de Unity. Este paquete consta de unos scripts los cuales facilitan el procesamiento de voz, ya que este tipo de programación es muy limitada en el software libre y no se contaba con la posibilidad de adquirir la versión paga.
- Los scripts utilizados para lograr los objetivos de este proyecto fueron, Word detection, SpectrumMicrophone, Example 3, Example 5.
 - Word detection: La principal tarea es guardar los comandos en un arreglo. Cuando el usuario dice un comando que esté en dicho arreglo, el programa lo detecta y busca la posición en donde se encuentre. Como respuesta envía un número entero el cual corresponde a la posición y así se sabrá que comando es el que el usuario está diciendo.

(Anexo B)

- SpectrumMicrophone : Lo que permite ésta programación es que el programa reconozca los micrófonos que el sistema tenga conectados en el momento de correr el juego.

(Anexo B)

- Example 3 : Es donde esta la programación de cada botón que compone el reconocimiento de voz. Esta ligado con el resto de scripts.

(Anexo B)

- Example 5 : Fue el único script del paquete que fue modificado. Ya que es aquí en donde se guardan los comandos que el usuario va a grabar y posteriormente el programa va a reconocer. Se le dio un nombre a cada Label para que el niño supiera en que casilla va cada palabra y así tener un orden en la programación.

(Anexo B)

El paquete generó problemas ya que al ser una programación muy limitada, no permitió libertad en el diseño del juego y no se pudo realizar algo más llamativo en la interfaz gráfica para los usuarios que en este caso son niños.

Pero el mayor inconveniente fue al momento de programar el puntaje que debe generar el juego al “destruir” una imagen. Los comandos que el usuario grababa se están ejecutando casi 1000 veces por segundo cuando hablaba y aún si el niño no está hablando la grabación de ruido de fondo también generaba comandos todo el tiempo. Si la orden era que al decir un comando el puntaje sumara un punto, lo que hacía era sumar todos los comandos que se estaban enviando por el reconocimiento de voz (más de mil por segundo) y el juego no duraba el tiempo suficiente, ya que la meta era llegar a 10 puntos en menos tiempo al propuesto en cada nivel (2 o 3 minutos).

Lo que se hizo para solucionar éste problema fue usar un método el cual ya está definido en Unity que es Update, él cual se ejecuta 80 veces por segundo. Se hizo un contador el cual se estuviera sumando cada segundo y una variable la cuál controlara la entrada y salida de cada comando. Cuando la variable fuera igual a cero podía entrar al comando específico que el usuario estuviera diciendo, luego la variable cambiaria a ser igual a 1 para así sacarla del comando y controlar la cantidad de comandos por segundo que ejecuta el programa.

```
int entro = 0;
int contador = 0;
void Update () {
    if (AudioWordDetection.ClosestIndex == 1 & entro == 0) {
        entro = 1;
        if(co.figura == 1){
            ama.renderer.enabled = false;
            ama.puntaje = int.Parse (ama.puntajes.text) + 1;
            if(ama.puntaje >= 10){
                Application.LoadLevel("SceneGanaste2");
            }
            ama.puntajes.text = ama.puntaje.ToString();
            ama.presionado = 1;
        }
    }
}
```

```
if (contador % 80 == 0) {
    entro = 0;
}
```

```
    contador ++;  
}
```

(Toda la programación en el Anexo B)

8. Presentación y análisis de resultados.

El videojuego fue probado por 3 niños de diferentes edades dentro del rango propuesto (5 a 8 años). Cada niño grabó las palabras correspondientes a los tres colores y a las 3 figuras geométricas necesarias para la correcta jugabilidad del videojuego.

Luego cada uno repitió 10 veces cada palabra para así dar el porcentaje de reconocimiento que tiene el software.

Se grabó en condiciones normales, en una sala sin ningún tipo de acondicionamiento acústico. El transductor que se usó para la interacción con el dispositivo fue el micrófono integrado de un computador portátil MacBook Pro.

A continuación se presentan las tablas de resultados donde el numero 1 representa que el niño acertó la palabra y donde 0 representa una falla.

Niño 1, edad: 8 años.

Niño 2, edad: 7 años.

Niño 3, edad: 5 años.





