

UNIVERSIDAD RICARDO PALMA
FACULTAD DE INGENIERÍA

PROGRAMA DE TITULACIÓN POR TESIS
ESCUELA PROFESIONAL DE INGENIERÍA ELECTRÓNICA



**DESARROLLO DE VISOR DE REALIDAD AUMENTADA EN
BASE A UN CONVERTOR MULTILINGÜE DE VOZ A
TEXTO PARA PERSONAS CON DISCAPACIDAD AUDITIVA**

TESIS
PARA OPTAR EL TÍTULO PROFESIONAL DE
INGENIERO ELECTRÓNICO

PRESENTADA POR

Bach. PACO MALPARTIDA, LUIS ALBERTO

Bach. HUAMÁN PEREDO, LEONCIO PABLO

Asesor: Dr. Ing. HUAMANÍ NAVARRETE, PEDRO FREDDY

LIMA - PERÚ

2021

DEDICATORIA

A mi madre Mercedes que siempre me ha estado apoyando con sus consejos y formado con valores. A mi hermana Mary Cielo que juntos nos hemos superado a las adversidades.

Luis Paco

A mi madre Dora por ser el motor de mi camino académico y a mi padre que en vida siempre me ha apoyado.

Leoncio Huamán

AGRADECIMIENTO

A nuestra casa de estudios La universidad Ricardo Palma, a La escuela de Ingeniería Electrónica y docentes que nos han formado profesionalmente en esta carrera.

Luis Paco y Leoncio Huamán

ÍNDICE GENERAL

RESUMEN.....	ix
ABSTRACT.....	x
INTRODUCCIÓN	xi
CAPÍTULO I: PLANTEAMIENTO Y DELIMITACIÓN DEL PROBLEMA	12
1.1. Formulación del problema.....	12
1.1.1. Problema General.....	12
1.1.2. Problemas Específicos.....	12
1.2. Objetivos.....	13
1.2.1. Objetivo General	13
1.2.2. Objetivos Específicos.....	13
1.3. Importancia y justificación.....	13
1.3.1. Importancia	13
1.3.2. Justificación.....	13
1.4. Limitaciones	14
CAPÍTULO II: MARCO TEÓRICO	15
2.1. Marco Histórico	15
2.2. Investigaciones relacionadas con el Tema.....	16
2.3. Estructura teórica y científica que sustenta el estudio.....	19
2.3.1 Visor de realidad aumentada	19
2.3.1.1 Hardware para el visor de realidad aumentada.....	20
2.3.1.2 Software para el visor de realidad aumentada	20
2.3.1.3 Sistema Óptico para el visor de realidad aumentada.....	20
2.3.2 Conversor multilingüe de voz a texto.....	22
2.3.2.1 Inteligencia artificial.....	22
2.3.2.2 Hardware para el conversor multilingüe de voz a texto	22
2.3.2.3 Software para el conversor multilingüe de voz a texto.....	23
2.3.2.4 Servicios de nube.....	23
2.3.2.5 Filtro.....	23
2.4. Definición de términos básicos	24
2.5. Diseño de la Investigación.....	24
2.5.1. Variables de investigación.....	24
2.5.2. Tipo y Método de investigación.....	24
2.5.3. Técnicas e Instrumentos de recolección de datos.....	25

2.5.4. Procedimiento para la recolección de datos.....	25
CAPÍTULO III: DESARROLLO DEL PROYECTO	26
3.1. Estructura del sistema.....	28
3.2. Desarrollo de los mecanismos de preprocesamiento de voz	32
3.2.1 Diseño teórico de filtro pasa-banda.....	32
3.2.2 Implementación de filtro digital pasa-banda en lenguaje Python.....	33
3.3. Integración de sistemas Edge computing	36
3.3.1 Implementación del filtro en Node-Red.....	36
3.3.2 Codificación de interfaces a gráfico mediante Node-Red	39
3.3.3 Orquestación con sistemas de servicios de nube para el reconocimiento de voz..	40
3.3.4 Orquestación con sistemas de servicios de nube para la traducción.....	44
3.3.5 Proyección de texto procesado.....	48
3.4. Implementación de sistema óptico	52
3.4.1 Diseño teórico de generación y tratamiento de imagen virtual.....	52
3.4.2 Diseño e impresión 3D de diseño de armazón del visor.....	53
3.4.3 Construcción e integración de armazón con sistemas de lentes.....	57
CAPÍTULO IV: PRUEBAS Y RESULTADOS	64
4.1. Detección de palabras más comunes según idioma.....	64
4.2. Detección de palabras con perturbación.....	72
4.3. Resumen de detecciones	81
4.4. Presupuesto	87
CONCLUSIONES.....	89
RECOMENDACIONES.....	90
REFERENCIAS BIBLIOGRÁFICAS.....	91
ANEXO	93

ÍNDICE DE FIGURAS

Figura N° 1: Tipo de lentes convergentes	21
Figura N° 2: Formacion de la imagen virtual	21
Figura N° 3: Flujo general del proyecto.....	27
Figura N° 4: Funcionamiento del visor en la comunicación del usuario con el hablante	28
Figura N° 5: Flujo técnico del proyecto.....	30
Figura N° 6: Flujo técnico de los servicios de nube IBM Watson	31
Figura N° 7: Diseño del filtro.....	32
Figura N° 8: Función de transferencia.....	33
Figura N° 9: Módulo del filtro elíptico pasa banda.....	33
Figura N° 10: Gráfico de la voz en el dominio del tiempo	34
Figura N° 11: Espectro en frecuencia del audio	35
Figura N° 12: Programación del filtro elíptico pasa banda.....	35
Figura N° 13: Espectro en frecuencia del audio filtrado.....	36
Figura N° 14: Almacenamiento de audios y filtros.....	37
Figura N° 15: Implementacion del filtro en Node-Red	37
Figura N° 16: Nodo de archivo	38
Figura N° 17: Nodo de ejecución del PreFilter.py	38
Figura N° 18: Nodo de ejecucion del MainFilter.py	39
Figura N° 19: Implementacion final del filtro digital en Node-Red.....	40
Figura N° 20: Servicio Speech to Text en la plataforma web en la nube IBM.....	41
Figura N° 21: Gestión del recurso Speech to Text.....	42
Figura N° 22: Nodo del servicio Speech to Text	43
Figura N° 23: Flujo parcial para el tratamiento de la voz en español.....	44
Figura N° 24: Texto procesado.....	44
Figura N° 25: Servicio Language Transalator en la plataforma web en la nube de IBM	45
Figura N° 26: Gestión del recurso Language Translator	45
Figura N° 27: Nodo Speech to Text para reconocer voz en inglés.....	46
Figura N° 28: Nodo de Language Translator.....	47
Figura N° 29: Flujo parcial para el tratamiento de la voz en inglés	48
Figura N° 30: Nodo de comuniación serial	48
Figura N° 31: Configuración de la comunicación serial.....	49

Figura N° 32: Flujo parcial del sistema para el tratamiento de la voz y la comunicación serial.....	50
Figura N° 33: Programación para proyectar el texto recibido.....	51
Figura N° 34: Texto proyectado en la pantalla OLED	52
Figura N° 35: Vista lateral y superior del brazo superior delantero del reflector	53
Figura N° 36: Vista lateral y superior del brazo delantero inferior del reflector	54
Figura N° 37: Vista lateral y superior del brazo de soporte superior del armazón	54
Figura N° 38: Vista lateral y superior del soporte del reflector.....	55
Figura N° 39: Vista lateral y superior del brazo de soporte inferior del armazón	55
Figura N° 40: Vista superior de los componentes del armazón	56
Figura N° 41: Vista diagonal desde la perspectiva interior.....	56
Figura N° 42: Vista diagonal desde la perspectiva exterior.....	57
Figura N° 43: Vista del Raspberry Pi 3B desde la perspectiva superior.....	57
Figura N° 44: Vista del micrófono y del Raspberry Pi 3B desde la perspectiva superior	58
Figura N° 45: Vista superior del interior del visor.....	59
Figura N° 46: Vista de la pantalla OLED.....	59
Figura N° 47: Vista del espejo utilizado	60
Figura N° 48: Vista del reflector	61
Figura N° 49: Vista del texto en objeto en la realidad.....	61
Figura N° 50: Vista del texto en objeto en la realidad.....	62
Figura N° 51: Vista del texto en objeto en la realidad.....	62
Figura N° 52: Vista del visor de realidad aumentada	63

ÍNDICE DE TABLAS

Tabla N° 1: Resultados usando el idioma español.....	64
Tabla N° 2: Resultados usando el idioma inglés	66
Tabla N° 3: Resultados usando el idioma francés	68
Tabla N° 4: Resultados usando el idioma portugués	71
Tabla N° 5: Resultados usando el idioma español con una perturbación.....	73
Tabla N° 6: Resultados usando el idioma inglés con una perturbación.....	75
Tabla N° 7: Resultados usando el idioma francés con una perturbación.....	77
Tabla N° 8: Resultados usando el idioma portugués con una perturbación.....	79
Tabla N° 9: Porcentaje de detección en idioma español	82
Tabla N° 10: Porcentaje de detección en idioma inglés.....	83
Tabla N° 11: Porcentaje de detección en idioma francés	83
Tabla N° 12: Porcentaje de detección en idioma portugués	84
Tabla N° 13: Porcentaje de detección en idioma español con una perturbación.....	84
Tabla N° 14: Porcentaje de detección en idioma inglés con una perturbación	85
Tabla N° 15: Porcentaje de detección en idioma francés I con una perturbación.....	85
Tabla N° 16: Porcentaje de detección en idioma portugués con una perturbación.....	86
Tabla N° 17: Promedio del porcentaje de detección de cada idioma	86
Tabla N° 18: Promedio del porcentaje de detección de cada idioma con una perturbación.....	86
Tabla N° 19: Costo de materiales.....	87

RESUMEN

En el presente proyecto de tesis se desarrolla un visor de realidad aumentada para personas con discapacidad auditiva que es capaz de convertir la voz en texto digital para poder ser leído por el usuario que esté utilizando el dispositivo mencionado. Con el fin que perturbaciones ajenas al rango de frecuencias de la voz humana no interfieran con el proceso, se implementó y utilizó un filtro digital pasa-banda elíptico. Luego, el proceso de conversión de voz a texto se da por medio de los servicios de inteligencia artificial en la nube IBM Watson Speech to text. Así como también, el reconocimiento de voz en los idiomas inglés, portugués y francés para su traducción al español posteriormente mediante el servicio IBM Watson Language Translator. La integración del filtro y servicios de nube se realizaron en la plataforma Node-Red desplegada en el Raspberry Pi 3B. El texto procesado se envió por comunicación serial al Arduino donde se implementó un código para la proyección del texto en una pantalla OLED. De esta manera, se permite a la persona con discapacidad auditiva poder leer el texto gracias al sistema óptico hace posible ver el texto por un reflector, con el fin de no obstaculizar la línea de visión entre el usuario del visor y el emisor del mensaje hablado. Finalmente, en los resultados se hicieron pruebas con las frases más comunes para cada idioma, se definieron fórmulas para el cálculo del porcentaje de detección con el fin de cuantificar el éxito del reconocimiento de palabras en los diferentes idiomas. De los cuatro idiomas configurados, las pruebas realizadas para el idioma francés tuvieron el mejor porcentaje de detección siendo del 100%, mientras en los demás idiomas en promedio superaron el 60%. De la misma manera, se obtuvieron pruebas exitosas respecto a la detección de palabras con perturbación ya que los porcentajes obtenidos para la detección con o sin perturbación llegaron a tener la misma cantidad porcentual.

Palabras claves: Realidad aumentada, filtro digital pasa-banda, IBM Watson, Speech to text, Language translator, Node-Red.

ABSTRACT

This thesis project develops an augmented reality viewer for hearing impaired people that is capable of converting voice into digital text to be read by the user who is using the aforementioned device. In order that disturbances outside the frequency range of the human voice do not interfere with the process, a digital elliptical band-pass filter was implemented and used. Then, the Speech to Text conversion process is done by means of the IBM Watson Speech to text artificial intelligence services in the cloud. As well as speech recognition in English, Portuguese and French for translation into Spanish later through the IBM Watson Language Translator service. The integration of the filter and cloud services were performed on the Node-Red platform deployed on the main processor. The processed text was sent by serial communication to the secondary processor where a code was implemented for the projection of the text on an OLED screen. In this way, the hearing-impaired person is allowed to read the text thanks to the optical system that makes it possible to see the text through a reflector, so as not to obstruct the line of sight between the user of the viewer and the sender of the spoken message. Finally, the results were tested with the most common phrases for each language, and formulas were defined for the calculation of the detection percentage in order to quantify the success of word recognition in the different languages. Of the four languages configured, the tests performed for French had the best detection percentage of 100%, while in the other languages the average was over 60%. In the same way, successful tests were obtained with respect to the detection of words with perturbation, since the percentages obtained for detection with and without perturbation reached the same percentage.

Keywords: Augmented reality, Digital band-pass filter, IBM Watson, Speech to Text, Language translator, Node-Red

INTRODUCCIÓN

La deficiencia auditiva es la imposibilidad de oír pudiendo ser una pérdida auditiva parcial o total, sus causas son por envejecimiento, hereditarias, enfermedades, mucha exposición al ruido, accidentes, entre otros. Por tal razón, las personas con estos problemas tienen dificultades para comunicarse más cuando están ante una multitud de gente, lo que genera una baja autoestima en ellos.

De esta manera, con el avance de las tecnologías computacionales, ha surgido la inteligencia artificial en plataformas virtuales como IBM Watson que facilitan el uso de reconocimiento gracias a los servicios de nube. Además, en los últimos años, la realidad aumentada ha podido integrarse en distintos sistemas para proponer soluciones que ayudan a grupo personas con discapacidades, siendo una herramienta de rehabilitación o de integración ya que puede reemplazar o enriquecer de manera virtual el ambiente donde se encuentra, así impactando de forma positiva en la manera de incluirlos socialmente.

Por lo cual en el presente proyecto de tesis se desarrolló un visor de realidad aumentada en base a un conversor multilingüe de voz a texto para las personas con discapacidad auditiva, de tal forma que puedan leer lo que les están hablando y así poder contribuir con su integración en la sociedad.

Así mismo para este proyecto de tesis se ha estructurado de la siguiente manera:

En el capítulo 1 se plantea la formulación del problema, así como también se define el objetivo general y específicos. Además, se explica la importancia, justificación y limitaciones.

En el capítulo 2 se elabora el marco teórico donde se comentan las investigaciones relacionadas con el tema, se explican las bases teóricas, así como también el diseño de investigación

En el capítulo 3 se desarrollan los mecanismos para el preprocesamiento de la voz, se integran los sistemas de Edge computing. Además, se describe la implementación del sistema óptico para el visor de realidad aumentada.

En el capítulo 4 se realizan las pruebas con las frases más comunes para cada idioma y así poder calcular el porcentaje de detección de cada una.

Finalmente, se redactan las conclusiones, recomendaciones y se describen las referencias bibliográficas.

CAPÍTULO I: PLANTEAMIENTO Y DELIMITACIÓN DEL PROBLEMA

1.1. Formulación del problema

Para más de 450 mil personas con discapacidad auditiva en Perú, según la Encuesta Nacional de Hogares (ENAHOG - 2019) realizada por el Instituto Nacional de Estadística e Informática - INEI, la comunicación con personas que no entienden el lenguaje de señas es un problema del día a día.

En el Perú no es común que este lenguaje se enseñe en instituciones educativas, entonces genera disconformidad, desentendimiento e incluso discriminación entre los hablantes. Ante ello, las personas con discapacidad auditiva recurren a diversas soluciones de alto costo que les permitan establecer un diálogo con las personas que quieren comunicarse para así mejorar su integración en la sociedad. Además, debido a la carencia en instituciones educativas hacia personas con problemas auditivos, el alumnado se ve obligado a acudir a otros centros de educación para reforzar la interpretación de lo dictado por los docentes en sus respectivas clases.

1.1.1. Problema General

¿Cómo se desarrollará el Visor de Realidad Aumentada en Base a un Con versor Multilingüe de Voz a Texto para Personas con Discapacidad Auditiva?

1.1.2. Problemas Específicos

- a) ¿Cómo se desarrollará el mecanismo del procesamiento de voz utilizando filtros digitales y el reconocimiento del mismo usando servicios de nube de inteligencia artificial IBM Watson?
- b) ¿Cómo se desarrollará la programación a nivel de código para proyectar en una pantalla OLED, con ayuda de un Arduino, la conversión multilingüe de voz a texto para personas con discapacidad auditiva?
- c) ¿Cómo se implementará el sistema óptico del Visor de Realidad Aumentada para que la persona con discapacidad auditiva pueda leer la conversión de voz a texto?

1.2. Objetivos

1.2.1. Objetivo General

Desarrollar un Visor de Realidad Aumentada en Base a un Conversor Multilingüe de Voz a Texto para Personas con Discapacidad Auditiva.

1.2.2. Objetivos Específicos

- a) Desarrollar el mecanismo del procesamiento de voz utilizando filtros digitales y el reconocimiento del mismo usando servicios de nube de inteligencia artificial IBM Watson.
- b) Desarrollar la programación a nivel de código para proyectar en una pantalla OLED, con ayuda de un Arduino, la conversión multilingüe de voz a texto para personas con discapacidad auditiva.
- c) Implementar el sistema óptico del Visor de Realidad Aumentada para que la persona con discapacidad auditiva pueda leer la conversión de voz a texto.

1.3. Importancia y justificación

1.3.1. Importancia

La importancia del desarrollo del Visor Multilingüe es que las personas con discapacidad auditiva puedan entender lo que se conversa alrededor de ellos. Esta alternativa de solución resulta eficaz ante cualquier tipo de gravedad del sentido auditivo ya que se apoya en el sentido ocular del usuario y no en el sentido defectuoso.

1.3.2. Justificación

El diseño físico del visor se logra gracias a microprocesadores de la tecnología de Edge Computing y sistemas ópticos para la proyección del texto transcrito, mientras que, para el diseño a nivel lógico, se utilizaron las cualidades de la Inteligencia Artificial mediante servicios de nube para realizar la conversión de un discurso hablado a texto escrito en diversos idiomas.

