

**ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL**

**Facultad de Ingeniería en Electricidad y Computación**

Detección de muletillas en exposiciones orales mediante la conversión  
de voz a texto

**PROYECTO INTEGRADOR**

Previo la obtención del Título de:

**Ingeniera en Ciencias de la Computación**

Presentado por:

Nicole Katherine Asqui Manzaba

Hayleen Ivilin Carrillo Carrillo

GUAYAQUIL - ECUADOR

Año: 2022

## **DEDICATORIA**

A quienes me han enseñado a través de consejos, momentos y canciones, que todo es posible. Que, sin soltar mi mano, me ayudaron a ser fuerte y confiar en mí hasta el último momento, brindándome el apoyo que necesitaba para lograr resultados y alcanzar metas.

En especial a mis padres, por ser mi motor de lucha y esperanza constante. A mi hermana por siempre motivarme a continuar hacia mis objetivos. A mis hermanas de vida, Josselyne e Ivanny, por ser esa guía y estabilidad en cada momento de mi vida. Y a cada uno de mis gatos, por brindarme paz, felicidad y ronroneos en mis horas de angustia.

**Hayleen Carrillo**

## **DEDICATORIA**

Dedico este proyecto de Tesis a Dios que ha estado conmigo todo el tiempo y me ha dado las fuerzas para poder cumplir con mis metas. A mi papá, por ser amigo, guiarme y brindarme su apoyo incondicional en todo momento. A mi mamá por animarme a continuar a perseguir mis sueños. A los amigos que hice durante este crecimiento de los cuales me sentiré eternamente agradecida de tenerlos en mi vida. A mi gato Kira que fue la mejor compañía que pude tener en mis momentos difíciles, y que desde el cielo me da las fuerzas para continuar. Y por último, pero no menos importante, dedico esta tesis a mi yo del pasado, lo hiciste, siéntete orgullosa de ello. Ingeniera.

**Nicole Asqui**

## **AGRADECIMIENTOS**

Mi más sincero agradecimiento a quienes me enseñaron a que si hay algo por lo que luchar, se debe batallar y ganar. A quienes me ayudaron a evitar crisis nerviosas, a confiar en mí misma, y a resistir un poco más. A mis amigos, con quien coincidí por mi suerte o casualidad, pero que hoy representan felicidad y motivación en mi camino. Sobre todo, a Maruk y Alejandra, que siempre me hicieron sentir como en familia, aunque me encontrara a kilómetros de la mía. A mi compañera Nicole que, a pesar de las dificultades, logramos culminar este proyecto y afianzar nuestra amistad. De manera especial, a mis padres y hermanas que, frente a diversos inconvenientes y muchas noches de desvelo, me ayudaron a continuar construyendo el camino hacia mis objetivos.

**Hayleen Carrillo**

## **AGRADECIMIENTOS**

Quiero agradecer a todas las personas que son parte de mi proceso de vida, en la universidad y fuera de ella. A toda mi familia y amigos que me dieron sus palabras de aliento y me motivaron a convertirme en la persona que soy ahora. A todos los llevo en mi corazón. Y quiero brindar un agradecimiento especial a mi amiga Hayleen por haber sido mi soporte, sin ella esto no hubiera sido posible.

**Nicole Asqui**

## DECLARACIÓN EXPRESA

“Los derechos de titularidad y explotación, nos corresponde conforme al reglamento de propiedad intelectual de la institución; *Nicole Katherine Asqui Manzaba* y *Hayleen Ivilin Carrillo Carrillo* damos nuestro consentimiento para que la ESPOL realice la comunicación pública de la obra por cualquier medio con el fin de promover la consulta, difusión y uso público de la producción intelectual”



Nicole Katherine Asqui  
Manzaba



Hayleen Ivilin Carrillo  
Carrillo

## **EVALUADORES**

**Ph.D. Lucia M. Villacres Falconi**

PROFESOR DE LA MATERIA

**Ph.D. Federico X. Dominguez Bonini**

PROFESOR TUTOR

## RESUMEN

Actualmente, los institutos universitarios reconocen la importancia de las habilidades comunicacionales en la educación de los estudiantes. Sin embargo, se ha priorizado la expresión escrita dentro de las aulas educativas. De esta manera, los estudiantes han tenido que desarrollar sus habilidades orales de manera natural, lo que repercute en su entorno académico y posteriormente profesional. Este proyecto responde a la necesidad de mejorar las habilidades orales mediante la implementación de un módulo de detección de muletillas integrado en el sistema de Retroalimentación Automática de Presentaciones (RAP) de ESPOL.

El proceso empieza cuando un estudiante realiza una presentación oral en el RAP, posteriormente se inicia la conversión del audio a texto utilizando Whisper OpenAI, luego, este resultado es procesado por el algoritmo de detección de muletillas; de inmediato, se generó un gráfico de barras que muestra cuáles han sido las muletillas detectadas en la exposición y su respectiva contabilización. De forma que el usuario pueda visualizarlo, se incluyó una nueva sección en el reporte del RAP con el gráfico generado y una recomendación para reducir el uso de muletillas.

Finalmente, se obtuvieron resultados positivos al realizar una correcta integración de esta propuesta en el sistema RAP, incluso, se determinó que las herramientas utilizadas permitieron una alta precisión en la transcripción y detección de muletillas, cumpliendo con el objetivo propuesto.

Con esto, se concluye que la implementación de este módulo brinda una retroalimentación más completa a los usuarios, incluso, logrando visualizar otras muletillas en común que tienen los usuarios.

**Palabras clave:** Retroalimentación automática de presentaciones, muletillas, comunicación oral, habilidades blandas

## **ABSTRACT**

*Nowadays, university institutes recognize the impact of communication skills in the training of students. However, priority has been given to written expression. In this way, students have had to develop their oral skills in a natural way, which has repercussions in their academic and professional scope. This project responds to the need to improve oral skills by implementing a detection module of word fillers integrated into ESPOL's Automatic Feedback of Presentations (RAP) system.*

*After an oral presentation was made in the RAP, this audio began the conversion to text using Whisper OpenAI, then, this result was processed by the algorithm for the detection of word fillers; after of this, a bar chart was generated showing which word fillers were detected in the presentation and their respective count. The user could visualize this result as a new section that was included in the RAP, it had the generated graph and a recommendation to reduce the use of word fillers.*

*Finally, positive results were obtained with a correctly integration of this module into the RAP system, even, we could determine that the tools that were used allowed a high accuracy in the transcription and detection of word fillers, thus fulfilling the proposed objective.*

*With this, it is concluded that the implementation of this module provides more complete feedback to the users, even managing to visualize other common word fillers that the users have in common.*

**Keywords:** *Automatic presentation feedback, filler words, oral communication, soft skills.*

# ÍNDICE GENERAL

EVALUADORES.....	7
RESUMEN.....	I
<i>ABSTRACT</i> .....	II
ÍNDICE GENERAL.....	III
ABREVIATURAS .....	V
ÍNDICE DE FIGURAS.....	VI
CAPÍTULO 1 .....	1
1.    Introducción .....	1
1.1    Descripción del problema .....	1
1.2    Justificación del problema.....	3
1.3    Objetivos.....	5
1.3.1    Objetivo General .....	5
1.3.2    Objetivos Específicos .....	5
1.4    Marco teórico .....	6
1.4.1    La comunicación y su importancia en el perfil de una persona .....	6
1.4.2    Herramientas de soporte al aprendizaje de la comunicación .....	7
1.4.3    Sistema de Retroalimentación Automática de Presentaciones (RAP) .....	7
1.4.4    Tecnologías de reconocimiento de audio.....	7
CAPÍTULO 2.....	10
2.    Metodología .....	10
2.1    Requerimientos de la solución.....	10
2.2    Alternativas de solución del problema .....	11
2.2.1    Arquitectura usando capa de Word Encodding y red LSTM.....	11
2.2.2    Arquitectura utilizando Bag of Words .....	13

2.2.3	Determinar muletillas con base a la frecuencia de ocurrencia en la transcripción .....	15
2.3	Diseño y desarrollo de la solución .....	15
2.3.1	Diseño del diagrama de bloques de la solución .....	15
2.3.2	Fases del desarrollo de la solución .....	16
2.4	Diseño y desarrollo de la solución .....	19
2.4.1	Herramientas.....	19
2.4.2	Diseño de aplicativo .....	20
CAPÍTULO 3.....		22
3.	Resultados y análisis .....	22
3.1	Validación del prototipo.....	22
3.1.1	Plan de pruebas de satisfacción .....	22
3.2	Resultados finales .....	23
3.2.1	Resultados de la validación del prototipo .....	23
3.2.2	Integración del módulo de detección de muletillas al RAP de ESPOL muletillas].....	27
CAPÍTULO 4.....		32
4.	Conclusiones Y Recomendaciones.....	32
Conclusiones .....		33
Recomendaciones .....		34
BIBLIOGRAFÍA.....		35
APÉNDICES .....		39

## **ABREVIATURAS**

ESPOL	Escuela Superior Politécnica del Litoral
RAP	Retroalimentación Automática de Presentaciones
CTI	Centro de Tecnologías de Información
AACU	Asociación de Colegios y Universidades Estadounidenses

## ÍNDICE DE FIGURAS

Figura 2.1 Arquitectura usando capa de Word Encodding y red LSTM .....	13
Figura 2.2 Arquitectura utilizando capa de Bag of Words .....	14
Figura 2.4 Diagrama de bloques de la solución propuesta .....	16
Figura 2.3 Diagrama de actividades del RAP de ESPOL con el módulo de detección de muletillas.....	20
Figura 2.5 Pretotipo de sección de detección de muletillas en reporte del RAP .....	21
Figura 3.1 Pretotipo final de resultados de detección de muletillas .....	23
Figura 3.2 Resultados de encuesta sobre la utilidad de la propuesta.....	24
Figura 3.3 Resultados de encuestas a usuarios sobre el entendimiento del pretotipo .....	24
Figura 3.4 Sugerencias de los usuarios para la nueva sección de detección de muletillas.....	25
Figura 3.5 Resultados sobre la inclusión de recomendaciones para la reducción de muletillas.....	26
Figura 3.6 Acciones que los usuarios tomarían para reducir el uso de las muletillas	27
Figura 3.7 Resultados de transcripción de exposición de prueba en el sistema RAP con el nuevo módulo de detección de muletillas .....	28
Figura 3.8 Cantidad de muletillas encontradas en la transcripción .....	29
Figura 3.9 Recepción del gráfico generado en el servidor del RAP de ESPOL.....	30
Figura 3.10 Inclusión de la sección de muletillas en el reporte final entregado al usuario .....	31

# CAPÍTULO 1

## 1. INTRODUCCIÓN

### 1.1 Descripción del problema

La comunicación oral forma parte de diversos procesos en los que participa el ser humano. Desde edad temprana las personas son instruidas para transmitir sus ideas y pensamientos hacia otros, a través de la comunicación. Por tanto, saber qué decir, cuándo y cómo, es una tarea muy importante para la educación [1].