Figura 71. Niños participantes del proyecto

Niño 1

Color/Aciertos										
Amarillo	1	0	1	1	0	1	1	1	1	1
Azul	1	1	1	1	1	1	1	1	1	1
Rojo	0	1	1	1	0	1	1	1	1	1
Circulo	1	1	1	1	1	1	1	1	1	1
Cuadrado	1	1	0	1	0	0	1	1	1	1
Triángulo	0	0	1	1	1	0	0	1	1	1

Tabla 12. Resultados niño 1, 8 años

Lo que se puede notar con este niño es que al tener mayor edad su pronunciación es más adecuada, se le facilitó mucho el ejercicio propuesto.

El software tuvo un porcentaje alto de reconocimiento en este caso, ya que las palabras quedaron bien grabadas y cada vez que el usuario las repitió, lo hizo muy parecido a la grabación inicial.

Amarillo: 80%

Azul: 100%

Rojo: 80%

Circulo: 100%

Triangulo: 60%

Niño 2

Color/Aciertos										
Amarillo	0	1	0	1	1	1	1	1	1	1
Azul	1	1	1	0	1	1	1	0	1	1
Rojo	0	0	0	1	1	1	1	1	1	0
Circulo	1	1	1	1	1	1	1	1	1	1
Cuadrado	1	1	1	1	1	1	0	0	1	1
Triangulo	1	0	0	1	0	0	1	1	1	1

Tabla 13. Resultados niño 2, 7 años

En este caso aun el software tiene una muy buena respuesta de reconocimiento de la palabra, aunque se nota un poco la disminución de aciertos. Esto debido a la pronunciación de algunas palabras, por ejemplo el niño aún no ha desarrollado bien la pronunciación de la "R" así que cada vez que intentaba decir rojo cambiaba de entonación, logrando confundir el reconocimiento.

Es necesario que la palabra quede correctamente grabada para que al momento de la comparación el sistema pueda identificar las palabras con mayor claridad.

Amarillo: 80%

Azul: 80%

Rojo: 60%

Circulo: 100%

Cuadrado: 80%

Triangulo: 60%

Niño 3

Color/Aciertos										
Amarillo	1	1	1	1	1	0	0	0	1	1
Azul	0	0	0	1	1	0	1	1	1	1
Rojo	0	0	1	1	1	1	0	0	1	1
Circulo	1	1	1	1	1	1	1	1	1	1
Cuadrado	1	1	1	1	0	1	0	1	1	1
Triangulo	0	0	1	1	0	0	0	1	1	1

Tabla 14. Resultados niño 3, 5 años

El niño de menor edad tuvo la mayor cantidad de errores en el reconocimiento. En el momento de grabar cambiaba la entonación y el nivel de su voz cada vez que pronunciaba las palabras, eso hace que al momento de ejecutar el software este no detecte las palabras con claridad, ya que deben ser pronunciadas lo más parecido posible a la grabación inicial.

Amarillo: 70%

Azul: 60%

Rojo: 60%
Circulo: 100%
Cuadrado: 80%
Triangulo: 50%

9. CONCLUSIONES

- A partir de la investigación previa al diseño del videojuego se puede notar el gran interés actual por la construcción de herramientas digitales más adecuadas para una alfabetización. El diseño del videojuego fue orientado hacia este propósito, desde su ambientación, creación de personajes y generando una curva de aprendizaje de acuerdo a las edades propuestas en el proyecto. Se logró una interacción con la máquina por medio de la voz con una serie de comandos específicos los cuáles le permiten al usuario superar cada reto propuesto en el videojuego. Aunque no se pudo medir el nivel educativo del juego.
- Se realizaron tres grabaciones con diferentes características como el lugar, el transductor y la persona que generó la señal de entrada para así conseguir lograr una caracterización y clasificación de cada señal grabada. El software Praat, que es una herramienta para el análisis de la voz, permitió determinar algunos parámetros que le dan una correcta funcionalidad al videojuego. Los comandos de voz escogidos son las seis palabras correspondientes a las tres figuras geométricas y a los tres colores que se mencionaron en el diseño del videojuego. Cada palabra se analizó y se pudo conocer los formantes de los cuales estaban compuestas, ayudando a explicar el funcionamiento y el porcentaje del reconocimiento de cada palabra una vez ejecutado.
- Al hacer las mediciones de los micrófonos que se utilizaron para las grabaciones de las palabras, se pudo notar que aunque el software de reconociendo tiene mejor respuesta entre mejores características de respuesta en frecuencia tenga el transductor, el sistema de micrófono integrado del MacBook Pro en el cuál fue probado el videojuego, alcanza a cumplir con las necesidades del programa si éste está en un ambiente con poco ruido y el usuario tiene una buena pronunciación de cada palabra.
- Para el desarrollo del videojuego Unity 3D es un motor de programación que facilita la creación de diferentes escenarios, haciendo posible el desarrollo de todo tipo juegos. En este caso se logró crear unos personajes y un ambiente adecuado para el propósito del videojuego. A

