



**UNIVERSIDAD
DEL AZUAY**

Facultad de Ciencias de la Administración

**Carrera de Ingeniería en Ciencias de la
Computación**

**COMPARATIVA DE PRECISIÓN/TIEMPO
ENTRE MODELOS DE SPEECH-TO-TEXT
BASADO EN AUDIOS DE CENTROS DE
COMANDO Y CONTROL**

**Trabajo de titulación previo a la obtención del
grado de Ingeniera en Ciencias de la
Computación**

Autora

Jacqueline Estefanía Sangolqui Gualpa

Director

Juan Fernando Lima Sigua

Cuenca – Ecuador

2024

DEDICATORIA

Este trabajo de titulación está dedicado, en primer lugar, a Di-s quien es mi guía y mi fortaleza en cada momento de mi vida y durante cada etapa de mi formación académica, su amor incondicional, su sabiduría infinita y su presencia constante han iluminado cada paso de mi camino, permitiéndome superar los desafíos y crecer como persona y como profesional, sin su gracia no habría logrado.

A ti, mamá, mi ejemplo, mi mentora, su esfuerzo y dedicación, me ha enseñado desde mis primeros años el valor del sacrificio y la importancia de luchar por mis sueños, por ser la inspiración más pura y sincera, por enseñarme que el éxito se construye con disciplina y amor, por brindarme siempre el apoyo incondicional que necesitaba para seguir adelante.

A mi hermano, quien con su amor y apoyo constante me han brindado la seguridad para avanzar con confianza, por creer en mí incluso en momentos en que yo misma dudaba, por las palabras de ánimo, su amor ha sido el motor que me impulsó a llegar hasta aquí y seguir adelante en cada desafío.

AGRADECIMIENTO

Con profundo agradecimiento, celebro hoy mi titulación, un logro que no habría sido posible sin el amor y apoyo incondicional de Di-s, mi familia y la invaluable enseñanza y apoyo de mi tutor Juan Fernando Lima y profesores. Gracias a cada uno de ustedes, este logro se ha hecho realidad.

Índice de Contenidos

DEDICATORIA	i
AGRADECIMIENTO.....	ii
Índice de Contenidos.....	iii
Índice de Tablas	iv
Índice de Figuras	v
RESUMEN.....	vi
ABSTRACT	vi
1. Introducción	1
1.1 Objetivos.....	1
1.1.1 Objetivo General.....	1
1.1.2 Objetivos Específicos	2
2. Revisión de Literatura	2
2.1 Marco Teórico	2
2.2 Estado del Arte	4
3. Materiales y Métodos.....	7
3.1 Compresión del negocio	8
3.2 Compresión de los datos.....	9
3.2.1 Datos recolectados	9
3.2.2 Análisis descriptivo de los datos.....	10
3.3 Preparación de los Datos	13
3.4 Modelado.....	15
3.4.1 Whisper.....	16
3.4.2 Chirp	17
3.5 Evaluación	17
4. Resultados y Discusión	18
4.1 Medias de modelos Whisper con y sin parámetros	18
4.2 Duración de los audios categorizada en cuartiles.....	21
4.3 Decibelios a escala completa (dBFS) de los audios categorizada en cuartiles.....	23
4.4 Género del alertante y tipo de alerta de los audios.....	25
4.5 Diarización	27
4.6 Whisper vs Chirp.....	28
4.6.1 Medias de WER y CER, según la duración de los audios categorizada en cuartiles.....	29
4.6.1 Medias de WER y CER, según los dBFS de los audios categorizada en cuartiles.....	31
4.6.3 Medias de tiempos de respuestas	33
5. Conclusiones y Recomendaciones	33
6. Referencias bibliográficas.....	34

Índice de Tablas

Tabla 1	Columnas de la base de datos de llamadas de emergencia	10
Tabla 2	Estadísticas por cada tipo de alerta de emergencia	10
Tabla 3	Prueba de normalidad de D'Agostino y Pearson	11
Tabla 4	Características de los modelos Whisper	16
Tabla 5	Parámetros de Whisper.....	18
Tabla 6	Fortalezas y debilidades de cada modelo.....	33