Actualmente, los institutos universitarios reconocen la importancia de las habilidades comunicacionales en la enseñanza de los estudiantes [2]. Sin embargo, se ha priorizado la expresión escrita por sobre la expresión oral dentro de las aulas educativas. De esta manera, los estudiantes han tenido que desarrollar sus habilidades orales de manera natural, sin un entrenamiento explícito o directo, por lo cual no todos llegan a un nivel óptimo [3], lo que repercute en su entorno académico y posteriormente profesional.

Estudios afirman que el desempeño académico y el trabajo en equipo se encuentra relacionados con sus competencias orales [4] [5] [6], por otro lado existen evidencias de que un mal desempeño está relacionado con ciertas conductas durante la transmisión del mensaje tales como deficiente fluidez, presencia de chasquidos, pausas inadecuadas, el uso de muletillas, entre otras [7], que repercuten incluso en la confianza de los estudiantes que posteriormente pueden presentar ansiedad [8].

Por otro lado, en el ámbito profesional según la encuesta realizada por Asociación de Colegios y Universidades Estadounidenses (AACU) a ejecutivos y gerentes de contratación de entre 15 habilidades laborales, la comunicación oral obtuvo el primer puesto por sobre el pensamiento crítico, la toma de decisiones éticas y el trabajo en equipo [9]. Otro estudio realizado

a 354 gerentes pertenecientes a diferentes industrias de entre 25 hasta 500 empleados, ponderaron la comunicación oral en primer lugar de una lista de 16 competencias que incluían otros elementos como comunicación escrita, informática básica y gestión del tiempo [10]. Estos resultados nos demuestran que la comunicación oral es significativamente más importante para los líderes de empresas que poder escribir bien, resolver problemas complejos o ser innovador.

Además, los resultados del primer estudio mostraron que 8 de cada 10 ejecutivos y 9 de cada 10 gerentes de contratación indican que los recién graduados necesitan de buenas habilidades orales cuando buscan un trabajo. Si bien el 80% de empresarios dijo que las habilidades comunicacionales son muy importantes solo el 40% dijo que los recién graduados estaban bien preparados para manejar las tareas que requieren esas habilidades [9].

Por ello las instituciones educativas implementan estrategias para mejorar estas destrezas en los futuros profesionales tales como cursos de entrenamiento, actividades públicas y técnicas que impliquen retroalimentación [6].

Las debilidades que presenta el estudiante para practicar durante sus exposiciones es que no siempre es posible contar con un grupo que posea las habilidades necesarias para reconocer las características de una buena presentación oral o que brinde soporte y retroalimentación precisa de su exposición.

Para afrontar este problema, existen herramientas que brindan retroalimentación automatizada de las exposiciones. Una de ellas, es el sistema de Retroalimentación Automática de Presentaciones (RAP) de la ESPOL, es un sistema tecnológico que ayuda a la construcción del aprendizaje comunicativo, simulando un entorno de audiencia para la exposición los estudiantes. Este sistema es capaz de analizar aspectos de

lenguaje no verbal tales como posturas, miradas, y diapositivas de presentación, y del lenguaje verbal tales como las pausas sonoras y el volumen [11].

Sin embargo, con lo establecido previamente, surge la necesidad de agregar un módulo de análisis de detección de muletillas como aspecto adicional en el análisis del lenguaje verbal que consolide el sistema y potencie el desarrollo de las habilidades orales en los estudiantes. A través de un estudio, se conoce sobre los impactos que tienen el uso de las muletillas en la comunicación, entre ellos, que afectan negativamente la comprensión del oyente [12], incluso, también deteriora la credibilidad de la persona y sus argumentos [13].

## **1.2 Justificación del problema**

Se considera una muletilla a la palabra o interjección que se repite de forma frecuente, ya sea por nerviosismo, inseguridad o escasez de vocabulario, pero que el emisor del discurso ignora. Esto logra distraer a la audiencia, convirtiendo al expositor en motivo de burla, por lo cual no permite una continuidad en la transmisión del mensaje, esto se representa a través del mal uso de sonidos o palabras que no aportan en su mensaje [14]. La transmisión clara y efectiva de un mensaje es importante para la total comprensión de los argumentos que explica el expositor, dado que, para convencer a la audiencia, es significativo transmitir un contenido preciso, con determinación y que no distraiga al público.

Sistemas como el RAP de ESPOL se tornan en una herramienta esencial para el desarrollo de las habilidades comunicacionales de estudiantes en pregrado. Este sistema analiza las exposiciones de los estudiantes y presenta una retroalimentación con criterios que califican las posturas, miradas, pausas sonoras, volumen y diapositivas final a través de una interfaz gráfica [11].

Como resultado, en los años que tiene funcionando en la ESPOL, el RAP ha ayudado a mejorar las habilidades comunicacionales de los estudiantes en un 5% después de su primer uso [15]. En las pruebas iniciales del RAP de ESPOL, los docentes también entregaron retroalimentación a los estudiantes, y muchos de estos resultados coincidían con lo entregado por el sistema [16]. Con esto, se determinó que el sistema ofrece recomendaciones similares a las impartidas por un profesor en una exposición presencial.

No obstante, un punto no favorable detectado en el sistema es la falta del análisis sobre el contenido de la presentación oral del estudiante. Por ejemplo, en el caso de que el estudiante tenga un buen dominio escénico, el sistema RAP determinaría que su presentación fue excelente, sin embargo, al no considerar el mensaje que se está transmitiendo, esa aseveración no estaría totalmente correcta, por lo que la lingüística también contribuye en la construcción de dicha comunicación.

Por tanto, se conoce que el sistema se encuentra parcialmente elaborado para el análisis verbal del mensaje transmitido sobre las exposiciones de los estudiantes. No obstante, es necesario que estos usuarios reciban retroalimentación completa a través de la evidencia de los errores cometidos durante su exposición, que les sirva para su retro inspección y continuo progreso.

Al implementar un módulo en el sistema RAP que ayude a los estudiantes a identificar cuáles y cuántas muletillas utilizan, permite que los estudiantes sean conscientes de las muletillas usadas durante su exposición y así puedan evitar usarlas en una siguiente presentación, mejorando paulatinamente en cada presentación futura.

Se conoce que el estudiante, al recibir una retroalimentación oportuna y casi inmediata, fomenta el desarrollo de sus destrezas y competencias [17],

lo cual mejora su estado académico y su perfil laboral aumentando así su nivel de competencia profesional. Además, el poseer una gran capacidad de comunicación representa un valor añadido para una persona y mejora la imagen de la Institución que la representa [2]. Por consiguiente, la universidad contaría con alumnos con una excelente formación técnica y buena comunicación oral, aumentando la excelencia académica.

### **1.3 Objetivos**

#### **1.3.1 Objetivo General**

Implementar un módulo de detección de muletillas dentro del sistema RAP de ESPOL mediante el uso de herramientas de conversión de voz a texto, con la finalidad de generar un reporte detallado de las muletillas utilizadas en las exposiciones de los estudiantes que les permita analizar y mejorar sus habilidades orales a través de la retroalimentación recibida.

#### **1.3.2 Objetivos Específicos**

- Detectar las muletillas utilizadas por los estudiantes en sus presentaciones orales.
- Identificar la frecuencia de uso de muletillas en las presentaciones orales que realicen los estudiantes.
- Generar un informe de retroalimentación para que el estudiante mejore sus presentaciones.

## **1.4 Marco teórico**

### **1.4.1 La comunicación y su importancia en el perfil de una persona**

La comunicación constituye una característica y una necesidad de las personas y de las sociedades [18]. El desarrollo de las habilidades comunicativas es indispensable en todos los ámbitos de interrelación humana, sobre todo en el ámbito académico y laboral en el cual nos centraremos en la presente investigación.

Entre los años 2020 y 2021 se realizó una investigación a estudiantes de la Escuela Superior Politécnica de Chimborazo sobre el talento comunicacional en la formación integral de los profesionales. Se observó que la competencia comunicativa destaca como una característica significativa para el crecimiento personal y logro de metas productivas [19].

Además, representan el talento comunicacional como un activo intangible importante en las relaciones personales y operativas de una empresa. Esto se debe a que la transmisión correcta de un mensaje permite generar acciones que facilitan o solucionan actividades de trabajo, individuales o de equipo dentro de la organización [19].

Asimismo, un artículo publicado por la Universidad EAN de Colombia destaca el impacto de esta competencia en el entorno laboral y como hoy en día los empleadores buscan profesionales con mejores habilidades comunicativas [20].

De acuerdo con lo mencionado, existen diferentes aspectos a considerar en la correcta transmisión de un mensaje al momento de exponer, desde los gestos, el contacto visual, la gesticulación, y como base para una buena comprensión del mensaje, la fluidez en el habla, la cual se ve obstaculizada cuando se hacen uso de muletillas.

