



Facultad de Ciencias de la Administración

**Carrera de Ingeniería de Sistemas y
Telemática**

**NOIREDECU911: METODOLOGÍA EN LA
TRANSCRIPCIÓN DE VOZ A TEXTO EN
AUDIOS DEL ECU 911**

**Trabajo de titulación previo a la obtención del
grado de Ingeniero en Sistemas y Telemática.**

Autor:

Angel Alberto Jiménez Sarango.

Director:

Ing. Marcos Patricio Orellana Cordero

Cuenca – Ecuador

Año

2023

DEDICATORIA

Este trabajo lo dedico a todas las personas que estuvieron presente durante mi tiempo en la universidad, a mi familia que es mi soporte para dar un paso más en mi vida, en especial a mi madre por ser mi fortaleza y mi ejemplo a seguir, quien a pesar de todas las dificultades estuvo ahí para apoyarme y animarme a seguir adelante y terminar con mis estudios.

A mi padre que estuvo ahí para ayudarme y brindarme su consejo.

A mi hermano que siempre estuvo presente cuando lo necesitaba y apoyarme a cumplir mi meta profesional.

A mis abuelitos que, aunque en la lejanía siempre me tuvieron presente en sus oraciones, en especial a mi abuelito Alberto que desde el cielo se que me cuida y apoya para que cumpla con todas mis metas, y sé que estaría orgulloso.

AGRADECIMIENTO

Agradezco al director del presente trabajo de titulación, Ing. Marcos Orellana, quien, con su conocimiento, guía y consejos, supo orientarme durante toda la elaboración de este proyecto.

Agradezco de igual manera, al grupo de investigación y desarrollo en informática LIDI, por todo su apoyo en la realización de este proyecto ya que sin ellos hubiera sido imposible la realización exitosa del mismo.

Agradezco a mi familia por todo el apoyo que me brindaron para cumplir con esta etapa de mi vida.

Agradezco también a todos los profesores que fueron parte de mi formación académica, por sus conocimientos y paciencia otorgada.

También mi sincero agradecimiento a mis amigos y compañeros, así como a toda la comunidad de la Universidad del Azuay que formaron parte de esta importante etapa de mi vida.

Índice de Contenidos

DEDICATORIA	i
AGRADECIMIENTO	ii
Índice de Contenidos	iii
Índice de Figuras	iv
Índice de tablas	v
Índice de Anexos	vi
Resumen:	vii
Abstract:	vii
1. Introducción	1
1.1 Objetivos	1
1.2 Marco teórico	1
2. Revisión de literatura	3
3. Métodos	4
3.1. Selección de los ASR	5
3.1.1. Revisión de los ASR candidatos	5
3.1.2. Generación de transcripciones	5
3.1.3. Evaluación de la transcripción	6
3.1.4 Análisis de los resultados	6
3.2. Selección de técnicas o metodologías de mejora del habla:	6
3.2.1. Selección de las técnicas candidatas	6
3.2.2. Procesamiento de los audios	6
3.2.3. Generación de transcripciones	6
3.2.4. Evaluación de la transcripción	6
3.2.5 Análisis de los resultados	6
3.3. Metodología de mejora del habla:	7
3.4. Evaluación de la metodología:	7
4. Resultados	7
5. Discusión	15
6. Conclusión	16
7. Referencias	17
8. Anexo	19

Índice de Figuras

Figura 1 Diagrama de la metodología diseñada	5
Figura 2 Diagrama de la sección de selección de los ASR	5
Figura 3 Diagrama de la sección de selección de las técnicas de mejora del habla	6
Figura 4 Metodología de mejora del habla	7
Figura 5 Resultados comparativa entre los ASR CMU-Sphinx, Google Speech-to-text y Amazon transcribe	7
Figura 6 Resultados promedios índice WER transcripciones para cada técnica de mejora del habla candidata.	8
Figura 7 Distribución de los audios al aplicar la técnica Log MMSE .	9
Figura 8 Distribución de los audios al aplicar la técnica Paso bajo.	9
Figura 9 Número de audios con cambios al aplicar la técnica Log MMSE.	10
Figura 10 Número de audios con cambios al aplicar la técnica Paso bajo.	10
Figura 11 Gráfico de frecuencia de las palabras no reconocidas por el ASR al aplicar las técnicas de mejora del habla.	11
Figura 12 Índices Silhouette para los audios con la técnica Log MMSE.	13
Figura 13 Índices Silhouette para los audios con la técnica Paso bajo.	13
Figura 14 Distribución de los clústeres en los audios con la técnica Log MMSE..	13
Figura 15 Distribución de los clústeres en los audios con la técnica Paso bajo.	14
Figura 16 Resultados de la aplicación del cuestionario TAM.	15

Índice de tablas

Tabla 1. Cantidad de audios resultantes de la prueba de las transcripciones	8
Tabla 2. Palabras con terminación -ito que no más se perdieron en las transcripciones.	11
Tabla 3. Palabras con terminación -ita que más se perdieron en las transcripciones.	12
Tabla 4. Sustantivos que más se perdieron en las transcripciones.	12
Tabla 5. Cuestionario para medir las variables con TAM	14
Tabla 6. Estadística de los resultados obtenidos del estudio TAM	14
Tabla 7. Estadística descriptiva para las variables basadas en la percepción del usuario	15
Tabla 8. Cadena de búsqueda automática	19
Tabla 9. Criterios de selección y exclusión de los ASR	22
Tabla 10. Criterios de selección y exclusión de las técnicas de mejora del habla	24

Índice de Anexos

Anexo 1 Protocolo de la revisión de literatura	17
Anexo 2: Criterios de selección y exclusión de los ASR	22
Anexo 3 Criterios de selección y exclusión de las técnicas y metodologías de mejora del habla.	24

Resumen:

Los sistemas de reconocimiento automático de voz (ASR) son herramientas que han ido ganando terreno en los últimos años debido a su aplicación y utilidad. Sin embargo, aún son susceptibles a la calidad de los audios que procesan. Una opción para mejorar la calidad es el procesamiento previo de los audios con técnicas denominadas como “mejora del habla”. El presente estudio se centra en desarrollar una metodología de mejora del habla en audios digitales del centro de comando y control ECU911, el mismo que disminuye el índice de error WER de las transcripciones generadas por los ASR. Los resultados de la prueba demostraron que el 40% los audios a los que se le aplicaron la técnica "paso bajo" mejoraron entre el 3% y 10%. Esto se debe a las características propias con las que cuentan los audios como la reverberación, pronunciación del hablante o el uso de regionalismos.

Palabras clave: filtro paso bajo, mejora del habla, modelo de aceptación tecnológica, reconocimiento automático de voz, tasa de error por palabra

Abstract:

Automatic speech recognition (ASR) systems are tools that have been gaining ground in recent years due to their application and usefulness. However, those are susceptible to the quality of the audio they process. An option to improve quality is the previous processing of the audios with "speech enhancement" techniques. This study focuses on developing a methodology for the improvement of speech in digital audios of the ECU911 command and control center, which decreases the WER error rate of the transcripts generated by the ASRs. The results showed that 40% of the audios to which the "low pass" technique was applied improved between 3% and 10%. This is due to the characteristics of the audios, such as reverberation, pronunciation of the speaker, or the use of regionalisms.

Keywords: automatic speech recognition, low pass filter, speech enhancement, technological acceptance model, word error rate



Este certificado se encuentra en el repositorio digital de la Universidad del Azuay, para verificar su autenticidad escanee el código QR

Este certificado consta de: 1 página