1.4. Limitaciones

- El multilingüismo del conversor dependerá de los idiomas disponibles de los servicios de inteligencia artificial tanto para la conversión de voz a texto como también la traducción del texto procesado; sin embargo, este trabajo se limita a 4 idiomas: español, inglés, francés y portugués debido a que son comunes en una conversación.
- El procesamiento de la voz percibida y su correspondiente reconocimiento se va a restringir a la señal más cercana al micrófono del dispositivo o a la de mayor amplitud.
- El proceso de entrenamiento de los modelos dentro del conversor de voz se realiza a través de servicios de nube, más no en una GPU ni en una CPU local, dado que no se cuenta con la factibilidad de procesamiento del hardware que se utiliza en la implementación del visor de realidad aumentada.
- La transcripción de voz a texto tiene un límite de 500 minutos de uso al mes, según la duración del audio a procesar acorde a las características del servicio utilizado en la nube.
- El proyecto se va a probar con 2 personas con discapacidad auditiva de la asociación de sordos del Perú

CAPÍTULO II: MARCO TEÓRICO

2.1. Marco Histórico

A través de los años se han propuesto distintas soluciones ante problemáticas relacionadas a pérdida de sentido auditivo. Entre las más conocidas contamos con los implantes cocleares como los audífonos médicos. Desde las creaciones de estas soluciones, se han implementado distintas mejoras con tecnologías específicas con el fin de incrementar la eficiencia en estos sistemas como también para obtener mayor satisfacción por parte de los usuarios. Debido a que estas soluciones implican el tratar con sonidos del ambiente como ruido, voces y sonidos externos se ha trabajado arduamente en el procesamiento de estas señales(sonidos). Diversas tecnologías han sido implementadas para solucionar problemáticas de accesibilidad. La realidad aumentada (R.A.) es una de ellas. Se ha estado aplicando la R.A. en sistemas de rehabilitación para personas con dificultades motoras.

Según Correa, A., De Assis, G. (2007). An augmented reality musical game for cognitive and motor rehabilitation, se desarrolló un sistema visual para gafas comerciales de R.A. junto a guantes con sensores con el fin de que los usuarios interactúen con juegos de simple entendimiento donde el objetivo es crear piezas musicales con los objetos virtuales en la realidad artificial. De esta manera, los usuarios con limitaciones motoras pueden entrenar sus habilidades cognitivas y tienen la oportunidad de desarrollar habilidades creativas.

En los últimos años se han tenido grandes avances en aplicaciones de realidad aumentada. La compañía de tecnología Epson junto al teatro inglés National Theater implementaron la aplicación de gafas de realidad aumentada para mostrar el texto de los libretos de las obras en el teatro. Es la primera aplicación tecnológica en un teatro dirigida a personas con discapacidades auditivas con el fin de tener más inclusión en su público. (Epson, 2019).

2.2. Investigaciones relacionadas con el Tema

Virkkunen, A. (2018). En la tesis titulada Automatic speech recognition for the hearing impaired in an augmented reality application, para obtener el título de Máster en ciencia de tecnología en Aalto University concluye que: la tesis logra abordar el proceso de implementación como la evaluación de una aplicación de asistencia para que las personas con discapacidad auditiva la utilicen en situaciones conversacionales. La aplicación propuesta utiliza la realidad aumentada (RA) para reunir una imagen del hablante y una transcripción automática de su discurso hablado, de modo que se minimiza la información que pierden los discapacitados auditivos al pasar de una a otra. Las principales aportaciones de este trabajo son la aplicación móvil de RA y la validación del enfoque mediante pruebas con usuarios. Además, el trabajo explica la configuración de un servidor de reconocimiento del habla remoto y escalable que utiliza software de código abierto existente para manejar el reconocimiento automático del habla moderno, computacionalmente pesado, que es difícil para los dispositivos móviles.

Comentario:

Esta tesis nos permite identificar distintas técnicas de la aplicación de realidad aumentada en sistemas ópticos, así como también, nos ayuda a definir la técnica adecuada para nuestra construcción del visor. Además, logran demostrar la utilización de servicios remotos para el reconocimiento de voz dentro sus sistemas, así logran reemplazar el trabajo de procesamiento en servicios locales sobre los dispositivos utilizados para su construcción.

Bano, S., Jithendra, P., Niharika, G., Sikhi, Y. (2020). Speech to Text Translation enabling Multilingualism. En la conferencia IEEE International Conference for Innovation in Technology (INOCON), concluye que: al implementar este modelo se ha aprendido cómo se utilizan los paquetes de SpeechRecognition para construir un modelo de traducción del habla. Cuanto más se utilice este tipo de paquetes más flexibilidad se obtendrá en el código y en la salida que se va a mostrar. Este modelo puede ser utilizado en cualquier propósito de traducción de voz a texto. Además, este modelo tiene grandes ventajas, una de ellas es que uno puede sobrevivir en lugares desconocidos donde no se conoce el idioma a hablar, pero con la ayuda de este modelo se puede traducir ese discurso regional a texto y también se puede utilizar en áreas

como las telecomunicaciones y multimedia. Además, este modelo es útil para proporcionar una comunicación eficaz entre el hombre y la máquina.

Comentario:

Este artículo nos ayuda a entender las distintas aplicaciones que tiene el uso de un conversor multilingüe de voz a texto mediante una interfaz gráfica en un software. Con esta aplicación, se logra que el usuario pueda interactuar con el sistema para la elección de los distintos idiomas configurados. Los distintos idiomas alcanzados fueron obtenidos por la personalización y parametrización de paquetes o librerías de SpeechRecognition.

Dabran, I., Avny, T., Singher, E., & Danan, H. B. (2017). Augmented Reality Speech Recognition for the Hearing Impaired. En IEEE International Conference on Microwaves, Antennas, Communications and Electronic Systems (COMCAS), se concluye que esta herramienta permite a las personas sordas y con deficiencias auditivas ver "subtítulos en vivo" de realidad aumentada en tiempo real mientras escuchan una charla sobre un tema específico, combinándolos con los gestos corporales del conferenciante del mundo real en el ambiente. Esta herramienta se puede utilizar en muchas instituciones como universidades, teatros y otros lugares, ayudando a las personas con discapacidad auditiva a comprender charlas u obras de teatro, por ejemplo, como cualquier otra persona. También concluyen que el sistema es fácil de implementar y esperan verlo en el futuro cercano sobre diversos visores ópticos de montaje en la cabeza.

Comentario: Este trabajo nos permite identificar que estos proyectos se pueden aplicar a múltiples escenarios de la vida cotidiana. También, Nos permite validar que la integración del reconocimiento de voz mediante servicios de nube con los dispositivos montables de realidad aumentada que existen en el mercado.

Mirzaei, M., Kan, P., & Kaufmann, H. (2020). EarVR: Using ear haptics in virtual reality for deaf and Hard-of-Hearing people. En IEEE transactions on visualization and computer graphics. En este artículo, se presenta y se evalúa EarVR como dispositivo montable para diferentes HMD-VR(visor de realidad virtual tipo casco). Actúa como asistente de las personas DHH (sordas o con dificultades de escucha) para localizar fuentes de sonido en el entorno de realidad virtual. Analiza la entrada

Sonidos 3D de un entorno de realidad virtual para localizar la dirección de la más cercana fuente de sonido para el usuario. Luego, proporciona la dirección al usuario utilizando dos vibromotores colocados en los oídos del usuario. EarVR ayuda a las personas DHH para completar ciertas tareas de realidad virtual que no pudieron terminar antes de. Además, mejora la experiencia de las personas con DHH al correr aplicaciones de realidad virtual relacionadas con el sonido. Al usar EarVR, las personas con DHH obtienen una experiencia de realidad virtual más cercana a la experiencia de las personas sin problemas auditivos. Los resultados de nuestras pruebas sugieren que EarVR ayuda a las personas con DHH a completar tareas de realidad virtual relacionadas con el sonido y también las alienta a usar y disfrutar de la tecnología de realidad virtual.

Comentario: Este proyecto muestra la integración de microprocesadores en visores de realidad aumentada a través de un computador con el fin de procesar y analizar los audios de las aplicaciones de realidad aumentada para reconocer la procedencia audio-espacial y así dar una referencia al usuario a través de vibradores adheridos al visor comercial.

Tapu, R., Mocanu, B., & Zaharia, T. (2019). DEEP-HEAR: A multimodal subtitle positioning system dedicated to deaf and hearing-impaired people. En IEEE. En este artículo, se presenta un novedoso sistema de posicionamiento dinámico de subtítulos, denominado DEEP-HEAR, diseñado específicamente para mejorar la experiencia de visualización y aumentar la accesibilidad de las personas sordas y con discapacidad auditiva a los documentos multimedia. El sistema propuesto explota conjuntamente algoritmos de visión por computadora y redes neuronales convolucionales profundas para detectar y reconocer al hablante activo. Como trabajo futuro, proyectan extender aún más la arquitectura DEEPHEAR para hacer frente a las bibliotecas de sonido a texto para vídeos donde el documento de subtítulos no está disponible con anticipación. Además, plantean realizar una evaluación de estudio de usuarios más completa en un conjunto de datos más grande con usuarios con problemas de audición.

Comentario: Este artículo desarrolla un software que permite que las personas con discapacidad auditiva puedan ver subtítulos de videos multimedia. Lo esencial de este proyecto es la utilización de algoritmos matemáticos para la separación de

material videográfico y material audible. De esta manera, se logra realizar un tratamiento directo del audio para poder caracterizar las frecuencias de las voces percibidas.

Mahamud, M. S., & Zishan, M. S. R. (2017). Watch IT: An assistive device for deaf and hearing impaired. En IEEE International Conference on Advances in Electrical Engineering (ICAEE). En este artículo concluye que su objetivo principal de este trabajo es ayudar a la persona con discapacidad auditiva a sentirse libre para comunicarse fácilmente con cualquier persona cercana y hacer su vida más cómoda. Arduino es la unidad de control principal de este dispositivo. Un dispositivo con forma de reloj está desarrollado para personas con discapacidad auditiva para que no tengan que sentir ningún tipo de problema para comunicarse. El Arduino fue programado de tal manera que nadie necesita tocar o cambiar ningún programa. La simulación se realizó antes de la implementación del circuito de hardware para asegurarse de que todos los componentes que se utilizan en el hardware funcionaran correctamente después de la implementación. Después de lograr el resultado de la simulación, se implementó el hardware.

Comentario: Del presente estudio, los autores logran utilizar una pantalla de pequeño tamaño para mostrar el resultado del procesamiento de señales de voz mediante servicio de nube alcanzados desde el equipo móvil del usuario. Con ello, logramos conocer sobre la integración de servicios de nube mediante acceso móvil con interacción de microprocesadores.

2.3. Estructura teórica y científica que sustenta el estudio

2.3.1 Visor de realidad aumentada

Según Almenara, J., Jimenez, F. (2016) El término de realidad aumentada se aplica a un tipo de realidad mixta formada por la integración coherente con la realidad física y en tiempo real de una capa de información digital que puede ser diversa (texto, símbolos, audio, vídeo y/u objetos tridimensionales) y con la que es posible la interacción, con el resultado de enriquecer o alterar la información de la realidad física en la que se integra. La Realidad mixta se constituye como una realidad híbrida en la que la percepción de lo físico se acompaña de la percepción de los elementos digitales mezclados. Esta mezcla

puede ser una superposición, una inclusión o una sustitución del entorno circundante o de fondo.

2.3.1.1 Hardware para el visor de realidad aumentada

Arduino: Según Artero, Ó. T. (2013), es una placa hardware libre que incorpora un microcontrolador programable y una serie de pines-hembra (los cuales están unidos internamente a las patillas de E/S del microcontrolador) que permiten conectar allí de forma muy sencilla y cómoda diferentes sensores y actuadores.

2.3.1.2 Software para el visor de realidad aumentada

Arduino IDE:

Según Evans B. (2007), El entorno de desarrollo integrado de Arduino (IDE) es una aplicación multiplataforma utilizada para escribir y cargar programas en placas compatibles con Arduino. La estructura básica del lenguaje de programación de Arduino basado en C++.

2.3.1.3 Sistema Óptico para el visor de realidad aumentada

Lente: Según Malacara, D. (2015), define que un lente es una placa de vidrio cuyas caras son por lo general esféricas y casi paralelas en el centro de ella. Consideremos un haz de rayos paralelos que inciden en una lente muy delgada. Si la lente hace que los rayos refractados converjan, se dice que la lente es convergente, y si hace que diverjan, que la lente es divergente. También se dice que una lente divergente es negativa y que una convergente es positiva.

En la figura 1 se pueden observar los tipos de lentes convergentes:

- Lente biconvexa: Ambas superficies son convexas (más gruesas en el centro).
- Lente plano convexa: Una superficie es plana y la otra convexa.
- Lente menisco: Una superficie es ligeramente cóncava y la otra convexa.

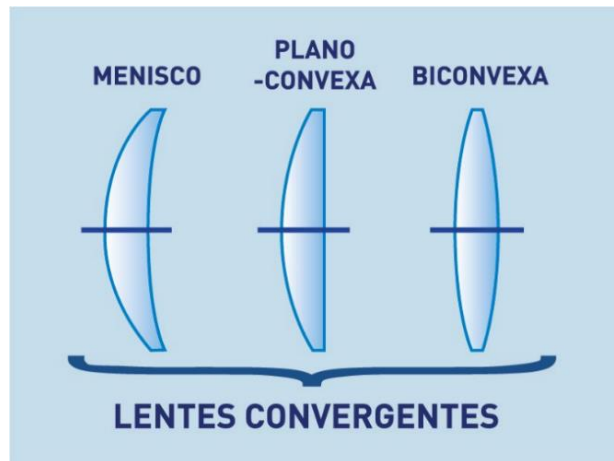


Figura N° 1: Tipos de lentes convergentes

Fuente: Malacara, D. (2015)

Imagen virtual: Según Nave R. (2000), se forma en la posición donde se cruzan las trayectorias de los rayos principales cuando se proyectan hacia atrás desde sus trayectorias más allá de la lente. Tal como se observa en la figura N° 2. Aunque una imagen virtual no forma una proyección visible en una pantalla, tiene una posición y tamaño definidos y puede ser "vista" o captada por el ojo, la cámara u otro instrumento óptico.

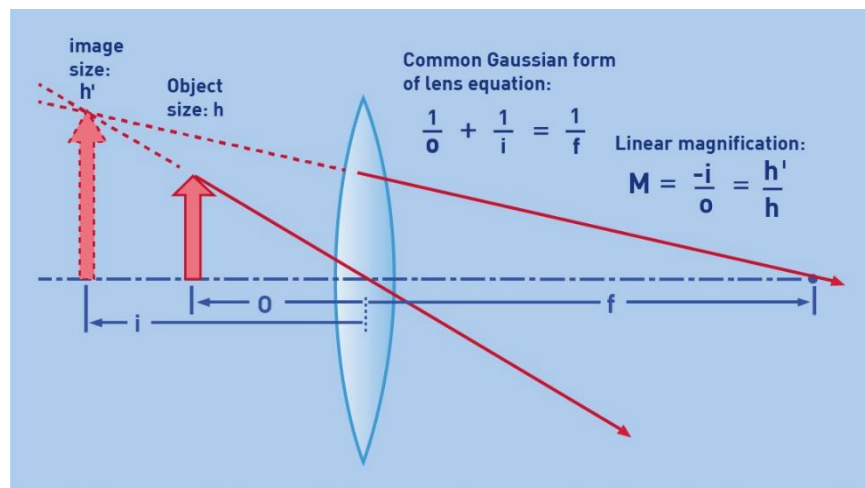


Figura N° 2.: Formación de la imagen virtual

Fuente: <http://hyperphysics.phy-astr.gsu.edu/hbase/geoopt/image4.html>

La ecuación del lente para la formación de una imagen virtual es:

$$\frac{1}{o} + \frac{1}{i} = \frac{1}{f} \quad \dots (1)$$

Donde:

o = distancia del objeto al centro del lente

i = distancia de la imagen al centro del lente

f = distancia focal

La magnificación es la relación entre el tamaño de la imagen y el tamaño del objeto, así es definida en la siguiente fórmula:

$$M = \frac{-i}{o} \quad \dots (2)$$

2.3.2 Conversor multilingüe de voz a texto

Según IBM (2020), en su documentación oficial de servicios de nube, el conversor de voz a texto equivalente al servicio IBM Watson Speech to Text proporciona prestaciones de transcripción de voz para las aplicaciones. El servicio de inteligencia artificial en la nube aprovecha el aprendizaje automático para combinar el conocimiento de gramática, estructura de lenguaje y composición de señales de audio y de voz para transcribir con precisión la voz humana. Tiene la cualidad de realizar actualizaciones continuamente y logra su perfección su transcripción a medida que recibe más conversación. Este servicio proporciona varias interfaces que permiten que se adapte a cualquier aplicación en la que la voz es la entrada y una transcripción textual es la salida.

2.3.2.1 Inteligencia artificial

Según Overton, J. (2018), la inteligencia artificial (IA) es cuando una máquina realiza una tarea que los seres humanos consideran interesante, útil y difícil de hacer. La actual ola de IA funciona utilizando modelos informáticos para simular un comportamiento inteligente. La IA puede impulsar el crecimiento económico al automatizar la mano de obra, mejorar la eficiencia y convertirse en una fuente de innovación de productos o servicios.

2.3.2.2 Hardware para el conversor multilingüe de voz a texto

Raspberry Pi 3B:

Según Rodríguez, E. (2018), Es un ordenador de bajo coste y tamaño

reducido, tanto es así que cabe en la palma de la mano, pero puedes conectarle un televisor y un teclado para interactuar con ella exactamente igual que cualquier otra computadora. Se puede usar en proyectos de electrónica y para tareas básicas que haría cualquier ordenador de sobremesa como navegar por internet, hojas de cálculo, procesador de textos, reproducir vídeo en alta definición.

2.3.2.3 Software para el conversor multilingüe de voz a texto

Node Red: Según Sancho P. (2020), es una herramienta de código libre (Open Source) construida en Node.js y que se encuadra en la familia de herramientas "flow-based programming (FBP) tools" ('Herramientas de programación basadas en flujos'). La programación basada en flujos es una forma de describir el comportamiento de una aplicación como una red de cajas negras o "nodos", como se les llama en Node-Red.

2.3.2.4 Servicios de nube

IBM Watson: La definición por parte de IBM (2016) es la siguiente: IBM Watson es la plataforma de inteligencia artificial (IA) de IBM. Utiliza tecnologías de lenguaje natural (NLP), la visión por ordenador y las tecnologías de aprendizaje automático para grandes cantidades de datos no estructurados. Asimismo, Watson permite crear aplicaciones cognitivas que ayudan a mejorar, escalar y acelerar la experiencia humana.

2.3.2.5 Filtro

Según Cogollos, S (2016), un Filtro es un elemento que discrimina una determinada frecuencia o gama de frecuencias de una señal eléctrica que pasa a través de él, pudiendo modificar tanto su amplitud como su fase. El propósito de los filtros es separar la información de interferencias, ruido y distorsión no deseada. Este elemento se modela con su función de transferencia.

2.4. Definición de términos básicos

2.4.1 Conversor multilingüe de voz a texto

Según Santiago, F., Singh, P. (2017) se define como un conversor del habla en texto legible según el idioma que el usuario especifique. El servicio es capaz de transcribir el habla de varios idiomas y formatos de audio a texto con baja latencia.

2.4.2 Procesamiento de lenguaje natural

Según Vásquez, A. (2009) consiste en la utilización de un lenguaje natural para comunicarnos con la computadora, debiendo ésta entender las oraciones que le sean proporcionadas, el uso de estos lenguajes naturales facilita el desarrollo de programas que realicen tareas relacionadas con el lenguaje o bien, desarrollar modelos que ayuden a comprender los mecanismos humanos relacionados con el lenguaje.