#### **1.4.2 Herramientas de soporte al aprendizaje de la comunicación**

Varios investigadores han logrado desarrollar herramientas que permiten brindar una retroalimentación sobre la presentación oral de los individuos que utilizan estos sistemas, las cuales analizan el lenguaje no verbal del expositor. Entre estas herramientas destacan:

- **Presentation Sensei:** Procesa una exposición en tiempo real, mostrando sugerencias de acuerdo con lo que se obtiene de los sensores (cámara y micrófono). Este sistema fue desarrollado con fines investigativos, y no para uso público [21].
- **Presentation Trainer:** Provee un feedback sobre el lenguaje verbal y no verbal observado/escuchado durante una exposición [22].

#### **1.4.3 Sistema de Retroalimentación Automática de Presentaciones (RAP)**

De acuerdo con las primeras pruebas realizadas en este sistema, los usuarios indicaron que este es de gran utilidad. Entre los aspectos positivos del RAP de ESPOLE destacan que ayuda a mitigar el miedo público y habilidades de comunicación, brinda una retroalimentación útil y permite una práctica adecuada [16]. Sumado a esto, los docentes que formaron parte de estas pruebas indicaron que el sistema los ayuda a evaluar adecuadamente a los estudiantes, porque no siempre es posible dar retroalimentación a cada estudiante de cada paralelo [16]. Es decir, este sistema permite optimizar el tiempo del estudiante y el docente, porque se obtienen resultados sin acordar una sesión presencial entre ellos.

#### **1.4.4 Tecnologías de reconocimiento de audio**

Con la finalidad de realizar un análisis del contenido del mensaje transmitido en las exposiciones que utilicen el sistema RAP, se investigaron

varias tecnologías capaces de convertir voz a texto. Estas permiten que los sistemas respondan de manera correcta y confiable a las voces humanas y brinden servicios útiles y valiosos. Entre algunas de ellas tenemos:

- Cloud Speech-to-Text de Google: ofrece una API capaz de reconocer más de 110 idiomas y dialectos. Además, los autores pueden escribir aplicaciones de dictado usando micrófonos, habilitando el comando y control de voz o escribiendo archivos de audio, entre muchos otros casos de uso. Por último, se puede realizar un cálculo de probabilidad para ver el factor de tono, pronunciación y velocidad del habla en el reconocimiento de voz [22].
- Coqui STT: Es un motor ASR basado en aprendizaje profundo con una API simple. Procesa un audio y lo convierte a texto, esto con una precisión alta, gracias al entrenamiento de modelos [23].
- CMU Sphinx: Posee un conjunto de herramientas útiles para la creación de modelos de machine learning enfocados a plataformas con pocos recursos y lograr reconocimiento del habla, esto hace que su uso se enfoque al desarrollo de aplicaciones prácticas en lugar de perseguir objetivos para la investigación. Sin embargo, la comunidad ha entrenado y publicado varios modelos compatibles con 16 idiomas y dialectos [24].
- DeepSpeech: Un motor desarrollado por Mozilla en Python 3 que usa un modelo de Machine Learning basado en la arquitectura de modelo end-to-end propuesta por Baidu. Actualmente solo ha lanzado modelos pre-entrenados con soporte para inglés y mandarín, sin embargo, provee la documentación suficiente para entrenar un modelo propio en varios idiomas incluido el español [25].

- SpeechBrain: Un conjunto de herramientas de IA conversacional basado en PyTorch, diseñado con el objetivo de acelerar y facilitar la investigación y desarrollo de muchas tecnologías del habla y lenguaje, entre ellas se encuentran el reconocimiento del habla, reconocimiento del hablante, mejora del habla e identificación del lenguaje. Cabe destacar que SpeechBrain se encuentra en su versión beta en la que ofrece un modelo pre-construido de traducción de voz con soporte a español, sin embargo, el soporte a español no está disponible en ninguno de sus modelos pre-construidos de speech-to-text [26].
- Whisper OpenAI: Modelo de reconocimiento del habla que ha sido entrenado con un gran conjunto de audios para su transcripción. Además, puede realizar el reconocimiento y traducción en otros idiomas diferentes al inglés. Para su entrenamiento utilizó 680000 horas de audio, donde 117000 horas correspondían a otros 96 idiomas. La alta precisión de este modelo fue conseguida sin la necesidad de técnicas de autosupervisión o autoentrenamiento [27].

A medida que se desarrolla la ciencia, dan paso a diferentes tecnologías como las vistas anteriormente, que nos permiten automatizar el trabajo de transcripción a través de diferentes implementaciones cada vez más eficientes y precisas.

# CAPÍTULO 2

## 2. METODOLOGÍA

### 2.1 Requerimientos de la solución

Previo al desarrollo de la solución, se realizaron las siguientes reuniones con varios involucrados en el problema, que se detallan a continuación:

- Cliente: aproximadamente cuatro horas durante dos sesiones, donde se explicó cómo estaba desarrollado el RAP actualmente, qué se requería implementar, cuáles son las razones e importancia de la búsqueda de esta solución.
- Estudiantes de ESPOL: una hora en una sesión, donde, a través de las opiniones de varios estudiantes se logró indagar sobre el uso de muletillas durante sus exposiciones, si son conscientes de ello, y cómo estas afectan a su vida cotidiana.
- Tutor: aproximadamente seis horas durante 4 sesiones. A lo largo de estas sesiones se conoció el estado previo del RAP, las investigaciones realizadas para las posibles soluciones al problema en cuestión, indicaciones generales y recomendaciones sobre los avances de la solución que se presentaban.

Luego de las sesiones, se presentaron y refinaron los requerimientos para la solución, mismos que se describen a continuación:

- El usuario podrá conocer cuáles han sido las muletillas más usadas que ha utilizado durante su presentación.
- El usuario conocerá la cantidad de muletillas que ha usado durante su exposición.

Por tanto, con el objetivo de lograr el cumplimiento de los requerimientos, se determinó que existen tres fases clave en el desarrollo de la solución:

- Conversión de audio a texto: dentro de las tecnologías de reconocimiento de audio, se utilizó Whisper. Esta herramienta permite la transcripción de las exposiciones orales realizadas en el RAP. Se identificó que existe una alta precisión en varios idiomas, como el español, logrando identificar correctamente cada palabra en el audio.
- Identificación de muletillas en el texto: es importante identificar cuándo una palabra es una muletilla durante una exposición. De acuerdo con la sección previa, se determinó que un factor a considerar es su repetición. A través de una investigación con los estudiantes del curso de la profesora Msc. Paquita Calderón Acosta se conoció la lista de las muletillas más usadas por los estudiantes. El algoritmo desarrollado implementa el reconocimiento de estas palabras en el texto transcrito.
- Visualización de resultados en un reporte: actualmente, el RAP presenta un reporte con aspectos previamente señalados como las posturas, miradas, pausas sonoras, volumen y diapositivas final, por lo que se incluye una sección detallada para el nuevo módulo implementado que muestra las muletillas más usadas durante la exposición. Finalmente, se muestra la recomendación dada por un especialista para evitar su uso.

## **2.2 Alternativas de solución del problema**

### **2.2.1 Arquitectura usando capa de Word Encodding y red LSTM**

Inicialmente, es necesario realizar una transcripción del audio de la exposición realizada en el RAP, lo que se desarrolla con ayuda de la herramienta Whisper. Se desea obtener la mayor precisión al momento de que el modelo transcriba a texto, debido a que la mayoría de las

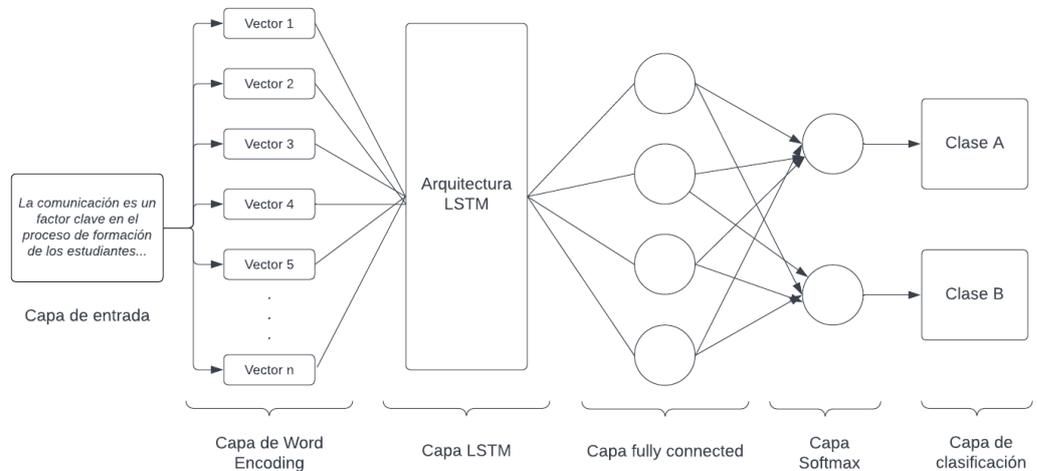
herramientas presentan mayor demanda en otros idiomas, y el entrenamiento en idioma español es menor.

Por otro lado, para el desarrollo del algoritmo se puede utilizar una arquitectura que permita analizar una secuencia y recordar ciertos datos importantes en un instante de tiempo futuro. La arquitectura que se ajustó a los requerimientos de la solución desarrollada es la Long Short-Term Memory (LSTM) o memoria a corto plazo.

Esta red neuronal, propuesta por Sepp Hochreiter y Jürgen Schmidhuber en 1997 [28], consiste en un conjunto de subredes conectadas de forma recurrente [29], que utiliza vectores de celdas de memoria y un conjunto de gates, o compuertas, de multiplicación para poder controlar la forma en la que la información es almacenada, olvidada y utilizada en la red [30].