demás de que es bastante útil para el análisis en tiempo real de señales, ayudando a que el procesamiento de voz funcionará correctamente.

- Se implementó el plug in de Word detection para la identificación de la voz, el cual resulto tener un porcentaje alto de reconocimiento. Aunque esto depende directamente del usuario, primero de la forma inicial cómo grabe la palabra, luego de que la pronuncie lo más parecido posible a la grabación inicial dentro de los niveles propuestos. Si esto se cumple el Software tiene un funcionamiento normal para el cumplimiento de los objetivos del juego.
- El porcentaje de reconocimiento de la palabra es directamente proporcional a la calidad del habla que tenga el usuario, de lograr el menor ruido de fondo y de la calidad de micrófono con la cual se está ejecutando el videojuego.
- El reconociendo de las seis palabras que se usaron para el videojuego la que mayor porcentaje de reconocimiento fue círculo con un 100%. Para el niño la pronunciación de esta palabra fue la que más facilidad le genero al intentar repetirla gran cantidad de veces de la misma manera y para el software, ya que los primeros formantes que la componen están en frecuencias medio –bajas, entre 200Hz y 400Hz lo cuál genero eficiencia en el reconocimiento.
- Triangulo fue la palabra con menor porcentaje de reconocimiento con 60% en las dos primeras pruebas y 50% de reconocimiento en la prueba con el niño de menor edad (5 años), primero por la composición de la palabra ya que no le es fácil al niño pronunciarla de una forma parecida un número de veces considerable. También es la palabra con mayor cantidad de vocales y por consiguiente con la mayor cantidad de formantes iguales a las del resto de palabras del juego, llegando a confundir el análisis del software.
- Entre menos edad tenga el usuario más difícil será la jugabilidad para él, ya que requiere de concentración tanto como para grabar las palabras, decirlas en el momento justo y con la entonación en la que en primer lugar fue grabada.
- Se logró un videojuego manejado a través de comandos de voz que cumpliera con los objetivos plantados, se requiere la optimización del reconocimiento generando espacios más adecuados para la utilización del juego.

10.RECOMENDACIONES

Si se desea continuar con el desarrollo de este trabajo se podrían realizar pruebas en condiciones diferentes, como recintos con diferentes características acústicas o con una variedad de micrófonos para luego ver la respuesta que puede llegar a tener el software.

Sería adecuado que a nivel educativo se hagan pruebas donde se pueda medir el nivel educativo que tiene el videojuego y si la interacción que se propone en este proyecto por medio del reconocimiento de voz es apropiada para aumentar el interés en los niños hacia el aprendizaje con esta herramienta.

También se podrían desarrollar más niveles donde se relacionen otro tipo de conceptos que puedan repasar los niños de primera infancia, dándole así un desarrollo a la educación actual, generando nuevas herramientas para los estudiantes.

Para mejorar el diseño y poder implementar varios temas con una mejor calidad y funcionalidad es adecuado desarrollar un software de reconocimiento propio que permita un libre desarrollo y así no generar tanto inconveniente ni limitar al programador.

