

UNIVERSIDAD POLITÉCNICA SALESIANA
SEDE CUENCA

CARRERA DE COMPUTACIÓN

*Trabajo de titulación previo a la
obtención del título de Ingeniera en
Ciencias de la Computación e Ingeniero
en Ciencias de la Computación*

PROYECTO TÉCNICO:

**“DISEÑO Y DESARROLLO DE UN PLUGIN PROTOTIPO INTELIGENTE PARA
SOPORTE DE SUBTITULADO EN LÍNEA PARA EL SISTEMA DE
GRABACIÓN Y TRANSMISIÓN DE VÍDEO POR INTERNET OPEN
BROADCASTER SOFTWARE”**

AUTORES:

DOMÉNICA ALEJANDRA MERCHÁN GARCÍA
ALEJANDRO SEBASTIÁN ENRIQUEZ MANCHENO

TUTOR:

ING. VLADIMIR ESPARTACO ROBLES BYKBAEV, PhD.

CUENCA - ECUADOR

2021

CESIÓN DE DERECHOS DE AUTOR

Nosotros, Doménica Alejandra Merchán García con documento de identificación N° 0104526520 y Alejandro Sebastián Enriquez Mancheno con documento de identificación N° 0605167949, expresamos nuestra voluntad y por medio del presente documento cedemos a la Universidad Politécnica Salesiana la titularidad sobre los derechos patrimoniales en virtud de que somos autores del trabajo de titulación: **“DISEÑO Y DESARROLLO DE UN PLUGIN PROTOTIPO INTELIGENTE PARA SOPORTE DE SUBTITULADO EN LÍNEA PARA EL SISTEMA DE GRABACIÓN Y TRANSMISIÓN DE VÍDEO POR INTERNET OPEN BROADCASTER SOFTWARE”**, mismo que ha sido desarrollado para optar por el título de: *Ingeniera en Ciencias de la Computación e Ingeniero en Ciencias de la Computación*, en la Universidad Politécnica Salesiana, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente.

En aplicación a lo determinado en la Ley de Propiedad Intelectual, en nuestra condición de autores nos reservamos los derechos morales de la obra antes citada. En concordancia, suscribimos este documento en el momento que hacemos entrega del trabajo final en formato digital a la Biblioteca de la Universidad Politécnica Salesiana.

Cuenca, octubre de 2021.



Doménica Alejandra Merchán García
C.I. 0104526520



Alejandro Sebastián Enriquez Mancheno
C.I. 0605167949

CERTIFICACIÓN

Yo, declaro que bajo mi tutoría fue desarrollado el trabajo de titulación: **“DISEÑO Y DESARROLLO DE UN PLUGIN PROTOTIPO INTELIGENTE PARA SOPORTE DE SUBTITULADO EN LÍNEA PARA EL SISTEMA DE GRABACIÓN Y TRANSMISIÓN DE VÍDEO POR INTERNET OPEN BROADCASTER SOFTWARE”**, realizado por Doménica Alejandra Merchán García y Alejandro Sebastián Enriquez Mancheno, obteniendo el *Proyecto Técnico*, que cumple con todos los requisitos estipulados por la Universidad Politécnica Salesiana.

Cuenca, octubre de 2021.



Vladimir Espartaco Robles Bykbaev, PhD.

C.I. 0300991817

DECLARATORIA DE RESPONSABILIDAD

Nosotros, Doménica Alejandra Merchán García con documento de identificación N° 0104526520 y Alejandro Sebastián Enriquez Mancheno con documento de identificación N° 0605167949, autores del trabajo de titulación: **“DISEÑO Y DESARROLLO DE UN PLUGIN PROTOTIPO INTELIGENTE PARA SOPORTE DE SUBTITULADO EN LÍNEA PARA EL SISTEMA DE GRABACIÓN Y TRANSMISIÓN DE VÍDEO POR INTERNET OPEN BROADCASTER SOFTWARE”**, certificamos que el total contenido del *Proyecto Técnico*, es de nuestra exclusiva responsabilidad y autoría.

Cuenca, octubre de 2021.



Doménica Alejandra Merchán García
C.I. 0104526520



Alejandro Sebastián Enriquez Mancheno
C.I. 0605167949

DEDICATORIA Y AGRADECIMIENTO

DEDICATORIA Y AGRADECIMIENTO

Primero dar gracias a Dios por darme la sabiduría para atravesar todos los obstáculos, por darme salud y una familia amorosa que siempre supo apoyarme para seguir adelante y luchar por mis metas. A mis padres, quienes fueron mi apoyo incondicional durante todos mis años de estudio y mis guías para continuar con todos mis proyectos, por presionarme a buscar siempre mis sueños y ayudarme a salir adelante. A mi padre por ser siempre mi amigo y escucharme cuando lo necesitaba, y a mi madre quien compartió conmigo en cada momento, luchó conmigo mis batallas y me dio fuerza para continuar; ustedes y mi hermano son mi orgullo e inspiración, y este trabajo es gracias a ustedes y para ustedes.

Gracias a quienes hicieron posible este trabajo y este nuevo triunfo, a mis maestros por su sabiduría y experiencia, y a mis amigos a quienes nunca les faltó una palabra de aliento.

Doménica Alejandra Merchán García

Agradezco a mis padres por ser los principales promotores de mis sueños, gracias a ellos por cada día confiar en mí y creer en mis capacidades, gracias a mi madre por compartir mis preocupaciones, tristezas y alegrías, por demostrarme que, tras un mal día, todo iba a estar bien con una sonrisa; a mi padre por siempre desear lo mejor para mi vida, gracias por cada consejo y por cada una de sus palabras que me guiaron durante mi vida.

A mis maestros, por su incansable batalla por crear un futuro mejor, gracias por apoyarme a cumplir esta meta. A mis amigos y compañeros, que me acompañaron a lo largo de todo este camino y siempre mostraron una mano amiga ante cualquier dificultad.

Gracias a Dios por la bendición de una familia amorosa y buena, que siempre me ha apoyado a cumplir mis objetivos.

Gracias a la vida por este nuevo triunfo, y a todas las personas que apoyaron en la realización de este trabajo.