Índice de Figuras

Figura 1	Meta-Modelo SPEM de la metodología CRISP DM.....	7
Figura 2	Modelo de gestión del ECU 911	9
Figura 3	Distribución de tiempos de duración por alerta.....	11
Figura 4	Tiempos de duración por Alerta.....	13
Figura 5	Meta-Modelo SPEM del modelado.....	15
Figura 6	Arquitectura Whisper.....	17
Figura 7	Análisis comparativo del WER en modelos Whisper ajustados y sin ajuste de parámetros.....	19
Figura 8	Análisis comparativo del CER en modelos Whisper ajustados y sin ajuste de parámetros.....	20
Figura 9	Análisis comparativo de tiempos de respuesta en modelos Whisper ajustados y sin ajuste de parámetros.....	21
Figura 10	WER categorizado por cuartiles del tiempo de duración del audio en segundos..	22
Figura 11	CER categorizado por cuartiles del tiempo de duración en segundos	23
Figura 12	WER categorizado por cuartiles de dBFS	24
Figura 13	CER categorizado por cuartiles de dBFS	25
Figura 14	Media de WER por género y categoría de alerta.....	26
Figura 15	Media de CER por género y categoría de alerta.....	27
Figura 16	Media de WER, según la duración de los audios categorizada en cuartiles	29
Figura 17	Media de CER, según la duración de los audios categorizada en cuartiles.....	30
Figura 18	Medias de WER, según los dBFS de los audios categorizada en cuartiles	31
Figura 19	Medias de CER, según los dBFS de los audios categorizada en cuartiles	32
Figura 20	Medias de tiempos de respuestas de cada modelo.....	33

COMPARATIVA DE PRECISIÓN/TIEMPO ENTRE MODELOS DE SPEECH-TO-TEXT BASADO EN AUDIOS DE CENTROS DE COMANDO Y CONTROL

RESUMEN

La investigación examina el desempeño de los modelos de reconocimiento de voz (speech-to-text) Whisper y Chirp en audios de emergencia del sistema ECU 911, analizando su precisión y velocidad en condiciones desafiantes, con profundos factores como la duración de los audios, niveles de decibelios, género de los alertantes, e incluso autore, el estudio revela los matices de cada modelo en situaciones reales a través de métricas de WER (Tasa de errores de palabra) y CER (Tasa de errores de caracteres), se compararon los modelos para descubrir cómo los parámetros avanzados de Whisper, como *best of* y *best size*, impulsan su rendimiento a nuevos niveles, los resultados son excelentes: el modelo Whisper LARGE optimizado con parámetros específicos y el modelo Chirp TELEPHONY alcanzaron un equilibrio entre precisión y velocidad, demostrando que una selección precisa de parámetros puede transformar los resultados de manera espectacular. Las pruebas, realizadas en un entorno que simula las condiciones operativas de emergencias con valores menores de WER de 0.2 y CER menores de 0.09, confirman que ambos modelos son altamente efectivos para mejorar la respuesta en emergencias, este estudio concluye que la combinación de precisión y velocidad en estos modelos de inteligencia artificial tiene un impacto real y positivo en la eficiencia de los sistemas de respuesta.

Palabras clave: Whisper Open IA, Chirp Google Cloud, speech-to-text, WER, CER, ECU 911.

ACCURACY/TIMING COMPARISON BETWEEN SPEECH-TO-TEXT MODELS BASED ON COMMAND AND CONTROL CENTRE AUDIOS

ABSTRACT

The research examines the performance of the Whisper and Chirp speech-to-text models on ECU 911 emergency audios, analyzing their accuracy and speed in challenging conditions, with profound factors such as audio duration, decibel levels, gender of alerters. The study reveals the nuances of each model in real-world situations through WER (Word Error Rate) and CER (Character Error Rate) metrics. The models were compared to discover how Whisper's advanced parameters, such as *best of* and *best size*, drive their performance to new levels, and the results are excellent: the Whisper LARGE model optimized with specific parameters and the Chirp TELEPHONY model achieved a balance between accuracy and speed, demonstrating that precise parameter selection can transform results dramatically. Tests, conducted in an environment that simulates emergency operating conditions with WER values less than 0.2 and CER less than 0.09, confirm that both models are highly effective in improving emergency response, this study concludes that the combination of accuracy and speed in these artificial intelligence models has a real and positive impact on the efficiency of response systems.

Keywords: Whisper Open IA, Chirp Google Cloud, speech-to-text, WER, CER, response time, ECU 911.