11. REFERENCIAS

- [1] A. L. Aldana Blanco y J. R. Piñeros Castaño, «Desarrollo e Implementación de un Algoritmo de Reconocimiento de Voz que permita seleccionar una Imagen a partir de un Banco de Nueve Fotografías utilizando Redes Neuronales,» Universidad de San Buenaventura, Bogotá D.C, Colombia, 2010.
- [2] O. Studios, «Los videojuegos como herramienta educativa y cultural de la sociedad, generadores de conocimientos e imaginarios,» Bogotá D.C, Colombia, 2012.
- [3] A. Calderón y A. López Astudillo, «Usos del videojuego “Rise of the Nations” en el proceso de enseñanza-aprendizaje en el programa de Ingeniería Industrial de la Universidad Icesi,» *CS No. 6*, pp. 101-128, 2010.
- [4] J. O. Montes de la Barrera, H. E. Hernández Riaño, J. M. López Pereira y J. Á. Chica Urzola, «IMPACTO DE LOS JUEGOS DIDÁCTICOS COMO HERRAMIENTA METODOLÓGICA EN EL APRENDIZAJE Y LA ENSEÑANZA DE LA INGENIERÍA INDUSTRIAL,» *Educación en Ingeniería*, pp. 37-48, 2010.
- [5] J. M. Karam Rozo, «CREATIVIDAD Y VIDEOJUEGOS: NUEVOS PARADIGMAS EN LA GENERACIÓN DE APRENDIZAJE AUTÓNOMO,» *Repertorio*, pp. 45-50, 2011.
- [6] J. D. Dúran Castañeda y N. R. Cely Garzón, «ANÁLISIS DE ELEMENTOS COMUNICATIVOS EN LOS VIDEOJUEGOS,» Universidad de la Sabana, Bogotá D.C, Colombia, 2011.
- [7] C. V. Rojas Rusinque, L. Gómez Sánchez y M. I. Guarnizo Pinto, «LOS VIDEOJUEGOS COMO OBJETOS DESARROLLADORES DE APRENDIZAJE,» Universidad de San Buenaventura, Facultad de Educación, Bogotá D.C, Colombia, 2014.
- [8] H. Sheikhzadeh y L. Deng, «Waveform-Based Speech Recognition Using Hidden Filter Models: Parameter Selection and Sensitivity to Power Normalization,» *IEEE Transactions on speech and audio processing*, 1994.
- [9] M. M. Cohen and D. W. Massaro, "What can Visual Speech Synthesis tell Visual Speech Recognition?," University of California, Santa Cruz, CA, 1994.
- [10] E. Morales Corral, «El uso de los videojuegos como recurso de aprendizaje en educación primaria y Teoría de la Comunicación,» Universidad Complutense, Madrid, 1995.
- [11] S. Anderson y D. Kewley-Port, «Evaluation of Speech Recognizers for Speech Training Applications,» *IEEE Transactions on speech and audio processing*, 1995.
- [12] D. G. Jaimieson, L. Deng, M. Price, V. Parsa y J. Till, «INTERACTION OF SPEECH DISORDERS WITH SPEECH CODERS: EFFECTS ON SPEECH INTELLIGIBILITY,» 1996.
- [13] S. Yeldener y J. H. Rieser, «A BACKGROUND NOISE REDUCTION TECHNIQUE BASED ON SINUSOIDAL SPEECH CODING SYSTEMS,»