Alejandro Sebastián Enriquez Mancheno

Resumen

En el presente trabajo se realiza una investigación sobre las dificultades de las personas con discapacidad auditiva en la nueva metodología de aprendizaje virtual, así mismo las herramientas que pueden mitigar este inconveniente y nuestro aporte por realizar una solución prototipo para estos casos. Esta investigación se llevó a cabo a través de docentes de la Universidad Politécnica Salesiana de las sedes Cuenca, Quito y Guayaquil quienes han trabajado con estudiantes que presentan esta discapacidad. Se encontró que para un nivel de discapacidad auditiva medio o bajo el cambio a una modalidad virtual no afectó a su rendimiento académico, por el contrario, el poder controlar el volumen o usar audífonos ayudó con el problema de audición. Por otro lado, para estudiantes con un nivel de discapacidad auditiva alto el cambio de modalidad resultó en un gran obstáculo para continuar con sus estudios. Como parte de una solución se presenta un modelo entrenado con una red neuronal recurrente capaz de realizar el reconocimiento del habla para su posterior transcripción y presentación a modo de subtítulos, generando una aplicación sencilla de aplicar en las plataformas de video conferencia por medio del software OBS. Se ejecutó un plan de experimentación en el que se recoge información de los docentes que pudieron visualizar el funcionamiento de la herramienta, obteniendo así su retroalimentación. Finalmente se generó un objeto de aprendizaje en el que se expone información importante sobre la herramienta, así como un manual de instalación, implementación y de uso sobre plataformas de conferencias virtuales.

Abstract

In this work, an investigation is carried out about the difficulties of people with hearing impairment, in the new methodology of virtual learning, in addition, the tools that can help to improve this inconvenience and our contribution to develop a prototype solution for these cases. This investigation was done in conjunction with teachers of the three Universidad Politécnica Salesiana seats, in Quito, Cuenca, and Guayaquil, who had experience working with students with this disability. It was found that for middle and low levels of hearing impairment the virtual modality has not affected to Its academic performance, conversely, the ability to control the volume or use headphones in classes helped to improve the auditory problems. On the other hand, for students with a high level of auditory problems the virtual modality has turned into a big obstacle to continue with Its studies. As part of the solution, a speech recognition model is developed using recurrent neural networks, which is capable of understanding and transcribe the audio to a text, who is presented as subtitles using a virtual camera in OBS studio. An experimentation plan is executed in which information is collected from teachers who could have an approach to the subtitulation tool and have feedback about it. Finally, it was generated a learning object in which information of the subtitulation program is exposed, as a manual of installation, implementation, and use in virtual video conferences software.

Índice de Contenido

1. Introducción	1
2. Antecedentes	1
2.1 Accesibilidad para personas con discapacidad auditiva en el Ecuador.....	1
3. Alcance del proyecto.....	2
4. Trabajo Relacionado	2
5. Metodología	3
5.1. Fundamentos de accesibilidad para discapacidad auditiva en clases virtuales	3
5.2. Diseño y desarrollo de un plugin inteligente para reconocimiento del habla	4
5.2.1. Data Set	4
5.2.2. Parámetros de entrada	5
5.2.3. Paralelismo	5
5.3. Validación del modelo generado	6
6. Resultados	7
6.1. Accesibilidad para personas con discapacidad auditiva durante las clases virtuales.....	7
A. Mecanismos de apoyo para estudiantes con discapacidad auditiva	7
B. Herramientas tecnológicas de apoyo para estudiantes con discapacidad auditiva	8
6.2. Desarrollo del plugin inteligente para el reconocimiento del habla y subtitulado automático	8
6.3. Objeto de Aprendizaje.....	11
6.4. Plan de experimentación	12
7. Discusión.....	13
8. Conclusiones y trabajo futuro	13
9. Referencias.....	14

1. Introducción

El intercambio de información por medio de la web ha evolucionado desde la Web 1.0, una red de conexiones de información estática, a la Web 2.0 que permitía relacionar tecnologías y medios sociales para conseguir transacciones bidireccionales, y finalmente, a la actual Web 3.0 que implementa técnicas de automatización e integración entre aplicaciones, caracterizándose por su contenido dinámico lleno de material audiovisual y archivos multimedia (Choudhury, 2014). Es muy común que esta clase de contenido acompañe o incluso reemplace a la información presentada en texto plano, poniendo en riesgo la inserción de personas con discapacidades a estas nuevas metodologías de información (Shiver & Wolfe, 2015).

Para evitar el crecimiento de esta brecha de acceso a la información, se han creado normativas para la accesibilidad web; sin embargo, debido a la rápida migración a una modalidad virtual causada por pandemia del COVID-19, estas normativas no han sido completamente implementadas para los nuevos medios de comunicación por medio de plataformas de videoconferencia.

Desde la pandemia, todos los aspectos de la vida humana se han visto afectados, especialmente la educación (Stefanile, 2020). Las medidas de seguridad adoptadas obligaron el cierre de las instituciones educativas, afectando hasta la fecha a casi la mitad de los estudiantes en todo el mundo (UNESCO, n.d.). La incorporación de las plataformas de videoconferencia surgió como una solución ante la necesidad de interconectar nuevamente a docentes y estudiantes, dando paso a nuevos mecanismos de enseñanza, pero también a nuevas necesidades (Karampidis, Trigoni, Papadourakis, Christofaki, & Escudeiro, 2021).

En el Ecuador se conoce que 66.722 personas presentan un grado de discapacidad auditiva, quienes podrían necesitar de estos mecanismos de accesibilidad para plataformas de videoconferencia (CONADIS, 2021). A este grupo de personas se debería añadir aquellas que tienen una disminución de su capacidad auditiva debido a la edad que, aunque no consten en las estadísticas, podrían requerir de adaptaciones similares.

2. Antecedentes

2.1. Accesibilidad para personas con discapacidad auditiva en el Ecuador

En el Ecuador se han realizado esfuerzos en temas de accesibilidad para la comunidad sorda. El derecho de las personas con discapacidad para comunicarse por medio de formas alternativas se reconoció inicialmente en la Constitución Política de la República del Ecuador en 1998. En 2008 se aprobó una nueva constitución, en donde se formula nuevamente el derecho a usar formas alternativas de comunicación (Oviedo, Carrera, & Cabezas, 2015). En el año 2012, la FENASEC¹ con el apoyo de la Vicepresidencia de la República del Ecuador, el Ministerio de Educación y el USAID², publicó el Diccionario Oficial de Lengua de Señas Ecuatoriana. Finalmente, para 2017 se registraron un total de ochenta y cinco instituciones de

¹ Federación Nacional de Sordos de Ecuador.