Como se muestra en la Figura 2.1, esta alternativa de solución propone el uso de una arquitectura con las siguientes capas:

- Capa de entrada: ingresará el texto a procesar.
- Capa de word embedding: los datos de entrada serán convertidos en una secuencia de valores numéricos (vectores), y se logrará capturar la semántica [31], infiriendo en que las palabras con significados similares tendrán vectores semejantes.
- Capa LSTM: se procesan las secuencias enviadas previamente para entrenar al modelo.
- Capa fully connected: utilizada para clasificar las entradas. Su salida es una función softmax.
- Capa Softmax: establece la probabilidad estimada de pertenencia a cada clase establecida del valor de análisis [31].
- Capa de clasificación: permitirá determinar la categoría a la que pertenece determina palabra o grupo de palabras del texto original [31].



**Figura 2.1 Arquitectura usando capa de Word Encoding y red LSTM**

### 2.2.2 Arquitectura utilizando Bag of Words

Al igual que la propuesta anterior, se pretende utilizar la herramienta Whisper, misma que facilitará la transcripción del audio a texto. Esto con la finalidad de obtener las palabras adecuadas, y hacer una correcta búsqueda de las respectivas muletillas.

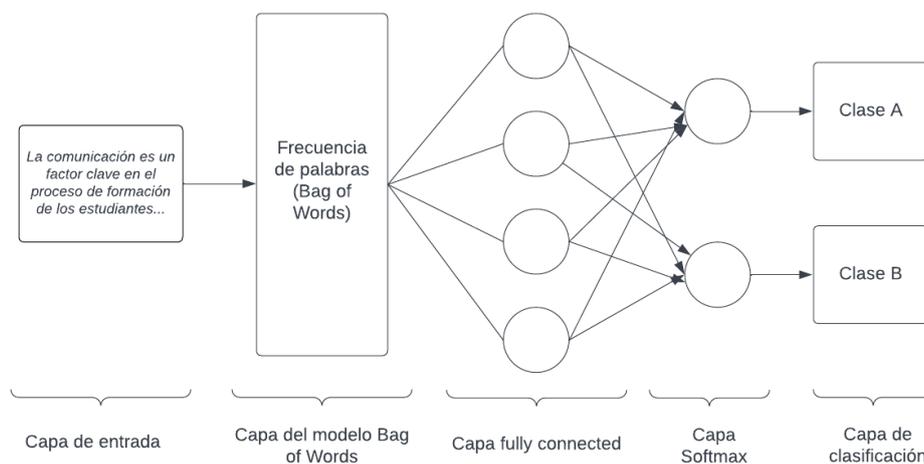
El objetivo de esta solución es determinar si el texto obtenido de la transcripción contiene muletillas o no, es decir, poder identificarlas y determinar la cantidad de estas. Para ello, se propone en lugar de usar una red de tipo LSTM, utilizar un modelo de bolsa de palabras, o bag of words, que permite extraer determinadas características (o palabras) y realizar un conteo de sus ocurrencias [32].

Con lo mencionado previamente, se analizaría una gran cantidad de texto (proveniente del audio de la exposición en el RAP), y se llevaría un registro de las apariciones de cada palabra existente en el texto, en donde nos enfocáramos en las que formen parte del listado de muletillas.

Luego de esto, se daría paso a una capa con una red de clasificación, en donde los parámetros de selección para que el texto sea clasificado como “texto con muletillas”, sería que contenga mínimo dos palabras del listado de muletillas con una cantidad total superior a cinco.

Entonces, el modelo sería entrenado para que logre clasificar el texto general en una de las dos clases (tiene muletillas o no). Al igual que en la alternativa anterior, se obtiene una probabilidad de ocurrencia de las clases, que se encuentra en una capa Softmax.

Finalmente, se tiene la capa de clasificación donde, con base a la probabilidad obtenida, se determina si el texto pertenece al grupo de que contiene muletillas, o no. Esto puede visualizarse en la Figura 2.2.



**Figura 2.2 Arquitectura utilizando capa de Bag of Words**

### **2.2.3 Determinar muletillas con base a la frecuencia de ocurrencia en la transcripción**

Esta alternativa tiene como objetivo lograr una alta precisión en el entrenamiento del modelo de transcripción del audio de las exposiciones del RAP de ESPOL. Para ello, se propone realizar un entrenamiento exhaustivo de un modelo que permita realizar la conversión de audio a texto, puesto que, sin esto, la detección de muletillas sería muy compleja de realizarlo.

En esta alternativa se propone utilizar la herramienta Whisper. Se entrena un modelo de reconocimiento de voz en idioma español para poder realizar una transcripción con una precisión alta. Esto será comparado con las transcripciones obtenidas de forma manual.

Luego de esto, se realizaría un análisis del texto obtenido, para realizar la detección de muletillas. El listado de muletillas con el que se debe de comparar las palabras de la transcripción es proporcionado por el área de Comunicación de ESPOL.

Con esta información obtenida, se presentaría en el reporte la cantidad de muletillas encontradas, y cuáles fueron las que presentaron un valor alto de ocurrencia. Sumado a esto, también se realizarían recomendaciones para futuras exposiciones del usuario.

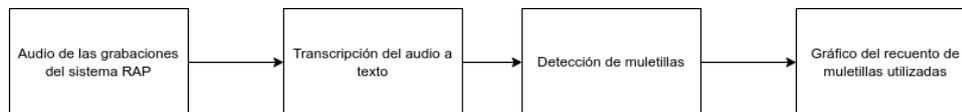
## **2.3 Diseño y desarrollo de la solución**

### **2.3.1 Diseño del diagrama de bloques de la solución**

De acuerdo con la metodología establecida previamente, el nuevo módulo del sistema RAP posee varias etapas, cuyo flujo se muestra en el diagrama de bloques de la **¡Error! No se encuentra el origen de la referencia..**

En este se indica que el módulo inicia con la obtención del audio de las grabaciones del sistema RAP, obteniendo un archivo con formato WAV. Luego de esto, este archivo pasa a ser transcrito por la herramienta Whisper OpenAI, siendo la salida de esta etapa la entrada para el algoritmo de detección de muletillas, en donde se determina el uso de muletillas durante la presentación, las mismas que son almacenadas en un archivo de formato JSON.

Finalmente, con este archivo se genera un gráfico de barras que muestre las muletillas detectadas, es decir, su salida es un archivo en formato JPG. Entonces, este es enviado al servidor RAP para que sea incluido en el reporte del usuario, que puede ser visualizado en la aplicación web del RAP de ESPOL.



**Figura 2.3 Diagrama de bloques de la solución propuesta**

## **2.3.2 Fases del desarrollo de la solución**

### **2.3.2.1 Fase 0: Encuesta sobre utilidad del nuevo módulo**

Previo a la realización del proyecto se realizó una encuesta a 20 estudiantes que anteriormente utilizaron el RAP. En la misma se realizaron preguntas sobre el diseño del gráfico con las muletillas detectadas y la utilidad del nuevo módulo que se desea implementar.

### **2.3.2.2 Fase 1: Preprocesamiento de los datos**

Con el propósito de medir la efectividad del algoritmo en la detección y descubrir algunas de las muletillas más frecuentes en los estudiantes, se realizó un estudio preliminar con 20 grabaciones, que consistió en la transcripción manual de estos audios.

Estos reportes permiten comprobar la efectividad del sistema escogido para la transcripción, siendo este Whisper. Además, permite detectar y comparar las muletillas comunes según el reporte generado por el área de Comunicación de ESPOL, en el cual detalla la siguiente lista con las muletillas más usadas por los estudiantes:

- Osea.
- Este.
- ¿Ya?
- Digamos.
- El cual, La cual.
- Mejor dicho.
- ¿Cómo es que es?
- Es decir.
- Pues.
- Bueno.
- Ok.
- Ajá.

Además, se añadieron al listado varias muletillas detectadas en las transcripciones manuales realizada de 20 audios de estudiantes, debido al constante uso por parte de estos usuarios. Los audios utilizados contenían exposiciones con una duración de aproximadamente 5 minutos.

### **2.3.2.3 Fase 2: Procesamiento de datos usando Whisper**

El sistema de transcripción Whisper OpenAI empieza recibiendo la muestra de audios aleatorios dentro de la base de datos del sistema RAP. Esta herramienta tiene la capacidad de transcribir los audios a texto dado que es un conjunto de herramientas de aprendizaje profundo

de código abierto para entrenar e implementar modelos de voz a texto [33].

Whisper es un modelo de reconocimiento de voz de última generación, es capaz de transcribir en varios idiomas y entre una de sus ventajas es que funciona bien aun cuando el audio tiene ruido de fondo [33]. Además, es una herramienta gratuita, lo que permite eliminar costos en su desarrollo [33].

Whisper, y su modelo en idioma español, fue entrenado con los audios de exposiciones realizadas en el RAP de ESPOL. Estos deben de cumplir ciertas especificaciones, que incluyen que deben ser en lenguaje español, con un sample rate de 16000 Hz, en formato WAV y con un canal de audio simple.

Se detectó que, al dividir el audio en fragmentos de 10 a 20 segundos, la precisión de transcripción aumentaba, por lo que se decidió dividir los audios de las presentaciones en varias secciones, sobre todo cuando existen pausas. Finalmente, se reúnen todas las transcripciones de los audios de toda la exposición, lo que se convertiría en la entrada de la siguiente fase.

Por último, para medir la efectividad de las transcripciones realizadas por Whisper OpenAI se compararon con las transcripciones manuales y se verificó que la herramienta tiene precisión con el idioma español.

#### **2.3.2.4 Fase 3: Identificación de muletillas en el texto**

Mediante el algoritmo desarrollado, el texto ingresado es comparado con cada palabra del listado descrito en la *Fase 1: Preprocesamiento de los datos*, junto con las nuevas muletillas detectadas en las transcripciones. Posteriormente, esta información es almacenada en un archivo de

formato JSON, donde se forma un diccionario cuya clave es el nombre de cada muletilla y el valor corresponde a su contabilización. Finalmente, a partir del archivo recibido se genera un gráfico de barras con el nombre de cada muletilla como eje horizontal y la contabilización de estas en el eje vertical.