- COMSAT Laboratories, Clarksburg, Maryland, USA, 2000.
- [14] D. O'shaughnessy, «Interacting With Computers by Voice: Automatic Speech Recognition and Synthesis,» IEEE, 2003.
- [15] C. Miranda-Palma, R. Camal, J. Cen Magaña, C. Gonzalez-Segura, S. Gonzalez-Segura, M. García y L. Narváez-Díaz, «Un Juego de Gravedad con Reconocimiento de Voz para Niños con Problemas de Lenguaje,» Universidad Autónoma de Yucatán, Yucatán, Mexico, 2004.
- [16] S. Grubesa, T. Grubesa y H. Domitrovic, «SPEAKER RECOGNITION METHOD COMBINING FFT, WAVELET FUNCTIONS AND NEURAL NETWORKS,» University of Zagreb, Croacia, 2005.
- [17] Z. Zeng, J. Tu, B. Pianfetti, M. Liu, T. Zhang, Z. Zhang, T. S. Huang y S. Levinson, «Audio-visual Affect Recognition through Multi-stream Fused HMM for HCI,» University of Illinois, Urbana, Champaign, 2005.
- [18] J. L. Oropeza Rodríguez, «Algoritmos y Métodos para el Reconocimiento de Voz en Español Mediante Sílabas,» Centro de Investigación en Computación-IPN, Mexico D.F., 2006.
- [19] C. González, C. Miranda, M. García y S. González, «Usabilidad en sistemas lúdicos infantiles con reconocimiento de voz como apoyo en la terapia de rehabilitación de niños con problemas de lenguaje,» Universidad Autónoma de Yucatán, Facultad de Matemáticas, Unidad Tizimín., Tizimín, Yuc. México., 2007.
- [20] S. Martín Moreno, F. J. Remesal Escalero y L. Rivera Rodríguez, «Videojuego Educativo para el Aprendizaje de SQL,» Universidad Complutense de Madrid, Madrid, 2007.
- [21] J. E. Jiménez y E. Rojas, «Efectos del videojuego Tradislexia en la conciencia fonológica y reconocimiento de palabras en niños disléxicos,» Universidad de la Laguna, Tenerife, 2008.
- [22] R. d. R. Camal Uc, «Uso de Reconocimiento de Voz en un Juego Electrónico para la Rehabilitación de Niños con el Problema de Lenguaje Dislalia,» Universidad de Yucatán, Tizimín, Yucatán, México, 2008.
- [23] F. Etxeberria Vlerdi, «VIDEOJUEGOS, CONSUMO Y EDUCACION,» *Revista Electrónica Teoría de la Educación. Educación y Cultura en la Sociedad de la Información*, pp. 11-28, 2008.
- [24] X. Wang, Y. Hao, D. Fu y C. Yuan, «Audio-Visual Automatic Speech Recognition for Connected Digits,» de *Second International Symposium on Intelligent Information Technology Application*, 2008.
- [25] P. Zwan y B. Kostek, «System for Automatic Singing Voice Recognition*,» *Audio Eng. Soc., Vol. 56, No. 9,* pp. 710-723, 2008.
- [26] O. Mayor, J. Bonada y J. Janer, «KALEIVOICECOPE: VOICE TRANSFORMATION FROM INTERACTIVE INSTALLATIONS TO VIDEO-GAMES,» AES 35th International Conference., London, UK, 2009.
- [27] P. Caloto Crespo, M. Moranchel Edras y Á. Ruiz Alonso, «Tecnologías de reconocimiento por voz y su aplicabilidad en videojuegos.,» Universidad Complutense de Madrid, Madrid, 2009.
- [28] P. S. Rajakumar, S. Ravi y R. M. Suresh, «Speech Enhancement Models Suited For Speech Recognition Using Composite Source And Wavelet Decomposition