² Agencia de los Estados Unidos para el Desarrollo Internacional.

educación especializada a nivel nacional que atienden a estudiantes con discapacidad auditiva y multidiscapacidad (Ministerio de Educación, 2020).

En el área de las TIC³ también se han implementado directrices de accesibilidad para el contenido web. En 2014 el Instituto Ecuatoriano de Normalización publicó en el Registro Oficial no. 171 la aprobación de la norma NTE INEN-ISO/IEC 40500 “Tecnología de la información – Directrices de accesibilidad para el contenido web del W3C (WCAG) 2.0 (ISO/IEC 40500:2012, IDT)” (Registro Oficial, 2014). Esta Norma Técnica Ecuatoriana es una traducción idéntica de la Norma Internacional ISO/IEC 40500:2012 Information technology - - W3C Web Content Accessibility Guidelines (WCAG) 2.0 publicado en 2008 (Luján, n.d.). Finalmente, en el Registro Oficial No. 686 de febrero del 2016 se publica el Reglamento Técnico Ecuatoriano RTE INEN 288 “ACCESIBILIDAD PARA EL CONTENIDO WEB” (Registro Oficial, 2016), que busca realizar el seguimiento del cumplimiento obligatorio de la norma NTE INEN-ISO/IEC 40500 para todos los sitios web ecuatorianos que presten un servicio público, con un nivel de conformidad A desde 2018, y un nivel de conformidad AA a partir del 2020 (INEN, 2016).

Dentro de este reglamento se incluyen normas de accesibilidad para contenido multimedia. En la Pauta 1.2 se establece que se deben ofrecer alternativas para medios basados en el tiempo, tales como audio o video. Los subtítulos son una alternativa para este tipo de contenido, y estos pueden ser pregrabados catalogándose con un nivel de conformidad A, o en tiempo real con un nivel de conformidad AA.

3. Alcance del proyecto

El objetivo de este trabajo es ser un punto de partida para la creación de un sistema de reconocimiento del habla y subtulado automático en español orientado a la academia. En esta investigación se presentan las recomendaciones para realizar un modelo por medio de aprendizaje por transferencia capaz de comprender el lenguaje técnico de una carrera en específico. El modelo presentado como base ha sido entrenado utilizando un corpus con 187.364 registros con un total de 324 horas de audio en español, distribuido de forma gratuita por Mozilla.

4. Trabajo Relacionado

Se realizó una investigación sobre las herramientas que ofrecen el servicio de subtulado automático. Las más destacadas se muestran en la Tabla 1:

³ Tecnologías de la Información y Comunicación.

Tabla 1: Herramientas para Reconocimiento del habla en español y subtulado automático.

Herramienta	Gratuita	Motor Propio	Open Source	Implementación en plataformas de Video Conferencia
Verbit	No	Si	No	Si
Google	No	Si	No	Si
WebCaptioner	Si	No, utiliza Google	No	Si
Amazon	No	Si	No	Si
IBM Watson	No	Si	No	Si
Speechnotes	Si	No, utiliza Google	No	No

Todas las herramientas revisadas ofrecen un servicio de reconocimiento del habla y subtulado automático en español; sin embargo, ninguna de ellas está orientada al lenguaje académico. La mayoría de estas herramientas requieren un pago para su uso, y aquellas que son gratuitas no son de código abierto o no permiten realizar un aprendizaje por transferencia⁴ sobre su modelo.

5. Metodología

El desarrollo de este proyecto se realizó en tres etapas: 1. Investigación sobre los fundamentos de accesibilidad para personas con discapacidad auditiva en plataformas de video conferencia durante las clases virtuales. 2. Diseño y desarrollo de un plugin que permita realizar reconocimiento del habla en español por medio de un modelo entrenado en una red neuronal. 3. En la tercera etapa se realiza la validación del modelo por medio de un análisis del nivel de precisión, y encuestas de percepción.

5.1. Fundamentos de accesibilidad para discapacidad auditiva en clases virtuales

En esta etapa se realizó un estudio sobre los principales problemas que tienen las personas con discapacidad auditiva al momento de utilizar plataformas de video conferencia durante las clases virtuales. Como muestra, se tomaron los casos de estudiantes de la Universidad Politécnica Salesiana que presentaban esta discapacidad. Al tratarse de información sensible, el contacto directo con los estudiantes no fue posible, por lo que se realizaron encuestas a los docentes y directores de las carreras en donde se encontraba matriculado por lo menos un estudiante con discapacidad auditiva. En total se conoce que en las sedes Cuenca, Quito y Guayaquil de la Universidad hay un total de 46 estudiantes con discapacidad auditiva matriculados en el período 58 (2021-2021) (Secretaría Técnica de Bienestar Estudiantil, 2021).

En las encuestas realizadas, se preguntó sobre el desempeño de los estudiantes con discapacidad auditiva durante las clases virtuales, si requerían de algún apoyo externo, si este apoyo lo utilizaban antes de la pandemia, y sobre qué mecanismos o herramientas han sido implementadas por los docentes para mejorar las clases virtuales y asegurar su accesibilidad.

⁴ El aprendizaje por transferencia es el proceso de volver a entrenar un modelo previamente entrenado, con un conjunto de datos relativamente pequeño al usado en el primer entrenamiento.

5.2. Diseño y desarrollo de un plugin inteligente para reconocimiento del habla

Para la generación del modelo base para el reconocimiento del habla y subtítulo automático se utilizó la Red Neuronal Recurrente (RNN) Deep Speech. Se escogió este tipo de red ya que presenta mejores resultados frente a sistemas de reconocimiento del habla tradicionales que implementan modelos ocultos de Markov, modelos acústicos o diccionarios de fonemas (Hannun, et al., 2014).

5.2.1. Data Set

Para el entrenamiento se realizó una investigación sobre data sets gratuitos que contengan grabaciones de audio de personas leyendo o hablando en español, y la transcripción de cada audio en texto. Los data sets más relevantes hallados en la investigación se muestran en la Tabla 2.