2.4.3 Implante Coclear

Federación de Asociaciones de Implantados Cocleares de España, 2015) El Implante Coclear es un transductor que transforma las señales acústicas en señales eléctricas que estimulan el nervio auditivo.

2.5. Diseño de la Investigación

2.5.1. Variables de investigación

Variable independiente: Conversor multilingüe de voz a texto.

Variable dependiente: Visor de texto en realidad aumentada.

2.5.2. Tipo y Método de investigación

El tipo de investigación es aplicada y tecnológica. El método de investigación es experimental debido a que se realiza el tratamiento de audios de la voz de hablantes, los cuales serán capturados mediante un micrófono dentro de un sistema que orquesta servicios de inteligencia artificial para la conversión de voz a texto y traducción los cuales serán alcanzados a través del internet gracias a los servicios de nube, de esta manera se podrá crear una imagen virtual a través de un sistema óptico para la visualización del texto procesado en un visor de realidad aumentada.

2.5.3. Técnicas e Instrumentos de recolección de datos

La técnica es la grabación de la voz análoga de una conversación mediante un micrófono con características específicas para luego ser procesada. Este micrófono será el instrumento de recolección de datos, el cual cuenta con una sensibilidad de $-47\text{dB} \pm 4\text{dB}$, con una respuesta de frecuencia de 100Hz-16KHz, una impedancia de $\leq 2.2\text{K}\Omega$, el voltaje de funcionamiento es de 4.5 V DC y una relación S/N de $> 67\text{dB}$.

2.5.4. Procedimiento para la recolección de datos

Para la presente investigación, se ha definido la intervención de 2 personas con discapacidad auditiva de la asociación de jóvenes sordos del Perú. Las personas involucradas deberán de estar en un lugar cerrado libre de sonidos de alto volumen para, con estas condiciones, utilizar el visor de realidad aumentada propuesto. Cada persona va a interactuar mediante una conversación con otra persona hablante, para así, lograr que el visor pueda procesar las voces percibidas y enviar el texto que será convertido y traducido hacia el lente del visor.

CAPÍTULO III: DESARROLLO DEL PROYECTO

Con el contexto explicado en el capítulo anterior de manera detallada, este capítulo presentó los diferentes procedimientos con los cuales se ha desarrollado el visor de realidad aumentada. En este capítulo, se explicó desde la forma más general del proyecto hasta los detalles de implementación. La sección 3.1 detalla una introducción sobre el proyecto, y sus casos de uso en diferentes escenarios, así mismo, explica sobre el funcionamiento técnico de manera lógica. La siguiente sección 3.2 cubre la explicación del desarrollo de los mecanismos de preprocesamiento de voz, desde el detalle teórico del diseño del filtro de voz utilizado hasta el almacenamiento de los audios procesados. En la sección 3.3, se explica sobre la integración de los sistemas Edge Computing con los distintos servicios de nube para el procesamiento relacionados con la Inteligencia artificial, según la necesidad. La última sección, 3.4 brinda el detalle de la implementación de los sistemas de lentes para la creación y reflexión de la imagen virtual creada a partir del texto procesado.

De esta manera, en la Figura N° 3, se muestra el diagrama de flujo general del proyecto de tesis planteado; donde se divide en bloques de funcionamiento respecto a la acción por parte del hablante, la funcionabilidad de procesamiento del visor y la recepción del resultado por parte del usuario. En la figura mencionada, se podrán visualizar los diferentes pasos necesarios para el procesamiento del visor de realidad aumentada; de esta manera, el visor logrará cumplir la tarea de comunicación entre el hablante y el usuario del visor. Se resaltan las siguientes tres etapas de funcionamiento: preprocesamiento de voz, integración de sistemas de Edge computing y el sistema óptico. Estas etapas serán explicadas y detalladas en las siguientes secciones de este capítulo.

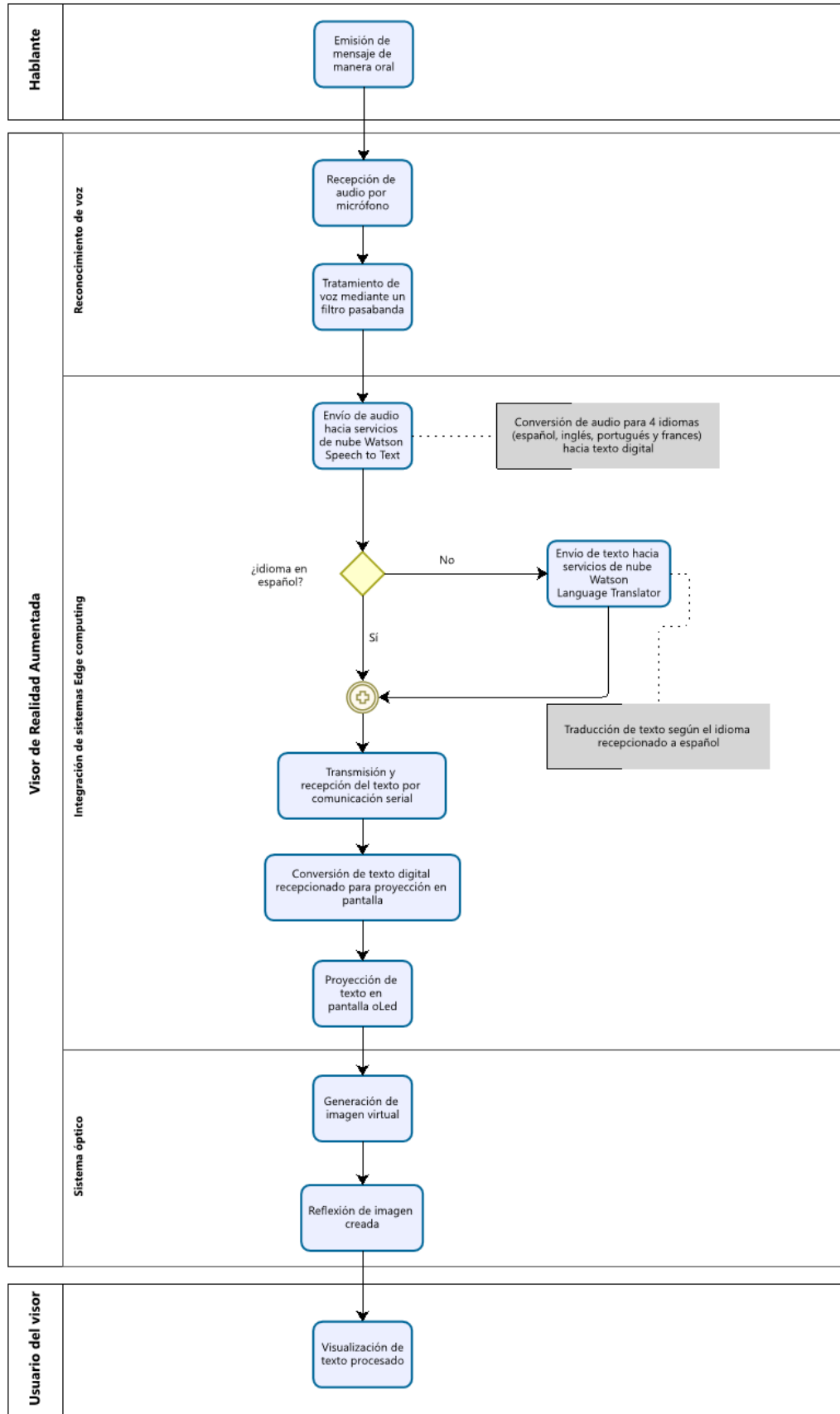


Figura N° 3: Flujo general del proyecto

Fuente: Elaboración propia.

3.1. Estructura del sistema

La estructura del funcionamiento para el visor de realidad aumentada está conformada por las siguientes tres entidades:

- El hablante, quien es la persona que emitirá un mensaje de manera hablada hacia el usuario del visor.
- El visor de realidad aumentada, cual realiza el trabajo de conversión del mensaje hablado a texto escrito para mostrarlo en el reflector del dispositivo.
- Usuario del visor, quien es la persona con discapacidad auditiva que podrá visualizar el mensaje de voz procesado en forma texto.
- En la figura N° 4 se aprecia la estructura del sistema.

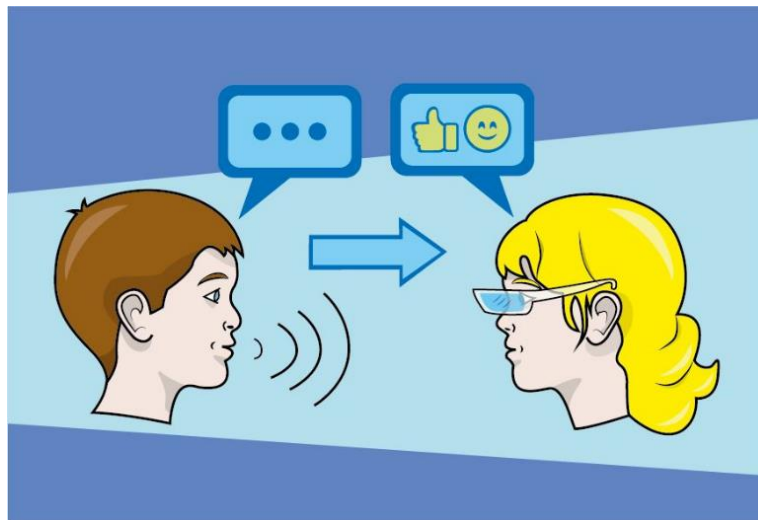


Figura N° 4: Funcionamiento del visor en la comunicación del usuario y el hablante
Fuente: Elaboración propia.

De las tres entidades mencionadas, la labor computacional fue ejecutada por el visor de realidad aumentada. La estructura del sistema del visor consiste de tres procedimientos identificados como los mecanismos de preprocesamiento de voz, la integración de sistemas de Edge computing y el sistema óptico.

El mecanismo de procesamiento de voz se compone de cinco actividades desplegadas en el Raspberry Pi 3B. Se inició con la activación mediante un panel táctil que muestra triggers como botones digitales, con ello, el usuario logró seleccionar el idioma a procesar. Luego de la activación del trigger, se habilitó la

recepción del audio mediante el micrófono integrado al dispositivo que captó la voz más cercana y de mayor volumen. Con el fin de afinar el audio percibido, se realizó la conversión de señal estéreo a señal de un solo canal, ya que se enfatizó el tratamiento de señales de voces y no de señales con frecuencias que podrían denotar un entorno multicanal. Una vez que la señal fue convertida en un audio de un solo canal, se procesó por un filtro digital con el fin de restringir el procesamiento de señales fuera del rango de frecuencia de la voz humana; con ello, se logró contar con una señal de voz lo más limpia posible. Se finalizó este primer procedimiento con el almacenamiento del audio filtrado de manera local para el procesamiento posterior.

La integración de sistemas de Edge computing es el bloque de funcionamiento del dispositivo que permitió realizar llamados a servicios de nube a través del Raspberry Pi 3B, el cual se conectó a una red de internet; para así, lograr acudir a servicios complejos de inteligencia artificial. Como primera actividad, se tiene un bloque de tarea que realizó la validación del idioma seleccionado donde se enviaron solicitudes hacia los servicios de nube de IBM Watson para el procesamiento de conversión de audio a texto, así como también para el procesamiento de traducción del texto percibido. Este bloque de validación del idioma es detallado más adelante ya que está compuesto de un subflujo de tareas. Una vez que el texto fue procesado por los servicios de nube, se realizó la transferencia mediante comunicación serial hacia el Arduino. Luego de la recepción, se procesó el texto para enviar señales que representaron el contenido del texto hacia una pantalla OLED que se situó internamente en el caparazón del visor.

Finalmente, el sistema óptico tiene la finalidad de formar imágenes virtuales a partir de la proyección del texto que ha sido procesado por los servicios de inteligencia artificial. Este fue el encargado de habilitar la visualización del texto hacia el usuario. Se realizó la reflexión de la imagen virtual mediante un espejo que se situó de manera diagonal frente a la pantalla OLED. Luego, esta imagen creada de la reflexión fue maximizada gracias a un lente convergente plano convexo. Esta imagen maximizada se proyectó en el reflector principal del visor que se situó frente a la vista del usuario.

En la figura N° 5, se aprecia la representación de los tres procesamientos indicados en los párrafos superiores.

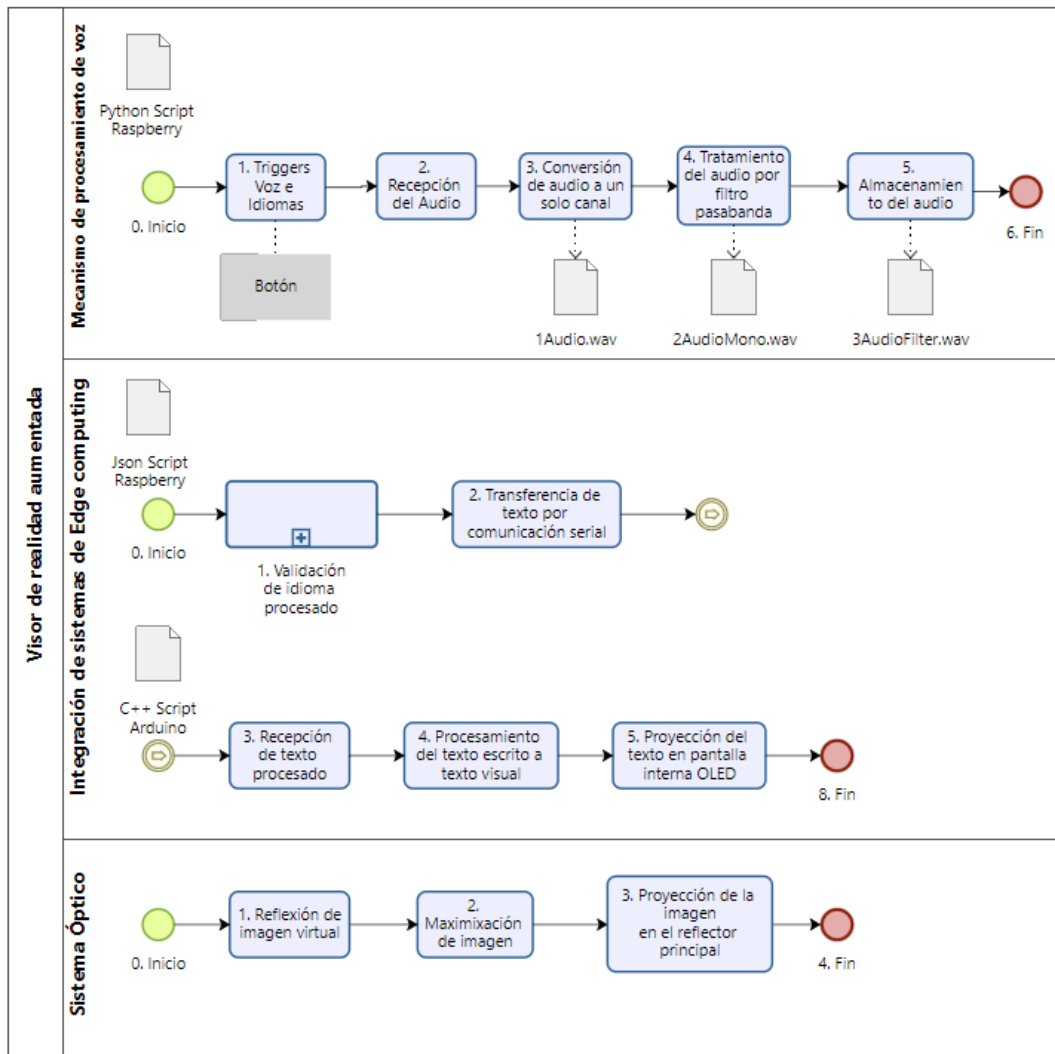


Figura N° 5: Flujo técnico del proyecto

Fuente: Elaboración propia

De la figura anterior N°5, en el bloque de la integración de sistemas de Edge computing se detalló el primer bloque de tarea que fue la validación de idioma procesado. En el flujo mostrado en la figura N°6, se muestran cuatro secuencias de actividades que refieren al procesamiento del texto de acuerdo al botón interactuado previamente, el cual desencadenó acciones según el idioma seleccionado. Para las tareas referentes a la primera línea de actividades, se visualizaron tres actividades para el idioma español. El trigger activado envió información hacia el servicio de Speech to Text configurado con el idioma español para luego trabajar en la respuesta del servicio de nube. Esta respuesta fue tratada para modificar y modelar la sintaxis del texto resultante. Luego, para la segunda y tercera línea de actividades, se visualizó una secuencia similar de actividades respecto al idioma español; sin

embargo, los servicios de Speech to Text fueron configurados según el idioma a tratar, en este caso para el inglés y francés respectivamente. Luego del modelamiento de la sintaxis, para estas dos líneas de flujo, se acudió al siguiente servicio de nube IBM Watson Language Translator para realizar la traducción del idioma seleccionado hacia el idioma español. Para la cuarta línea de actividad del presente bloque de validación, se configuró el servicio de Speech to Text para el idioma portugués. Finalmente, luego del modelamiento, se acudió a dos servicios de traducción. Debido a que los servicios de nube elegidos no contaron con una traducción directa de portugués a español, se integró secuencialmente la traducción de portugués a inglés, y luego, se procesó la traducción de inglés a español.