- Model,» de *International Conference on Signal and Image Processing*, 2010.
- [29] H. del Castillo, S. Cortés, A. B. García Varela, N. Monjelat y G. Nogueiras, «Videojuegos Comerciales y Aprendizaje Escolar Análisis de las creencias del alumnado de Educación Secundaria Obligatoria,» Grupo Imágenes, Palabras e Ideas UAH y Electronic Arts España, 2010.
- [30] M. Marín, «Los videojuegos y su impacto en el aprendizaje,» Cátedra Telefónica de la Universidad de Deusto, Bilbao, Bizkaia, España, 2011.
- [31] R. Ahmed, R. Gil Pita, D. Ayllón y L. Álvarez, «Speech Source Separation Using a Multi-Pitch Harmonic Product Spectrum Based Algorithm,» 130th Convention AES, London, UK, 2011.
- [32] K. Drossos, A. Floros, K. Agavanakis, N. A. Tatlas y N. G. Kanellopoulos, «Emergency Voice/Stress-level Combined Recognition for Intelligent House Applications,» 132nd Convention AES, Budapest, Hungria, 2012.
- [33] F. I. Revuelta Domínguez y J. Guerra Antequera, «¿Qué aprendo con videojuegos? Una perspectiva de meta-aprendizaje del videojugador,» *Revista de Educación a Distancia. Número 33*, pp. 1-25, 2012.
- [34] M. S. Hawley, S. P. Cunningham, P. D. Green, P. Enderby, R. Palmer, S. Sehgal y P. O'Neill, «VOICE-INPUT VOICE-OUTPUT COMMUNICATION AID FOR PEOPLE WITH SEVERE SPEECH IMPAIRMENT,» *IEEE TRANSACTIONS ON NEURAL SYSTEMS AND REHABILITATION ENGINEERING, VOL. 21, NO. 1*, pp. 23-31, 2013.
- [35] T. Becker, Y. Solewicz, G. Jardine y S. Gfrörer, «Comparing Automatic Forensic Voice Comparison Systems under Forensic Conditions,» AES 46th International Conference,, Denver,USA, 2012.
- [36] P. N. Petkov, G. Eje Henter y W. Bastiaan Kleijn, «MAXIMIZING PHONEME RECOGNITION ACCURACY FOR ENHANCED SPEECH INTELLIGIBILITY IN NOISE,» de *TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 21, NO. 5,*, 2013.
- [37] A. Kupryjanow y A. Czyzewski, «Real-Time Speech Signal Segmentation Methods,» de *J. Audio Eng. Soc., Vol. 61, No. 7/8*, 2013.
- [38] «MANUAL DE PRAAT EN ESPAÑOL,» *Fonoaudiología, Fonética y Fonología*, p. 6, 2007.
- [39] P. L. Galindo Riaño, «Introducción al conocimiento de la voz,» Dpto. de Lenguajes y Sistemas Informáticos, Universidad de Cádiz, España, 1996.
- [40] M. García y G. Pazmiño, «LA TRANSFORMADA DE FOURIER,» universidad tecnológica Equinoccial, Quito, Ecuador, 2008.
- [41] B. Valenciano Martínez, «Speech Enhancement using Kalman filtering,» Área de Tratamiento de Voz y Señales Dpto. de Ingeniería Informática Escuela Politécnica Superior Universidad Autónoma de Madrid , Madrid, 2008.
- [42] A. Castellanos, «Universidad de Sonora, Departamento de física,» [En línea]. Available:
http://info.fisica.uson.mx/arnulfo.castellanos/archivos_html/quesonredneu.htm.
 [Último acceso: 2013].
- [43] «Apple,» [En línea]. Available:
http://support.apple.com/kb/SP649?viewlocale=es_ES.

12. BIBLIOGRAFÍA

- [1] H. Sheikhzadeh y L. Deng, «Waveform-Based Speech Recognition Using Hidden Filter Models: Parameter Selection and Sensitivity to Power Normalization,» IEEE Transactions on speech and audio processing, 1994.
- [2] M. M. Cohen and D. W. Massaro, "What can Visual Speech Synthesis tell Visual Speech Recognition?," University of California, Santa Cruz, CA, 1994.
- [3] E. Morales Corral, «El uso de los videojuegos como recurso de aprendizaje en educación primaria y Teoría de la Comunicación,» Universidad Complutense, Madrid, 1995.
- [4] S. Anderson y D. Kewley-Port, «Evaluation of Speech Recognizers for Speech Training Applications,» IEEE Transactions on speech and audio processing, 1995.
- [5] D. G. Jaimieson, L. Deng, M. Price, V. Parsa y J. Till, «INTERACTION OF SPEECH DISORDERS WITH SPEECH CODERS: EFFECTS ON SPEECH INTELLIGIBILITY,» 1996.
- [6] S. Yeldener y J. H. Rieser, «A BACKGROUND NOISE REDUCTION TECHNIQUE BASED ON SINUSOIDAL SPEECH CODING SYSTEMS,» COMSAT Laboratories, Clarksburg, Maryland, USA, 2000.
- [7] D. O'shaughnessy, «Interacting With Computers by Voice: Automatic Speech Recognition and Synthesis,» IEEE, 2003.
- [8] C. Miranda-Palma, R. Camal, J. Cen Magaña, C. Gonzalez-Segura, S. Gonzalez-Segura, M. García y L. Narváez-Díaz, «Un Juego de Gravedad con Reconocimiento de Voz para Niños con Problemas de Lenguaje,» Universidad Autonoma de Yucatán, Yucatán, Mexico, 2004.
- [9] S. Grubesa, T. Grubesa y H. Domitrovic, «SPEAKER RECOGNITION METHOD COMBINING FFT, WAVELET FUNCTIONS AND NEURAL NETWORKS,» University of Zagreb, Croacia, 2005.
- [10] Z. Zeng, J. Tu, B. Pianfetti, M. Liu, T. Zhang, Z. Zhang, T. S. Huang y S. Levinson, «Audio-visual Affect Recognition through Multi-stream Fused HMM for HCI,» University of Illinois, Urbana, Champaign, 2005.
- [11] J. L. Oropeza Rodriguez, «Algoritmos y Métodos para el Reconocimiento de Voz en Español Mediante Sílabas,» Centro de Investigación en Computación-IPN, Mexico D.F., 2006.
- [12] C. González, C. Miranda, M. García y S. González, «Usabilidad en sistemas lúdicos infantiles con reconocimiento de voz como apoyo en la terapia de rehabilitación de niños con problemas de lenguaje,» Universidad Autónoma de Yucatán, Facultad de Matemáticas, Unidad Tizimín., Tizimín, Yuc. México., 2007.
- [13] S. Martín Moreno, F. J. Remesal Escalero y L. Rivera Rodríguez, «Videojuego Educativo para el Aprendizaje de SQL,» Universidad Complutense de Madrid, Madrid, 2007.