Tabla 2: Resumen de los Data Sets investigados para el entrenamiento del modelo base.

Data Set	Número de registros	Horas de audio grabado
Kaggle - Spanish Single Speaker Speech Dataset	3 273	24
OpenSLR - Crowdsourced high-quality Argentinian Spanish speech data set	5 739	n.d.
Ciempies - LibriVox Spanish Corpus	36 338	73
Mozilla - Common Voice Spanish	187 364	324

Se conoce que un mayor número de horas de grabación conseguirá un mejor entrenamiento de la red, por lo que se optó por utilizar el Data Set de Mozilla, Common Voice. Este data set se ha dividido en tres grupos:

Tabla 3: Grupos para entrenamiento, evaluación y pruebas.

Grupo	Número de archivos	Porcentaje
Train	158 157	84%
Dev	14 622	8%
Test	14 584	8%

El grupo Train es utilizado para el entrenamiento de la red neuronal para la construcción del modelo. La red neuronal recibe como entrada el espectrograma de la voz y obtiene como salida una letra del abecedario disponible. Así, el objetivo de la red neuronal es convertir un espectrograma en una secuencia de probabilidades de caracteres.

El grupo Dev es utilizado durante el entrenamiento para evaluar que la red no caiga en un *overfitting*, es decir, que se sobreentrene la red. Este grupo detiene el entrenamiento si detecta que la red se está sobre ajustando.

El grupo Test se utilizó para validar el error de las predicciones obtenidas por el modelo entrenado en la red neuronal. Las medidas de validación se explican en la Sección 5.3.

5.2.2. Parámetros de entrada

Para el entrenamiento del modelo es necesario definir algunos parámetros de entrada. Primero se debe definir el abecedario que se va a utilizar. En la Tabla 4 se muestra la lista de parámetros usados para el entrenamiento del modelo.

Tabla 4: Parámetros utilizados para el entrenamiento del modelo.

Parámetro	Definición
load_cudnn	Sirve para cargar funcionalidades extra en GPU
automatic_mixed_precision	Utilizando el entrenamiento con GPU hace que se use precisión mixta (FP32 y FP16) según corresponda
alphabet_config_path	Es la ruta al archivo en el que se encuentran los caracteres que se necesitan para el texto del idioma
checkpoint_dir	Son puntos de control que se almacenan en el directorio elegido, sirven para validar el mejor modelo y no solo el último
train_file	Archivo en formato '.csv' que tiene las rutas a los audios y su respectiva transcripción que se utilizará para el entrenamiento
dev_files	Archivo en formato '.csv' que tiene las rutas a los audios y su respectiva transcripción que se utilizará para la validación
test_files	Archivo en formato '.csv' que tiene las rutas a los audios y su respectiva transcripción que se utilizará para las pruebas
train_batch_size	Define el número de muestras que serán tomadas para cada iteración del entrenamiento, así no toma toda la muestra sino, una porción al azar de esta.
n_hidden	Es la cantidad de capas que se utilizarán para la creación de la red neuronal
epochs	El número de iteraciones que se realizarán en la creación de la red neuronal.

5.2.3. Paralelismo

Ya que para el entrenamiento se realizará un procesamiento de grandes volúmenes de datos, es preferible utilizar una GPU. Se recomienda utilizar un ordenador con tarjeta NVIDIA de por lo menos 8GB de VRAM. En caso de que el entrenamiento se hiciera únicamente sobre la CPU el entrenamiento tomaría mucho más tiempo en completarse.

El computador en el que se realizó el entrenamiento de la red neuronal posee las siguientes características:

Tabla 5: Características del computador utilizado durante el entrenamiento de la red.

Sistema Operativo	Ubuntu 18.04.5 LTS
Máquina	ASUS TUF GAMING X570-PLUS
CPU	8 core AMD Ryzen 7 3700X
GPU	Nvidia GeForce RTX 2060 SUPER/PCIe/SSE2
Nvidia Drivers	4.6.0 NVIDIA 470.57.02
Audio	Advanced Micro Devices [AMD]

Para el uso de la GPU se instaló CUDA 11.2, que permite ejecutar el entrenamiento en cómputo paralelo.

5.3. Validación del modelo generado

Los parámetros utilizados para la comprobación de la eficacia del modelo son WER (Word Error Rate), CER (Character Error Rate) y Loss.

WER: Representa la tasa de error por palabra, donde se analiza la cantidad de inserciones, modificaciones y cambios en el texto original para transformarla al resultado deseado, dependiendo de la diferencia que hay entre las palabras tendrá un valor u otro. Esta medición es sobre todo válida cuando se realiza la implementación de un scorer. Aplica la siguiente fórmula

$$WER = \frac{sust + elim + ins}{palab}$$

donde

sust = cantidad de cambios realizados

elim = cantidad de eliminados

ins = cantidad de palabras insertadas

palab = cantidad de palabras en el texto

CER: Representa la tasa de error por caracter. Trabaja sustituyendo, eliminando e insertando caracter por caracter en el resultado y realizando una división para la cantidad de caracteres existentes.

$$CER = \frac{(sust + elim + ins)}{carac}$$

donde

sust = cantidad de cambios realizados

elim = cantidad de eliminados

ins = cantidad de caracteres insertados

carac = cantidad total de caracteres

Loss: es el cálculo del error de la predicción resultante de la red neuronal. Mientras más bajo sea este valor significa que las predicciones que realiza son más precisas.

6. Resultados

6.1. Accesibilidad para personas con discapacidad auditiva durante las clases virtuales

De acuerdo con el Ministerio de Educación, las instituciones de educación deberán garantizar el derecho a la educación a los habitantes del territorio ecuatoriano y su acceso universal, elaborando y ejecutando las adaptaciones necesarias para la inclusión y permanencia dentro del sistema educativo de personas con discapacidades (Espinoza, 2013). Es decir, las instituciones educativas deben proveer medios para asegurar que estudiantes con discapacidades puedan acceder de forma exitosa a los contenidos ofertados.