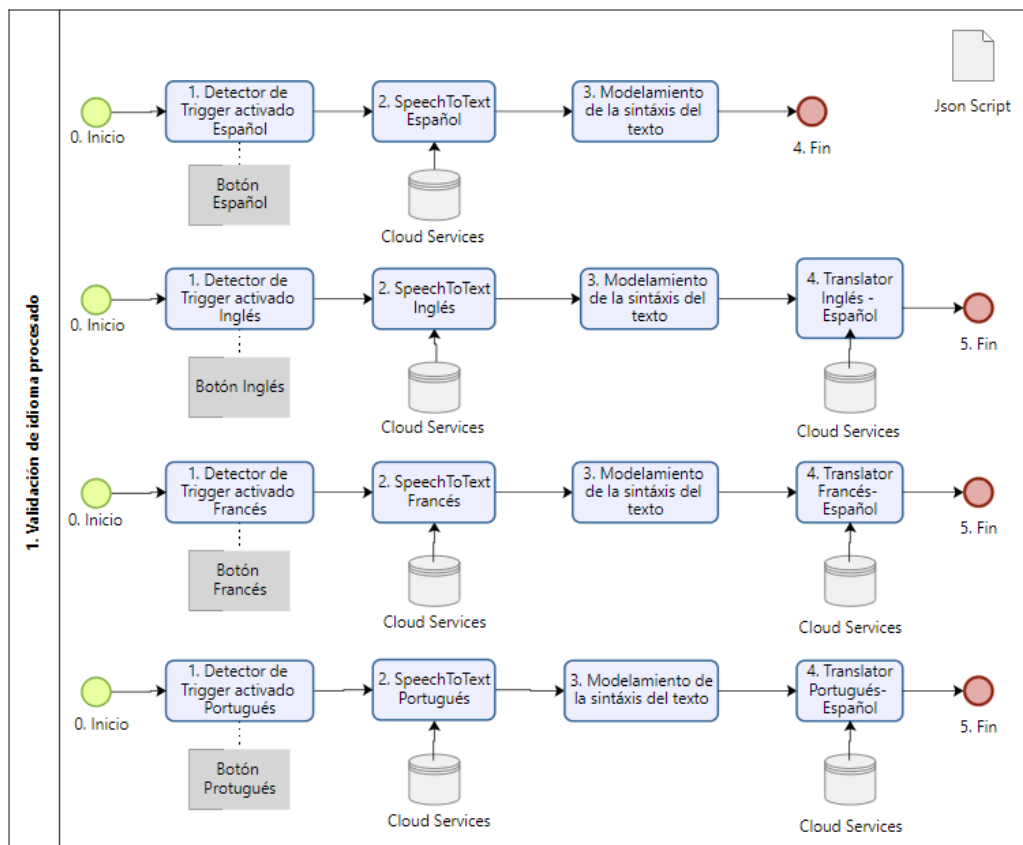


Figura N° 6: Flujo técnico de los servicios de nube IBM Watson

Fuente: Elaboración propia

En las siguientes secciones, se detalla el desarrollo técnico de los tres procedimientos especificados: el mecanismo de procesamiento de voz, la integración de sistemas de Edge computing, y el sistema óptico.

3.2. Desarrollo de los mecanismos de preprocesamiento de voz

3.2.1 Diseño teórico de filtro pasa-banda

En el bloque del procesamiento de la voz se utilizó un filtro digital pasa banda para filtrar la frecuencia de la voz comprendida entre los 50Hz – 3KHz. De esta manera, se evitan frecuencias ajenas al rango denominadas como perturbaciones, que puedan interferir en la conversión de la voz a texto.

Para el diseño del filtro digital pasa banda se utilizó el filtro elíptico ya que tiene una mejor respuesta en frecuencia, y se caracteriza por tener una banda de transición estrecha acercándose al filtro ideal.

Se desarrolló en el software Matlab y a continuación se puede observar en la figura N° 7 la programación, y en la figura N° 8 la función de transferencia del filtro digital pasa banda de orden 8:

```
58 %% Filtro eliptico pasa banda
59
60 - Fc=[50 3000];
61 - Fs=8000
62 - n=4;
63 - [num,den]=ellip(n,1,40,Fc/(Fs/2),'bandpass');
64 - sys=tf(num,den,1)
65 - [h,w]=freqz(num,den);
66 - plot(w,20*log10(abs(h)))
67
68 - xlabel("Frecuencia [KHz]");
69 - ylabel("Magnitud [dB]");
```

Figura N° 7: Diseño del filtro

Fuente: Elaboración propia


```
sys =
-----
0.3141 z^8 - 0.09591 z^7 - 1.059 z^6 + 0.0785 z^5 + 1.525 z^4 + 0.0785 z^3 - 1.059 z^2 - 0.09591 z + 0.3141
-----
z^8 - 2.137 z^7 + 0.5553 z^6 + 0.3488 z^5 + 1.562 z^4 - 1.47 z^3 + 0.1361 z^2 - 0.2251 z + 0.2302
```

Figura N° 8: Función de transferencia

Fuente: Elaboración propia

Seguidamente, se graficó la respuesta en frecuencia del módulo del filtro, tal como se observa en la figura N° 9:

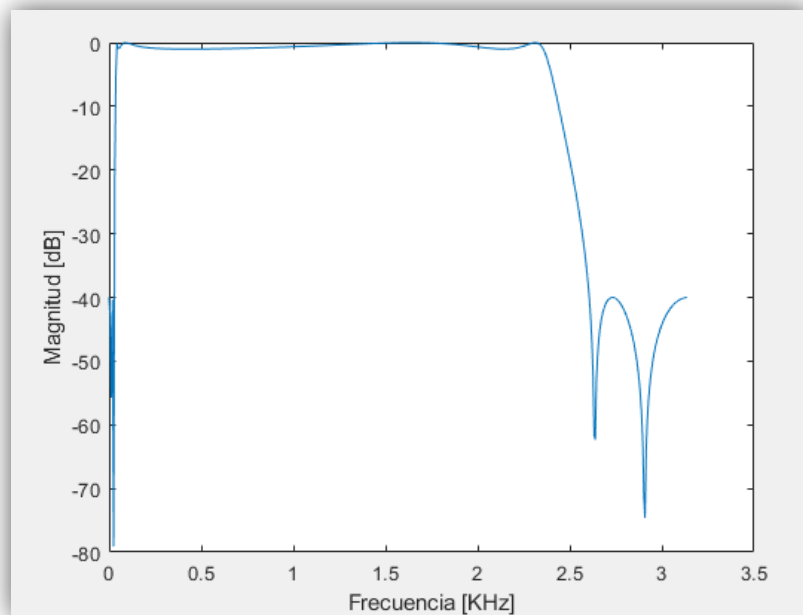


Figura N° 9: Módulo del filtro elíptico pasa banda

Fuente: Elaboración propia

3.2.2 Implementación de filtro digital pasa-banda en lenguaje Python

Luego, se procedió a realizar la programación en Python para implementar el filtro digital.

Se realizaron pruebas inyectando un audio con perturbaciones. A continuación, el procedimiento:

Se convirtió el audio de estéreo a mono y luego se cargó para poder ser procesado tal como se observa en la figura N ° 10, también se graficó el audio en el dominio del tiempo.

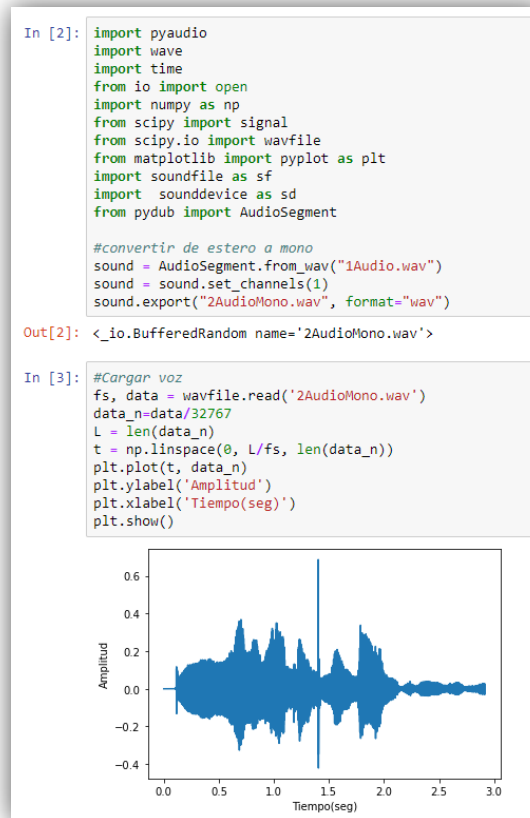


Figura N° 10: Grafico de la voz en el dominio del tiempo

Fuente: Elaboración propia

Se observa en la figura N ° 11 el gráfico del espectro en frecuencia del audio con perturbaciones:

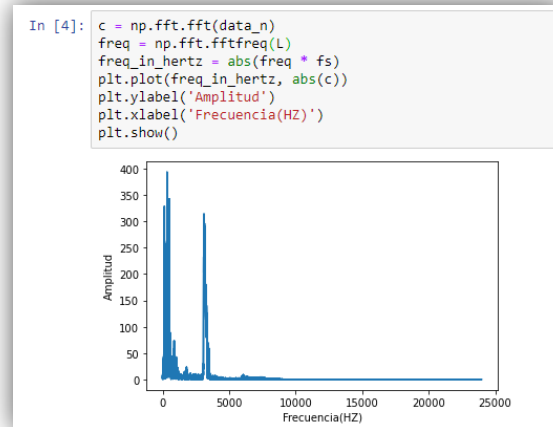


Figura N° 11: Espectro en frecuencia del audio

Fuente: Elaboración propia

Se implementó el algoritmo para el filtro elíptico pasa banda y se filtra el audio tal como se observa en la figura N° 12:

```
In [5]: #Filtro
sos = signal.ellip(8, 1, 40, [50, 3000], 'bandpass', fs=fs, output='sos')
filtered = signal.sosfilt(sos, data_n)
```

Figura N° 12: Programación del filtro elíptico pasa banda

Fuente: Elaboración propia

Se observa en la figura N° 13 el gráfico del espectro en frecuencia del audio filtrado, y se puede apreciar que la señal de ruido de frecuencia alta fue suprimida:

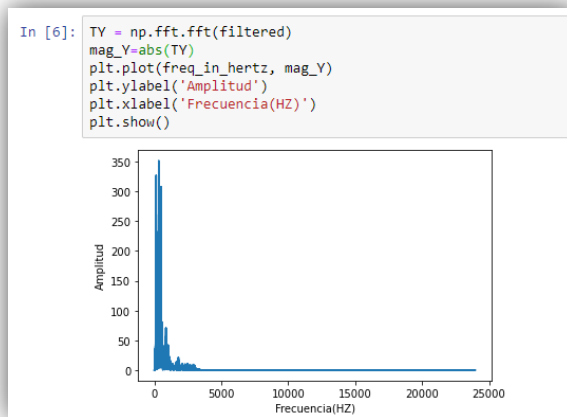


Figura N° 13: Espectro en frecuencia del audio filtrado

Fuente: Elaboración propia

3.3. Integración de sistemas Edge computing

3.3.1 Implementación del filtro en Node-Red

Para el almacenamiento del audio tratado se definió de la siguiente manera:

- El archivo “1Audio.wav” es el audio grabado por el micrófono en estéreo.
- El archivo “2AudioMono.wav” es el audio convertido de estéreo a mono (1 canal).
- El archivo “3AudioFilter.wav” es el audio filtrado y listo para pasar por el Speech to Text.

La programación en Python explicada en la subsección anterior, fue separado de la siguiente manera:

- El archivo “Prefilter.py” es la programación para convertir el audio de estéreo a mono
- El archivo “MainFilter.py” es la programación para filtrar el audio.

En la figura N° 14 se muestra la ubicación del almacenamiento de los audios y el archivo de la programación del filtro, mostrados en una carpeta de la memoria del Raspberry Pi 3B; de esta forma, los archivos quedaron listos para su ejecución.

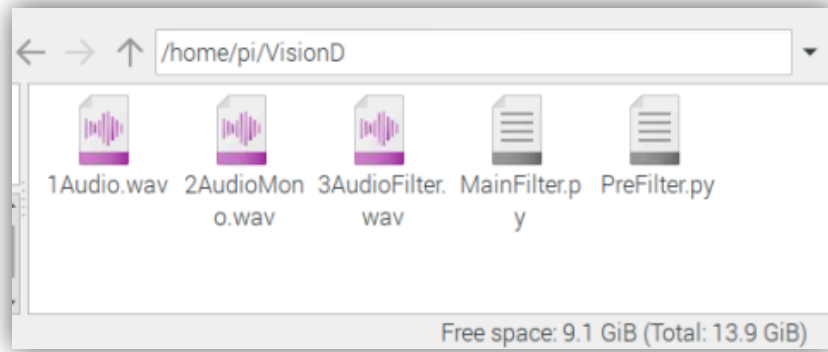


Figura N° 14: Almacenamiento de audios y filtros

Fuente: Elaboración propia

Además, se introdujo el código del filtro en lenguaje Python en la herramienta Node-Red. Para ello, se definió el flujo del código.

En la figura N° 15 se observa el proceso de filtrado del audio conformado por los nodos: Audio original, Prefilter y MainFilter.

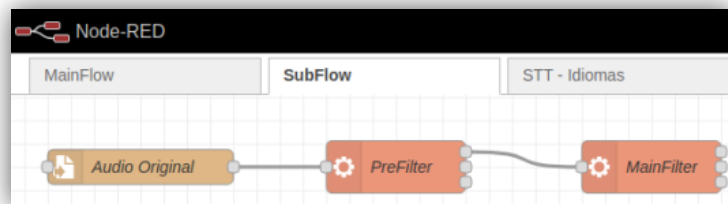


Figura N° 15: Implementación del filtro en Node-Red

Fuente: Elaboración propia

En la figura N° 16 se observa que la configuración del primer nodo guardó el audio grabado en la ruta previamente definida, la acción fue la sobrescritura del archivo ya que por defecto el nodo guarda el archivo de audio para procesarlo posteriormente. La salida simplemente activó la ejecución del siguiente nodo.

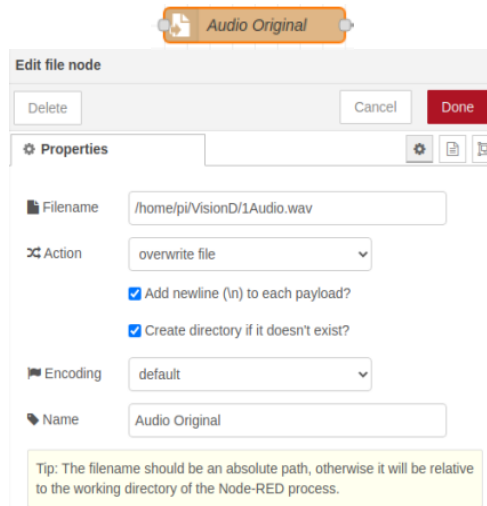


Figura N° 16: Nodo de archivo

Fuente: Elaboración propia

En la figura N° 17 se observa que el segundo nodo ejecutó el código PreFilter.py y que convirtió el audio de estéreo a mono, para luego almacenarlo en la ruta previamente definida.

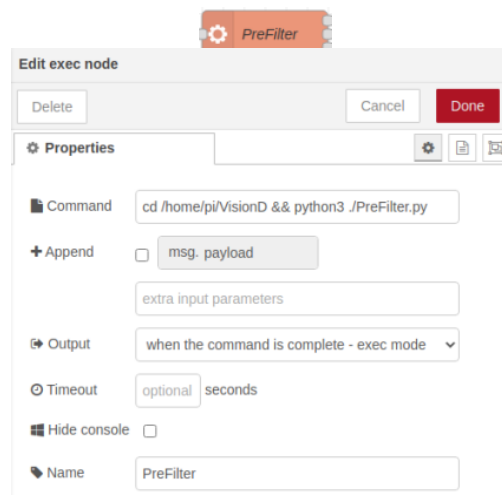


Figura N° 17: Nodo de ejecución del PreFilter.py

Fuente: Elaboración propia

En la figura N° 18 se observa que el tercer nodo ejecutó el código MainFilter.py y que filtró el audio, para luego almacenarlo en la ruta previamente definida.

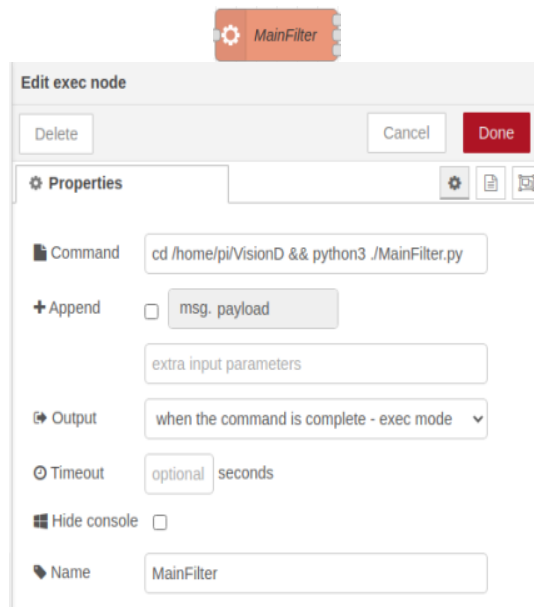


Figura N° 18: Nodo de ejecución del MainFilter.py

Fuente: Elaboración propia

3.3.2 Codificación de interfaces a gráfico mediante Node-Red

Para grabar la voz se utilizó un nodo que reconoce el micrófono y empezó a grabar al presionar el botón. Se utilizó un nodo que guardó el audio grabado en la ruta mencionada anteriormente. Se utilizaron los nodos de función para ejecutar los códigos de programación Python para filtrar la voz. En la figura N° 19 se puede observar la implementación final del filtro en el Node-Red.

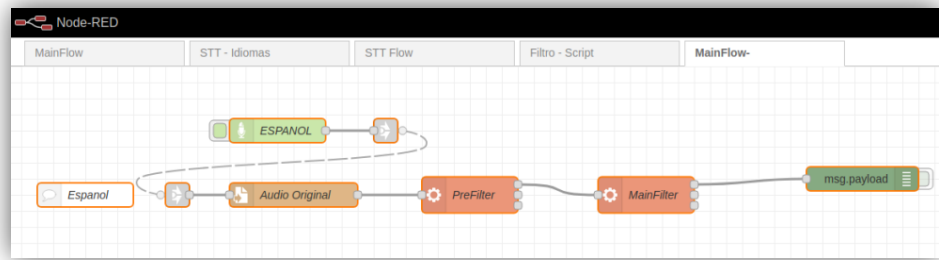


Figura N° 19: Implementación final del filtro digital en Node-Red

Fuente: Elaboración propia

3.3.3 Orquestación con sistemas de servicios de nube para el reconocimiento de VOZ

De manera resumida, el objetivo de la utilización de los servicios de reconocimiento de voz es llegar a realizar la conversión de las señales de voz hablada, al equivalente del texto en forma escrita. Esta acción es realizada de una manera natural para las personas con sentido auditivo sin afectaciones severas; sin embargo, esta tarea es compleja en simular desde un ambiente computacional. El realizar la conversión de voz a texto mediante la computación requiere de altos niveles de procesamiento de algoritmos de inteligencia artificial, los cuales son entrenados para detectar las señales de voz según las palabras percibidas en un específico idioma. Dicho procesamiento es llevado al Raspberry Pi 3B, el cual llega a consumir muchos recursos ya que un sistema de reconocimiento de voz requiere que las especificaciones técnicas del procesador sean de alto rango. Adicionalmente, el mencionado entrenamiento de inteligencia artificial requiere el entendimiento de varios componentes fonéticos y lingüísticos como por ejemplo el sustantivo, verbo, predicados y adjetivos; con este entendimiento, se logra comprender la integridad de dicho mensaje de voz procesado.

Este procesamiento puede ser invocado mediante servicios de nube con el fin de no realizar el procesamiento complejo de manera local en el Raspberry Pi 3B. Estos servicios son alcanzados gracias a la conectividad de internet del Raspberry Pi 3B hacia servicios de inteligencia artificial en

la nube pública. Para el presente proyecto, se acudió a los servicios de IBM en su plataforma de inteligencia artificial IBM Watson.