- [14] J. E. Jiménez y E. Rojas, «Efectos del videojuego Tradislexia en la conciencia fonológica y reconocimiento de palabras en niños disléxicos,» Universidad de la Laguna, Tenerife, 2008.
- [15] R. d. R. Camal Uc, «Uso de Reconocimiento de Voz en un Juego Electrónico para la Rehabilitación de Niños con el Problema de Lenguaje Dislalia,» Universidad de Yucatán, Tizimín, Yucatán, México, 2008.
- [16] F. Etxeberria Vlerdi, «VIDEOJUEGOS, CONSUMO Y EDUCACION,» *Revista Electrónica Teoría de la Educación. Educación y Cultura en la Sociedad de la Información*, pp. 11-28, 2008.
- [17] X. Wang, Y. Hao, D. Fu y C. Yuan, «Audio-Visual Automatic Speech Recognition for Connected Digits,» de *Second International Symposium on Intelligent Information Technology Application*, 2008.
- [18] P. Zwan y B. Kostek, «System for Automatic Singing Voice Recognition*,» *Audio Eng. Soc., Vol. 56, No. 9,* pp. 710-723, 2008.
- [19] O. Mayor, J. Bonada y J. Janer, «KALEIVOICECOPE: VOICE TRANSFORMATION FROM INTERACTIVE INSTALLATIONS TO VIDEO-GAMES,» AES 35th International Conference, London, UK, 2009.
- [20] P. Caloto Crespo, M. Moranchel Edras y Á. Ruiz Alonso, «Tecnologías de reconocimiento por voz y su aplicabilidad en videojuegos,» Universidad Complutense de Madrid, Madrid, 2009.
- [21] A. L. Aldana Blanco y J. R. Piñeros Castaño, «Desarrollo e Implementación de un Algoritmo de Reconocimiento de Voz que permita seleccionar una Imagen a partir de un Banco de Nueve Fotografías utilizando Redes Neuronales,» Universidad de San Buenaventura, Bogotá D.C, Colombia, 2010.
- [22] P. S. Rajakumar, S. Ravi y R. M. Suresh, «Speech Enhancement Models Suited For Speech Recognition Using Composite Source And Wavelet Decomposition Model,» de *International Conference on Signal and Image Processing*, 2010.
- [23] H. del Castillo, S. Cortés, A. B. García Varela, N. Monjelat y G. Nogueiras, «Videojuegos Comerciales y Aprendizaje Escolar Análisis de las creencias del alumnado de Educación Secundaria Obligatoria,» Grupo Imágenes, Palabras e Ideas UAH y Electronic Arts España, 2010.
- [24] O. Studios, «Los videojuegos como herramienta educativa y cultural de la sociedad, generadores de conocimientos e imaginarios,» Bogotá D.C, Colombia, 2012.
- [25] A. Calderón y A. López Astudillo, «Usos del videojuego “Rise of the Nations” en el proceso de enseñanza-aprendizaje en el programa de Ingeniería Industrial de la Universidad Icesi,» *CS No. 6*, pp. 101-128, 2010.
- [26] J. O. Montes de la Barrera, H. E. Hernández Riaño, J. M. López Pereira y J. Á. Chica Urzola, «IMPACTO DE LOS JUEGOS DIDÁCTICOS COMO HERRAMIENTA METODOLÓGICA EN EL APRENDIZAJE Y LA ENSEÑANZA DE LA INGENIERÍA INDUSTRIAL,» *Educación en Ingeniería*, pp. 37-48, 2010.
- [27] M. Marín, «Los videojuegos y su impacto en el aprendizaje,» Cátedra Telefónica de la Universidad de Deusto, Bilbao, Bizkaia, España, 2011.
- [28] R. Ahmed, R. Gil Pita, D. Ayllón y L. Álvarez, «Speech Source Separation Using a