La Universidad Politécnica Salesiana ha realizado esfuerzos para mejorar la calidad de la educación para personas con discapacidades. Los estudiantes con discapacidad pueden acudir a docentes o al departamento de Bienestar Universitario para solicitar apoyo durante las clases. Para casos de estudiantes con discapacidad auditiva o problemas de audición este apoyo consiste en la asignación de tutores, la creación de material escrito disponible en la plataforma virtual AVAC, o asignar un intérprete que sirva de mediador entre el estudiante y el docente. Estas soluciones funcionarían para mantener un nivel adecuado de aprendizaje; sin embargo, desde la pandemia y el traspaso de las clases a una modalidad virtual ha causado un impacto negativo en algunos estudiantes. Si bien para estudiantes con un porcentaje de discapacidad medio o bajo, la modalidad virtual no ha representado un mayor problema, para otros estudiantes con un alto nivel de discapacidad auditiva esta nueva modalidad ha sido un obstáculo más para continuar con sus estudios.

Se realizó una investigación donde participaron docentes y directores de las carreras donde está matriculado por lo menos un estudiante con discapacidad auditiva. Del total de docentes entrevistados, el 30% manifestó que sus estudiantes no necesitaban ningún apoyo extra para las clases, puesto que pueden comprender el contenido dictado por el docente subiendo el volumen o mediante el uso de audífonos. El 70% restante indicó que desde el inicio de las clases virtuales los estudiantes han tenido problemas para comprender las materias, especialmente las técnicas. Para contrarrestar estos problemas se realizaron capacitaciones a los docentes de la Universidad para brindar apoyo mediante diferentes mecanismos y herramientas:

A. Mecanismos de apoyo para estudiantes con discapacidad auditiva

Algunos de los mecanismos implementados en la Universidad Politécnica Salesiana son:

- **Generar contenido accesible:** Los docentes reciben capacitaciones sobre cómo generar material accesible para las clases virtuales. Este material debería evitar el uso de oraciones largas y complejas. En su lugar se recomienda usar términos simples y resúmenes para explicar el contenido.

- **Texto alternativo para archivos multimedia:** El uso de archivos multimedia es muy común dentro de las clases; sin embargo, requieren de ajustes para que estos sean accesibles. Para los casos de discapacidad auditiva los docentes deberán proveer de texto explicativo o subtítulos para contenido basado en el tiempo como archivos audio y video.
- **Tutorías:** Los estudiantes que requieran apoyo extra para las clases pueden solicitar un tutor para las materias que requieran una explicación más detallada. En la Universidad existen varios proyectos de tutorías realizadas por docentes o por grupos de estudiantes. Esta clase de soporte, entre estudiantes, es una iniciativa implementada en las tres sedes, que busca generar redes de estudiantes que apoyen a otros en temas que se les dificulta durante las clases.

B. Herramientas tecnológicas de apoyo para estudiantes con discapacidad auditiva

De acuerdo con las encuestas realizadas a los docentes, las herramientas tecnológicas que se han implementado como apoyo para estudiantes con discapacidad auditiva durante las clases presenciales y virtuales son:

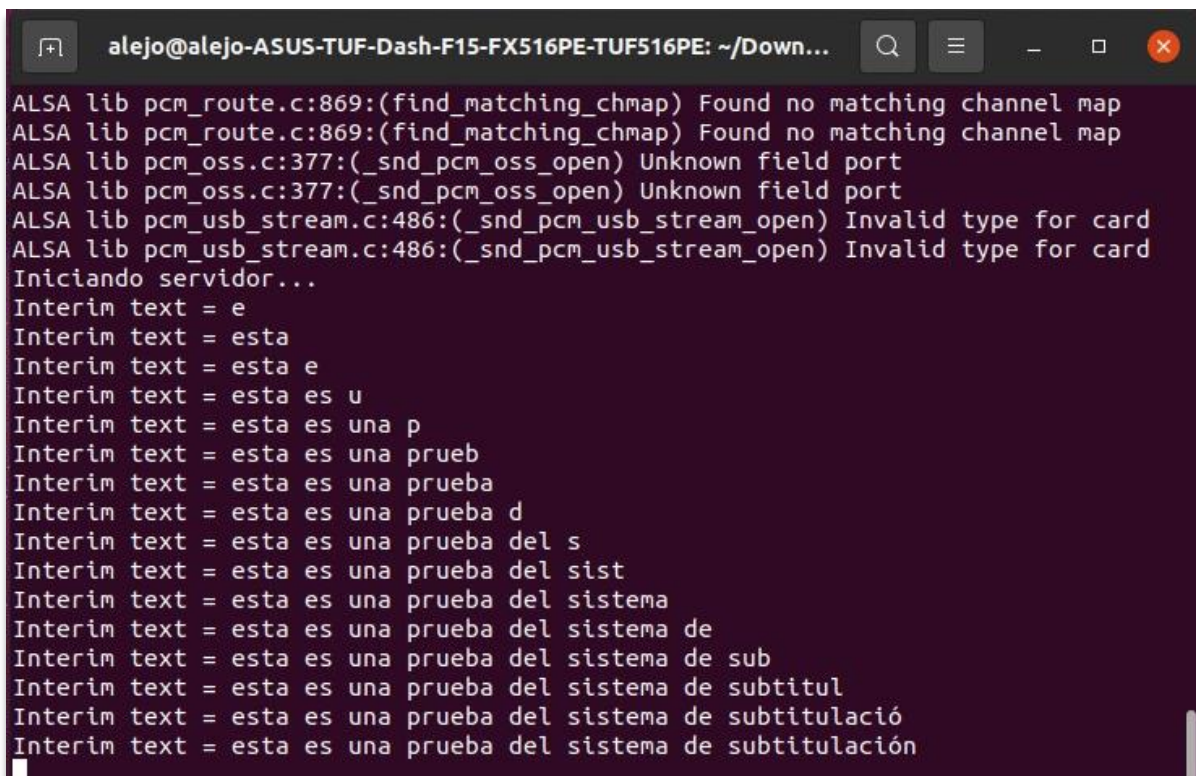
- **Apps para dispositivos móviles:** Durante las clases presenciales algunos estudiantes hacían uso de aplicaciones gratuitas para dispositivos móviles (Android) que les permitía realizar subtulado automático. De acuerdo con las encuestas, estas herramientas no tienen un buen desempeño durante las clases virtuales, puesto que el audio del computador llega con ruido y la frecuencia de la voz cambia, provocando que el reconocimiento no sea adecuado.
- **Subtitulado manual:** La Universidad Politécnica Salesiana hace uso de la herramienta Zoom para dictar las clases de forma virtual. Esta herramienta permite agregar subtítulos de forma manual en donde el docente debe escribir los subtítulos al mismo tiempo que dicta la clase. Esta forma de subtulado también permite que otro participante realice la transcripción de los subtítulos durante la clase.
- **Herramientas de subtulado automático:** Existen varias herramientas que permiten realizar reconocimiento del habla para subtulado automático. La mayoría de estas herramientas son privativas o no están disponibles para el español. Las herramientas que sí reconocen el español y son gratuitas no dan información sobre el entrenamiento realizado, ni dejan disponible el modelo implementado para realizar un aprendizaje por transferencia.