Dentro de esta plataforma, se encontró el servicio llamado IBM Watson Speech to Text. Este servicio transcribe audio a texto para habilitar las prestaciones de transcripción de voz en diferentes aplicaciones. La habilitación de este servicio se realiza mediante el ingreso en la página web de la nube de IBM. La figura N° 20 muestra la visualización del servicio Speech to Text en la plataforma web.

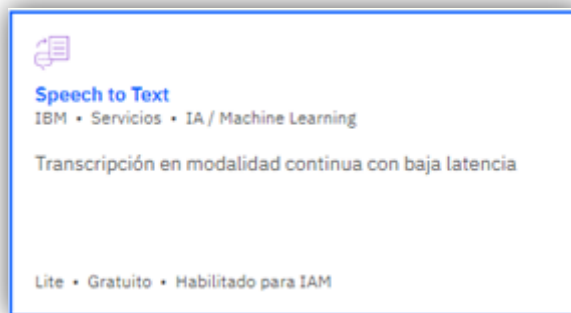


Figura N° 20: Servicio Speech to Text en la plataforma web de la nube de IBM

Fuente: Elaboración propia

La creación del servicio se dirige a la figura N° 21 que muestra la pantalla de gestión del recurso Speech to Text en la plataforma web. En la figura mostrada, se aprecia la sección de credenciales donde se mostraron dos valores esenciales para la invocación de estos servicios de inteligencia artificial: la clave de API y la URL del servicio.

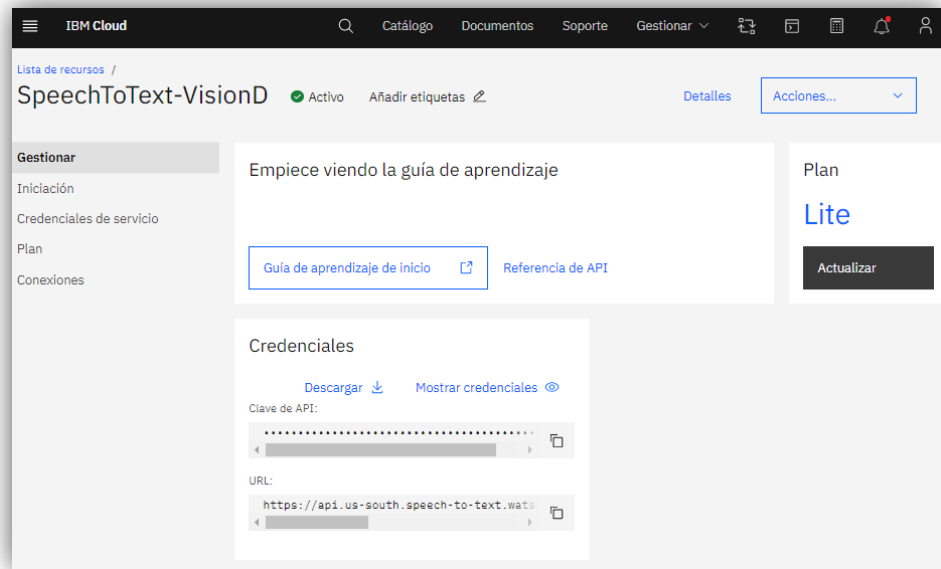


Figura N° 21: Gestión del recurso Speech to Text

Fuente: Elaboración propia

Para la integración del servicio en la nube en el Raspberry Pi 3B, fue necesario definir los nodos a utilizar para la invocación al servicio creado previamente (IBM Watson Speech to Text). En la interfaz de Node-Red, se agregó el nodo de Speech to Text de la paleta de IBM Watson services; este nodo se utilizó para reconocer voz en español. Para poder establecer la conexión con los servicios de nube, se tuvo que configurar los parámetros del api key y la url que se obtuvieron anteriormente tal como se observa en la figura N° 22.

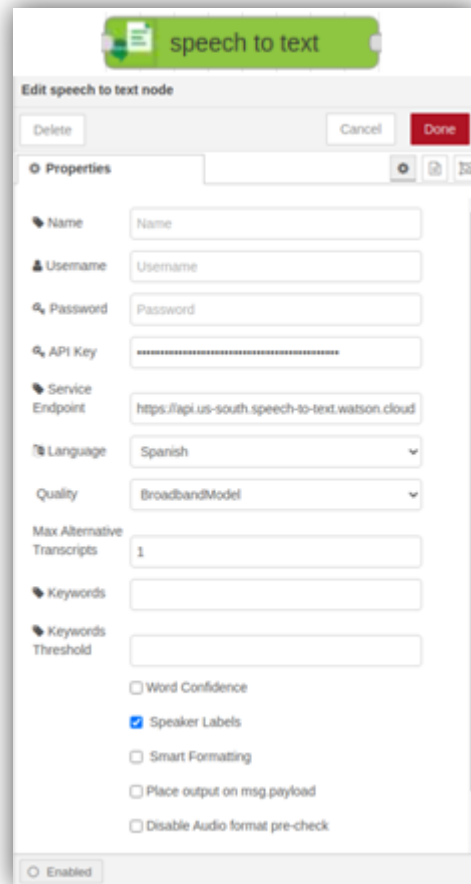


Figura N° 22: Nodo de Speech to Text

Fuente: Elaboración propia

La salida del nodo dió como resultado una sentencia json el cual fue tratado para extraer el texto de la voz procesada. De esta forma, se contó con el resultado esperado que es el texto referente al audio que ha sido procesado por el servicio de nube.

En la figura N° 23, se observa el flujo de manera parcial para el tratamiento de la voz en español. Se aprecia la activación del micrófono para el idioma español. Luego, se aprecian los bloques que corresponden al pre-procesamiento de la voz con los filtros de audios. Con ello, se ingresó al bloque de Speech to Text para realizar la conversión de audio a texto, y finalmente este fue el bloque que extrajo netamente el texto de la respuesta del servicio de nube.

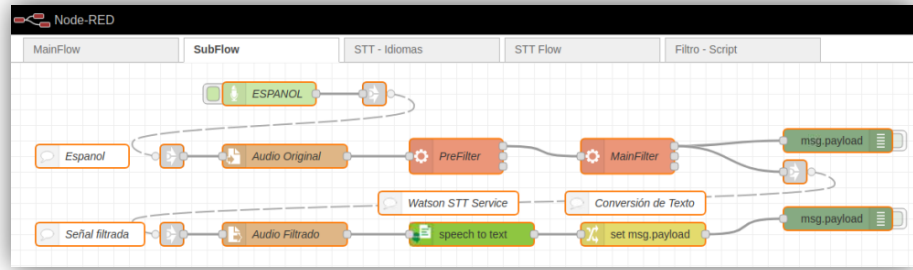


Figura N° 23: Flujo parcial para el tratamiento de la voz en español

Fuente: Elaboración propia

En la figura N° 24 se aprecia la conversión de voz a texto por el flujo anterior que se ha explicado.

```
9/15/2021, 8:45:09 PM node: b028b47c6e6abc15
msg.payload : string[4]
"holá"
```

Figura N° 24: Texto procesado

Fuente: Elaboración propia

3.3.4 Orquestación con sistemas de servicios de nube para la traducción

En esta sección, se detalla el proceso de traducción de manera similar al bloque de conversión de voz a texto. Como se mencionó, el procesamiento de entrenamiento de algoritmos para traducción tampoco fue ejecutado desde el Raspberry Pi 3B. Este procesamiento se llevó a cabo mediante el llamado de servicios de nube.

En la figura N° 25, se aprecia el IBM Watson Language Translator que es el servicio encargado de la traducción de textos en varios idiomas según el sitio web de IBM. Para esta aplicación, se invocó el servicio para el tratamiento de los textos en idioma inglés, francés y portugués.

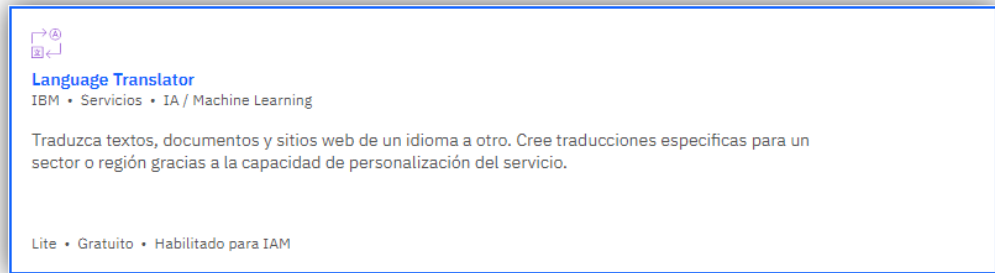


Figura N° 25: Servicio Language Translator en la plataforma web de la nube de IBM

Fuente: Elaboración propia

La creación del servicio se dirige a la figura N° 26 que muestra la pantalla de gestión del recurso Language Translator en la plataforma web. En la figura mostrada, se aprecia la sección de credenciales de la clave de API y la URL del servicio.

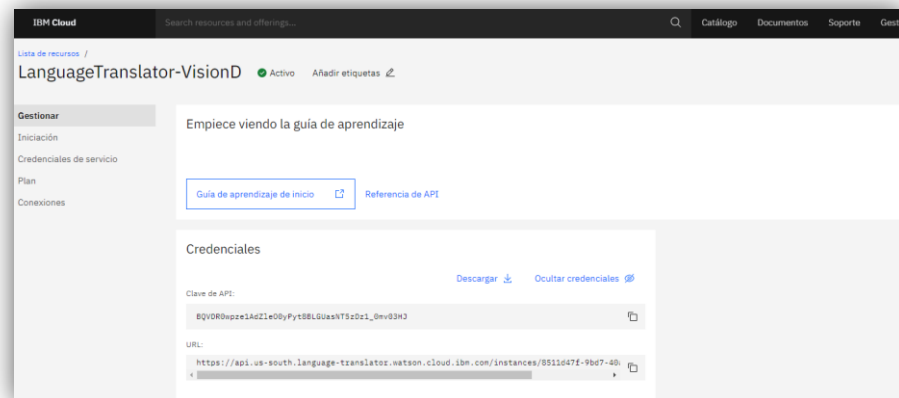


Figura N° 26: Gestión del recurso Language Translator

Fuente: Elaboración propia

Para poder traducir el texto, previamente agregó el nodo de Speech to Text para reconocer voz en inglés, y se configuró el api key y la url para poder

realizar la integración con los servicios de nube tal como se observa en la figura N° 27.

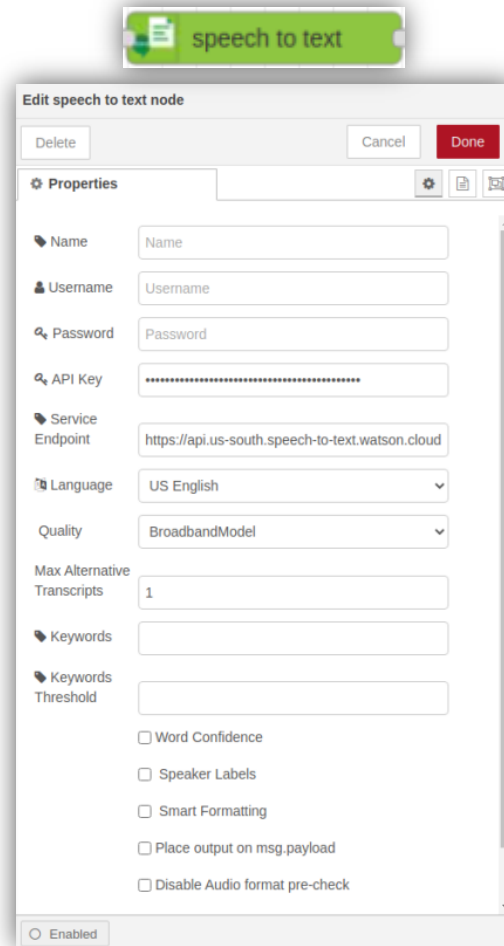


Figura N° 27: Nodo Speech to text para reconocer en voz en inglés

Fuente: Elaboración propia

En la figura N° 28, se observa el nodo de Language Translator para traducir el texto de inglés a español. Además, la conexión con los servicios de nube se tuvo que configurar los parámetros del api key y la url que se obtuvieron anteriormente. De esta manera, la salida dio como resultado el mensaje traducido.

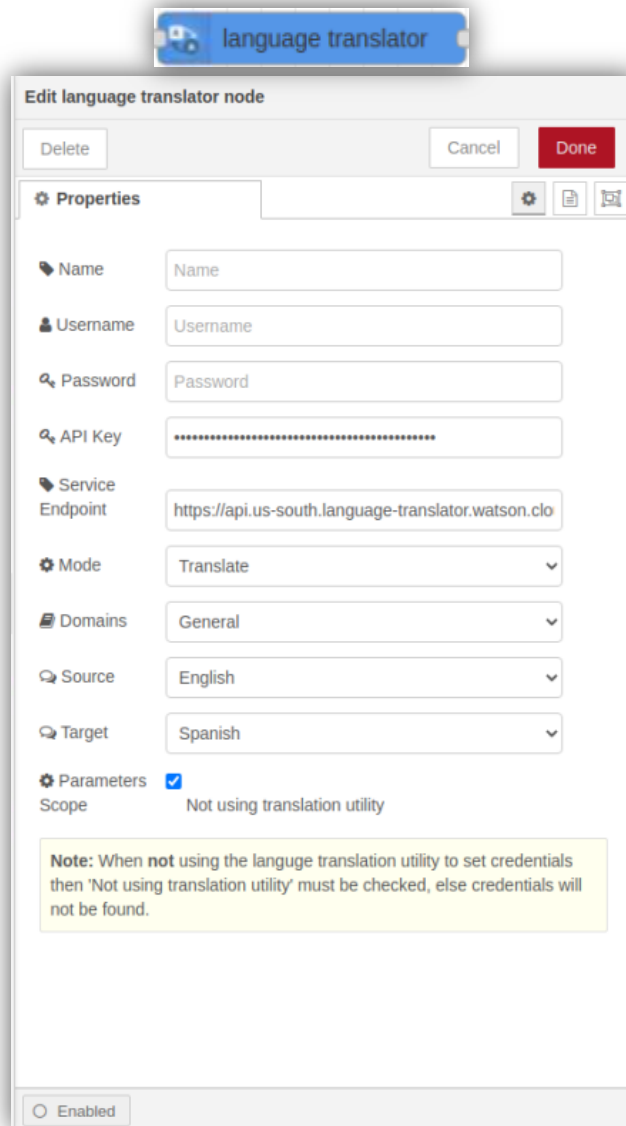


Figura N° 28: Nodo de language translator

Fuente: Elaboración propia

En la figura N° 29, se observa el flujo explicado anteriormente, con la diferencia que se agregó el nodo Speech to Text para reconocer voz en inglés y el nodo de Language Translator para la traducción de inglés a español.

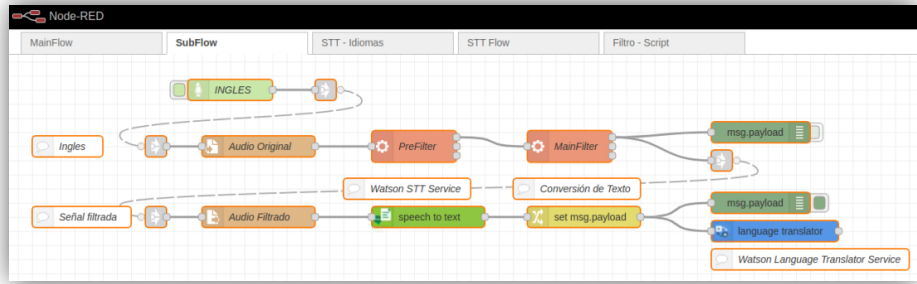


Figura N° 29: Flujo parcial para el tratamiento de la voz en inglés

Fuente: Elaboración propia

De la misma manera, se realizó un flujo similar para reconocer la voz en francés y portugués, y luego se acudió al servicio de IBM Watson Language Translator para traducir el texto a español.

3.3.5 Proyección de texto procesado

Una vez que se procesó la voz por los servicios de nube anteriormente mencionados, se obtuvo el texto digital de manera local en el Raspberry Pi 3B. Para la transmisión de este texto, se utilizó un nodo que envió el texto por comunicación serial al Arduino. En la figura N° 30 se aprecia el nodo que permitió la comunicación serial.

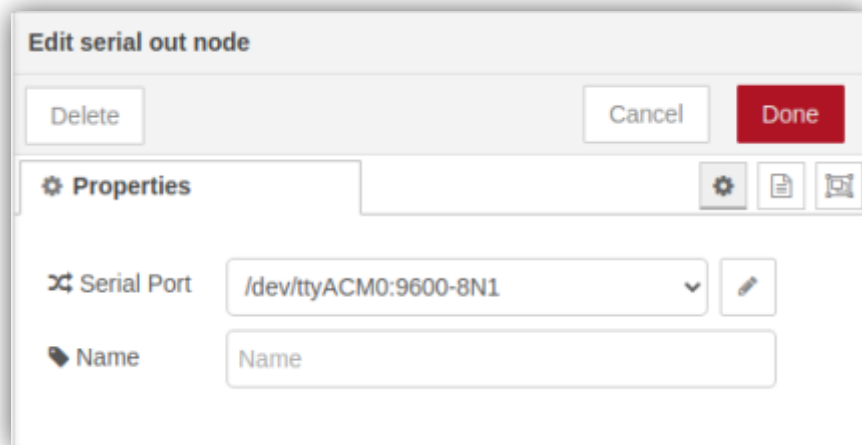


Figura N° 30: Nodo de comunicación serial

Fuente: Elaboración propia

En la figura N° 31 se definieron los parámetros de la comunicación serial, el valor del parámetro baud rate de origen debe ser igual al valor del destino. Los parámetros DTR, RTS, CTS y DSR se dejaron por defecto ya que solo son para alimentar dispositivos a través de los pines.

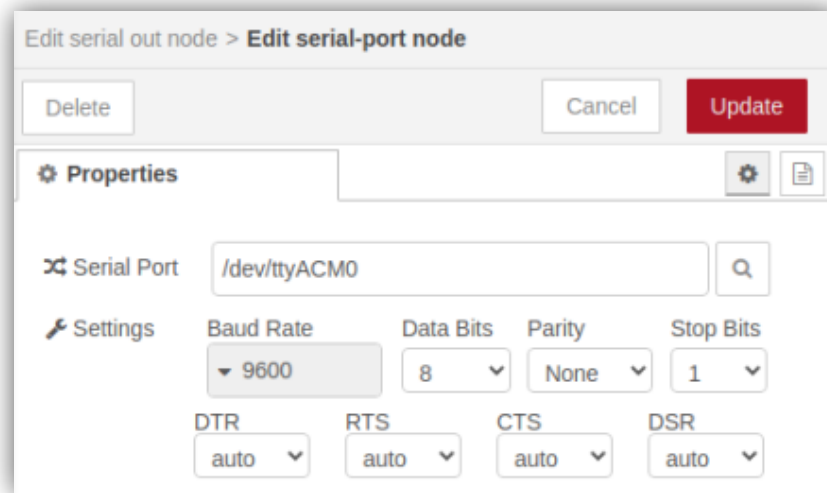


Figura N° 31: Configuración de la comunicación serial

Fuente: Elaboración propia

Para los flujos parciales, en este caso para el flujo del idioma francés, se agregó el nodo de comunicación serial para establecer la comunicación entre el Raspberry Pi 3B y el Arduino. Este flujo parcial es mostrado en la figura N° 32.

