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 Guallpa

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

Índice de Tablas

Índice de Figuras

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.