6.2. Desarrollo del plugin inteligente para el reconocimiento del habla y subtulado automático

Posterior a la revisión del funcionamiento práctico de la herramienta OBS Studio, se desarrolló un algoritmo en JavaScript que permite realizar la conexión al servidor de transcripción de audio a texto y capture los resultados para ser presentados mediante la herramienta OBS. Este plugin hace de cliente para establecer una conexión con el servidor, que

para esta implementación se encuentran de manera local (cliente y servidor en el mismo equipo).

La implementación del servidor se realizó mediante la herramienta Deepspeech, que, utilizando el modelo base entrenado en esta investigación, sirve de motor de reconocimiento de voz, recibiendo una entrada de audio y devolviendo un texto en el idioma español. Esta implementación se desarrolló utilizando un script en el lenguaje de programación de Python, el cual abre una conexión por WebSockets que permite a la máquina recibir y enviar paquetes de información a los clientes que se conecten. Posteriormente, en el mismo script, se establecen los parámetros necesarios para el motor de reconocimiento de voz, así como la dirección local del modelo que se va a utilizar y la fuente de audio que debe recibir. El funcionamiento está adaptado para que devuelva los resultados conforme va procesando, permitiendo que, al recibir un audio, éste sea procesado y devuelva el texto resultante en tiempo real (de acuerdo con la normativa de accesibilidad web). El resultado es transmitido al cliente que estableció previamente una conexión, y el plugin se encarga de ordenar y mostrar estos datos en un archivo local que posteriormente es utilizado para el subtulado con la herramienta OBS.



```
alejo@alejo-ASUS-TUF-Dash-F15-FX516PE-TUF516PE: ~/Down...
ALSA lib pcm_route.c:869:(find_matching_chmap) Found no matching channel map
ALSA lib pcm_route.c:869:(find_matching_chmap) Found no matching channel map
ALSA lib pcm_oss.c:377:(_snd_pcm_oss_open) Unknown field port
ALSA lib pcm_oss.c:377:(_snd_pcm_oss_open) Unknown field port
ALSA lib pcm_usb_stream.c:486:(_snd_pcm_usb_stream_open) Invalid type for card
ALSA lib pcm_usb_stream.c:486:(_snd_pcm_usb_stream_open) Invalid type for card
Iniciando servidor...
Interim text = e
Interim text = esta
Interim text = esta e
Interim text = esta es u
Interim text = esta es una p
Interim text = esta es una prueb
Interim text = esta es una prueba
Interim text = esta es una prueba d
Interim text = esta es una prueba del s
Interim text = esta es una prueba del sist
Interim text = esta es una prueba del sistema
Interim text = esta es una prueba del sistema de
Interim text = esta es una prueba del sistema de sub
Interim text = esta es una prueba del sistema de subtitul
Interim text = esta es una prueba del sistema de subtitulació
Interim text = esta es una prueba del sistema de subtítulo
```

Ilustración 1: Interfaz de la consola del servidor



Ilustración 2: Navegador con el plugin implementado

Utilizando la herramienta OBS se exportó una escena con los componentes necesarios precargados y organizados en el orden necesario para el subtitulado, estos siendo principalmente el acceso a la cámara del equipo, el contenedor para el plugin y una captura de pantalla. Al momento de cargar esta escena en la herramienta OBS, se debe cargar el archivo resultado del plugin en el componente predefinido, este archivo contiene el texto recopilado del servidor con la transcripción del audio a texto. Cuando se carga este archivo y se actualiza el componente el texto se empieza a mostrar en pantalla, haciendo un desplazamiento del texto cuando se llene el espacio del componente.