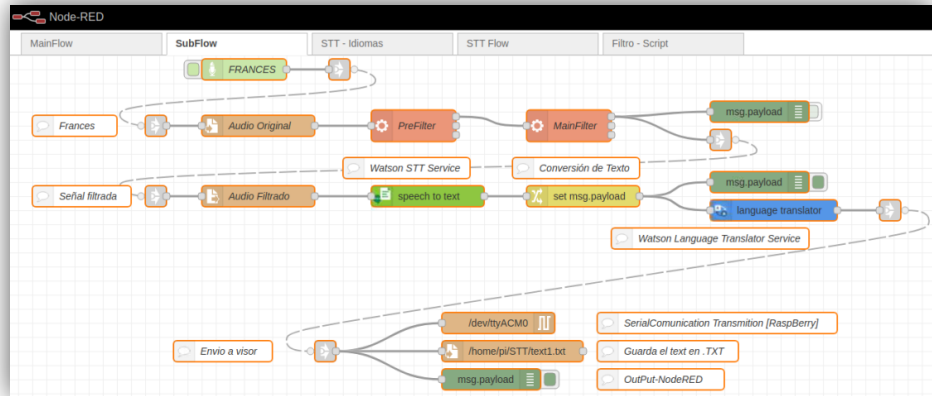


Figura N° 32: Flujo parcial del sistema para el tratamiento de la voz y la comunicación serial

Fuente: Elaboración propia

Respecto al flujo total, fue desarrollado como un conjunto de flujos parciales configurados para los diferentes idiomas. Estos flujos tuvieron estructuras similares al flujo mostrado en la figura anterior. Este flujo final se muestra en el anexo.

Finalmente, para la recepción del texto en el Arduino, se desarrolló la programación para agrupar cada carácter y mostrar el mensaje en la pantalla OLED. A continuación, en la figura N° 33 se observa el código.



```
RxArduino Arduino 1.8.15
Archivo Editar Programa Herramientas Ayuda

RxArduino
#include <Adafruit_SSD1306.h>
#include <Wire.h>
Adafruit_SSD1306 display;
String estado;

void setup() {

    Serial.begin(9600); // Inicializamos el puerto serie
    display.begin(SSD1306_SWITCHCAPVCC, 0x3C);
    display.clearDisplay ();
    display.display ();
}

void loop() {

    while (Serial.available())
    {
        char c = Serial.read(); //Lee el dato entrante y lo almacena en una variable tipo char
        estado += c;           //Crea una cadena tipo String con los datos entrates
    }
    if (estado.length() > 0) //Se verifica que la cadena tipo String tenga un largo mayor a cero
    {
        display.setTextColor(WHITE); //Color blanco para el texto
        display.setTextSize(1); // Tamaño del texto
        display.setCursor(0,20); // Se fija las coordenadas
        display.print(estado); // Se imprime el texto en la pantalla
        display.display();
        estado=""; // Limpia la variable para recibir nuevos datos
        display.clearDisplay ();
    }
}
```

Figura N° 33: Programación para proyectar el texto recibido

Fuente: Elaboración propia

Luego, el texto fue proyectado en la pantalla OLED, tal como se muestra en la figura N° 34.

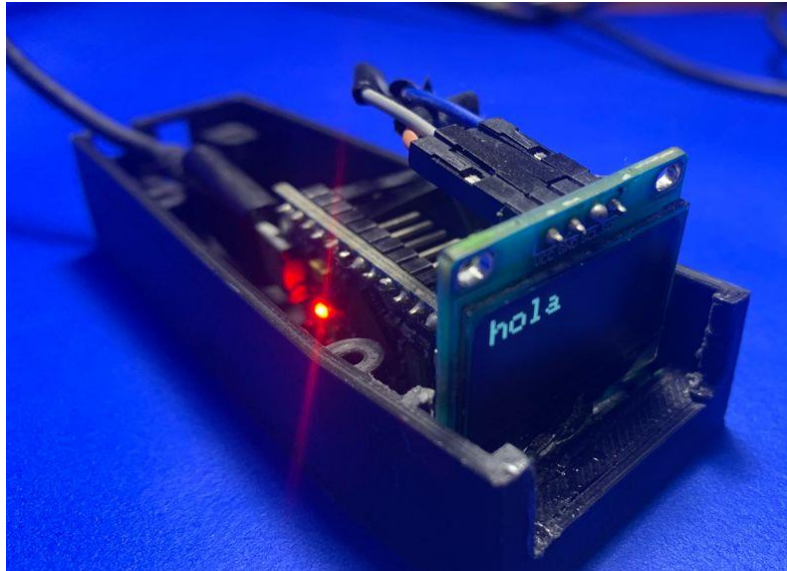


Figura N° 34: Texto proyectado en la pantalla OLED

Fuente: Elaboración propia

3.4. Implementación de sistema óptico

3.4.1 Diseño teórico de generación y tratamiento de imagen virtual

La pantalla OLED fue reflejada en un pequeño espejo donde los rayos incidieron en un lente plano-convexo formando una imagen virtual de la pantalla OLED más grande respecto al real, ya que la distancia del OLED respecto al lente fue menor que su distancia focal.

Cálculo:

La distancia focal del lente plano convexo: $f = 13 \text{ cm}$

Distancia del OLED respecto al lente plano convexo: $o = 7.65 \text{ cm}$

De la fórmula (1)

$$\frac{1}{7.65} + \frac{1}{i} = \frac{1}{13}$$

$$i = -18.59 \text{ cm}$$

De la fórmula (2)

$$M = \frac{-(-18.59 \text{ cm})}{7.65 \text{ cm}}$$

$M = 2.43$

Se tuvo una imagen virtual del OLED mayor respecto a la real (2.43 más grande).

3.4.2 Diseño e impresión 3D de diseño de armazón del visor

En las figuras N° 35 al N° 42, consecutivamente se representa el diseño de las partes del visor que posteriormente sirvieron para realizar la impresión. El diseño se realizó mediante el software Blender. En ellos, se aprecia las medidas que van acorde al tamaño del visor utilizado.

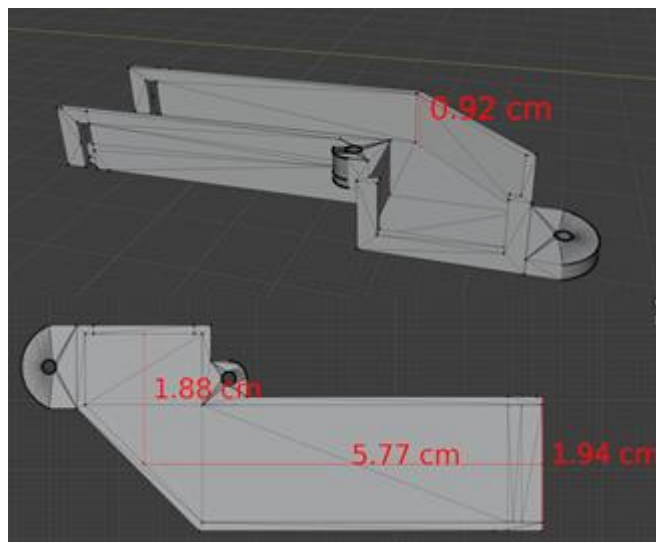


Figura N° 35: Vista lateral y superior del brazo superior delantero del reflector

Fuente: Elaboración propia

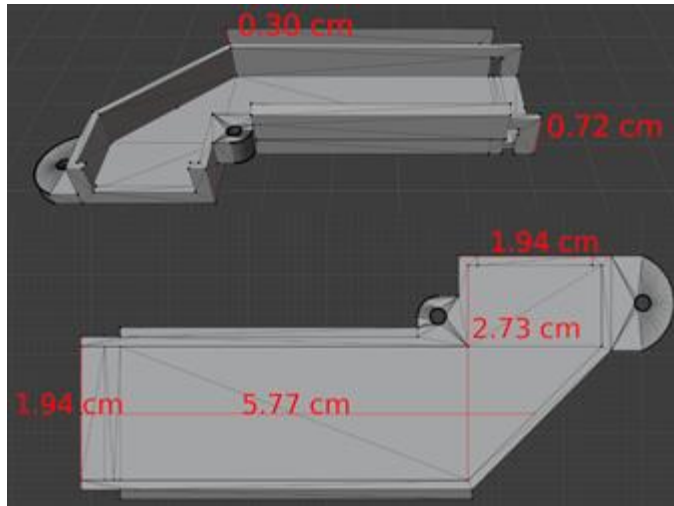


Figura N° 36: Vista lateral y superior del brazo delantero inferior del reflector

Fuente: Elaboración propia

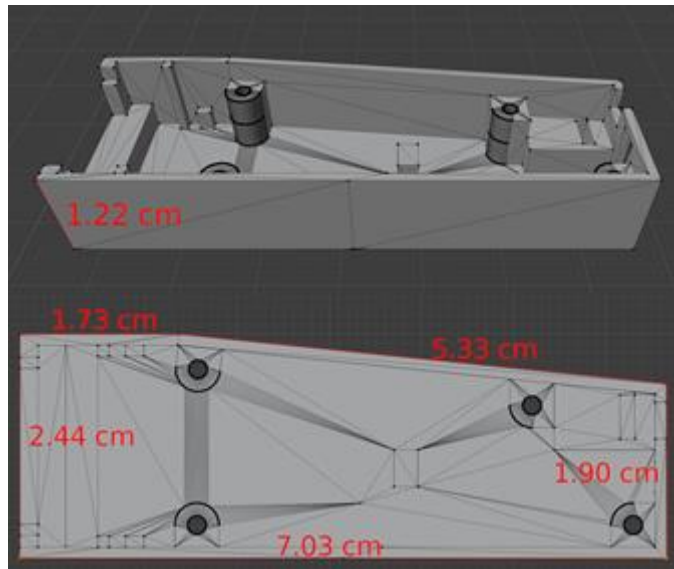


Figura N° 37: Vista lateral y superior del brazo de soporte superior del armazón

Fuente: Elaboración propia

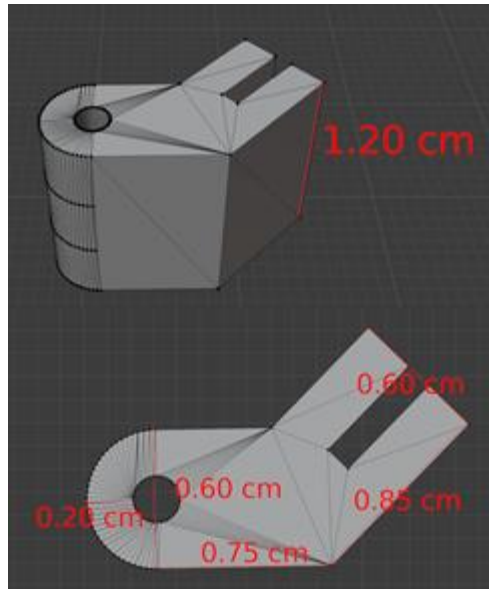


Figura N° 38: Vista lateral y superior del soporte del reflector

Fuente: Elaboración propia

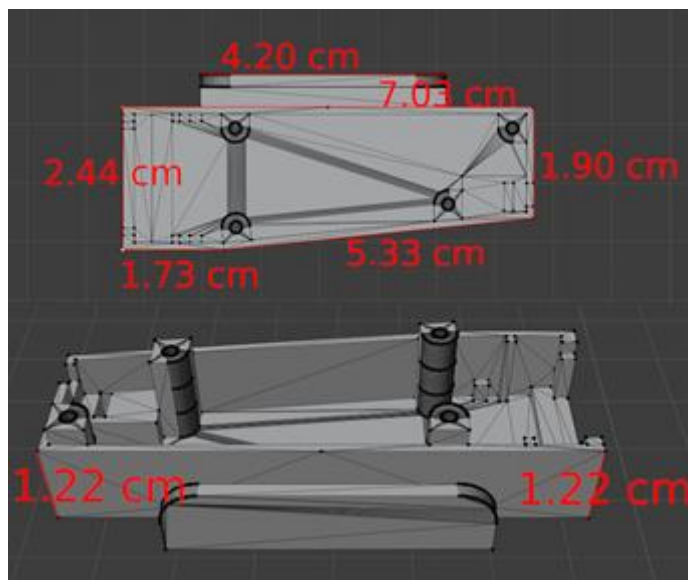


Figura N° 39: Vista lateral y superior del brazo de soporte inferior del armazón

Fuente: Elaboración propia

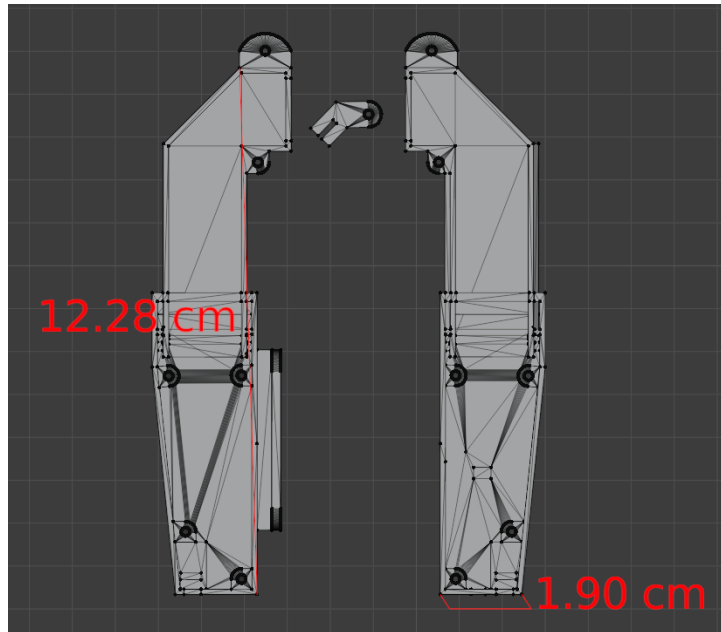


Figura N° 40: Vista superior de los componentes del armazón

Fuente: Elaboración propia



Figura N° 41: Vista diagonal desde la perspectiva interior

Fuente: Elaboración propia

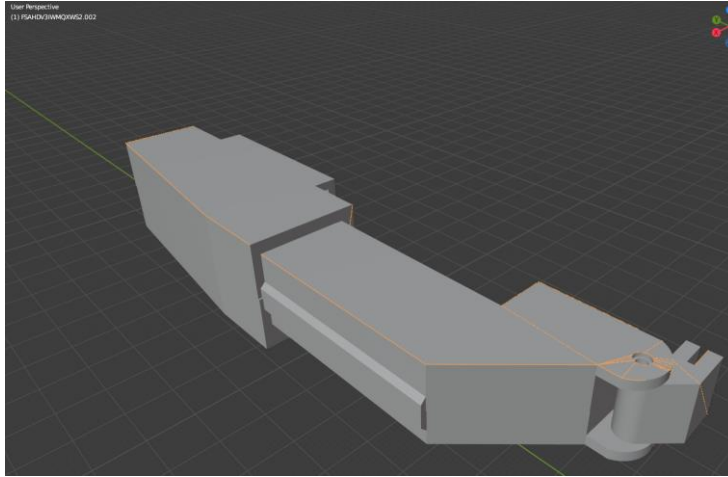


Figura N° 42: Vista diagonal desde la perspectiva exterior

Fuente: Elaboración propia

3.4.3 Construcción e integración de armazón con sistemas de lentes

El eje central de procesamiento del visor fue el Raspberry Pi 3B. Así como se aprecia en la figura N° 43, se conectó el Raspberry Pi 3B con dos dispositivos mediante los puertos USB. El Raspberry Pi 3B fue configurado con el sistema de codificación de Node-Red. Respecto a la conexión en sus puertos, con uno de ellos, se realizó la captura de las voces mediante un micrófono, mientras que con el otro puerto se contó con la conexión hacia el Arduino con el fin del envío del texto procesado.



Figura N° 43: Vista del Raspberry Pi 3B desde la perspectiva superior

Fuente: Elaboración propia

En la figura N° 44, se aprecia el micrófono el cuál realizó la captura de las señales de voz. Esta gráfica manifiesta la conexión directa del micrófono con el Raspberry Pi 3B.



Figura N° 44: Vista del micrófono y del Raspberry Pi 3B desde la perspectiva superior

Fuente: Elaboración propia

De acuerdo con las medidas obtenidas en el subtítulo anterior, se colocó la pantalla OLED, el espejo y el lente plano convexo en el armazón de la siguiente manera como se observa en la figura N° 45. Es este el bloque interior del caparazón del visor que tiene como interconector la conexión serial entre el Raspberry Pi 3B y el Arduino.



Figura N° 45: Vista superior del interior del visor

Fuente: Elaboración propia

En la figura N° 46, se aprecia la impresión del texto procesado y proyectado en una pantalla OLED dentro del armazón del visor. El texto proyectado fue mostrado línea a línea de acuerdo al mensaje que procesó. Cabe resaltar que para las últimas palabras que sobrepasaron la línea fueron cortadas para continuar la impresión de la palabra en la siguiente línea.

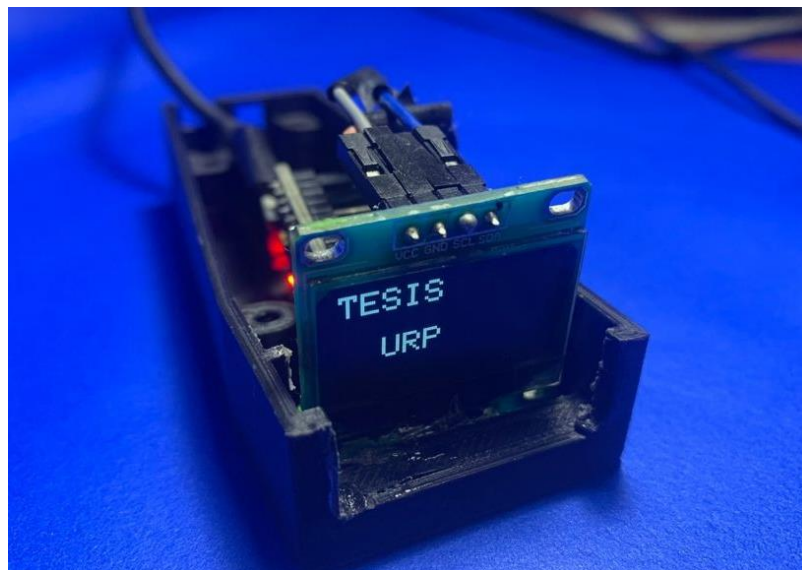


Figura N° 46: Vista de la pantalla OLED

Fuente: Elaboración propia

En la figura N° 47, se visualizó el mensaje en la pantalla. Este mensaje fue reflejado en un espejo que se situó de manera diagonal. Como se aprecia, el mensaje fue visualizado de forma invertida horizontalmente, pero se visualizó correctamente el mensaje por completo.