#### **2.3.2.5 Fase 4: Inclusión de resultados en el reporte**

La información enviada en la *Fase 3: Identificación de muletillas en el texto* es escuchada en el servidor de RabbitMQ, intermediario de mensajes o administrador de colas [34] que utiliza el sistema RAP. Finalmente, a través de la interfaz gráfica se muestra a cada estudiante la nueva sección de muletillas. Este apartado indica mediante un gráfico de barras las muletillas más utilizadas durante la exposición. De esta manera, el usuario puede visualizar cuáles han sido y cuántas veces ha hecho uso de ellas.

## **2.4 Diseño y desarrollo de la solución**

### **2.4.1 Herramientas**

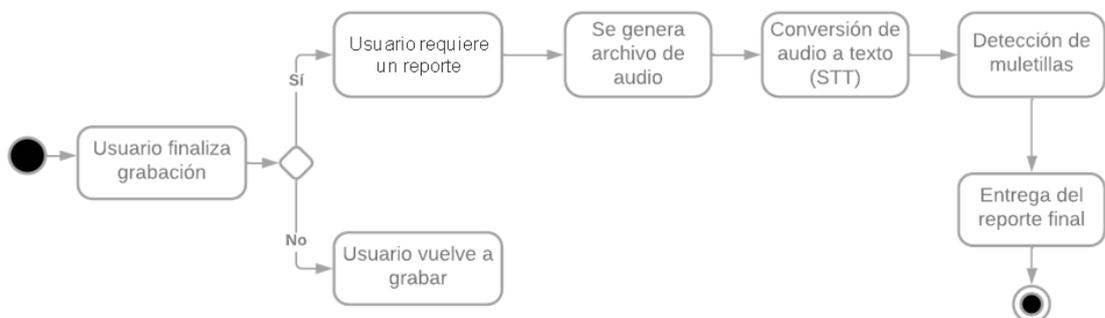
- Whisper OpenAI: Es un modelo de ASR (Automatic Speech Recognition) el cual está entrenado con 680,000 horas de distintos audios en múltiples lenguajes y acentos. Sirve la transcripción de voz a texto [33].
- Python (versión 3.8): Es uno de los lenguajes más populares para el desarrollo de software porque es sencillo de implementar código [35].
- Pydub (versión 0.25): librería de Python utilizada para trabajar con los audios en formato WAV.
- Matplotlib: es una librería de Python especializada en la creación de gráficos en dos dimensiones.

- Requests: Es una librería de Python que sirve para realizar solicitudes HTTP.
- RabittMQ: es un software donde se definen las colas a las que se conectan las aplicaciones para transferir un mensaje o mensajes [34].

## 2.4.2 Diseño de aplicativo

### 2.4.2.1 Diagrama de actividades.

En el sistema RAP de ESPOL, cuando el usuario finaliza la grabación, y desea obtener el reporte de su presentación, el sistema procede a generar un archivo de audio. Posteriormente, con ayuda de Whisper, este audio se convierte en texto, para poder empezar con la identificación de muletillas.

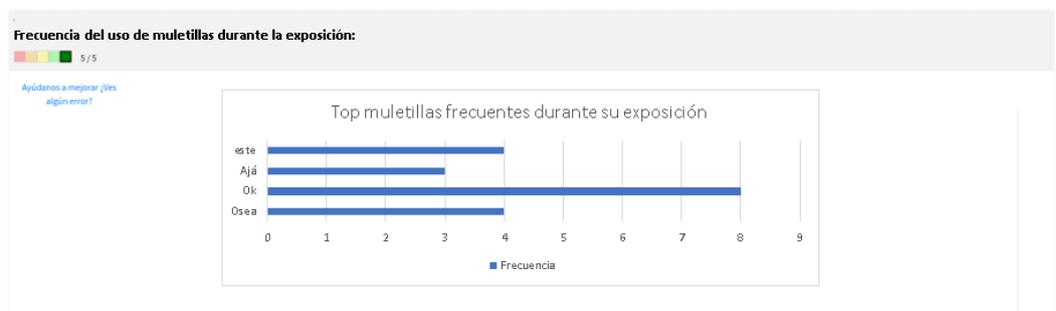


**Figura 2.4 Diagrama de actividades del RAP de ESPOL con el módulo de detección de muletillas**

### 2.4.2.2 Pretotipo de la solución.

El sistema RAP luego de realizar el análisis de la presentación oral del expositor, presenta un reporte en el perfil de cada usuario. Las muletillas encontradas, y su respectiva cantidad, serán incluidas como una nueva sección de este sistema. Como se puede visualizar en la Figura 2.5, esta parte del reporte se compone de:

- Frecuencia del uso de muletillas durante la exposición: se presenta un gráfico que indique las muletillas más utilizadas durante la presentación oral. Aquí también se indica la cantidad de cada muletilla usada.
- Recomendación: se provee un consejo o recomendación al usuario para que logre generar impacto en futuras presentaciones.



**Figura 2.5 Pretotipo de sección de detección de muletillas en reporte del RAP**

# CAPÍTULO 3

## 3. RESULTADOS Y ANÁLISIS

La integración del nuevo módulo de detección de muletillas en el sistema RAP de ESPOL se incluye como una nueva sección dentro del reporte final del usuario, lo que fue logrado correctamente. No obstante, para validar que lo entregado era lo requerido por el usuario, previamente se realizó una encuesta a varios estudiantes que han hecho uso del RAP recientemente, con la finalidad de conocer si el resultado generado por el nuevo módulo les resultaría útil en una integración futura, incluso, mencionar las recomendaciones que tienen respecto al prototipo presentado.

### 3.1 Validación del prototipo

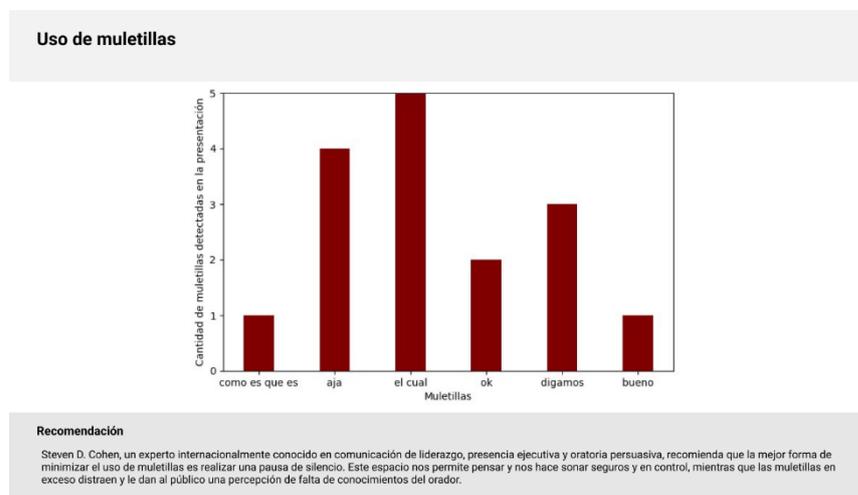
Para validar el prototipo diseñado se realizaron pruebas de usuario. A través del desarrollo de un plan de pruebas de satisfacción, se encuestó a los usuarios para conocer su opinión sobre el prototipo, cómo se sienten respecto a los resultados presentados, y encontrar sugerencias de utilidad para que sean incluidas en la integración final.

#### 3.1.1 Plan de pruebas de satisfacción

Para realizar esta encuesta se convocaron a 20 estudiantes de ESPOL, de la carrera Ingeniería en Computación, quienes habían utilizado el sistema RAP durante los meses de diciembre de 2022 y enero de 2023.

De forma general, la encuesta inició con la presentación del prototipo mostrado en la Figura 3.1. Luego de un análisis a este, se pretendía determinar la opinión de los usuarios sobre los siguientes aspectos:

- Utilidad de la inclusión de una sección de detección de las muletillas en el reporte.
- Correcto entendimiento del gráfico para el usuario.
- Interés del usuario en recomendaciones para disminuir el uso de muletillas.
- Cambios que el usuario realizaría en la nueva sección.

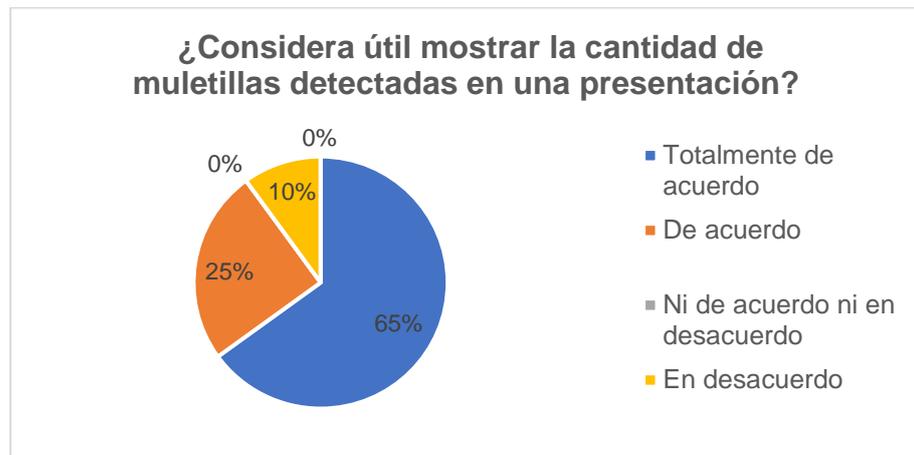


**Figura 3.1 Pretotipo final de resultados de detección de muletillas**

## 3.2 Resultados finales

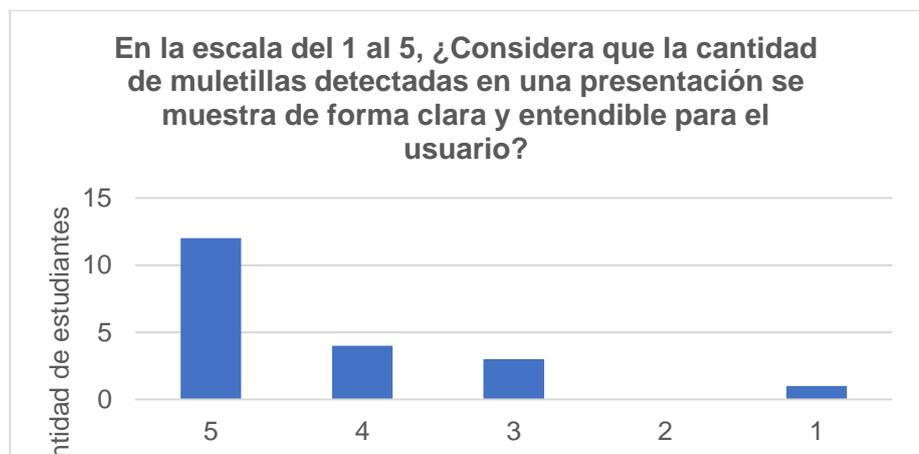
### 3.2.1 Resultados de la validación del pretotipo

Como se puede visualizar en la Figura 3.2, el 90% de los usuarios indican que es útil mostrar la cantidad de muletillas detectadas en una presentación. A través de esto, se podría considerar que la inclusión del módulo de detección de muletillas en el RAP de ESPOL resulta beneficiosa.