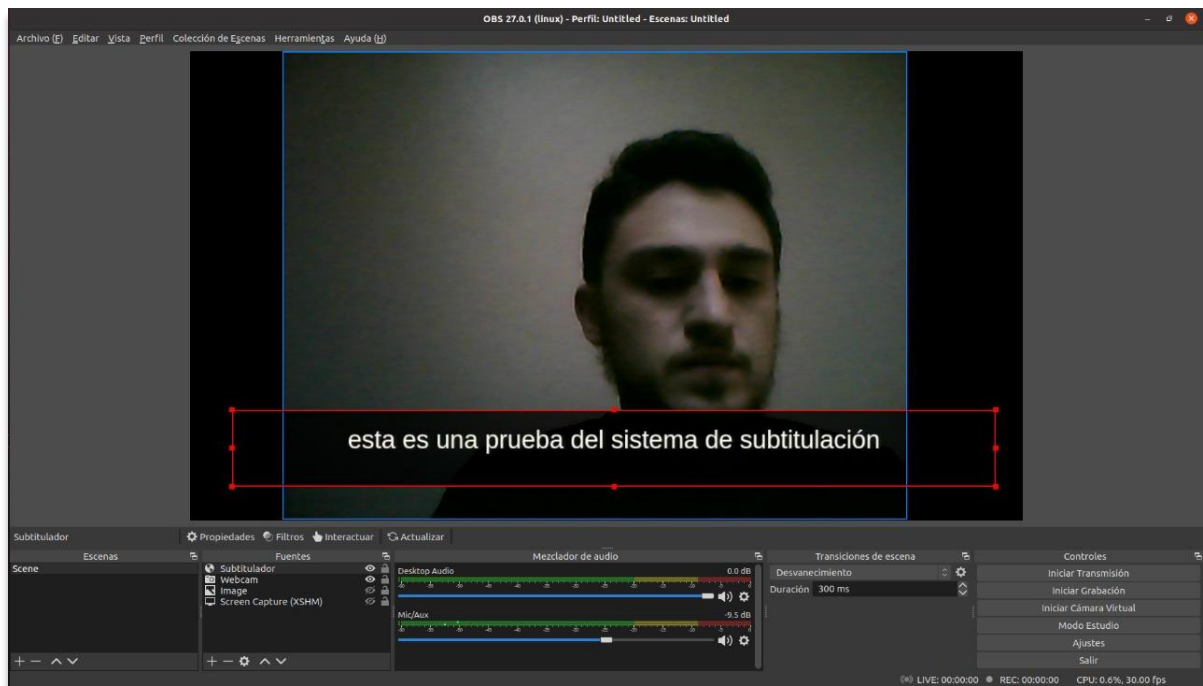


Ilustración 3: Escena generada en OBS

La cámara virtual es implementada con la herramienta OBS, que permite utilizar la escena definida anteriormente como una cámara montada en nuestro computador. A partir de la creación de esta cámara virtual, se puede utilizar en plataformas de videoconferencia únicamente cambiando en la configuración la cámara que se desea que muestre.



Ilustración 4: Implementación de cámara virtual en Zoom

6.3. Objeto de Aprendizaje

Utilizando la herramienta eXeLearning⁵ se generó un objeto de aprendizaje que tiene como objetivo ilustrar el proceso a realizar para aplicar la herramienta de subtítulo y cómo se debe realizar su aplicación en el programa OBS, así como los pasos a seguir para la activación de la cámara virtual y su implementación en la plataforma de videoconferencia Zoom. Se realizó la carga de este objeto de aprendizaje a la nube por medio de Github.

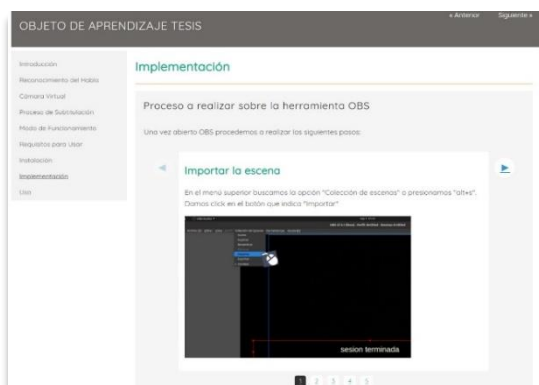


Ilustración 5: Objeto de Aprendizaje A



Ilustración 6: Objeto de Aprendizaje B

⁵ eXeLearning es una herramienta de código abierto que permite generar contenido educativo y presentarlo en formato web.

6.4. Plan de experimentación

El nivel de precisión del reconocimiento realizado por nuestro modelo se midió mediante el Word Error Rate, Character Error Rate y Loss. En la Tabla 6 se muestran los resultados obtenidos.

Tabla 6: WER, CER y Loss obtenido durante el testing del modelo generado.

Texto Original	Texto Detectado	WER	CER	Loss
“el formato se basó en una serie similar que se hizo en israel”	“el formato se basó en una serie similar que se hizo en israel ”	0.00	0.016	8.129
“los trabajos de reconstrucción del monumento duraron dos años”	“los trabajos de reconstrucción del monumento duraron dos años ”	0.00	0.016	7.872
“tuvieron un hijo en común”	“puvieron un ijor común”	0.60	0.20	19.54
“tienen ascendencia irlandesa y siria”	“tiene asendencia landesa y siria”	0.60	0.114	19,41
“fortalecimiento institucional”	“fortele imentes de tus enan”	2.50	0.482	59.62
“república de irak”	“publica del gacefue río y en intraconja carande”	2.666	2.117	107.2

Mediante la utilización de un formulario en línea se realizó la implementación del plan de experimentación, en la que se tuvo la participación de docentes de la universidad, los cuales pudieron observar el funcionamiento del sistema y opinar al respecto, dejando como resultado las siguientes conclusiones:

- El 100% de los docentes entrevistados consideran que la Universidad Politécnica Salesiana debe implementar herramientas tecnológicas de accesibilidad para personas con discapacidad.
- Las herramientas de detección del habla y subtitulado automático pueden mejorar la enseñanza durante las clases virtuales ya que ayuda en la comprensión del material, y sirve como herramienta de accesibilidad para personas con discapacidad auditiva o problemas en la audición.
- El sistema de subtitulado propuesto podría implementarse durante las clases virtuales, después de mejorar su precisión, ortografía y predicción de palabras.
- De acuerdo con la percepción de los entrevistados, el modelo realiza un subtitulado con una precisión del 72.5%.
- Es necesario que el modelo de reconocimiento del habla y subtitulado automático utilizado durante las clases virtuales sea capaz de comprender acentos y tecnicismos de las asignaturas, así como comprender el idioma español de manera fluida.

7. Discusión

El comportamiento del modelo depende de la cantidad de datos con el que fue entrenado. Para lograr una comprensión fluida la documentación sugiere realizar un entrenamiento con 5mil horas de audio o más. Puesto a que el modelo propuesto tiene un entrenamiento con únicamente 324 horas de audio, ya que no se encontraron data sets con un mayor número de horas gratuito y en español, se esperaba que su funcionamiento no sea el más fluido o acertado.

Para mejorar la precisión también se recomienda implementar un modelo de lenguaje que debe ser generado en base al ambiente en el que el sistema de reconocimiento será utilizado. Es decir, si el modelo será utilizado en un ambiente académico, este modelo del lenguaje deberá ser generado a partir de frases comunes del ámbito académico. Este modelo ayuda con la predicción de palabras al momento del reconocimiento. Por ejemplo, si decimos “una red” para un lenguaje natural las palabras que tengan más probabilidad de aparecer serían “una red de pesca”; sin embargo, si se trata de un lenguaje académico en el área de la computación lo más probable es que lo siguiente en la frase sea “una red neuronal”. Para generar este modelo es necesario utilizar las frases necesarias para generar un vocabulario con TODAS las palabras que se esperaría sean utilizadas en una conversación. Ya que la cantidad de frases necesarias para generar un modelo de lenguaje válido es muy alta, no pudo ser implementado en este prototipo. Para tener una idea, un vocabulario de 495 000 palabras fue generado con 200 millones de frase.