Figura N° 47: Vista del espejo utilizado

Fuente: Elaboración propia

En la figura N° 48, se aprecia el reflector principal por medio del cual se pudo ver el texto de la imagen virtual creada por el sistema de óptico. Este reflector es el que está frente a la línea de visión del usuario.

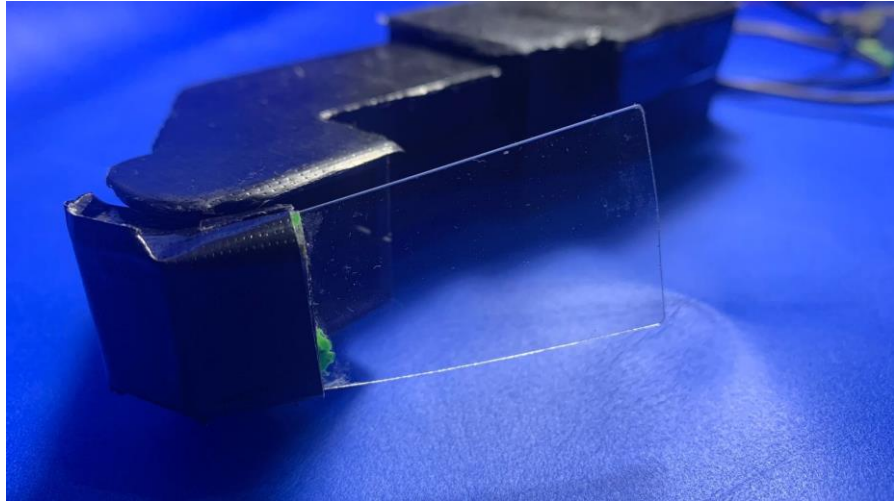


Figura N° 48: Vista del reflector

Fuente: Elaboración propia

En las figuras N° 50, 51 y 52 se aprecia el punto de vista del usuario respecto a la visualización del texto mediante el reflector, y este texto es mostrado en la realidad visual del usuario en diferentes ángulos.



Figura N° 49: Vista de texto en objeto en la realidad

Fuente: Elaboración propia



Figura N° 50: Vista de texto en objeto en la realidad

Fuente: Elaboración propia



Figura N° 51: Vista de texto en objeto en la realidad

Fuente: Elaboración propia

Luego de la integración de los diferentes bloques de procesamiento y los dispositivos utilizados se logró concretar la implementación del dispositivo unificado, así como se aprecia en la figura N°52.



Figura N° 52: Vista del visor de realidad aumentada

Fuente: Elaboración propia

CAPÍTULO IV: PRUEBAS Y RESULTADOS

En este capítulo se muestran las tres pruebas y sus respectivos resultados. En la primera prueba se consideró la detección de las palabras más comunes en los distintos idiomas. La segunda prueba constó en utilizar las mismas palabras con perturbaciones de fondo para poder observar la eficiencia del filtro. Posteriormente se realizó el resumen de detecciones de los 4 idiomas para obtener el porcentaje de detección de cada frase.

4.1. Detección de palabras más comunes según idioma

Se analizaron varias páginas en internet respecto a las frases más comunes y se eligió una de ellas, ya que está relacionado al apoyo para una persona que va a realizar un viaje.

Se usaron 6 frases para cada idioma. En la primera columna, se muestra el texto procesado en el Node-Red, y en la segunda columna, se muestra el texto recibido en la pantalla OLED. A continuación, en la tabla N° 1 se muestran los resultados obtenidos usando las frases en español, en este caso, los hablantes son los autores de la presente tesis debido al manejo del idioma hablado:

Tabla N° 1: Resultados usando el idioma español

Español	
Buenos días	
<pre>9/4/2021, 2:09:28 PM node: b7d0660375991e29 msg.payload : string[12] "buenos días "</pre>	
¿Cómo te llamas?	


<p>9/4/2021, 2:11:53 PM node: b7d0660375991e29 msg.payload : string[15] "cómo te llamas "</p>	
<p>Me llamo Luis</p>	
<p>9/4/2021, 2:13:42 PM node: b7d0660375991e29 msg.payload : string[9] "no lo es "</p>	
<p>¿Cómo estás?</p>	
<p>9/4/2021, 2:20:05 PM node: b7d0660375991e29 msg.payload : string[10] "qué estás "</p>	
<p>Gusto en conocerte</p>	
<p>9/4/2021, 2:22:58 PM node: b7d0660375991e29 msg.payload : string[19] "gusto en conocerte "</p>	
<p>Adiós</p>	



Fuente: Elaboración Propia

En la primera fila de cada frase se puso el texto procesado en el idioma que se habló, luego en la primera columna se muestra el texto traducido, y en la segunda columna el texto que se proyecta en la pantalla OLED. En la tabla N° 2 se muestran los resultados obtenidos usando las frases en inglés. Los hablantes son los autores de la presente tesis debido al manejo del idioma hablado

Tabla N° 2: Resultados usando el idioma inglés

Inglés	
Good morning	
<pre>9/4/2021, 2:29:33 PM node: 467fef2ec790214e msg.payload : string(13) "good morning "</pre>	
<pre>9/4/2021, 2:29:36 PM node: b7d0660375991e29 msg.payload : string(12) "Buenos días "</pre>	
What's your name	
<pre>9/4/2021, 2:33:26 PM node: 467fef2ec790214e msg.payload : string(17) "what's your name "</pre>	

<p>9/4/2021, 2:33:28 PM node: b7d0660375991e29 msg.payload : string[19] "¿Cuál es tu nombre "</p>	
<p>My name is Luis</p>	
<p>9/4/2021, 2:36:32 PM node: 467fef2ec790214e msg.payload : string[18] "my name is Louise "</p>	
<p>9/4/2021, 2:36:36 PM node: b7d0660375991e29 msg.payload : string[20] "mi nombre es Louise "</p>	
<p>How are you</p>	
<p>9/4/2021, 2:41:06 PM node: 467fef2ec790214e msg.payload : string[12] "how are you "</p>	
<p>9/4/2021, 2:41:08 PM node: b7d0660375991e29 msg.payload : string[9] "¿Cómo te "</p>	
<p>It is nice to meet you</p>	
<p>9/4/2021, 2:53:08 PM node: 467fef2ec790214e msg.payload : string[23] "it is nice to meet you "</p>	





Fuente: Elaboración Propia

En la tabla N° 3 se muestran los resultados obtenidos usando las frases en francés. El hablante en este caso ha sido una persona externa quien maneja el idioma francés de manera nativa.

Tabla N° 3: Resultados usando el idioma francés

Francés
<pre>Bonjour 9/7/2021, 9:30:22 PM node: 57e550a31a16b149 msg.payload : string[8] "bonjour "</pre>

<p>9/7/2021, 9:30:24 PM node: b7d0660375991e29 msg.payload : string[5] "hola "</p>	
<p>Comment vous appelez-vous?</p>	
<p>9/7/2021, 9:33:34 PM node: 57e550a31a16b149 msg.payload : string[26] "comment vous appelez vous "</p>	
<p>9/7/2021, 9:33:36 PM node: b7d0660375991e29 msg.payload : string[20] "cómo se llama usted "</p>	
<p>Je m'appelle Luis</p>	
<p>9/7/2021, 9:35:50 PM node: 57e550a31a16b149 msg.payload : string[19] "je m'appelle luisse "</p>	
<p>9/7/2021, 9:35:52 PM node: b7d0660375991e29 msg.payload : string[16] "Me llamo luisse. "</p>	
<p>Comment allez-vous?</p>	
<p>9/7/2021, 9:39:37 PM node: 57e550a31a16b149 msg.payload : string[19] "comment allez vous "</p>	

<p>9/7/2021, 9:39:39 PM node: b7d0660375991e29 msg.payload : string[14] " cómo va usted "</p>	
<p>Enchanté</p>	
<p>9/7/2021, 9:42:28 PM node: 57e550a31a16b149 msg.payload : string[9] "enchanté "</p>	
<p>9/7/2021, 9:42:30 PM node: b7d0660375991e29 msg.payload : string[10] " encantado "</p>	
<p>Adieu</p>	
<p>9/7/2021, 9:49:17 PM node: 57e550a31a16b149 msg.payload : string[6] " adieu "</p>	
<p>9/7/2021, 9:49:18 PM node: b7d0660375991e29 msg.payload : string[6] " adiós "</p>	

Fuente: Elaboración Propia

En la tabla N° 4 se muestran los resultados obtenidos usando las frases en portugués. Los hablantes son los autores de la presente tesis debido al manejo del idioma hablado.

Tabla N° 4: Resultados usando el idioma portugués

Portugués	
Bom dia	
<p>9/7/2021, 8:02:16 PM node: f11c7c2bce98ee13 msg.payload : string[8] "bom dia "</p>	
<p>9/7/2021, 8:02:18 PM node: b028b47c6e6abc15 msg.payload : string[12] "Buenos días "</p>	
Como se chama?	
<p>9/7/2021, 8:05:14 PM node: f11c7c2bce98ee13 msg.payload : string[14] "como se chama "</p>	
<p>9/7/2021, 8:05:17 PM node: b7d0660375991e29 msg.payload : string[14] "cómo se llama "</p>	
Chamo-me Luis	
<p>9/7/2021, 8:23:02 PM node: f11c7c2bce98ee13 msg.payload : string[17] "eu me chamo luis "</p>	
<p>9/7/2021, 8:23:05 PM node: b7d0660375991e29 msg.payload : string[14] "Me llamo luis "</p>	
Você como vai?	
<p>9/7/2021, 9:20:17 PM node: f11c7c2bce98ee13 msg.payload : string[14] "você como vai "</p>	

<p>9/7/2021, 9:20:20 PM node: b7d0660375991e29 msg.payload : string[20] "usted como usted va "</p>	
<p>Muito prazer</p>	
<p>9/7/2021, 9:23:26 PM node: f11c7c2bce98ee13 msg.payload : string[13] "muito prazer "</p>	
<p>9/7/2021, 9:23:28 PM node: b7d0660375991e29 msg.payload : string[10] "muy grato "</p>	
<p>Adeus</p>	
<p>9/7/2021, 9:25:28 PM node: f11c7c2bce98ee13 msg.payload : string[7] "a deus "</p>	
<p>9/7/2021, 9:25:31 PM node: b7d0660375991e29 msg.payload : string[8] "el dios "</p>	

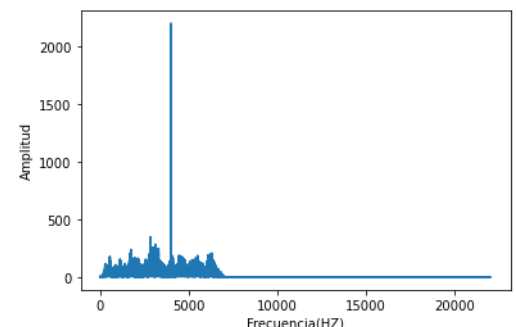
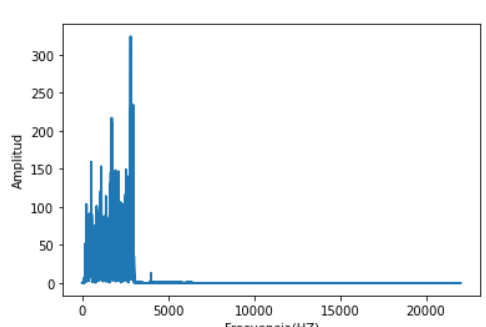


Fuente: Elaboración Propia

4.2. Detección de palabras con perturbación

Para esta prueba se grabó la voz usando las mismas frases de la sección anterior, pero con una perturbación intencional de 4KHz fuera del rango de la frecuencia de la voz. Del mismo modo, se graficó cada espectro de frecuencia de voz grabado para poder observar la señal original y la filtrada.

En la primera fila se pudo observar que, en la imagen de la derecha, la perturbación de 4KHz ha sido suprimida por el filtro. De la misma forma, se muestra el texto procesado en el Node-Red y el texto recibido en la pantalla OLED. A continuación, en la tabla N° 5 se muestran los resultados obtenidos usando las frases en español con perturbaciones.

Tabla N° 5: Resultados usando el idioma español con una perturbación

Español	
Buenos días	
	
<pre>9/10/2021, 8:41:42 PM node: b028b47c6e6abc15 msg.payload : string[11] "buenos días"</pre>	
¿Cómo te llamas?	
<pre>9/10/2021, 8:43:11 PM node: b028b47c6e6abc15 msg.payload : string[14] "cómo te llamas"</pre>	
Me llamo Luis	

<p>9/10/2021, 8:30:56 PM node: b028b47c6e6abc15 msg.payload : string[8] "no lo es"</p>	
<p>¿Cómo estás?</p>	
<p>9/10/2021, 8:44:19 PM node: b028b47c6e6abc15 msg.payload : string[9] "qué estás"</p>	
<p>Gusto en conocer te</p>	
<p>9/10/2021, 8:35:03 PM node: b028b47c6e6abc15 msg.payload : string[18] "gusto en conocer te"</p>	
<p>Adiós</p>	
<p>9/10/2021, 8:39:24 PM node: b028b47c6e6abc15 msg.payload : string[5] "adiós"</p>	

Fuente: Elaboración Propia

En la tabla N° 6 se muestran los resultados obtenidos usando las frases en inglés con una perturbación intencional de 4KHz.

Tabla N° 6: Resultados usando el idioma inglés con una perturbación

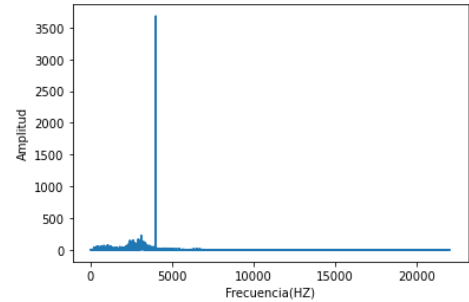
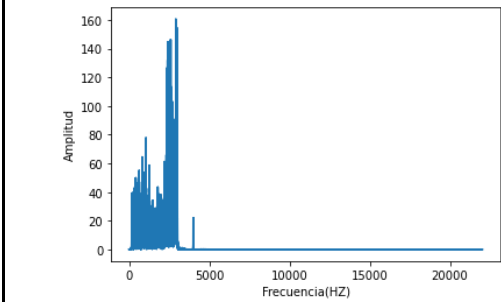


Inglés	
Good morning	
 <p>Amplitud</p> <p>Frecuencia(HZ)</p>	 <p>Amplitud</p> <p>Frecuencia(HZ)</p>
<pre>9/10/2021, 8:11:20 PM node: b028b47c6e6abc15 msg.payload : string[12] "good morning"</pre>	
What's your name	
<pre>9/10/2021, 8:49:42 PM node: 467fef2ec790214e msg.payload : string[16] "what's your name"</pre>	
My name is Luis	

<p>9/10/2021, 8:53:05 PM node: 467fef2ec790214e msg.payload : string[17] "my name is Louise"</p>	
<p>How are you</p>	
<p>9/10/2021, 8:54:29 PM node: 467fef2ec790214e msg.payload : string[11] "how are you"</p>	
<p>It is nice to meet you</p>	
<p>9/10/2021, 8:55:24 PM node: 467fef2ec790214e msg.payload : string[22] "it is nice to meet you"</p>	
<p>Bye Bye</p>	
<p>9/10/2021, 8:56:45 PM node: 467fef2ec790214e msg.payload : string[7] "bye bye"</p>	

Fuente: Elaboración Propia

En la tabla N° 7 se muestran los resultados obtenidos usando las frases en francés con una perturbación intencional de 4KHz.

Tabla N° 7: Resultados usando el idioma francés con una perturbación

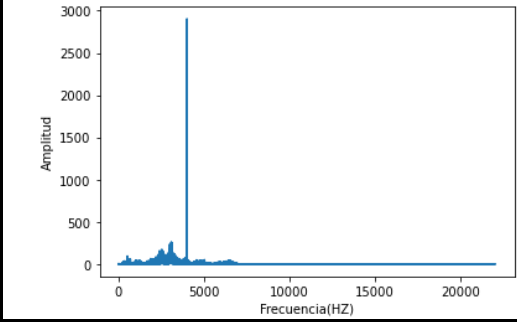
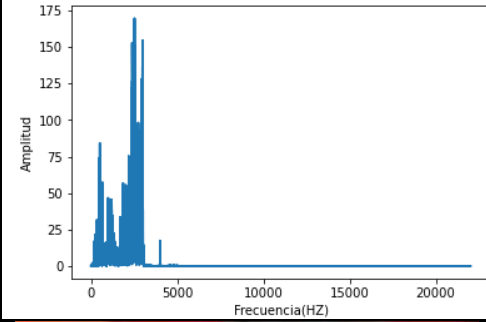


Francés	
Bonjour	
	
<pre>9/10/2021, 9:00:18 PM node: 57e550a31a16b149 msg.payload : string[7] "bonjour"</pre>	
Comment vous appelez-vous?	
<pre>9/10/2021, 9:03:55 PM node: 57e550a31a16b149 msg.payload : string[25] "comment vous appelez vous"</pre>	
Je m'appelle Luis	

<p>9/10/2021, 9:10:10 PM node: 57e550a31a16b149 msg.payload : string[18] "je m'appelle luise"</p>	
<p>Comment allez-vous?</p>	
<p>9/10/2021, 9:12:02 PM node: 57e550a31a16b149 msg.payload : string[18] "comment allez vous"</p>	
<p>Enchanté</p>	
<p>9/10/2021, 9:13:58 PM node: 57e550a31a16b149 msg.payload : string[8] "enchanté"</p>	
<p>Adieu</p>	
<p>9/10/2021, 9:15:18 PM node: 57e550a31a16b149 msg.payload : string[5] "adieu"</p>	

Fuente: Elaboración Propia

En la tabla N° 8 se muestran los resultados obtenidos usando las frases en portugués con una perturbación intencional de 4KHz.