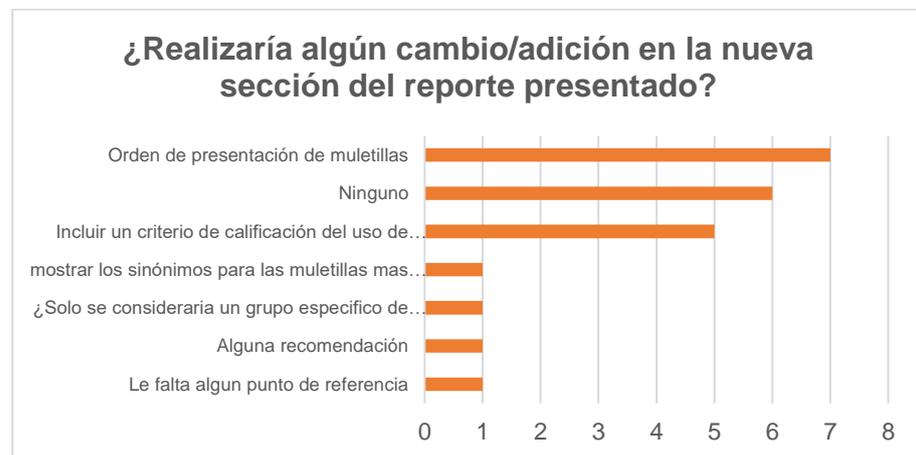
**Figura 3.2 Resultados de encuesta sobre la utilidad de la propuesta**

Por otra parte, a través de la Figura 3.3, se puede observar que para el 80% de los usuarios encuestados la presentación del gráfico que muestra la cantidad de muletillas detectadas es bastante clara y entendible. No obstante, al 20% de usuarios colocaron notas más bajas, como 3 y 1. Es decir, es necesario realizar cambios para que el usuario logre entender completamente la nueva sección en el reporte del RAP. Esta dificultad podría deberse al orden de presentación del conteo de cada muletilla encontrada, puesto que no presentaban un orden definido.



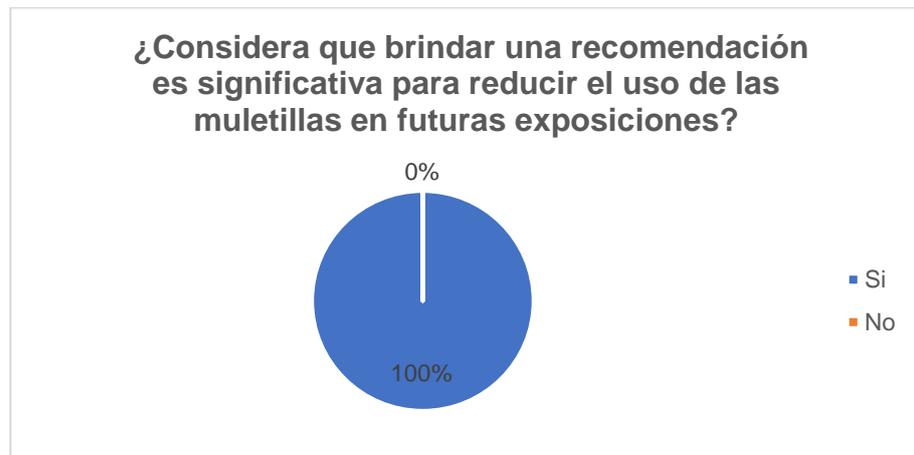
**Figura 3.3 Resultados de encuestas a usuarios sobre el entendimiento del prototipo**

Pese a que los usuarios indicaron que les resulta útil la inclusión del nuevo módulo, a nivel de presentación en el reporte también indicaron varias sugerencias como se muestra en la Figura 3.4. La más notable es que debería existir un orden de presentación en las muletillas. En este caso, se debe considerar realizar un reordenamiento según las más utilizadas por el usuario en su presentación. Por otra parte, también se indicó que se podría incluir un criterio de calificación en el uso de muletillas, tal y como se la realiza en otras secciones del reporte (antes de la inclusión del módulo de detección de muletillas).



**Figura 3.4 Sugerencias de los usuarios para la nueva sección de detección de muletillas**

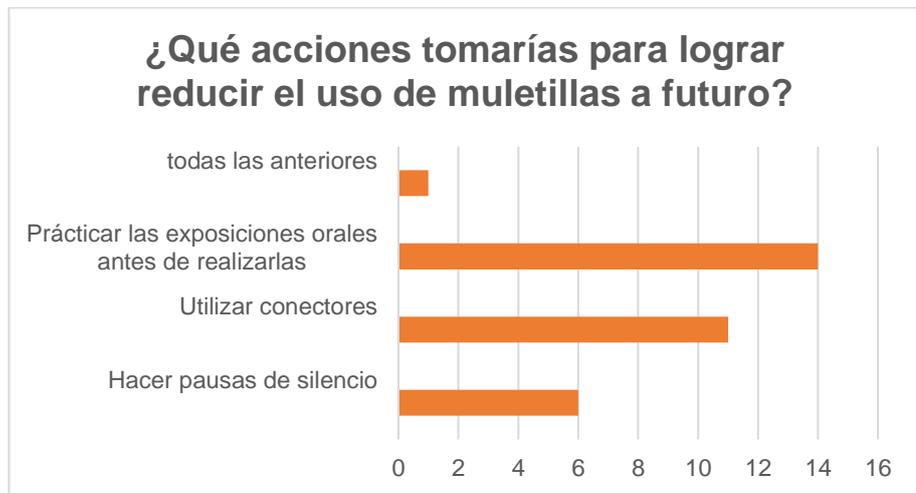
Por otro lado, como se visualiza en la Figura 3.5, se obtuvo una respuesta unánime que nos permite conocer que a los usuarios les parece significativa la inclusión de una recomendación brindada por un experto para reducir el uso de las muletillas, esto es importante para nuestro proyecto porque nos indica que los usuarios tendrán en cuenta y aplicarán la recomendación dada en sus futuras exposiciones.



**Figura 3.5 Resultados sobre la inclusión de recomendaciones para la reducción de muletillas**

Finalmente, según los resultados mostrados en la Figura 3.6, obtuvimos que la mayoría de los usuarios tomarían la acción de practicar las exposiciones orales antes de realizarlas, lo cual es un indicativo significativo para el uso del RAP, dado que es recomendable que el usuario al detectar las muletillas que usa durante su exposición realice pruebas constantes que le permita conocer si continúa haciendo uso de ellas.

Por otra parte, algunos usuarios mencionaron que el uso de conectores también es un método viable, por lo que se debería considerar implementar aquella sugerencia para una versión futura del proyecto. Por último, aquellos que seleccionaron la pausa de silencio como una acción que reduciría el uso de muletillas da un indicativo de que la recomendación realizada por el experto se considera importante para reducir su uso.



**Figura 3.6 Acciones que los usuarios tomarían para reducir el uso de las muletillas**

### **3.2.2 Integración del módulo de detección de muletillas al RAP de ESPOL muletillas|**

De acuerdo con la metodología planteada, se lograron resultados bastante precisos en la transcripción y detección de muletillas gracias a las herramientas utilizadas. Para esto, se desarrollaron múltiples pruebas luego de la integración del nuevo módulo, y de esta forma, verificar que se detecten las muletillas que se mencionaban en la presentación oral.

Las pruebas fueron desarrolladas en el sistema RAP del Centro de Tecnologías de Información (CTI) de ESPOL. Durante este proceso, se realizaron exposiciones de prueba que duraban un minuto, donde se presentaron temas netamente de pruebas, como oraciones de saludo y descripción de lo que se estaba realizando, incluyendo varias muletillas a lo largo de estas pruebas.

En este caso, se presentarán los resultados de una presentación de prueba realizada, sin embargo, cabe destacar que los resultados de las otras pruebas desarrolladas presentaron niveles similares de precisión en su transcripción y detección de muletillas.

### 3.2.2.1 Resultados de presentación de prueba realizada en el sistema RAP del CTI de ESPOL

Luego de que la presentación haya sido realizada, de forma automática, se procesaba el audio de esta. Entonces, se daba paso al análisis realizado por el nuevo módulo, mismo que, aproximadamente, toma un minuto y medio, en caso de que la presentación sea de máximo un minuto.