8. Conclusiones y trabajo futuro

La búsqueda de la accesibilidad dentro de la educación es una tarea de gran esfuerzo y trabajo constante que deberá continuarse indefinidamente a medida que las tecnologías de comunicación e información sigan avanzando. El aprovechar las tecnologías disponibles para disminuir la brecha al acceso a una educación de calidad para personas con discapacidades debe ser un foco de atención para todas las instituciones educativas.

Este trabajo es una base de investigación y desarrollo de una herramienta que puede mejorar la calidad de la educación para cientos de estudiantes. Aunque existan otras herramientas similares, el enfoque hacia la educación es lo que le da un valor agregado para continuar con este proyecto y mejorarlo hasta conseguir un ambiente educativo accesible para todos.

En este informe se presenta una discusión sobre cómo mejorar el modelo planteado. En trabajos futuros se recomienda realizar un aprendizaje por transferencia utilizando audios tomados de clases virtuales, videos académicos explicativos, etc. El data set para este nuevo entrenamiento no requiere ser tan grande, puesto a que se parte de un modelo que ya comprende el español; aunque como se explicó en secciones anteriores mientras más horas de entrenamiento tenga el modelo su fluidez será mejor. Se recomienda también generar un data set de frases comunes utilizadas en la academia. Lo ideal será generar un conjunto de datos propio para cada área de conocimiento dentro de la academia (ej. Ciencias de la computación, eléctrica y electrónica, medicina, etc.).

9. Referencias

- Choudhury, N. (2014). World Wide Web and Its Journey from Web 1.0 to Web 4.0. *International Journal of Computer Science and Information Technologies*, 5(6), 8096-8100. Retrieved 2021, from <https://ijcsit.com/docs/Volume%205/vol5issue06/ijcsit20140506265.pdf>
- CONADIS. (2021). *Estadísticas de Discapacidad*. Retrieved from Consejo Nacional para la Igualdad de Discapacidades: <https://www.consejodiscapacidades.gob.ec/estadisticas-de-discapacidad/>
- Espinoza, A. (2013, 8 15). *Acuerdo No 295-13*. Retrieved 2021, from Ministerio de Educación: https://educacion.gob.ec/wp-content/uploads/downloads/2013/08/ACUERDO_295-13.pdf
- Hannun, A., Case, C., Casper, J., Catanzaro, B., Diamos, G., Elsen, E., . . . others. (2014). Deep speech: Scaling up end-to-end speech recognition. *arXiv preprint arXiv:1412.5567*.
- INEN. (2016, 01 18). *Servicio Ecuatoriano de Normalización*. Retrieved from RESOLUCIÓN No. 16 008: <https://www.normalizacion.gob.ec/buzon/reglamentos/RTE-288.pdf>
- Karampidis, K., Trigoni, A., Papadourakis, G., Christofaki, M., & Escudeiro, N. (2021). Difficulties and Disparities to Distance Learning During Covid-19 Period for Deaf Students--A Proposed Method to Eradicate Inequalities. *International Conference on Intelligent Tutoring Systems* (pp. 3-7). Springer. doi:10.1007/978-3-030-80421-3_1
- Lau, R. W., Yen, N. Y., Li, F., & Wah, B. (2014). Recent development in multimedia e-learning technologies. *World Wide Web*, 17(2), 189-198. doi:10.1007/s11280-013-0206-8
- Lawson, T., Comber, C., Gage, J., & Cullum-Hanshaw, A. (2010). Images of the future for education? Videoconferencing: A literature review. *Technology, pedagogy and education*, 19(3), 295-314. doi:10.1080/1475939X.2010.513761
- Luján, S. (n.d.). *Accesibilidad Web*. Retrieved 2021, from Ecuador: <http://accesibilidadweb.dlsi.ua.es/?menu=ecuador>
- Ministerio de Educación. (2020). Retrieved 2021, from MODELO EDUCATIVO NACIONAL BILINGÜE BICULTURAL PARA PERSONAS CON DISCAPACIDAD AUDITIVA: <https://educacion.gob.ec/wp-content/uploads/downloads/2020/02/Modelo-Educativo-Bilingue-Bicultural-para-Personas-con-Discapacidad-Auditiva.pdf>
- Oviedo, A., Carrera, X., & Cabezas, R. (2015). *Ecuador, atlas sordo*. Retrieved 2021, from Cultura Sorda: <https://cultura-sorda.org/ecuador-atlas-sordo/>
- Registro Oficial. (2014). Retrieved 2021, from Regostro Oficial No 171: <https://lexis.ueb.edu.ec/WebTools/LexisFinder/ImageVisualizer/ImageVisualizer.aspx?id=0F3102E51180FCE46F650E34F7289BD893097DD9&type=RO&pagenum=17>

- Registro Oficial. (2016). Retrieved 2021, from Registro Oficial No 686:
<https://www.derechoecuador.com/registro-oficial/2016/02/registro-oficial-no-686---miercoles-10-de-febrero-de-2016>
- Ruiz, B., & Sánchez, J. P. (2017). Al servicio de la inserción social de las personas con discapacidad sensorial: el Centro Español del Subtitulado y la Audiodescripción. *Panorama Social*, 143-158. Retrieved from https://www.funcas.es/wp-content/uploads/Migracion/Articulos/FUNCAS_PS/026art12.pdf
- Secretaría Técnica de Bienestar Estudiantil. (2021). *Listado de estudiantes con discapacidad auditiva*. Cuenca: Universidad Politécnica Salesiana.
- Shiver, B., & Wolfe, R. (2015). Evaluating Alternatives for Better Deaf Accessibility to Selected Web-Based Multimedia. *The 17th International ACM SIGACCESS Conference*, (pp. 231-238). doi:10.1145/2700648.2809857
- Stefanile, A. (2020). The transition from classroom to Zoom and how it has changed education. *Journal of social science research*, 33-40. doi:10.24297/jssr.v16i.8789
- UNESCO. (n.d.). *UNESCO*. Retrieved 2021, from Education: From disruption to recovery: <https://en.unesco.org/covid19/educationresponse>