Tabla N° 8: Resultados usando el idioma portugués con una perturbación

Portugués	
Bom dia	
	
<pre>9/10/2021, 9:23:02 PM node: f11c7c2bce98ee13 msg.payload : string[7] "bom dia"</pre>	
Como se chama?	
<pre>9/10/2021, 9:25:16 PM node: f11c7c2bce98ee13 msg.payload : string[13] "como se chama"</pre>	
Eu me chamo Luis	

<p>9/10/2021, 9:27:02 PM node: f11c7c2bce98ee13 msg.payload : string[16] "eu me chamo luis"</p>	
<p>Você como vai?</p>	
<p>9/10/2021, 9:29:28 PM node: f11c7c2bce98ee13 msg.payload : string[13] "voce como vai"</p>	
<p>Muito prazer</p>	
<p>9/10/2021, 9:32:01 PM node: f11c7c2bce98ee13 msg.payload : string[12] "muito prazer"</p>	
<p>Adeus</p>	
<p>9/10/2021, 9:33:08 PM node: f11c7c2bce98ee13 msg.payload : string[6] "a deus"</p>	

Fuente: Elaboración Propia

4.3. Resumen de detecciones

Para la presente sección, se ha evaluado el procesamiento de las voces percibidas respecto al resultado mostrado al usuario mediante el reflector. Se tomó en consideración diversos factores relacionados al texto que fue procesado, con el fin de evaluar las diferentes características de las frases definidas en la etapa de pruebas de detecciones en la sección anterior.

El objetivo de cuantificar la evaluación del resultado del visor fue obtener un porcentaje de la detección total. En este porcentaje se consideró la cantidad de caracteres y la cantidad de palabras que contiene cada frase definida asimismo como cada frase resultante; es decir, las frases que han sido procesadas por el sistema propuesto. Además, se considera un valor binario que hace referencia al entendimiento del mensaje, el cual es un criterio de nosotros los autores de esta tesis respecto a la afirmación y la negación de entendimiento del mensaje de la frase. Se definieron las siguientes fórmulas para calcular los porcentajes de detección para el idioma español:

$$p1 (\%) = \frac{p2+p3}{2} \dots (3)$$

$$p2 (\%) = \left(100 - \left|\frac{n1-n11}{n1}\right| \times 100\right) \times e1 \dots (4)$$

$$p3 (\%) = \left(100 - \left|\frac{n2-n22}{n2}\right| \times 100\right) \times e1 \dots (5)$$

Donde:

p1 = Porcentaje de detección total

p2 = Porcentaje de detección de caracteres

p3 = Porcentaje de detección de palabras

n1 = Número de caracteres de las frases definidas

n2 = Número de palabras de las frases definidas

n11 = Número de caracteres de las frases procesadas

n22 = Número de palabras de las frases procesadas

e1 = Entendimiento del mensaje

En la tabla N° 9 se aplicó la fórmula (4) para el número de caracteres de las frases establecidas y las frases procesadas. Posteriormente, se aplicó la fórmula (5) para el número de palabras de las frases establecidas y las frases procesadas. Luego de haber calculado ambos porcentajes de detección se obtuvo el promedio mediante la fórmula (3).

Tabla N° 9: Porcentaje de detección en idioma español

Frases definidas en español			Frases procesadas				p2 (%)	p3(%)	p1(%)
Frase establecida	n1	n2	Frase resultante	n11	n22	e1			
Buenos días	11	2	buenos días	11	2	1	100	100	100
¿Cómo te llamas?	16	3	como te llamas	14	3	1	87.5	100	93.75
Me llamo Luis	13	3	no lo es	8	3	0	0	0	0
¿Cómo estás?	12	2	que estas	9	2	0	0	0	0
Gusto en conocerte	18	3	gusto en conocerte	18	3	1	100	100	100
Adiós	5	1	adios	5	1	1	100	100	100

Fuente: Elaboración Propia

Para calcular el porcentaje de detección de los 3 idiomas restantes solo se utilizó la cantidad de palabras y no la de caracteres ya que el servicio de traducción puede interpretar y mostrar más palabras similares a la frase esperada. Por lo tanto, la cantidad de caracteres no influye en la comparación.

Se definió la siguiente fórmula para calcular el porcentaje de detección para idiomas diferentes al español, es decir, para las frases que hayan sido procesadas por sistemas de traducción:

$$p (\%) = \left(100 - \left| \frac{n-n'}{n} \right| \times 100 \right) \times e \quad \dots (6)$$

Donde:

p = Porcentaje de detección total

n = Número de palabras de las frases definidas

n' = Número de palabras de las frases procesadas

e = Entendimiento del mensaje

En la tabla N° 10 se aplicó la formula (6) para el número de palabras de las frases esperadas y las frases resultantes, con ello se obtuvo el porcentaje de detección.

Tabla N° 10: Porcentaje de detección en idioma inglés

Frases definidas en inglés	Frases esperadas		Frases traducidas			p(%)
	Frase en español	n	Frase resultante	n'	e	
Good morning	Buenos días	2	Buenos días	2	1	100
What's your name	¿Cuál es tu nombre?	4	Cual es tu nombre	4	1	100
My name is Luis	Mi nombre es Luis	4	mi nombre es Louise	4	1	100
How are you	¿Cómo estás?	2	Como te	2	0	0
Nice to meet you	Es agradable el conocerte.	4	es agradable encontrarse con usted	5	1	75
Goodbye	Adiós	1	adios	1	1	100

Fuente: Elaboración Propia

En la tabla N° 11 el porcentaje de detección en el idioma francés fue del 100% en todas las frases.

Tabla N° 11: Porcentaje de detección en idioma francés

Frases definidas en francés	Frases esperadas		Frases traducidas			p(%)
	Frase en español	n	Frase resultante	n'	e	
Bonjour	Hola	1	hola	1	1	100
Comment vous appelez-vous?	¿Como se llama usted?	4	como se llama usted	4	1	100
Je m'appelle Luis	Me llamo Luis	3	Me llamo luise	3	1	100
Comment allez-vous?	¿Cómo está usted?	3	como va usted	3	1	100
Enchanté	Encantado	1	encantado	1	1	100
Au revoir	Adiós	1	adios	1	1	100

Fuente: Elaboración Propia

En la tabla N° 12 se muestra el porcentaje de detección de las frases en portugués.

Tabla N° 12: Porcentaje de detección en idioma portugués

Frases definidas en portugués	Frases esperadas		Frases traducidas			p(%)
	Frase en español	n	Frase resultante	n'	e	
Bom dia	Buenos días	2	Buenos días	2	1	100
Como se chama?	¿Cómo te llamas?	3	como se llama	3	1	100
Chamo-me Luis	Me llamo Luis	3	Me llamo Luis	3	1	100
Como vai?	¿Cómo estas?	2	Usted como va usted	4	0	0
Muito prazer	Mucho gusto	2	muy grato	2	1	100
Adeus	Adiós	1	el dios	2	0	0

Fuente: Elaboración Propia

De la misma manera se aplicaron las fórmulas (4), (5) y (3) respectivamente para el idioma español con una perturbación intencional de 4 KHz. En la tabla N° 13 se muestra el porcentaje de detección en el idioma mencionado.

Tabla N° 13: Porcentaje de detección en idioma español con una perturbación

Frases definidas en español			Frases procesadas				p2 (%)	p3(%)	p1(%)
Frase establecida	n1	n2	Frase resultante	n11	n22	e1			
Buenos días	11	2	buenos dias	11	2	1	100	100	100
¿Cómo te llamas?	16	3	como te llamas	14	3	1	87.5	100	93.75
Me llamo Luis	13	3	no lo es	8	3	0	0	0	0
¿Cómo estás?	12	2	que estas	9	2	0	0	0	0
Gusto en conocerte	18	3	gusto en conocerte	18	3	1	100	100	100
Adiós	5	1	adios	5	1	1	100	100	100

Fuente: Elaboración Propia

De la misma manera se aplicó la formula (6) para los 3 idiomas restantes. En la tabla N° 14 se muestra el porcentaje de detección en idioma inglés con una perturbación intencional de 4 KHz.

Tabla N° 14: Porcentaje de detección en idioma inglés con una perturbación

Frases definidas en inglés	Frases esperadas		Frases traducidas			p(%)
	Frase en español	n	Frase resultante	n'	e	
Good morning	Buenos días	2	Buenos días	2	1	100
What's your name	¿Cuál es tu nombre?	4	Cual es tu nombre	4	1	100
My name is Luis	Mi nombre es Luis	4	mi nombre es Louise	4	1	100
How are you	¿Cómo estás?	2	Como te	2	0	0
Nice to meet you	Es agradable el conocerte.	4	es agradable encontrarse con usted	5	1	75
Goodbye	Adiós	1	adios	1	1	100

Fuente: Elaboración Propia

En la tabla N° 15 se muestra el porcentaje de detección en idioma francés con una perturbación intencional de 4KHz.

Tabla N° 15: Porcentaje de detección en idioma francés con una perturbación

Frases definidas en francés	Frases esperadas		Frases traducidas			p(%)
	Frase en español	n	Frase resultante	n'	e	
Bonjour	Hola	1	hola	1	1	100
Comment vous appelez-vous?	¿Como se llama usted?	4	como se llama usted	4	1	100
Je m'appelle Luis	Me llamo Luis	3	Me llamo luisse	3	1	100
Comment allez-vous?	¿Cómo está usted?	3	como va usted	3	1	100
Enchanté	Encantado	1	encantado	1	1	100
Au revoir	Adiós	1	adios	1	1	100

Fuente: Elaboración Propia

En la tabla N° 16 se muestra el porcentaje de detección en idioma portugués con una perturbación intencional de 4 KHz.

Tabla N° 16: Porcentaje de detección en idioma portugués con una perturbación

Frases definidas en portugués	Frases esperadas		Frases traducidas			p(%)
	Frase en español	n	Frase resultante	n'	e	
Bom dia	Buenos días	2	Buenos días	2	1	100
Como se chama?	¿Cómo te llamas?	3	como se llama	3	1	100
Chamo-me Luis	Me llamo Luis	3	Me llamo Luis	3	1	100
Como vai?	¿Cómo estas?	2	Usted como va usted	4	0	0
Muito prazer	Mucho gusto	2	muy grato	2	1	100
Adeus	Adiós	1	el dios	2	0	0

Fuente: Elaboración Propia

En las tabla N° 17 se calculó el promedio del porcentaje de detección de cada idioma, y en la tabla N° 18 se realizó el mismo cálculo de cada idioma con la perturbación intencional de 4 KHz.

Tabla N° 17: Promedio del porcentaje de detección de cada idioma

Frases procesadas	Promedio del porcentaje de detección (%)
Frases en español	65.63
Frases en inglés	79.17
Frases en francés	100
Frases en portugués	66.67

Fuente: Elaboración Propia

Tabla N° 18: Promedio del porcentaje de detección de cada idioma con una perturbación intencional de 4 KHZ

Frases procesadas	Promedio del porcentaje de detección (%)
Frases en español	65.63
Frases en inglés	79.17
Frases en francés	100
Frases en portugués	66.67

Fuente: Elaboración Propia

Los resultados sin y con perturbación fueron iguales, por lo cual el proceso de filtrado ha sido efectivo. Todos los promedios superaron el 60% en la detección de las frases.

Se obtuvieron porcentajes de detección de valor cero en frases en español, inglés y portugués debido a que la respuesta del servicio de conversión de voz a texto requiere una alta precisión en la pronunciación y entonación. Se pronunciaron las frases predefinidas en una entonación y rapidez promedio con respecto a una conversación coloquial cotidiana. Los servicios de nube tanto el conversor de voz a texto como el servicio de traducción tienen un foco de procesamiento centralizado en sistemas de inteligencia artificial, esto conlleva a adaptarse a los mecanismos y las mejores prácticas del uso de esta tecnología. Un punto importante de trabajar con estos sistemas es que, a más datos de entrada, se obtienen mejores resultados respecto a la predicción resultante como también a su autoaprendizaje. Es decir, a mayores palabras contenidas en una oración, el resultado de la conversión o traducción será más preciso y por ende se tendrá un mayor entendimiento del mensaje.

Por otro lado, para el idioma francés, se logró obtener el 100% de detección en todas las frases debido a que la pronunciación a comparación de los otros idiomas fue óptima por parte del hablante nativo, lo que llevo a una mejor traducción.

4.4. Presupuesto

En la tabla N° 19, se detalla el costo de los materiales que se utilizaron para realizar el proyecto de tesis.

Tabla N° 19: Costo de materiales

Dispositivos	Precio (S/.)
Raspberry Pi 3B	150
Pantalla táctil para Raspberry Pi 3B	90
Pantalla OLED	25
Arduino	32

Lunas	5
Impresión de armazón	30
Micrófono USB	25
Adiciones	50
Maqueta externa	20
TOTAL	427

Fuente: Elaboración Propia

CONCLUSIONES

1. Se desarrollaron los mecanismos de reconocimiento de voz empezando por el filtro digital pasa banda elíptico de orden 8 con frecuencias de corte 50 - 3000 Hz. Para ello, se realizó el diseño teórico del filtro en el Matlab tal como se representó en la figura N° 9; posteriormente, dicho filtro fue implementado utilizando el lenguaje de programación Python, y luego se realizaron las pruebas en voces con perturbación intencional con el fin de llevarlas a un proceso de filtrado, tal como se mostró en la figura N° 11. De la misma forma, para la conversión de voz a texto se utilizaron los servicios de inteligencia artificial en la nube de IBM Watson Speech to Text, y se pudo observar la correcta conversión de la misma tal como se observó en la figura N° 24. También, se podrá visualizar el proceso de conversión de la voz y traducción mediante el servicio IBM Watson Language Translator para los idiomas inglés, francés y portugués en las tablas N° 2, N°3 y N°4 respectivamente.
2. Se desarrolló la programación a nivel de código para poder mostrar el texto recibido del Raspberry Pi 3B hacia el Arduino conectado con la pantalla OLED tal como se observó en la figura N° 33. En la figura N° 34 se pudo visualizar el texto en la pantalla OLED.
3. Se desarrolló el cálculo teórico de la imagen virtual para la implementación del sistema óptico como se pudo observar en el capítulo 3, sección 3.4.1, se diseñó e integró el armazón del visor con el sistema de lentes tal como se describió en el capítulo 3, subtítulo 3.4.3. Asimismo, se pudo leer la conversión de voz a texto en el visor de realidad aumentada tal como se observó en la figura N° 49. Es así que, los resultados entregaron un promedio por encima del 60% en la detección de las frases, por lo cual el acierto de la conversión de voz a texto en diferentes idiomas se considera buena para poder establecer una conversación con la persona con discapacidad auditiva.

RECOMENDACIONES

1. Se recomienda hablar a un ritmo lento y pausado para que se tenga un mejor procesamiento en el servicio de Speech to Text de reconocimiento de voz.
2. La conectividad para los servicios de nube tiene que realizarse mediante una conexión a internet, es decir que el Raspberry Pi 3B debe contar con una conexión a internet inalámbrica preferentemente.
3. Se tiene que usar la entonación en caso sea una pregunta para la interpretación del servicio de conversión de voz según el idioma seleccionado.

REFERENCIAS BIBLIOGRÁFICAS

- Aguado, D., Andersen, T., Avetisyan, A., Budnik, J., Criveti, M., Doroiman, A., ... & Szumczyk, S. (2016). *A practical approach to cloud IaaS with IBM SoftLayer: Presentations guide*. IBM Redbooks.
- Artero, Ó. T. (2013). *ARDUINO. Curso práctico de formación*. RC Libros.
- Bano, S., Jithendra, P., Niharika, G. L., & Sikhi, Y. (2020, November). *Speech to Text Translation enabling Multilingualism*. In 2020 IEEE International Conference for Innovation in Technology (INOCON) (pp. 1-4). IEEE.
- Cogollos Borrás, S. (2016). *FUNDAMENTOS DE LA TEORÍA DE FILTROS. Colección Manual de referencia*.
- Correa, A. G. D., De Assis, G. A., do Nascimento, M., Ficheman, I., & de Deus Lopes, R. (2007, September). *Genvirtual: An augmented reality musical game for cognitive and motor rehabilitation*. In 2007 *Virtual Rehabilitation* (pp. 1-6). IEEE.
- Dabran, I., Avny, T., Singher, E., & Danan, H. B. (2017, November). *Augmented reality speech recognition for the hearing impaired*. In 2017 *IEEE International Conference on Microwaves, Antennas, Communications and Electronic Systems (COMCAS)* (pp. 1-4). IEEE.
- Epson. (2019) National Theatre uses Moverio to provide captioning for hearing impaired. Recuperado en junio 25, 2021 de <https://www.epson.co.uk/insights/casestudy/national-theatre-uses-moverio-to-provide-captioning-for-hearing-impaired>
- Evans, B. W. (2007). *Arduino programming notebook* (Vol. 1). Brian Evans.
- Federación de Asociaciones de Implantados Cocleares de España. (2015). *¿Qué es un ¿Implante Coclear?* Recuperado en mayo 25, 2021, de http://implantecoclear.org/index.php?option=com_content&view=article&id=76&Itemid=82
- García, J. M. S. (2016). *Realidad aumentada: tecnología para la formación*. Pixel-Bit, *Revista de Medios y Educación*, (49), 241-243.
- IBM. (2020). *Speech to Text*. Recuperado en mayo 29, 2021 de <https://cloud.ibm.com/docs/speech-to-text?topic=speech-to-text-about>
- Mahamud, M. S., & Zishan, M. S. R. (2017, September). *Watch IT: An assistive device for deaf and hearing impaired*. In 2017 4th International Conference on Advances in Electrical Engineering (ICAEE) (pp. 556-560). IEEE.
- Malacara, D. (2015). *Óptica básica*. Fondo de cultura económica.

- Mirzaei, M., Kan, P., & Kaufmann, H. (2020). EarVR: Using ear haptics in virtual reality for deaf and Hard-of-Hearing people. *IEEE transactions on visualization and computer graphics*, 26(5), 2084-2093.
- Nielsen, L. B. (1989, May). A computer controlled digital master hearing aid. In *IEEE International Symposium on Circuits and Systems*, (pp. 1291-1294). IEEE.
- Overton, J. (2018). *Artificial Intelligence: The Simplest Way*. O'Reilly Media.
- Rodriguez, E. De cero a maker: todo lo necesario para empezar con Raspberry Pi. Recuperado en mayo 27, 2021, de <https://www.xataka.com/makers/cero-maker-todo-necesario-para-empezar-raspberry-pi>
- Santiago, F., Singh, P., & Sri, L. (2017). *Building Cognitive Applications with IBM Watson Services: Volume 6 Speech to Text and Text to Speech*. IBM Redbooks.
- Tapu, R., Mocanu, B., & Zaharia, T. (2019). DEEP-HEAR: A multimodal subtitle positioning system dedicated to deaf and hearing-impaired people. *IEEE Access*, 7, 88150-88162.
- Techedge. (2020) *Fundamentos de node red*. Recuperado en mayo 29, 2021 de <https://www.techedgegroup.com/es/blog/fundamentos-node-red>
- Vásquez, A. C., Quispe, J. P., & Huayna, A. M. (2009). Procesamiento de lenguaje natural. *Revista de investigación de Sistemas e Informática*, 6(2), 45-54.
- Viikki, I., Kiss, I., & Tian, J. (2001, May). Speaker-and language-independent speech recognition in mobile communication systems. In *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 01CH37221)* (Vol. 1, pp. 5-8). IEEE.
- Virkkunen, A. (2018). *Automatic speech recognition for the hearing impaired in an augmented reality application*. Aalto University

ANEXO