Posteriormente, con el propósito de verificar que se estaba realizando correctamente el proceso de transcripción, se estableció que se pueda visualizar la transcripción de la presentación de prueba en el servidor del sistema RAP del CTI, como se observa en la Figura 3.7. Y, de acuerdo con lo expuesto, se determina que el nivel de precisión en la transcripción es alto, a comparación de otras herramientas mencionadas.

```
[x] Done. extrayendo el archivo
/home/rsp/Documents/rsp_filler/filler_service/audios/1/554/wav
/home/rsp/Documents/rsp_filler/filler_service/audios/1/554
Using the whisper STT module
Loading model
Loading model took 3.1304574413225058s
Transcribing audio file
/root/.miniconda3/envs/filler/lib/python3.7/site-packages/whisper/transcribe.py:78: UserWarning: FP16 is not supported on CPU; using FP32 instead
  warnings.warn("FP16 is not supported on CPU; using FP32 instead")
Transcribing the audio took 11.914856549862027s
Final transcription:
Buenos días, voy a hacer una presentación muy corta, espero que sean unos 15 segundos de grabación máximo. Vamos a ver cómo queda el video final y nos vamos a hacer una presentación. Gracias
```

Figura 3.7 Resultados de transcripción<sup>1</sup> de exposición de prueba en el sistema RAP con el nuevo módulo de detección de muletillas

---

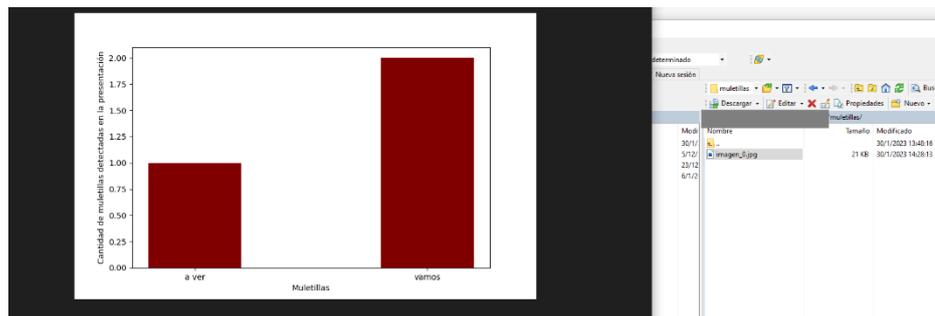
<sup>1</sup> El resultado de la transcripción es: *Buenos días, voy a hacer una presentación muy corta, espero que sean unos 15 segundos de grabación máximo. Vamos a ver cómo queda el video final y nos vamos a hacer una presentación final. Gracias*

Sumado a esto, también se puede visualizar las muletillas con su respectivo conteo. En el caso de que alguna de ellas no sea utilizada durante la presentación, simplemente se le asigna con un conteo de 0. Esto se debe a que se buscan coincidencias de las muletillas con base a un listado establecido previamente. Los resultados de muletillas encontradas en esta prueba realizada se pueden observar con mayor detalle en la Figura 3.8.

```
Common repetitions counter: {
  "cómo es que es": 0,
  "a fin de cuentas": 0,
  "de eso se trata": 0,
  "cómo se llama": 0,
  "cómo se dice": 0,
  "mejor dicho": 0,
  "es decir": 0,
  "el cual": 0,
  "la cual": 0,
  "o sea": 0,
  "me entienden": 0,
  "me comprenden": 0,
  "de hecho": 0,
  "en plan": 0,
  "quiero decir": 0,
  "me explico": 0,
  "a ver": 1,
  "ok": 0,
  "ajá": 0,
  "digamos": 0,
  "bueno": 0,
  "este": 0,
  "pues": 0,
  "ya": 0,
  "entonces": 0,
  "lógicamente": 0,
  "ciertamente": 0,
  "obviamente": 0,
  "verdad": 0,
  "sabes": 0,
  "vamos": 2,
  "venga": 0,
  "vale": 0,
  "ojo": 0
}
<Response [200]>
```

**Figura 3.8 Cantidad de muletillas encontradas en la transcripción**

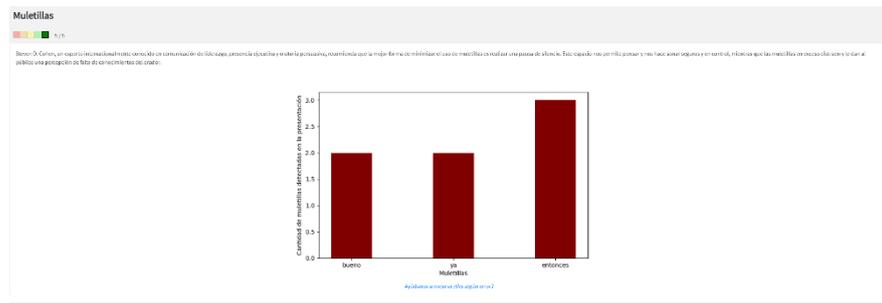
Mientras se realiza este proceso, también se elabora un gráfico que muestre solo las muletillas utilizadas y la cantidad encontrada en la transcripción. Esto, es enviado al servidor general del RAP de ESPOL, para que sea incluido dentro del reporte del usuario, en este caso, al usuario de prueba. Lo mencionado puede ser visualizado en Figura 3.9.



**Figura 3.9 Recepción del gráfico generado en el servidor del RAP de ESPOL**

Luego de las pruebas realizadas, se logró determinar que la integración del módulo fue exitosa, por lo que se inició el análisis de una presentación de un estudiante, cuya duración era máximo de 5 minutos. En este caso, el análisis de detección de muletillas duró aproximadamente 7 minutos y medio, por lo que se evidencia que, de acuerdo con los minutos de exposición, al módulo de detección de muletillas le tomará analizar aproximadamente 1.5 veces la misma cantidad de minutos utilizados para la presentación.

Finalmente, como parte de la integración, la nueva sección de detección de muletillas fue incluida en el reporte final entregado por el sistema RAP. Para esto, se consideraron las indicaciones y sugerencias brindadas por los usuarios, como el orden de presentación en el gráfico de las muletillas e incluir una recomendación del usuario, lo que se puede visualizar en Figura 3.10.



**Figura 3.10 Inclusión de la sección de muletillas en el reporte final entregado al usuario**

Pese a que esta integración aún se encuentra en fase pruebas, se determina que funciona de acuerdo con los requerimientos solicitados, incluyendo recomendaciones por parte de usuarios que pueden ayudar en el mejoramiento de esta herramienta.

# CAPÍTULO 4

## 4. CONCLUSIONES Y RECOMENDACIONES

Actualmente, la comunicación oral es un factor diferenciador y positivo en los perfiles de los estudiantes y profesionales técnicos, por lo que, mejorar este tipo de habilidades implica muchas ventajas. El sistema RAP ha ayudado a muchos estudiantes a lograr visualizar cómo es su comportamiento comunicacional, tanto en aspectos de su lenguaje verbal tal como el volumen y pausas, como del no verbal tal como la postura, miradas y análisis de las diapositivas. Sin embargo, previamente se identificó que no realizaba un análisis a las muletillas utilizadas por los usuarios, lo que implicaba un análisis parcial del sistema en los criterios de lenguaje verbal.

A través del desarrollo de este trabajo se logró potenciar al sistema RAP, incluso, gracias a la herramienta de conversión de voz a texto, Whisper, se logró obtener transcripciones de las presentaciones orales en poco tiempo y con una precisión alta, permitiendo encontrar de forma sencilla las muletillas utilizadas. Pese a esto, inicialmente también fue necesario realizar transcripciones manuales y verificar que estas coincidan con el texto transcrito por la herramienta utilizada.

Como se mencionó en la literatura, existen algunas opciones frente al uso del RAP, pero estas, no son tan completas o rápidas como este sistema. Muchas veces se necesita de personas especializadas para brindar algún tipo de retroalimentación, o no existen herramientas tecnológicas especializadas, en nuestro país o idioma.

Entonces, con base a lo realizado a lo largo de este proyecto, se presentan las principales conclusiones obtenidas, y varias recomendaciones para futuros trabajos.

## Conclusiones

- La implementación del módulo de detección de muletillas brinda una retroalimentación más completa a los usuarios sobre sus exposiciones orales, puesto que, ahora, también se incluirá un análisis a otro criterio del lenguaje verbal, como es la detección de muletillas.
- Se identificaron varias muletillas (indicadas en el listado de la metodología) dentro de las exposiciones orales del RAP, incluso, a través de la transcripción manual, se identificaron nuevas muletillas en común que utilizan los estudiantes, por lo que fueron incluidas en el listado de muletillas inicial.
- Se logró determinar la cantidad de cada muletilla utilizada por el usuario. Estas son mostradas en un gráfico de barras, para que se visualicen de forma clara, y se distinga cuáles son las muletillas que más se presentaron dentro de su exposición oral.
- Las sugerencias del usuario previo a la integración del módulo de detección de muletillas del sistema RAP fueron de gran utilidad, puesto que permitieron construir un mejor resultado para futuros usuarios.
- La integración del módulo de detección de muletillas cumplió con los requerimientos y expectativas de los involucrados, por ello, se puede concluir que el objetivo de este proyecto fue alcanzado con éxito.

## **Recomendaciones**

- Se puede mejorar el funcionamiento del algoritmo de detección de muletillas con el uso de técnicas de inteligencia artificial y una obtención de datos mayor. Esto con el objetivo de detectar y reconocer el contexto de las muletillas, y establecer con mayor precisión cuándo es una muletilla o no.
- Como se realizó dentro del proyecto, se recomienda realizar más pruebas y detección de nuevas muletillas en común que utilizan los usuarios. Todo esto, se puede lograr a través de la transcripción de los audios de quienes utilicen este sistema, y llevar un registro de cuáles son palabras con tendencia a ser muletillas, y poder agregarlas al listado, detectándolas en presentaciones futuras.
- Es posible realizar otros estudios relacionados a la búsqueda de errores frecuentes que cometen los estudiantes en las presentaciones orales, como la detección de palabras comodín. Esto se puede realizar con la ayuda de la transcripción generada de la exposición, y, a través de un algoritmo, identificar cuándo son utilizadas este tipo de palabras.
- Los estudiantes pueden utilizar los recursos del RAP continuamente, siguiendo la recomendación planteada por el experto sobre las pausas de sonido, para que de esta manera puedan evidenciar un progreso en la reducción del uso de muletillas.