- Multi-Pitch Harmonic Product Spectrum Based Algorithm,» 130th Convention AES, London, UK, 2011.
- [29] J. M. Karam Rozo, «CREATIVIDAD Y VIDEOJUEGOS: NUEVOS PARADIGMAS EN LA GENERACIÓN DE APRENDIZAJE AUTÓNOMO,» *Repertorio*, pp. 45-50, 2011.
- [30] J. D. Dúran Castañeda y N. R. Cely Garzón, «ANÁLISIS DE ELEMENTOS COMUNICATIVOS EN LOS VIDEOJUEGOS,» Universidad de la Sabana, Bogotá D.C, Colombia, 2011.
- [31] K. Drossos, A. Floros, K. Agavanakis, N. A. Tatlas y N. G. Kanellopoulos, «Emergency Voice/Stress-level Combined Recognition for Intelligent House Applications,» 132nd Convention AES, Budapest, Hungria, 2012.
- [32] F. I. Revuelta Domínguez y J. Guerra Antequera, «¿Qué aprendo con videojuegos? Una perspectiva de meta-aprendizaje del videojugador,» *Revista de Educación a Distancia. Número 33*, pp. 1-25, 2012.
- [33] M. S. Hawley, S. P. Cunningham, P. D. Green, P. Enderby, R. Palmer, S. Sehgal y P. O'Neill, «VOICE-INPUT VOICE-OUTPUT COMMUNICATION AID FOR PEOPLE WITH SEVERE SPEECH IMPAIRMENT,» *IEEE TRANSACTIONS ON NEURAL SYSTEMS AND REHABILITATION ENGINEERING, VOL. 21, NO. 1*, pp. 23-31, 2013.
- [34] T. Becker, Y. Solewicz, G. Jardine y S. Gfrörer, «Comparing Automatic Forensic Voice Comparison Systems under Forensic Conditions,» AES 46th International Conference,, Denver,USA, 2012.
- [35] P. N. Petkov, G. Eje Henter y W. Bastiaan Kleijn, «MAXIMIZING PHONEME RECOGNITION ACCURACY FOR ENHANCED SPEECH INTELLIGIBILITY IN NOISE,» de *TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 21, NO. 5,*, 2013.
- [36] A. Kupryjanow y A. Czyzewski, «Real-Time Speech Signal Segmentation Methods,» de *J. Audio Eng. Soc., Vol. 61, No. 7/8*, 2013.
- [37] C. V. Rojas Rusinque, L. Gómez Sánchez y M. I. Guarnizo Pinto, «LOS VIDEOJUEGOS COMO OBJETOS DESARROLLADORES DE APRENDIZAJE,» Universidad de San Buenaventura, Facultad de Educación, Bogotá D.C, Colombia, 2014.
- [38] «MANUAL DE PRAAT EN ESPAÑOL,» *Fonoaudiología, Fonética y Fonología*, p. 6, 2007.
- [39] P. L. Galindo Riaño, «Introducción al conocimiento de la voz,» Dpto. de Lenguajes y Sistemas Informáticos, Universidad de Cádiz, España, 1996.
- [40] M. García y G. Pazmiño, «LA TRANSFORMADA DE FOURIER,» universidad tecnológica Equinoccial, Quito, Ecuador, 2008.
- [41] B. Valenciano Martínez, «Speech Enhancement using Kalman filtering,» Área de Tratamiento de Voz y Señales Dpto. de Ingeniería Informática Escuela Politécnica Superior Universidad Autónoma de Madrid , Madrid, 2008.
- [42] A. Castellanos, «Universidad de Sonora, Departamento de física,» [En línea]. Available: http://info.fisica.uson.mx/arnulfo.castellanos/archivos_html/quesonredneu.htm. [Último acceso: 2013].