# BIBLIOGRAFÍA

- [1] J. Santillán-Aguirre, *La importancia de la comunicación oral y escrita en el siglo XXI*, Riobamba, Chimborazo: Casa Editora del Polo, 2022.
- [2] J. A. Barrio del Campo y Á. Barrio Fernández, «Análisis de la habilidad comunicativa docente y pautas de actuación. Atraer la atención hablando: un reto para la enseñanza universitaria,» *Revista INFAD de Psicología*, vol. 1, nº 1, pp. 73-84, Abril 2018.
- [3] M. J. Ruiz Muñoz, «El desarrollo de competencias orales y escritas en el marco del espacio europeo de educación superior (ees). Reflexiones, propuestas y experiencias en el grado en publicidad y relaciones públicas,» España, 2012.
- [4] R. Mowbray y L. B. Perry, «Improving lecture quality through training in public speaking,» 21 Octubre 2013. [En línea]. Available: <https://www.tandfonline.com/doi/abs/10.1080/14703297.2013.849205>.
- [5] R. Martínez , C. González, . P. Campoy, Á. García Sánchez y M. Ortega-Mier, «Do classes in cooperative classrooms have a positive influence on creativity and teamwork skills for engineering students?,» 2014. [En línea]. Available: <https://dialnet.unirioja.es/servlet/articulo?codigo=7377523>.
- [6] M. Maldonado, A. García, J. Armada Crespo, F. Alós y E. Moreno Osella, «Competencia oral y ansiedad: entrenamiento y eficacia en estudiantes universitarios,» *Revista Latina de Comunicación Social*, nº 80, pp. 401-434, 11 Julio 2022.
- [7] C. Mancuso y R. G. Miltenberger, «Using habit reversal to decrease filled pauses in public speaking,» 2 Noviembre 2015. [En línea]. Available: <https://onlinelibrary.wiley.com/doi/10.1002/jaba.267>.
- [8] P. Charoensukmongkol, «The role of mindfulness in reducing English language anxiety among Thai college students,» *International Journal of Bilingual Education and Bilingualism*, vol. 22, nº 4, pp. 414-427, 19 Mayo 2019.

- [9] . C. Gewertz, «What Do Employers Want in a New Hire? Mostly, Good Speaking Skills,» 28 Agosto 2018. [En línea]. Available: <https://www.edweek.org/teaching-learning/>.
- [10] T. G. Weldy y . M. L. Icenogle, «A Managerial Perspective: Oral Communication Competency Is Most Important for Business Students in the Workplace Jeanne D. Maes,» 16 Septiembre 2016. [En línea]. Available: <https://journals.sagepub.com/doi/10.1177/002194369703400104>.
- [11] X. Ochoa, F. Domínguez, B. Guamán, R. Maya, G. Falcones y J. Castells, *The RAP System: Automatic Feedback of Oral Presentation Skills Using Multimodal Analysis and Low-Cost Sensors*, New York: Proceedings of the 8th International Conference on Learning Analytics and Knowledge, 2018, pp. 360-364.
- [12] E. Duvall, A. Robbins, T. Graham y D. Scott, *Exploring Filler Words*, Provo: Brigham Young University, 2014, pp. 35-50.
- [13] M. Decastro, «Mindful Presenter,» 20 Mayo 2022. [En línea]. Available: <https://mindfulpresenter.com/the-problem-with-filler-words-when-presenting-and-how-to-control-your-use-of-them>. [Último acceso: Noviembre 2022].
- [14] M. Jiménez, «Uso excesivo de muletillas y comodines en las exposiciones científicas orales,» *MEDISAN*, vol. 19, nº 4, pp. 541-554, 2015.
- [15] X. Ochoa y F. Dominguez, «Controlled evaluation of a multimodal system to improve oral presentation skills in a real learning setting,» *British Journal of Educational Technology*, vol. 51, pp. 1615-1630, 2020.
- [16] F. Dominguez, X. Ochoa, D. Zambrano, K. Camacho y J. Castells, «Scaling and Adopting a Multimodal Learning,» *IEEE Transactions on Learning Technologies*, vol. 14, pp. 400-414, Junio 2021.
- [17] P. Romo, «Didáctica y evaluación de la comunicación oral,» *Anales de la Universidad Central del Ecuador*, vol. 1, nº 373, pp. 329-341, 25 Octubre 2015.
- [18] J. Rosales, B. Caparros, I. Molina y Alonso, *Habilidades sociales*, Primera ed., Madrid: McGraw-Hill Interamericana de España, 2013.
- [19] L. D. Flores Rivera, M. Morocho Amaguaya y M. P. Naula Barros, «Valoración del talento comunicacional en la formación integral del profesional técnico,» *Revista de*

*Investigación en Ciencias de la Educación, Horizontes*, vol. 6, nº 24, pp. 906-918, Junio 2022.

- [20] Universidad Ean, «¿Cuáles son las habilidades blandas más demandadas del mercado?: Ean Universidad,» 11 Junio 2021. [En línea]. Available: <https://universidadean.edu.co/noticias/cuales-son-las-habilidades-blandas-mas-demandadas-del-mercado>. [Último acceso: Noviembre 2022].
- [21] K. Kurihara, M. Goto, J. Ogata, Y. Matsusaka y T. Igarashi, «Presentation sensei: a presentation training system using speech and image processing,» *Proceedings of the 9th international conference on Multimodal interfaces*, pp. 358-365, 2007.
- [22] N. Anggrani, A. Kurniawan, L. Kesuma Wardhami y N. Hakiem, *Speech Recognition Application for the Speech*, vol. 16, Jakarta: Telkomnika, 2018.
- [23] R. Morais, «STT Fast, Lean and Ubiquitous: Coqui,» 5 Diciembre 2019. [En línea]. Available: <https://coqui.ai/blog/stt/deepspeech-0-6-speech-to-text-engine>. [Último acceso: Noviembre 2022].
- [24] X. Huang, F. Alleva, H.-W. Hon, M.-Y. Hwang, K.-F. Lee y R. Rosenfeld, *The SPHINX-II speech recognition system: an overview*, vol. 7, 1993, pp. 137-148.
- [25] A. Hannun, C. Case, J. Casper, B. Catanzaro, G. Diamos, E. Elsen, R. Prenger, S. Satheesh, S. Sengupta, A. Coates y A. Y. Ng, *Deep Speech: Scaling up end-to-end speech recognition*, arXiv, 2014.
- [26] M. Ravanelli, T. Parcollet, P. Plantinga, A. Rouhe, S. Cornell, L. Lugosch, C. Subakan, N. Dawalatabad, A. Heba, J. Zhong, J.-C. Chou, S.-L. Yeh, S.-W. Fu, C.-F. Liao, E. Rastorgueva y G, *SpeechBrain: A General-Purpose Speech Toolkit*, arXiv, 2021.
- [27] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey y I. Sutskever, *Robust Speech Recognition via Large-Scale Weak Supervision*, arXiv, Ed., 2022.
- [28] S. Hochreiter y J. Schmidhuber, *Long Short-Term Memory*, vol. 9, Neural Comput, 1997, pp. 1735-1780.
- [29] C. Mosquera, G. Nápoles y G. Van Houdt, «A review on the Long Short-Term Memory Model,» *Artificial Intelligence Review*, vol. 53, 1 Diciembre 2020.

- [30] L. Yao y Y. Guan, «An Improved LSTM Structure for Natural Language Processing,» *2018 IEEE International Conference of Safety Produce Informatization (IICSPI)*, pp. 565-569, 10 Diciembre 2018.
- [31] The MathWorks, Inc., «Classify Text Data Using Deep Learning,» 2022. [En línea]. Available: <https://la.mathworks.com/help/deeplearning/ug/classify-text-data-using-deep-learning.html>. [Último acceso: 25 Noviembre 2022].
- [32] W. Qader, M. Ammen y B. Ahmed, *An Overview of Bag of Words;Importance, Implementation, Applications, and Challenges*, 2019, pp. 200-204.
- [33] A. Merdler, «Deploy OpenAI Whisper as a Cloud Product: Lightning AI,» 8 Diciembre 2022. [En línea]. Available: <https://lightning.ai/pages/community/tutorial/deploy-openai-whisper/>. [Último acceso: 20 Enero 2023].
- [34] L. Johansson, «Part 1 RabbitMQ for beginners - What is RabbitMQ?: CloudAMPQ,» 23 Septiembre 2019. [En línea]. Available: <https://www.cloudamqp.com/blog/part1-rabbitmq-for-beginners-what-is-rabbitmq.html>. [Último acceso: Enero 2023].
- [35] A. Visus, «¿Para qué sirve Python? Razones para utilizar este lenguaje de programación: ESIC,» Octubre 2020. [En línea]. Available: <https://www.esic.edu/rethink/tecnologia/para-que-sirve-python>. [Último acceso: Enero 2023].

# APÉNDICES

# APÉNDICE A

## Encuesta de experiencia de usuario del módulo de RAP

**1. ¿Considera útil mostrar la cantidad de muletillas detectadas en una presentación?**

- Totalmente de acuerdo
- De acuerdo
- Ni de acuerdo ni en desacuerdo
- En desacuerdo
- Totalmente en desacuerdo

**2. En la escala del 1 al 5, ¿Considera que, en el gráfico previo, la cantidad de muletillas detectadas en una presentación se muestra de forma clara y entendible para el usuario?**

- 5
- 4
- 3
- 2
- 1

**3. ¿Realizaría algún cambio/adición en la nueva sección del reporte presentado?**

- Ninguno
- Orden de presentación de muletillas (ascendente o descendente)
- Incluir un criterio de calificación del uso de muletillas
- Otro:

**4. ¿Considera que brindar una recomendación (como la que se mostró en la imagen previamente) es significativa para reducir el uso de las muletillas en futuras exposiciones?**

- Sí
- No

**5. ¿Qué acciones tomarías para lograr reducir el uso de muletillas a futuro?**

- Hacer pausas de silencio
- Utilizar conectores
- Practicar las exposiciones orales antes de realizarlas
- Otro: