

**PONTIFICIA UNIVERSIDAD  
CATÓLICA DEL PERÚ**

**Escuela de Posgrado**



**Construcción de recursos para la detección y  
clasificación automática de disfluencias producidas  
por tartamudez en Español**

Tesis para obtener el grado académico de Maestro en Informática  
con mención en Ciencias de la Computación que presenta:

*Daniel Alonso Cabrera Díaz*

**Asesor:**

*Marco Antonio Sobrevilla Cabezudo*

Lima, 2024


## Informe de Similitud

Yo, **Marco Antonio SOBREVILLA CABEZUDO**, docente de la Escuela de Posgrado de la Pontificia Universidad Católica del Perú, asesor de la tesis titulada “Construcción de recursos para la detección y clasificación automática de disfluencias producidas por tartamudez en Español” de el autor **Daniel Alonso Cabrera Díaz**, dejo constancia de lo siguiente:

- El mencionado documento tiene un índice de puntuación de similitud de 08%. Así lo consigna el reporte de similitud emitido por el software *Turnitin* el 22/07/2024.
- He revisado con detalle dicho reporte y la tesis, y no se advierte indicios de plagio.
- Las citas a otros autores y sus respectivas referencias cumplen con las pautas académicas.

Lugar y fecha:

San Miguel, 31 de Julio de 2024.

Apellidos y nombres del asesor: <b>SOBREVILLA CABEZUDO, Marco Antonio</b>	
DNI: <b>46299018</b>	Firma 
ORCID: <b>0000-0001-7625-9914</b>	

## **Dedicatoria**

A Sophie, mi amada esposa, por haberme dado su amor, comprensión y apoyo incondicional día y noche en todo este camino. A mis padres, quienes me enseñaron que la vida está repleta de retos superables y que es esencial levantarse tras cada caída.



## **Agradecimientos**

A mi asesor, el Dr. Marco Sobrevilla, por su valioso apoyo constante, paciencia y guía a lo largo de todo el proceso de esta investigación. A la Asociación Peruana de Tartamudez y al Centro Especializado en Tartamudez por su ayuda en el entendimiento de la correcta anotación de eventos de tartamudez. A los profesores Arturo Oncevay e Ian Paul Brossard por sus comentarios y sugerencias que ayudaron a mejorar el presente trabajo.



## Resumen

Esta tesis abordó el desarrollo de recursos computacionales para la detección y clasificación de disfluencias de tartamudez en español, cubriendo desde la recolección y anotación de audios hasta la implementación de un modelo de aprendizaje automático y estrategias de aumento de datos. Se recolectaron audios en español de cinco participantes con tartamudez, conformes a los estándares del dataset SEP-28K y con apoyo de dos especialistas en tartamudez. Aunque la naturaleza controlada de las grabaciones limitó la diversidad de disfluencias observadas, estos audios proporcionaron una base sólida para el desarrollo del modelo. El modelo presentado se basó en el modelo DisfluencyNet. Este modelo fue pre entrenado utilizando wav2vec 2.0 XLSR53 aprovechando su robusta base de datos multilingüe. El modelo demostró su capacidad para identificar y clasificar disfluencias en español, aunque su rendimiento fue inferior comparado con modelos equivalentes en inglés. Esta diferencia subraya la necesidad de más datos. Para mejorar la detección de disfluencias, se implementaron dos estrategias de aumento de datos. La primera incluyó variaciones de pitch, adición de reverberación y ruido blanco, duplicando efectivamente la cantidad de datos disponibles. Aunque esta estrategia mejoró el recall en ciertas disfluencias, los resultados en precisión y F1 fueron mixtos. La segunda estrategia, mediante clonación de voz con el modelo XTTS-v2, generó nuevos audios que emulaban disfluencias naturales, como prolongaciones y bloqueos. Aunque mejoró el recall, especialmente en rondas posteriores de aumento de datos, la precisión y F1 continuaron siendo desafiantes. Futuras investigaciones se enfocarán en expandir la anotación de disfluencias en contextos de habla espontánea y procesar los audios restantes del corpus inicial para explorar mejoras en la clasificación y detección de disfluencias. Además, se explorarán métodos avanzados de clonación de voz y otras técnicas de modificación de audios para enriquecer los datasets y mejorar los modelos de detección y clasificación de disfluencias.

**Palabras clave:** Tartamudez, Detección de Disfluencias, *Machine Learning*, *Deep Learning*, *Automatic Speech Recognition*, *Data Augmentation*.

## Abstract

This thesis focused on the development of computational resources for the detection and classification of stuttering disfluencies in Spanish, spanning from the collection and annotation of audio data to the implementation of a machine learning model and data augmentation strategies. Audios in Spanish from five participants with stuttering were collected, adhering to the SEP-28K dataset standards and supported by two specialists in stuttering. Although the controlled nature of the recordings limited the diversity of observed disfluencies, these audios provided a solid foundation for the model development. The model was based on the DisfluencyNet and pre-trained using wav2vec 2.0 XLSR53, leveraging its robust multilingual database. The model demonstrated its ability to identify and classify disfluencies in Spanish, though its performance was inferior compared to similar models in English, highlighting the need for more data. To enhance disfluency detection, two data augmentation strategies were implemented. The first involved pitch variations, reverberation addition, and white noise, effectively doubling the available data. Although this strategy improved recall for certain disfluencies, precision and F1 results were mixed. The second strategy, using voice cloning with the XTTS-v2 model, generated new audios that emulated natural disfluencies, such as prolongations and blocks. While it enhanced recall, particularly in later rounds of data augmentation, precision and F1 continued to be challenging. Future research will focus on expanding the annotation of disfluencies in spontaneous speech contexts and processing the remaining audios from the initial corpus to explore improvements in classification and detection of disfluencies. Additionally, advanced voice cloning methods and other audio modification techniques will be explored to enrich the datasets and enhance the detection and classification models.

**Keywords:** Stuttering, Disfluency Detection, Machine Learning, Deep Learning, Automatic Speech Recognition, Data Augmentation.

# Índice

<b>Resumen</b>	<b>i</b>
<b>Abstract</b>	<b>ii</b>
<b>Índice</b>	<b>iii</b>
<b>Capítulo I - Introducción</b>	<b>1</b>
1.1 Planteamiento del problema	2
1.2 Objetivos	2
1.2.1 Objetivo Principal	2
1.2.2 Objetivos Específicos	2
1.3 Resultados Esperados (Resultados por cada objetivo)	3
<b>Capítulo II - Marco Teórico</b>	<b>4</b>
2.1 Tartamudez	4
2.2 Aprendizaje Automático o <i>Machine Learning</i>	5
2.3 Aumento de Datos o <i>Data Augmentation</i>	6
2.4 Reconocimiento Automático del Habla o <i>Automatic Speech Recognition</i>	6
2.4 Detección de Actividad de Voz o <i>Voice Activity Detection</i>	7
<b>Capítulo III - Trabajos Relacionados</b>	<b>8</b>
3.1 Preguntas de Investigación	8
3.2 Metodología	8
3.3 Resultados	9
3.3.1 ¿Qué métodos/técnicas/algoritmos/enfoques han sido usados en la detección y clasificación de disfluencias producidas por tartamudez?	10
3.3.2 ¿Cuáles han sido las limitaciones actuales para la detección de disfluencias producidas por tartamudez?	12
3.3.3 ¿De qué manera se ha realizado la construcción de recursos para la detección y clasificación de disfluencias producidas por tartamudez?	12
3.3.4 ¿Qué enfoque ha producido mejores resultados en los estudios con niños y con adultos?	13
3.3.5 ¿Existen enfoques que pueden ser utilizados en más de un idioma?	13
3.3.6 ¿Existen enfoques que han intentado usar modelos en otros idiomas o para otras tareas para mejorar el desempeño en una lengua diferente?	14
<b>Capítulo IV - Construcción del Dataset</b>	<b>15</b>
4.1 Objetivo de la Anotación	15
4.2 Proceso de Anotación	15

4.2.1 Selección de recursos fuente	15
4.2.2 Procesamiento del dataset	16
4.2.3 Anotación del dataset	20
4.3 Resultados de la anotación	21
<b>Capítulo V - Modelo de detección y clasificación</b>	<b>24</b>
5.1 Descripción del modelo	24
5.2 División del Dataset	25
5.3 Métricas de evaluación	26
5.4 Estrategias para lidiar con el desbalanceamiento de datos	26
5.4.1 Downsampling	26
5.4.2 Aumento de datos por modificación de audios	27
5.4.3 Aumento de datos con clonación de voz y reglas	28
5.5 Resultados de los experimentos	29
5.5.1 Resultados Preliminares	29
5.5.2 Resultados de Downsampling	30
5.5.3 Resultados de Aumento de datos	31
<b>Capítulo VI - Conclusiones</b>	<b>41</b>
<b>Capítulo VII - Trabajos Futuros</b>	<b>43</b>
<b>Referencias</b>	<b>44</b>
<b>Anexos</b>	<b>49</b>

## Capítulo I - Introducción

La tartamudez es un desorden del habla caracterizado por disfluencias tales como repeticiones de sonidos, sílabas o palabras, prolongaciones de sonidos y bloqueos, que son pausas involuntarias en el habla. Aunque las personas que tartamudean son conscientes de lo que desean decir, estos patrones disruptivos de habla dificultan la fluidez en su expresión. Este trastorno puede estar acompañado de movimientos físicos involuntarios, como parpadeos o temblores de labios, que son comportamientos aprendidos de manera inconsciente para intentar reducir la severidad de la tartamudez (Craig, 2000). Además, la tartamudez puede tener un impacto significativo en la calidad de vida, afectando las relaciones interpersonales, el rendimiento laboral y la capacidad para aprovechar oportunidades.

Generalmente, la tartamudez comienza en la infancia, afectando aproximadamente al 2.5% de los niños menores de cinco años (Simmons, 2017). Aunque muchos niños superan este trastorno durante la adolescencia, alrededor del 1% de la población adulta continúa experimentando tartamudez, lo que puede repercutir de manera considerable en los ámbitos académico, social y profesional (Schoen, 2017). Las personas con un nivel de tartamudez moderado a severo pueden percibir su condición negativamente, lo cual a su vez puede derivar en problemas psicológicos como ansiedad, angustia, y un sentimiento de impotencia, disminuyendo sus oportunidades laborales y reduciendo su calidad de vida.

Por estas razones, la comprensión y manejo de la tartamudez es prioritario, siendo la infancia la etapa más oportuna para intervenir, con el fin de evitar las complicaciones mencionadas. El tratamiento involucra a profesionales de diversas disciplinas, incluyendo la patología del habla y lenguaje, la audiolología y la psicología, que colaboran para determinar las mejores estrategias de detección y tratamiento de este trastorno (Singh & Kumar, 2022).

En este contexto, la herramienta *Overall Assessment of the Speaker's Experience of Stuttering* (OASES) es un recurso valioso para los terapeutas. Disponible en varios idiomas, OASES facilita el diagnóstico, la calificación para

terapia, la evaluación de la eficacia del tratamiento y la documentación del progreso del paciente. Además, permite recopilar antecedentes, tomar decisiones informadas y evaluar los resultados post-tratamiento para asegurar un seguimiento adecuado. OASES se adapta a diferentes edades con versiones específicas: OASES-S para niños de 7 a 12 años con 60 ítems, OASES-T para adolescentes de 13 a 17 años con 80 ítems y OASES-A para adultos de 18 años en adelante con 100 ítems. Esta herramienta es esencial para entender la percepción del hablante sobre su tartamudez, las reacciones del paciente, las dificultades comunicativas diarias y el impacto general en su calidad de vida.

## **1.1 Planteamiento del problema**

La evaluación de la tartamudez, además de contar con la estructura de OASES, se realiza típicamente a través de una sesión de entrevista, donde un especialista observa y registra la severidad y los tipos de disfluencias presentes. Esta tarea, tradicionalmente manual, exige que el especialista esté especialmente atento para identificar y anotar las disfluencias observadas. No obstante, los avances en tecnología ofrecen la posibilidad de automatizar este proceso de anotación, permitiendo que los especialistas se concentren en otros aspectos relevantes del diagnóstico y tratamiento de la tartamudez (Alharbi, 2018).

Por ello nos hacemos la pregunta, ¿de qué manera se pueden construir recursos que den soporte a la tarea de los profesionales en la detección de tartamudez?

## **1.2 Objetivos**

### **1.2.1 Objetivo Principal**

Construir recursos computacionales que den soporte a la tarea de detección y clasificación automática de disfluencias producidas por tartamudez en español.

### **1.2.2 Objetivos Específicos**

- OE1: Recolectar audios de personas con tartamudez en español.

- OE2: Implementar un modelo de detección y clasificación de disfluencias producidas por personas con tartamudez para el español.
- OE3: Implementar una estrategia de aumento de datos para mejorar el desempeño de los métodos utilizados.

### **1.3 Resultados Esperados (Resultados por cada objetivo)**

OE:1:

- Audios producidos por personas con tartamudez en español.
- Preprocesamiento siguiendo las indicaciones del conjunto de datos SEP-28K
- Anotación de disfluencias por parte de los especialistas.

OE2:

- Modelo que permita detectar y clasificar disfluencias producidas por tartamudez.

OE3:

- Estrategia que permita generar data con disfluencias artificialmente en base a data sin disfluencias.
- Estrategia basada en técnicas de aumento de datos.

## Capítulo II - Marco Teórico

En el presente capítulo explica los principales conceptos asociados a la presente tesis.

### 2.1 Tartamudez

La tartamudez, por lo general, es conocida como un trastorno que implica una falta de fluidez en la comunicación oral. Aunque es común observar cierto grado de disfluencia en niños que están aprendiendo las bases del lenguaje, se convierte en un “problema” cuando persiste y se consolida con el tiempo (Seyfeddinipur, 2006). Más allá de afectar la fluidez del habla, este trastorno puede generar consecuencias secundarias significativas, como el desarrollo de miedo a hablar en público. Esto lleva a que las personas afectadas eviten situaciones de expresión oral o experimenten angustia, lo que puede reducir sus capacidades intelectuales y oratorias (Rodríguez, 2000).

Existen varios tipos de disfluencias, siendo las más comunes las repeticiones, prolongaciones, pausas e interjecciones. Las repeticiones implican la repetición de sílabas o palabras completas (por ejemplo: "pe-pe-pero", "est-est-esto"). Las prolongaciones se refieren a alargar la sílaba de una palabra (por ejemplo: "gaaaaaalleta"). Las pausas son silencios innecesarios que ocurren entre palabras o sílabas, con una duración inusualmente prolongada (por ejemplo: *\*pausa\** juguete, a*\*pausa\**bierto). Las interjecciones, también conocidas como muletillas, se presentan de manera innecesaria (por ejemplo: "ahm") (Ambrose y Yairi, 1999).

Adicionalmente, la tartamudez se clasifica en tartamudez del desarrollo y tartamudez adquirida. La última suele asociarse a lesiones cerebrales y se diferencia de la discinesia facial, que implica movimientos físicos involuntarios durante el habla. La tartamudez del desarrollo, o disfemia, es un trastorno del desarrollo del habla que afecta temporalmente el control cerebral sobre el habla. Desde una perspectiva genética, se ha identificado que un hombre con

tartamudez tiene un 22% de probabilidades de tener un hijo con el mismo trastorno y un 9% si son hijas (Manaut-Gil, 2005).

En la actualidad, Tichenor y Yaruss (2019) buscaron la definición en adultos con tartamudez. En base a ello, indican que la tartamudez es la sensación de pérdida de control del habla que puede o no mostrar exteriormente. La persona con tartamudez también experimenta reacciones afectivas o cognitivas que puede conducir a consecuencias negativas en su entorno, esto dado a que la percepción del hablante, en este caso la persona con tartamudez, tiende a depender mucho de la reacción que ellos creen que pueda tener el oyente.

## **2.2 Aprendizaje Automático o *Machine Learning***

Machine Learning, o aprendizaje automático, es una rama de la inteligencia artificial que se enfoca en desarrollar técnicas que permitan a las máquinas aprender a partir de datos. El objetivo principal es permitir que las computadoras manejen y analicen grandes volúmenes de información de manera eficiente y automática, extrayendo patrones y tomando decisiones basadas en la experiencia previa sin intervención humana directa. Esta disciplina utiliza una variedad de algoritmos para abordar y resolver problemas complejos en múltiples áreas, particularmente en campos que requieren el manejo extensivo de datos como las matemáticas y la programación (Mahesh, 2020). Existen 4 tipos de modelos de aprendizaje automático en Machine Learning (Raj, 2019), los cuales son: supervisado, no supervisado, semi supervisado y de refuerzo. El aprendizaje supervisado trabaja con instancias suministradas externamente, es utilizado en casos de clasificación donde se busca que el sistema tenga la capacidad de realizar las clasificaciones que se le ha ordenado. El aprendizaje no supervisado recibe una entrada sin etiquetar y en base a eso se puede un resultado del cual se tiene cierta comprensión, básicamente consiste en descubrir patrones en un conjunto de observaciones. El aprendizaje semisupervisado, por otra parte, se utiliza en modelos donde el conjunto de datos cuenta con etiquetas incompletas. En estos casos, el conjunto de entrenamiento cuenta con pocas instancias con etiquetas, lo que puede producir mejoras en la

precisión del aprendizaje. Por último, el aprendizaje por refuerzo requiere una retroalimentación de recompensa simple para que el modelo aprenda el comportamiento, esto permite maximizar el rendimiento. Gracias a sus capacidades, Machine Learning es ampliamente utilizado para optimizar procesos, realizar predicciones precisas y mejorar las decisiones en una diversidad de contextos laborales (Mahesh, 2020).

### **2.3 Aumento de Datos o *Data Augmentation***

*Data Augmentation* es una técnica utilizada para ampliar el volumen de datos de entrenamiento sin necesidad de recolectar nuevos datos. Esta estrategia es especialmente valiosa en el contexto del aprendizaje automático, donde un conjunto de datos más grande puede mejorar significativamente el rendimiento del modelo. Aunque *Data Augmentation* todavía no está ampliamente explorada debido a los desafíos que implica generar datos sintéticos que se asemejen de manera convincente a datos reales, el interés en su aplicación ha crecido exponencialmente, especialmente en el campo del Procesamiento del Lenguaje Natural (Natural Language Processing o NLP por sus siglas en inglés). Esta creciente atención está directamente relacionada con el desarrollo y expansión del área, que busca soluciones innovadoras para enriquecer los datasets y mejorar la eficacia de los algoritmos en diversas aplicaciones (Feng et al., 2021).

### **2.4 Reconocimiento Automático del Habla o *Automatic Speech Recognition***

El Reconocimiento Automático del Habla (ASR, por sus siglas en inglés) es un área de la Inteligencia Artificial que busca permitir a las computadoras interpretar y convertir la voz humana en texto. Generalmente, un sistema ASR captura la entrada de audio, usualmente a través de un micrófono, y la somete a una serie de modelos o algoritmos para identificar patrones de habla. Esta compleja cadena de procesamiento analiza el audio y lo transcribe en texto, facilitando así la interacción entre humanos y máquinas. ASR es utilizado en una amplia gama de aplicaciones, desde asistentes personales virtuales hasta sistemas de atención al cliente automatizados (Lai, Karat, y Yankelovich, 2008, como se cita en Levis y Suvorov, 2012).

## **2.4 Detección de Actividad de Voz o *Voice Activity Detection***

*Voice Activity Detection* (VAD) es una tarea computacional que busca determinar la presencia o ausencia de habla en una señal de audio. VAD es fundamental en una variedad de aplicaciones que requieren procesamiento de audio eficiente, tales como la medición y control de la sonoridad, optimización de diálogos, codificación de audio perceptual, monitoreo de transmisiones y mezcla a ciegas de fuentes sonoras. Esta tecnología funciona mediante la extracción y análisis de características de la señal de audio para diferenciar segmentos de habla de aquellos sin habla. Los métodos utilizados incluyen la detección y clasificación a través de aprendizaje supervisado, y la discriminación entre habla y música. VAD también juega un papel crucial en aplicaciones como el reconocimiento automático de habla (ASR), Voice over IP (VoIP), entre otras (Uhle y Bäckström, 2017).



## Capítulo III - Trabajos Relacionados

Este capítulo se centra en explorar investigaciones previas relevantes para la presente tesis. En la sección 3.1, se examinan las preguntas de investigación que han guiado estudios similares en el campo. La sección 3.2 se centra en las metodologías empleadas en estas investigaciones, analizando las técnicas y enfoques que han resultado ser efectivos. Finalmente, en la sección 3.3, se discuten los resultados y conclusiones derivados de estos trabajos.

### 3.1 Preguntas de Investigación

1. ¿Qué métodos/técnicas/algoritmos/enfoques han sido usados en la detección y clasificación de disfluencias producidas por tartamudez?
2. ¿Cuáles han sido las limitaciones actuales para la detección de disfluencias producidas por tartamudez?
  - a. ¿Ha sido una limitación de tamaño de dataset?
  - b. ¿Ha sido una limitación del propio método?
3. ¿De qué manera se ha realizado la construcción de recursos para la detección y clasificación de disfluencias producidas por tartamudez?
4. ¿Qué enfoque ha producido mejores resultados en los estudios con niños y con adultos?
5. ¿Existen enfoques que pueden ser utilizados en más de un idioma?
6. ¿Existen enfoques que han intentado usar modelos en otros idiomas o para otras tareas para mejorar el desempeño en una lengua diferente?

### 3.2 Metodología

La metodología empleada en esta investigación para la recolección de artículos relacionados con el estado del arte es la revisión sistemática. Para responder a las preguntas de investigación planteadas en la sección anterior, se siguieron tres criterios de búsqueda específicos. Estos criterios, que se detallarán a continuación, están diseñados para asegurar la exhaustividad y relevancia de la selección de artículos, permitiendo una evaluación crítica y sistemática de la literatura existente en el campo de estudio.

### **Q1: Enfoques en general**

*("stuttering detection" OR "stammering detection" OR "stuttering classification" OR "stammering classification" OR ""stuttering detection" OR "stammering identification") AND ("natural language processing" OR "NLP") AND ("algorithm" OR "method" OR "technique" OR "approach")*

### **Q2: Tartamudez para niños**

*("stuttering detection" OR "stammering detection" OR "stuttering classification" OR "stammering classification" OR ""stuttering detection" OR "stammering identification") AND ("natural language processing" OR "NLP") AND "children"*

### **Q3: Construcción de recursos**

*("stuttering detection" OR "stammering detection" OR "stuttering classification" OR "stammering classification" OR ""stuttering detection" OR "stammering identification") AND ("dataset building" OR "dataset construction" OR "corpus building" OR "corpus construction" OR "corpus annotation" OR "dataset annotation")*

Criterios de inclusión:

- Año mínimo: 2018
- Número de páginas (sin contar bibliografía): 5 mínimo
- No surveys, no revisiones de literatura, sólo contribuciones

### **3.3 Resultados**

En la Figura 1 se presenta la cantidad de trabajos incluidos en la Revisión de Literatura desde 2018 hasta 2023, totalizando 21 trabajos.

## Cantidad de trabajos en la Revisión de Literatura

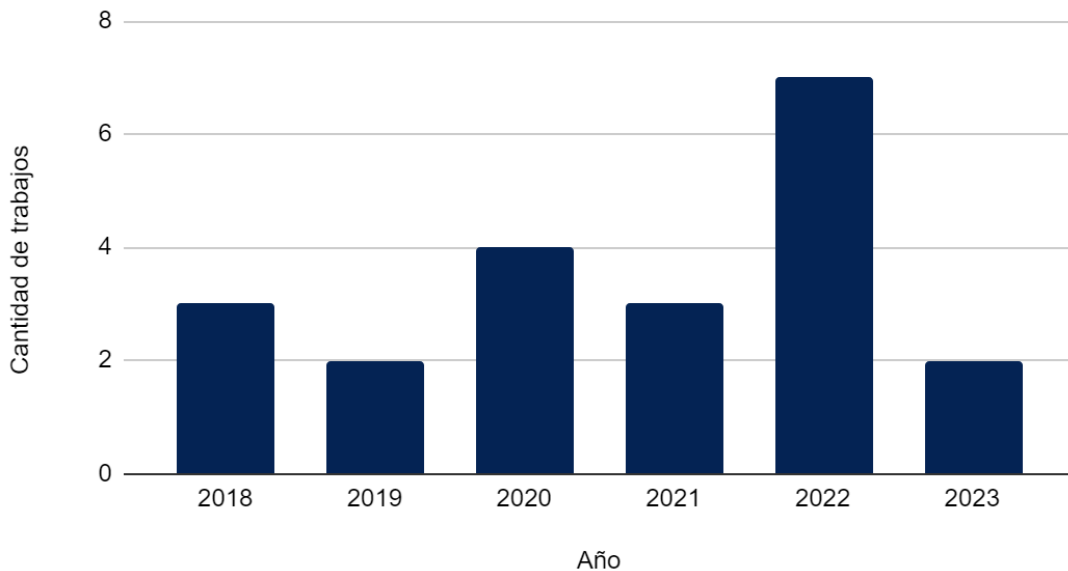


Figura 1. Gráfica de la relación de trabajos utilizados en la Revisión de Literatura por año.

En esta sección se responden a las preguntas de investigación propuestas en la revisión sistemática de la literatura:

### 3.3.1 ¿Qué métodos/técnicas/algoritmos/enfoques han sido usados en la detección y clasificación de disfluencias producidas por tartamudez?

Se identificaron varios métodos prominentes utilizados en la detección y clasificación de disfluencias. Entre los más destacados se encuentran *Support Vector Machine* (SVM) (Hosseini et al., 2018; Arbajian et al., 2019; Riad et al., 2020; Mishra et al., 2021; Jouaiti y Dauthenhahn, 2022), conocido por su robustez en tareas de clasificación; *Time Delay Neural Network* (TDNN) (Kourkounakis et al., 2021; Sheikh et al. 2021; Nöth y Riedhammer 2023; Sheikh et al. 2023), que es efectiva para el análisis de secuencias temporales; y *Bidirectional Long Short-Term Memory* (Bi-LSTM) (Alharbi et al., 2020; Gupta et al. 2020; Riad et al., 2020; Bayerl et al. 2023; Sheikh et al. 2023), que captura información de secuencias tanto en direcciones futuras como pasadas, mejorando así la comprensión contextual del habla. Además, *Convolutional Neural Network* (CNN) (Kourkounakis et al., 2021; Mohapatra et al., 2022,

Murugan et al., 2022; Shah y Nautiyal; 2022; Bayerl et al. 2023) y *Long Short-Term Memory* (LSTM) (Kourkounakis et al., 2020; Kourkounakis et al., 2021; Bayerl et al. 2023) también han sido ampliamente utilizadas, la primera por su capacidad para reconocer patrones espaciales en datos y la segunda por su habilidad para aprender dependencias a largo plazo en datos de series temporales. Estos métodos representan algunas de las técnicas más avanzadas en el campo del procesamiento del habla y han sido aplicadas con éxito en estudios previos para abordar las complejidades de las disfluencias.

Algoritmo	Frecuencia
SVM	6
TDNN	4
Bi-LSTM	4
CNN	3
LSTM	2

Tabla 1: Métodos más utilizados en la detección y clasificación de disfluencias

Por ejemplo, Jouaiti y Dauthenhahn (2022) usaron *Support Vector Machines* (SVM) como clasificador principal en su estudio sobre la tartamudez, complementado con Random Forest para mejorar el manejo de la clasificación multi etiqueta, un área donde SVM por sí solo puede presentar limitaciones. Esta combinación de técnicas logró resultados favorables en la métrica F1-score, especialmente en las categorías de repetición de palabras, interjecciones y habla fluida. Sin embargo, los autores reportaron que la técnica no fue efectiva para la categoría de repetición de sonidos, donde no se observaron mejoras significativas en ninguna de las métricas evaluadas por este método.

### **3.3.1.1 ¿Qué *features* o características han sido utilizadas en el modelamiento del problema de clasificación de disfluencias?**

En la extracción de características para la detección y clasificación de tartamudez, el uso de *Mel Frequency Cepstral Coefficients* (MFCC) ha sido particularmente recurrente. Esta técnica comienza con la captura del audio, que posteriormente se procesa para extraer características significativas antes de ser introducido en el modelo de clasificación. Según Sheikh et al. (2021), MFCC es considerado la mejor opción para extraer características del habla en el campo de la tartamudez, debido a su efectividad en capturar las cualidades esenciales del habla que son relevantes para identificar disfluencias.

### **3.3.2 ¿Cuáles han sido las limitaciones actuales para la detección de disfluencias producidas por tartamudez?**

Según Sheikh et al. (2021), una de las principales restricciones es la falta de atención suficiente hacia los métodos de detección de disfluencias. Adicionalmente, Kourkounakis et al. (2021) identifican la complejidad de detectar disfluencias debido a la gran cantidad de clases y subclases existentes, así como la influencia de factores variables como el género, la edad, el acento y el lenguaje, que pueden alterar significativamente los resultados de detección. Otra limitación crítica, señalada por Mishra et al. (2021), es la escasez de datasets en idiomas distintos al inglés, lo que restringe la aplicación y evaluación de modelos de detección de tartamudez en contextos lingüísticos más amplios.

### **3.3.3 ¿De qué manera se ha realizado la construcción de recursos para la detección y clasificación de disfluencias producidas por tartamudez?**

Los estudios recientes destacan la viabilidad de utilizar datasets públicos y adaptarlos para mejorar los resultados en investigaciones específicas. Por ejemplo, Jouaiti y Dauthenhahn (2022) adaptaron el dataset FluencyBank para su estudio sobre la tartamudez, seleccionando específicamente a 32 adultos que presentaban este trastorno. Además, complementaron su investigación con la utilización de otro dataset llamada UCLASS, de la cual seleccionaron 128 audios

que incluían grabaciones tanto de adultos como de niños. Esta estrategia de selección y adaptación de datasets públicos permite a los investigadores ajustar los recursos disponibles a las necesidades concretas de sus estudios, optimizando así la relevancia y precisión de los resultados obtenidos.

Por otro lado, Mohapatra et al. (2022) utilizaron, además de FluencyBank, la base de datos open-source SEP-28K, que incluye 28,000 clips de audio en inglés procedentes de ocho podcasts diferentes de personas con tartamudez. Los clips fueron anotados para eventos de tartamudez por tres anotadores, lo que permitió el posterior procesamiento y clasificación de las disfluencias producidas por tartamudez.

### **3.3.4 ¿Qué enfoque ha producido mejores resultados en los estudios con niños y con adultos?**

Con niños, el enfoque más exitoso fue el implementado por Gupta et al. (2020), quienes emplearon la técnica de Bi-LSTM para identificar las disfluencias asociadas a la tartamudez. Ellos obtuvieron con Bi-LSTM precisión de 96.18%, recall igual a 96.31%, f-score de 96.01% y accuracy de 96.67%.

En cambio, con adultos, los mejores resultados fueron reportados por Jouaiti y Dauthenhahn (2022). En su investigación, utilizaron una combinación de los modelos SVM y Random Forest para clasificar las disfluencias en individuos con tartamudez utilizando los datasets FluencyBank y UCLASS. Esta metodología permitió una clasificación efectiva de las disfluencias, destacándose por su precisión y robustez. Por ejemplo, en FluencyBank, obtuvieron accuracy de 92.23%, 96.16, 92.06, 94.83, 91.28, 92.48 para *Repetición de palabra*, *Repetición de sonido*, *Interjección*, *Prolongación*, *Bloqueo*, *Fluente* respectivamente.

### **3.3.5 ¿Existen enfoques que pueden ser utilizados en más de un idioma?**

Desde la revisión de literatura, se identificaron cuatro estudios que destacan por el uso de modelos aplicados en múltiples idiomas. Sheikh et al. (2023) emplearon *Time Delay Neural Network* y Bi-LSTM, demostrando la viabilidad de trabajar con 12 idiomas diferentes. Por su parte, Bayerl et al. (2023)

extendieron la aplicación de sus modelos más allá del inglés, incluyendo el alemán y otras lenguas adicionales. Finalmente, Shah y Nautiyal (2022) implementaron sus investigaciones en el idioma chino, contribuyendo a la diversificación lingüística en el estudio de las disfluencias.

### **3.3.6 ¿Existen enfoques que han intentado usar modelos en otros idiomas o para otras tareas para mejorar el desempeño en una lengua diferente?**

En Shah y Nautiyal (2022) desarrollaron un modelo que clasifica las disfluencias producidas por tartamudez en el idioma chino y usaron el mismo modelo propuesto para el idioma inglés.



## **Capítulo IV - Construcción del Dataset**

### **4.1 Objetivo de la Anotación**

El objetivo principal de la anotación en este estudio fue etiquetar clips de audio con diversos tipos de tartamudez, incluyendo bloqueos, prolongaciones, interjecciones, repeticiones de palabras y sonidos, entre otros. Una vez completada la fase de etiquetado, se procedió a entrenar los modelos destinados a la detección de disfluencias generadas por tartamudez.

### **4.2 Proceso de Anotación**

#### **4.2.1 Selección de recursos fuente**

Los audios analizados en este trabajo provinieron de siete participantes miembros de la Asociación Peruana de Tartamudez, involucrados en un estudio anterior. Durante el periodo comprendido entre el 27 de mayo y el 5 de junio de 2020, estos participantes realizaron grabaciones leyendo noticias seleccionadas de los meses de mayo y junio de 2020, extraídas de los siguientes diarios peruanos: El Comercio, La República, Perú21 y Gestión. Las noticias abarcaron temas de las secciones informativa, política, económica y de actualidad, detalladas en el Anexo A.

Cada participante recibió 10 noticias diarias para leer y grabar en voz alta, produciendo así 10 audios por día, con una duración promedio de 3 minutos cada uno. La duración varió según la severidad de la tartamudez de cada participante. Para preservar la confidencialidad, los participantes fueron mantenidos en anonimato, y no se recopilaban datos adicionales como nombres, direcciones o documentos de identidad.

Al finalizar el periodo de grabación, se recolectaron un total de 696 audios dado que una de las participantes no pudo cumplir con 4 grabaciones. Debido a la baja calidad de los audios de dos participantes, el conjunto final se redujo a 500 audios útiles, sumando un total de 21.92 horas de grabaciones.

#### 4.2.2 Procesamiento del dataset

La anotación de los audios se llevó a cabo siguiendo las directrices establecidas en la construcción del dataset SEP-28K, según lo descrito por Lea et al. (2021). Los autores anotaron clips de audio de 3 segundos empleando un conjunto de clases de disfluencias. La elección de este intervalo de tiempo se debe a los retos asociados con la identificación de las disfluencias del habla, que a menudo presentan ambigüedades. Para manejar estas ambigüedades, implementaron una metodología de evaluación basada en intervalos de tiempo (Valente et al., 2015, como se citó en Lea et al., 2021), determinando los clips de audio de 3 segundos. Dado que los autores usaron audios de podcasts de tartamudez (largos), tuvieron que implementar una estrategia para extraer los clips. Específicamente, resaltaron que los eventos de disfluencias producidas por tartamudez aparecían antes, durante o después de una pausa por lo que usaron un detector de actividad vocal (VAD) para realizar la extracción de los intervalos de 3 segundos cerca de las pausas.

Originalmente, los audios presentaban diversas tasas de muestreo. Con el fin de homogeneizar estos datos y alinearnos con los estándares utilizados en investigaciones previas sobre detección de disfluencias, como las de Mohapatra et al. (2022), la tasa de muestreo de los audios se estandarizó a 16 kHz. En general, se realizaron los siguientes pasos:

- **Subida de audios a Zenodo:** Se crearon repositorios individuales en Zenodo para cada participante, como se muestra en la Figura 2. Cada repositorio contiene todos los audios de un participante, facilitando el acceso integral a los mismos, tal como se ilustra en la Figura 3.

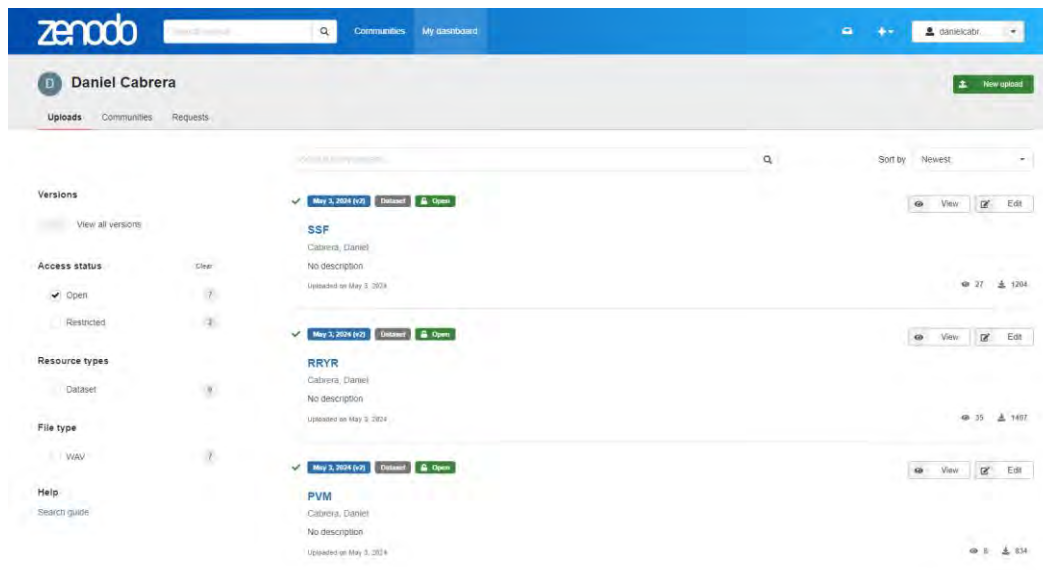


Figura 2. Lista de repositorios en Zenodo

Files (456.2 MB)		
Name	Size	Download all
2020-05-27_03_54_34.wav md5:1158ad1589d123daf7444bf4e32aed85	3.0 MB	Preview Download
2020-05-27_03_55_03.wav md5:5ad769514964e3866662f6b4204cc59	4.0 MB	Preview Download
2020-05-27_03_55_25.wav md5:99a90636d7b55e71925c6428bd4fae90	3.9 MB	Preview Download
2020-05-27_03_55_51.wav md5:d2126d2336d237e6ec2d2032a0c00090	3.7 MB	Preview Download
2020-05-27_03_56_16.wav md5:a15ca18c6c8d56b1577a831b5f1efdf	4.4 MB	Preview Download
2020-05-27_06_06_55.wav md5:49ce896a0c811152ef20eab33f937d	3.9 MB	Preview Download
2020-05-27_10_23_37.wav md5:0654bae53f5c7d45c0bb77d8fcbf46	3.7 MB	Preview Download

Figura 3. Ejemplo de audios en uno de los repositorios de Zenodo

- Creación de archivo csv basado en SEP-28K para descargar audios:**  
 Inspirado en el modelo de SEP-28K, se generó un archivo CSV que cataloga los audios. Este archivo, visualizado en la Figura 4, incluye el ID del participante de origen, el ID del audio de origen y el enlace a los audios almacenados en Zenodo.

ATMA	2020-06-03_05_34_19	<a href="https://zenodo.org/records/11111351/files/2020-06-03_05_34_19.wav">https://zenodo.org/records/11111351/files/2020-06-03_05_34_19.wav</a>
ATMA	2020-06-03_05_34_27	<a href="https://zenodo.org/records/11111351/files/2020-06-03_05_34_27.wav">https://zenodo.org/records/11111351/files/2020-06-03_05_34_27.wav</a>
ATMA	2020-06-03_05_38_59	<a href="https://zenodo.org/records/11111351/files/2020-06-03_05_38_59.wav">https://zenodo.org/records/11111351/files/2020-06-03_05_38_59.wav</a>
ATMA	2020-06-03_05_39_25	<a href="https://zenodo.org/records/11111351/files/2020-06-03_05_39_25.wav">https://zenodo.org/records/11111351/files/2020-06-03_05_39_25.wav</a>
ATMA	2020-06-03_05_40_18	<a href="https://zenodo.org/records/11111351/files/2020-06-03_05_40_18.wav">https://zenodo.org/records/11111351/files/2020-06-03_05_40_18.wav</a>
ATMA	2020-06-03_05_40_44	<a href="https://zenodo.org/records/11111351/files/2020-06-03_05_40_44.wav">https://zenodo.org/records/11111351/files/2020-06-03_05_40_44.wav</a>
ATMA	2020-06-03_05_41_07	<a href="https://zenodo.org/records/11111351/files/2020-06-03_05_41_07.wav">https://zenodo.org/records/11111351/files/2020-06-03_05_41_07.wav</a>
ATMA	2020-06-03_05_41_22	<a href="https://zenodo.org/records/11111351/files/2020-06-03_05_41_22.wav">https://zenodo.org/records/11111351/files/2020-06-03_05_41_22.wav</a>
ATMA	2020-06-03_05_41_41	<a href="https://zenodo.org/records/11111351/files/2020-06-03_05_41_41.wav">https://zenodo.org/records/11111351/files/2020-06-03_05_41_41.wav</a>
ATMA	2020-06-03_05_42_02	<a href="https://zenodo.org/records/11111351/files/2020-06-03_05_42_02.wav">https://zenodo.org/records/11111351/files/2020-06-03_05_42_02.wav</a>
HEL R	2020-06-03_04_07_02	<a href="https://zenodo.org/records/11111486/files/2020-06-03_04_07_02.wav">https://zenodo.org/records/11111486/files/2020-06-03_04_07_02.wav</a>
HEL R	2020-06-03_04_08_12	<a href="https://zenodo.org/records/11111486/files/2020-06-03_04_08_12.wav">https://zenodo.org/records/11111486/files/2020-06-03_04_08_12.wav</a>
HEL R	2020-06-03_04_19_24	<a href="https://zenodo.org/records/11111486/files/2020-06-03_04_19_24.wav">https://zenodo.org/records/11111486/files/2020-06-03_04_19_24.wav</a>
HEL R	2020-06-03_04_19_38	<a href="https://zenodo.org/records/11111486/files/2020-06-03_04_19_38.wav">https://zenodo.org/records/11111486/files/2020-06-03_04_19_38.wav</a>
HEL R	2020-06-03_04_43_30	<a href="https://zenodo.org/records/11111486/files/2020-06-03_04_43_30.wav">https://zenodo.org/records/11111486/files/2020-06-03_04_43_30.wav</a>
HEL R	2020-06-03_04_44_17	<a href="https://zenodo.org/records/11111486/files/2020-06-03_04_44_17.wav">https://zenodo.org/records/11111486/files/2020-06-03_04_44_17.wav</a>
HEL R	2020-06-03_06_16_33	<a href="https://zenodo.org/records/11111486/files/2020-06-03_06_16_33.wav">https://zenodo.org/records/11111486/files/2020-06-03_06_16_33.wav</a>
HEL R	2020-06-03_06_16_51	<a href="https://zenodo.org/records/11111486/files/2020-06-03_06_16_51.wav">https://zenodo.org/records/11111486/files/2020-06-03_06_16_51.wav</a>
HEL R	2020-06-03_06_33_23	<a href="https://zenodo.org/records/11111486/files/2020-06-03_06_33_23.wav">https://zenodo.org/records/11111486/files/2020-06-03_06_33_23.wav</a>
HEL R	2020-06-03_06_33_12	<a href="https://zenodo.org/records/11111486/files/2020-06-03_06_33_12.wav">https://zenodo.org/records/11111486/files/2020-06-03_06_33_12.wav</a>
LHCM	2020-06-03_01_38_01	<a href="https://zenodo.org/records/11111521/files/2020-06-03_01_38_01.wav">https://zenodo.org/records/11111521/files/2020-06-03_01_38_01.wav</a>
LHCM	2020-06-03_01_39_20	<a href="https://zenodo.org/records/11111521/files/2020-06-03_01_39_20.wav">https://zenodo.org/records/11111521/files/2020-06-03_01_39_20.wav</a>
LHCM	2020-06-03_01_39_32	<a href="https://zenodo.org/records/11111521/files/2020-06-03_01_39_32.wav">https://zenodo.org/records/11111521/files/2020-06-03_01_39_32.wav</a>
LHCM	2020-06-03_06_07_37	<a href="https://zenodo.org/records/11111521/files/2020-06-03_06_07_37.wav">https://zenodo.org/records/11111521/files/2020-06-03_06_07_37.wav</a>
LHCM	2020-06-03_06_06_46	<a href="https://zenodo.org/records/11111521/files/2020-06-03_06_06_46.wav">https://zenodo.org/records/11111521/files/2020-06-03_06_06_46.wav</a>
LHCM	2020-06-03_06_09_17	<a href="https://zenodo.org/records/11111521/files/2020-06-03_06_09_17.wav">https://zenodo.org/records/11111521/files/2020-06-03_06_09_17.wav</a>
LHCM	2020-06-03_06_19_12	<a href="https://zenodo.org/records/11111521/files/2020-06-03_06_19_12.wav">https://zenodo.org/records/11111521/files/2020-06-03_06_19_12.wav</a>
LHCM	2020-06-03_06_21_00	<a href="https://zenodo.org/records/11111521/files/2020-06-03_06_21_00.wav">https://zenodo.org/records/11111521/files/2020-06-03_06_21_00.wav</a>
LHCM	2020-06-03_06_20_24	<a href="https://zenodo.org/records/11111521/files/2020-06-03_06_20_24.wav">https://zenodo.org/records/11111521/files/2020-06-03_06_20_24.wav</a>
LHCM	2020-06-03_06_36_12	<a href="https://zenodo.org/records/11111521/files/2020-06-03_06_36_12.wav">https://zenodo.org/records/11111521/files/2020-06-03_06_36_12.wav</a>
LPJM	2020-06-03_09_28_57	<a href="https://zenodo.org/records/11111524/files/2020-06-03_09_28_57.wav">https://zenodo.org/records/11111524/files/2020-06-03_09_28_57.wav</a>
LPJM	2020-06-03_09_31_45	<a href="https://zenodo.org/records/11111524/files/2020-06-03_09_31_45.wav">https://zenodo.org/records/11111524/files/2020-06-03_09_31_45.wav</a>
LPJM	2020-06-03_09_35_06	<a href="https://zenodo.org/records/11111524/files/2020-06-03_09_35_06.wav">https://zenodo.org/records/11111524/files/2020-06-03_09_35_06.wav</a>
LPJM	2020-06-03_09_41_10	<a href="https://zenodo.org/records/11111524/files/2020-06-03_09_41_10.wav">https://zenodo.org/records/11111524/files/2020-06-03_09_41_10.wav</a>

Figura 4. Archivo csv “tesis\_audios.csv”

- **Detección de actividad de habla para los audios:** Se utilizó un sistema de Detección de Actividad Vocal Silero<sup>1</sup> para identificar los segmentos de habla en cada audio, siguiendo la metodología de Lea et al. (2021). Este sistema determina con precisión los puntos de inicio y fin de los segmentos que contienen habla.
- **Determinación de los segmentos de 3 segundos:** se empleó VAD para seleccionar aleatoriamente un punto dentro del rango inicial y final de

<sup>1</sup> Disponible en <https://github.com/snakers4/silero-vad.git>

donde el VAD detectó habla para marcar el inicio de cada nuevo segmento de 3 segundos. El fin de cada segmento se fijó sumando 3 segundos al punto de inicio establecido.

- **Extracción de clips de audio:** A partir de los puntos de inicio y fin especificados para cada clip en la Figura 5, se procesaron los audios para extraer segmentos de tres segundos. Este procedimiento generó un total de 28,005 clips de audio, los cuales acumularon una duración conjunta de 23.18 horas. Es importante señalar que la duración total de los clips superó la duración total de los audios originales grabados. Esto se debe a que muchos de los clips comparten fragmentos de audio, resultando en que algunos clips contienen contenido casi idéntico con variaciones mínimas en los puntos de inicio y fin.
- **Creación del esquema de anotación basado en SEP-28K para extraer clips de audios:** Como se ilustra en la Figura 5, se desarrolló un esquema de anotación basado en la estructura de SEP-28K. Utilizando los segmentos de 3 segundos previamente definidos, se elaboró el archivo CSV "tesis\_segmentos.csv", cuya estructura se detalla en la Figura 5. Este archivo incluye información sobre el ID del participante, el ID del audio, el ID del clip, los puntos de inicio y fin de cada segmento de audio, y las etiquetas correspondientes.

Show	Epid	ClipId	Start	Stop	Unsure	PoorAudioQuality	Prolongation	Block	SoundRep	WordRep	DifficultToUnderstand	Interjection	NoStuttered	NaturalPause	Music	NoSpeech
ATMA	2020-06-03_05_34_19	0	5274	53274	0	0	0	0	0	0	0	0	0	0	0	0
ATMA	2020-06-03_05_34_19	1	33999	81999	0	0	0	0	0	0	0	0	0	0	0	0
ATMA	2020-06-03_05_34_19	2	80035	128035	0	0	0	0	0	0	0	0	0	0	0	0
ATMA	2020-06-03_05_34_19	3	122092	170092	0	0	0	0	0	0	0	0	0	0	0	0
ATMA	2020-06-03_05_34_19	4	190695	238695	0	0	0	0	0	0	0	0	0	0	0	0
ATMA	2020-06-03_05_34_19	5	233616	281616	0	0	0	0	0	0	0	0	0	0	0	0
ATMA	2020-06-03_05_34_19	6	298282	346282	0	0	0	0	0	0	0	0	0	0	0	0
ATMA	2020-06-03_05_34_19	7	324697	372697	0	0	0	0	0	0	0	0	0	0	0	0
ATMA	2020-06-03_05_34_19	8	410704	458704	0	0	0	0	0	0	0	0	0	0	0	0
ATMA	2020-06-03_05_34_19	9	418063	466063	0	0	0	0	0	0	0	0	0	0	0	0
ATMA	2020-06-03_05_34_19	10	454002	502002	0	0	0	0	0	0	0	0	0	0	0	0
ATMA	2020-06-03_05_34_19	11	527792	575792	0	0	0	0	0	0	0	0	0	0	0	0
ATMA	2020-06-03_05_34_19	12	570692	618692	0	0	0	0	0	0	0	0	0	0	0	0
ATMA	2020-06-03_05_34_19	13	584370	632370	0	0	0	0	0	0	0	0	0	0	0	0
ATMA	2020-06-03_05_34_19	14	589766	637766	0	0	0	0	0	0	0	0	0	0	0	0
ATMA	2020-06-03_05_34_19	15	608474	656474	0	0	0	0	0	0	0	0	0	0	0	0
ATMA	2020-06-03_05_34_19	16	680681	728681	0	0	0	0	0	0	0	0	0	0	0	0
ATMA	2020-06-03_05_34_19	17	727585	775585	0	0	0	0	0	0	0	0	0	0	0	0
ATMA	2020-06-03_05_34_19	18	762333	810333	0	0	0	0	0	0	0	0	0	0	0	0
ATMA	2020-06-03_05_34_19	19	783002	831002	0	0	0	0	0	0	0	0	0	0	0	0
ATMA	2020-06-03_05_34_19	20	821804	869804	0	0	0	0	0	0	0	0	0	0	0	0
ATMA	2020-06-03_05_34_19	21	842807	890807	0	0	0	0	0	0	0	0	0	0	0	0
ATMA	2020-06-03_05_34_19	22	869704	917704	0	0	0	0	0	0	0	0	0	0	0	0
ATMA	2020-06-03_05_34_19	23	879802	927802	0	0	0	0	0	0	0	0	0	0	0	0
ATMA	2020-06-03_05_34_19	24	928835	976835	0	0	0	0	0	0	0	0	0	0	0	0
ATMA	2020-06-03_05_34_19	25	962983	1010983	0	0	0	0	0	0	0	0	0	0	0	0
ATMA	2020-06-03_05_34_19	26	997137	1045137	0	0	0	0	0	0	0	0	0	0	0	0
ATMA	2020-06-03_05_34_19	27	1013786	1061786	0	0	0	0	0	0	0	0	0	0	0	0
ATMA	2020-06-03_05_34_19	28	1071486	1119486	0	0	0	0	0	0	0	0	0	0	0	0
ATMA	2020-06-03_05_34_19	29	1130944	1178944	0	0	0	0	0	0	0	0	0	0	0	0
ATMA	2020-06-03_05_34_19	30	1248349	1296349	0	0	0	0	0	0	0	0	0	0	0	0
ATMA	2020-06-03_05_34_19	31	1269591	1317591	0	0	0	0	0	0	0	0	0	0	0	0

Figura 5. Archivo csv "tesis\_segmentos.csv"

### 4.2.3 Anotación del dataset

Las anotaciones de los clips de audio estuvieron basadas en la metodología de Lea et al. (2021) considerando las etiquetas descritas en la Tabla 2.

Etiqueta	Definición
<i>Prolongación</i>	Sílabas alargadas (por ejemplo: S[ssss]ilencio)
<i>Bloqueo</i>	Pausas o silencios producidos por tartamudez
<i>Repetición de sonido</i>	Repetición de sílabas (por ejemplo: Ayer [ma-ma-ma-]manejé temprano)
<i>Repetición de palabra</i>	Repetición de la misma palabra o frase (por ejemplo: Ayer comí [comí] ligero)
<i>Sin palabras tartamudeadas</i>	Indica que no se dio ninguna de las anteriores
<i>Interjección</i>	Palabras o sílabas de relleno (por ejemplo: este, um, eh, entre otros.)

Tabla 2. Clases relacionadas a la tartamudez en el archivo de anotaciones.

Adicionalmente, se incluyeron otras etiquetas, como se detalla en la Tabla 3.

Etiqueta	Definición
<i>Inseguro</i>	Cuando el anotador no está seguro de su anotación
<i>Baja calidad de audio</i>	Cuando es difícil de entender por la calidad de audio
<i>Difícil de entender</i>	Cuando es difícil de entender el habla
<i>Pausa natural</i>	Pausas que no se consideran bloqueos u otro tipo de tartamudez
<i>Música</i>	Música o ruido de fondo
<i>Sin habla</i>	No hay habla, puede ser silencio o solo ruido de fondo.

Tabla 3. Clases adicionales en el archivo de anotaciones

Para facilitar el acceso, se cargaron en Google Drive los clips de audio de 3 segundos y se creó un Google Sheet con enlaces a cada clip, como se muestra en la Figura 6. Inicialmente, tres personas participaron en este proceso, incluyendo a dos especialistas en tartamudez y al autor de esta tesis. Los

especialistas capacitaron al autor en las técnicas de anotación mediante una prueba inicial con 120 clips para asegurar la correcta anotación de cada tipo de tartamudez. Posteriormente, se seleccionaron al azar 4265 clips de audio, distribuyendo 853 clips para cada uno de los cinco participantes que grabaron los audios.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Show	Epid	ClipId	Start	Stop	Unsure	PoorAudi	Prolonga	Block	SoundRep	WordRep	DifficultTot	Interjer	NoStutt	NaturalPa	Mu	NoSpee
2	ATMA	2020-05-27_03_54_34	14	711326	759326	0	0	0	0	0	0	0	0	0	1	0	0
3	ATMA	2020-05-27_03_54_34	19	909896	957896	0	0	0	0	0	0	0	0	0	1	0	0
4	ATMA	2020-05-27_03_54_34	20	947555	995555	0	0	0	0	0	0	0	0	0	1	0	0
5	ATMA	2020-05-27_03_54_34	22	1035183	1083183	1	0	0	0	0	0	0	0	0	0	0	0
6	ATMA	2020-05-27_03_55_03	1	124688	172688	0	0	0	0	0	0	0	0	0	1	0	0
7	ATMA	2020-05-27_03_55_03	9	672613	720613	0	0	1	0	0	0	0	0	0	0	0	0
8	ATMA	2020-05-27_03_55_03	14	905588	953588	0	0	0	0	0	0	0	0	0	1	0	0
9	ATMA	2020-05-27_03_55_03	24	1315648	1363648	0	0	0	1	0	0	0	0	0	0	0	0
10	ATMA	2020-05-27_03_55_03	31	1598942	1646942	0	0	0	0	0	0	0	0	0	1	0	0
11	ATMA	2020-05-27_03_55_25	1	143438	191438	0	0	0	0	0	0	0	0	0	1	0	0
12	ATMA	2020-05-27_03_55_25	8	566418	614418	0	0	0	0	0	0	0	0	0	1	0	0

Figura 6. Ejemplo de Google Sheet proporcionado al anotador

En cuanto a la estructura del Google Sheet, esta se basó en el archivo CSV "SEP-28K\_labels.csv" de Lea et al. (2021), mostrado en la Figura 6. El Sheet incluye las siguientes columnas: "Show" (el participante), "Epid" (ID del audio), "ClipID" (ID de cada clip generado), "Start" (inicio del audio) y "Stop" (fin del mismo), seguido de las clases relacionadas con la tartamudez, presentadas en la Tabla 2 y 3.

Estas clases deben ser marcadas con un "1" si el anotador considera que se presentan en el clip, o con un "0" en caso contrario. Es relevante destacar que un mismo clip puede contener más de un tipo de tartamudez.

### 4.3 Resultados de la anotación

Como se mencionó anteriormente, los tres anotadores realizaron un ejercicio inicial de anotación con 120 clips para asegurar que nuestras anotaciones fueran consistentes y alineadas. Para evaluar la concordancia, se utilizó el Kappa de Fleiss, un índice conocido en el ámbito psicológico que evalúa la confiabilidad del acuerdo entre evaluadores (Falotico y Quatto, 2014). En la Tabla 4 se muestra el Kappa de Fleiss para cada clase de tartamudez entre los tres anotadores. El autor de esta tesis es el anotador 1, mientras que los

anotadores 2 y 3 son especialistas en tartamudez. Se realizó una comparación por pares entre el anotador 1 y el anotador 2, el anotador 1 y el anotador 3, y el anotador 2 y el anotador 3, así como una evaluación conjunta de todos los anotadores.

Etiqueta	Antr1_Antr2	Antr1_Antr3	Antr2_Antr3	Antr1_Antr2_Antr3
Inseguro	0,4894	0,4783	0,6089	0,5478
Baja calidad de audio	0,0000	0,4979	0,4979	0,4986
Prolongación	0,6997	0,6497	0,6870	0,6797
Bloqueo	0,6652	0,6296	0,7321	0,6708
Repetición de sonido	0,7944	0,8197	0,8095	0,8077
Repetición de palabra	0,4937	0,6936	0,4958	0,5944
Difícil de entender	0,4958	0,6321	0,4894	0,5629
Sin palabras tartamudeadas	0,8632	0,8074	0,8295	0,8338
Pausa natural	0,4714	0,4643	0,6321	0,5029

Tabla 4: Relación de Kappa de Fleiss comparado entre los 3 anotadores.

En el análisis de las clases de tartamudez, al evaluar la fiabilidad entre el anotador 1 y los anotadores 2 y 3 individualmente, se observó una fiabilidad sustancial en la mayoría de las clases, con valores que oscilan entre 0.6497 y 0.8197. Sin embargo, se identificaron dos clases, Bloqueo y Repetición de palabra, donde la fiabilidad no fue alta. Para la clase Bloqueo, se registró una fiabilidad sustancial tanto entre el anotador 1 y el anotador 2 como entre el anotador 1 y el anotador 3. En el caso de Repetición de palabra, se observó una fiabilidad moderada entre el anotador 1 y el anotador 2, mientras que entre el anotador 1 y el anotador 3 la fiabilidad fue sustancial.

Dado este contexto, se solicitó a los anotadores 2 y 3 que revisaran las anotaciones correspondientes a estas etiquetas en los 4265 clips anotados por

el anotador 1, de los cuales 1010 fueron catalogados como Bloqueo y 142 como Repetición de palabra. Cada especialista revisó 505 clips etiquetados como Bloqueo y 71 como Repetición de palabra para asegurar la consistencia en la anotación. Tras esta revisión, realizaron las correcciones necesarias para mejorar la precisión de las anotaciones realizadas. En la Tabla 5 se observa la distribución total de cada etiqueta de tartamudez identificada entre los 4265 audios. Es importante recalcar que por cada clip de audio pudo haber más de una etiqueta anotada.

Etiqueta	Cantidad	Porcentaje de Distribución (%)
Prolongación	549	12,87%
Bloqueo	424	9,94%
Repetición de sonido	484	11,35%
Repetición de palabra	103	2,42%
Sin palabras tartamudeadas	2688	63,02%
Interjección	17	0,40%
Inseguro	67	1,37%
Baja calidad de audio	3	0,06%
Difícil de entender	26	0,53%
Pausa natural	317	6,49%
Música	156	3,19%
Sin habla	52	1,06%
<b>Total</b>	<b>4886</b>	<b>100,00%</b>

Tabla 5: Distribución total de etiquetas anotadas.

## Capítulo V - Modelo de detección y clasificación

### 5.1 Descripción del modelo

Para la extracción de características de los clips, se empleó el modelo WAV2VEC 2.0 XLSR53 (Conneau et al., 2021). En Mohapatra et al. (2022), mientras que WAV2VEC 2.0 Base se usa en tareas de menor demanda computacional y en entornos donde la latencia no es crítica, WAV2VEC 2.0 Large se entrena con un dataset mucho más amplio, ofreciendo mayor precisión en comparación con el modelo Base. El modelo presentado en Mohapatra et al. (2022) se entrenó en un solo idioma. Por esta razón, en este estudio se optó por utilizar XLSR53, debido a su amplio entrenamiento en entornos multilingües. Para la clasificación de la tartamudez, se adoptó un modelo basado en DisfluencyNet, también mencionado en Mohapatra et al. (2022). Los componentes principales de la arquitectura del modelo, mostrada en la Figura 7, incluyen capas de convolución 2D con max-pooling y capas completamente conectadas, como se detalla en la Figura 7. Las capas convolucionales utilizan la función de activación ReLU para modelar las no linealidades y aplican un dropout con una probabilidad del 0.5. La salida de la última capa de convolución es aplanada y alimentada a capas completamente conectadas que emplean una función de activación ReLU con fugas. Se usaron tres capas de convolución 2D con un tamaño de núcleo de (3,3). Después de procesar la salida de la tercera capa convolucional, la estructura de la capa sigue un diseño similar a un codificador, reduciendo la cantidad de salidas en cada capa posterior. La clasificación se optimiza mediante la pérdida de *Binary cross-entropy*.

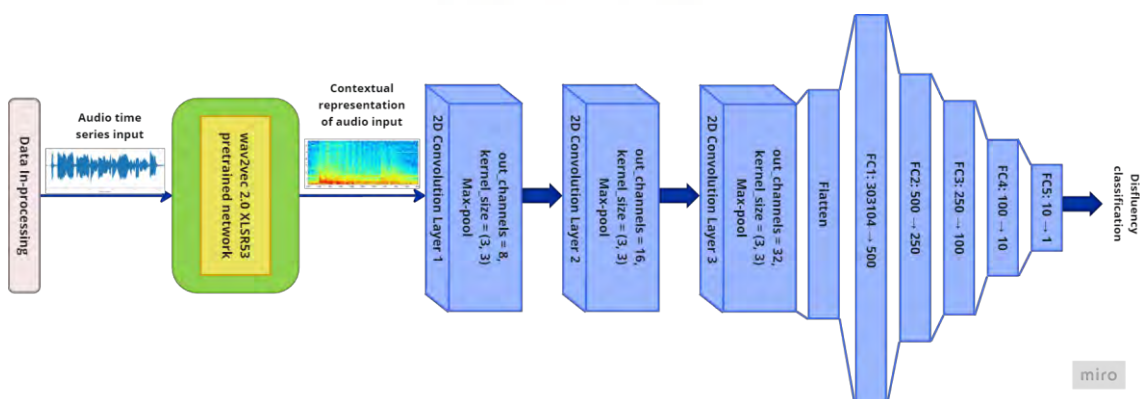


Figura 7. Arquitectura del modelo

## 5.2 División del Dataset

Como se había establecido previamente, el total de audios disponibles para este estudio era de 4265, cada uno segmentado en clips de tres segundos para el análisis del modelo. No obstante, se excluyeron las interjecciones de la anotación debido a que constituyen menos del 0.40% del corpus, lo que las hace poco significativas para el modelo. Adicionalmente, se descartaron 25 audios que no cumplían con la duración mínima requerida de tres segundos, resultando así en 4223 clips aptos para la anotación.

Para la organización del dataset en conjuntos de entrenamiento, validación y prueba, se realizaron dos cortes específicos. Inicialmente, se separó el 34% del total de los datos para el conjunto de entrenamiento. El 66% restante se dividió en partes iguales para los conjuntos de validación y prueba, asignando el 50% de este segmento a cada uno.

Los cortes de datos se llevaron a cabo individualmente por cada participante, asignando a cada uno su propio conjunto de entrenamiento, validación y prueba. Posteriormente, se agruparon todos los conjuntos de cada categoría para establecer el número total de clips en cada uno. En detalle, el conjunto de entrenamiento se compuso de 1427 clips de audio, como se muestra en la Tabla 6. De manera similar, los conjuntos de validación y prueba contaron con 1407 y 1389 clips de audio, respectivamente.

Etiqueta	Cantidad	Porcentaje de Distribución (%)
Prolongación	184	12,89%
Bloqueo	144	10,09%
Repetición de sonido	159	11,14%
Repetición de palabra	36	2,52%
Sin palabras tartamudeadas	912	63,91%
Audios descartados	-8	-0,56%
<b>Total</b>	<b>1427</b>	<b>100,00%</b>

Tabla 6: Distribución de audios para datos de entrenamiento por cada etiqueta de tartamudez utilizado para el modelo.

### 5.3 Métricas de evaluación

Para validar nuestros resultados, se eligieron las métricas de *Precision*, *Recall*, y F1 score, los cuales se calculan de la siguiente manera:

$$Precision = \frac{TP}{TP + FP}, Recall = \frac{TP}{TP + FN}, F1\ score = \frac{2}{(1/Precision) + (1/Recall)} \quad (1)$$

Donde TP, TN, FP, FN son Verdadero Positivo (True Positive), Verdadero Negativo (True Negative), Falso Positivo (False Positive), Falso Negativo (False Negative) respectivamente.

### 5.4 Estrategias para lidiar con el desbalanceamiento de datos

Como se puede ver en la Tabla 6, existe un gran desbalance entre la cantidad de clips que no tienen ninguna palabra con disfluencias (Sin palabras tartamudeadas). Por esta razón, se experimentó con algunas estrategias de balanceamiento de datos.

#### 5.4.1 Downsampling

Posteriormente al entrenamiento inicial, se realizó un proceso de downsampling en el conjunto de datos de entrenamiento para equilibrar las clases. Este ajuste, consistió en reducir la cantidad de muestras de la clase mayoritaria, específicamente las etiquetadas como '*Sin palabras tartamudeadas*', hasta alcanzar una cantidad equivalente a las muestras de clases con

disfluencias. Este procedimiento fue implementado con el objetivo de minimizar los sesgos asociados a la desproporción inicial de las clases y mejorar la capacidad del modelo para detectar y clasificar las disfluencias de manera más efectiva. El enfoque de *downsampling* pretende proporcionar un entrenamiento más equitativo y representativo, facilitando así la evaluación comparativa del modelo bajo condiciones de mayor equidad entre las clases de disfluencias y no disfluencias.

#### 5.4.2 Aumento de datos por modificación de audios

Dado el uso de grandes datasets en estudios previos, como los de Mohapatra et al. (2022) que emplearon FluencyBank para validación y SEP-28K para entrenamiento, el estudio empleó la técnica de upsampling para aumentar la representación de la clase minoritaria, buscando equilibrar el conjunto de datos mediante el aumento de datos de esta clase específica. Para abordar esto, se implementó una estrategia de aumento de datos, inspirada en la metodología de Mohapatra et al. (2023). Se emplearon herramientas avanzadas como WavAugment y TorchAudio para enriquecer nuestros datos.

Las técnicas de aumento de datos aplicadas incluyeron:

- **Pitch Shifting:** Modificación de las características vocales mediante ajustes en la tonalidad. Se variaron los tonos en un rango de -300 a 300, permitiendo una diversidad de modulaciones vocales que imitan variaciones naturales en la voz humana.
- **Reverb Addition:** Añadido de efectos de reverberación para simular diferentes ambientes acústicos. Se ajustaron los parámetros de reverberación para alterar el tamaño aparente de la "habitación", con valores aleatorios entre 0 y 100.
- **Additive White Noise:** Incorporación de ruido de fondo para aumentar la robustez del modelo frente a condiciones acústicas no ideales.

Se diseñaron dos estrategias específicas de aumento de datos:

1. La primera estrategia aplicaba ajustes de Pitch y Reverb a cada audio.

2. La segunda estrategia combinaba Pitch, Reverb y Additive White Noise para una simulación más compleja de entornos acústicos.

Estas dos estrategias se utilizaron para la clase minoritaria (es decir, las disfluencias por tartamudez). En cada aumento de datos, se aplicó de manera aleatoria una de estas dos estrategias para cada audio.

#### 5.4.3 Aumento de datos con clonación de voz y reglas

Otra estrategia para aumentar los datos de la clase minoritaria consistió en aumentar los datos clonando las voces de los participantes de manera que se generen nuevos audios con disfluencias. De esta manera, se utilizó el modelo de clonación de voz XTTS-v2 (Casanova, et al, 2024)<sup>2</sup>, 1 clip de 3 segundos para cada participante y un conjunto de textos de noticias que no fueron anotados en el presente trabajo, con los cuales se generaron los nuevos audios.

En experimentos preliminares, se percibió que la clonación de voz solamente emula el tono de voz de los participantes y no las potenciales disfluencias al hablar, por tanto, se tuvo que realizar algunas modificaciones en los textos de manera que puedan generarse audios con disfluencias. Las reglas que aplican las modificaciones son descritas a continuación:

- *Prolongación*: Se agregaron repeticiones de las letras consonantes “m”, “n”, “l” y “s”. Estas consonantes fueron escogidas siguiendo el criterio de los expertos. De esta forma, sentencias como “Muchos jugadores estuvieron de acuerdo” fueron cambiadas a “Mmmmmmmmuchos jugadores estuvieron de acuerdo”.
- *Bloqueo*: En este caso, se agregaron audios sin sonido en cualquier punto de los audios generados previamente. Los audios sin sonido tuvieron una duración entre 0.75 y 2 segundos.
- *Repetición de palabra*: Se seleccionaron palabras de manera aleatoria las cuales fueron aumentadas 2 ó 3 veces. De esta forma, la posible sentencia generada podría ser “Muchos jugadores jugadores estuvieron de acuerdo”.

---

<sup>2</sup> Disponible en <https://huggingface.co/coqui/XTTS-v2>

- *Repetición de sonido*: Se seleccionaron palabras de manera aleatoria y se extrajeron las sílabas de las mismas. Luego se seleccionó una sílaba de manera aleatoria la cual fue repetida entre 2-5 veces. Por ejemplo, una posible sentencia generada podría ser “Muchos ju-ju-ju-gadores estuvieron de acuerdo”.

Con estas modificaciones en los textos, se procedió a generar los nuevos audios usando el modelo de clonación de voz por cada participante. Sin embargo, dado que el dataset está compuesto por clips de 3 segundos, se tuvo que extraer clips de 3 segundos que incluyeran las disfluencias insertadas. Para conseguir esto, se usó la herramienta de alineamiento audio-texto *timething*<sup>3</sup> con la cual se encontraron los puntos en el audio en que ocurrieron las disfluencias generadas. Una vez encontrado algún punto con disfluencia, se procedió a identificar los puntos de inicio y fin con duración de 3 segundos en total.

## 5.5 Resultados de los experimentos

### 5.5.1 Resultados Preliminares

Se clasificaron cuatro tipos de disfluencias asociadas con la tartamudez utilizando el modelo presentado. Los resultados, como se detallan en la Tabla 7, incluyen las métricas de *precision*, *recall* y F1 para la evaluación de prueba, y para cada evento (presencia o ausencia de tartamudez) dentro de cada tipo de disfluencia.

De manera general, para todos los tipos de disfluencias, el modelo demostró una alta efectividad en detectar las instancias donde no hay tartamudez. Los valores de *precision* oscilaron entre 0.83 y 0.96, los de *recall* entre 0.73 y 1.00, y los de F1 entre 0.78 y 0.98.

Sin embargo, para las instancias donde sí hay disfluencias, los valores para cada métrica fueron más bajos, siendo particularmente notorios en la categoría de "*Repetición de palabra*". Este fenómeno puede atribuirse a la escasa disponibilidad de datos para este tipo de disfluencia en comparación con los datos de no tartamudez.

---

<sup>3</sup> Disponible en <https://github.com/feldberlin/timething>

Train_Data		Test		
Etiqueta	IsStuttered	Precision	Recall	F1
Prolongación	No	0,83	0,73	0,78
	Sí	0,18	0,30	0,23
Bloqueo	No	0,86	0,81	0,83
	Sí	0,14	0,19	0,16
Repetición de sonido	No	0,85	0,79	0,82
	Sí	0,17	0,22	0,19
Repetición de palabra	No	0,96	1,00	0,98
	Sí	0,00	0,00	0,00

Tabla 7: Resultados de las disfluencias en los audios preprocesados.

### 5.5.2 Resultados de Downsampling

Los resultados de este ajuste se pueden observar detalladamente en la Tabla 8, donde se presentan las métricas de rendimiento post-downsampling. Para las disfluencias de prolongación y bloqueo y repetición de sonido, se registró un incremento en el recall, indicando que el modelo, tras el balance de datos, logró identificar más casos verdaderos de estas disfluencias, mientras que la precisión se mantuvo estable, la reducción más baja fue en la precisión de bloqueo de 0.14 a 0.12. Sin embargo, para las repeticiones de palabras, no se notaron mejoras; el modelo no logró identificar correctamente este tipo de disfluencia después del ajuste del balance de datos.

Train_Data		Test		
Etiqueta	IsStuttered	Precision	Recall	F1
Prolongación	No	0,82	0,51	0,63
	Sí	0,17	0,46	0,24
Bloqueo	No	0,84	0,36	0,50

	Sí	0,12	0,57	0,20
Repetición de sonido	No	0,85	0,44	0,58
	Sí	0,16	0,58	0,25
Repetición de palabra	No	0,96	1,00	0,98
	Sí	0,00	0,00	0,00

Tabla 8: Resultados de las disfluencias en los audios preprocesados después de downsampling.

### 5.5.3 Resultados de Aumento de datos

Para ampliar el conjunto de datos disponible para entrenamiento, se consideró aumentar los datos en cuatro ocasiones de la clase minoritaria. Para el primer aumento, aplicamos la fórmula descrita en la ecuación (2), que consiste en sumar los datos originales de entrenamiento (*train\_data*) con una primera tanda de datos aumentados (*data\_augmentation\_train\_data*).

A partir del segundo aumento, utilizamos la fórmula en la ecuación (3) para cada iteración subsiguiente. Esta fórmula implica añadir a cada nuevo conjunto de datos aumentados (*data\_augmentation\_train\_data*) al total acumulado de datos de entrenamiento aumentados previamente (Aumento  $x_n$ ). De este modo, cada nueva ronda de aumento incluye los datos generados en las rondas anteriores, permitiendo así una acumulación progresiva del volumen de datos para el entrenamiento:

$$\text{Aumento } x_1 = \text{train\_data} + \text{data\_augmentation\_train\_data} \quad (2)$$

$$\text{Aumento } x_{n-1} = \text{Aumento } x_n + \text{data\_augmentation\_train\_data} \quad (3)$$

Este enfoque iterativo garantiza que cada ciclo de aumento enriquezca la diversidad y cantidad del conjunto de datos, lo que es crucial para la mejora continua del modelo en tareas de detección y clasificación de disfluencias. En la Figura 8 se muestran los incrementos en el número de datos de entrenamiento

para cada tipo de disfluencia tras aplicar las estrategias de aumento de datos. La tabla detalla cómo aumentaron progresivamente las cantidades de datos desde el conjunto inicial hasta el cuarto aumento para disfluencias como prolongación, bloqueo, repetición de sonido y repetición de palabra.

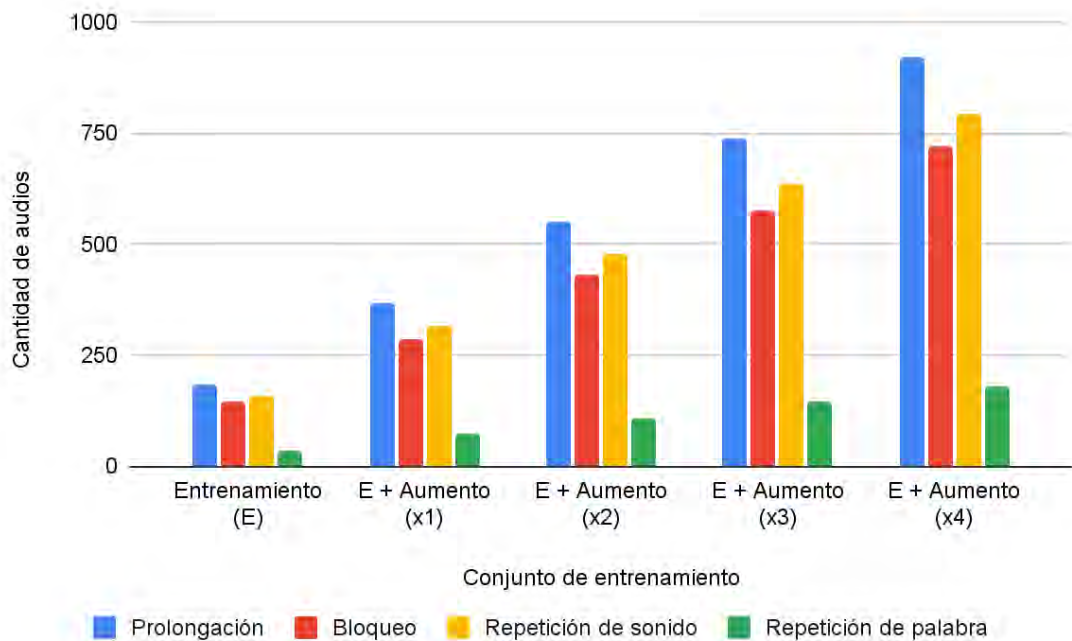


Figura 8: Relación del conjunto inicial de entrenamiento y los cuatro aumentos de datos.

### 5.5.3.1 Resultados de estrategia basada en modificación de audios

Para la primera experimentación del aumento de datos, se aplicó las dos estrategias de aumento de datos detallados en la sección 5.4.2 por separado para los datos de entrenamiento y así visualizar con qué estrategia se obtienen mejores resultados.

No obstante, al igual que en la primera experimentación, en todos los resultados para todos los tipos de disfluencias, los resultados de precisión continúan siendo bajos. Los resultados de la primera estrategia varían mostrando mejoras para algunas disfluencias. En cambio, los resultados de la

segunda estrategia, registran resultados no tan favorables, indicando que la primera estrategia es más efectiva para manejar todos los tipos de disfluencia.

Según los datos en la Figura 9, el modelo demostró eficacia al clasificar instancias sin disfluencias. Sin embargo, en las instancias con disfluencias, especialmente prolongaciones, la primera estrategia mostró mejoras en el recall durante el segundo aumento, alcanzando valores de 0.69 en prueba.

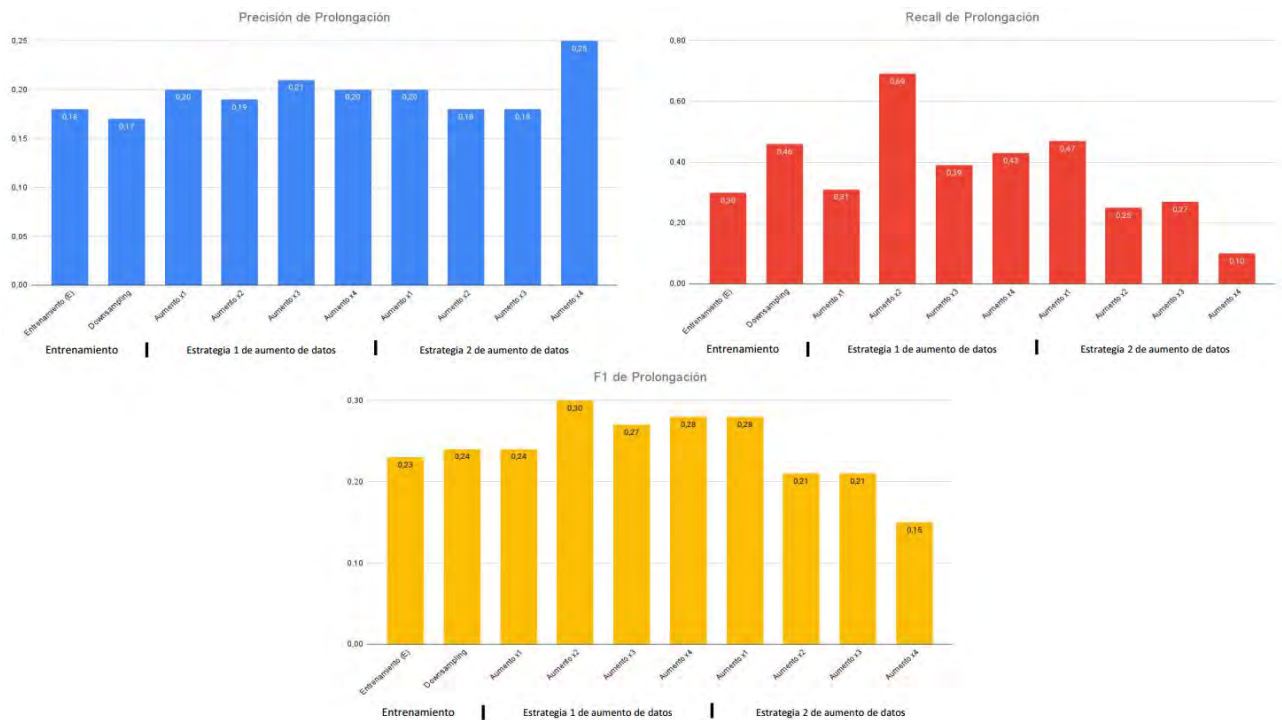


Figura 9: Resultados de prueba de las estrategias de aumento de datos por separado para Prolongación

En la Figura 10 se observa que con la estrategia 1, aunque hay mejoras en el recall que alcanza hasta 0.46 en la cuarta ronda de aumento, los valores de precisión y F1 siguen siendo bajos, aunque mejoran levemente. Esto indica que, aunque el modelo está detectando un mayor número de casos de bloqueo, también está cometiendo errores en la clasificación. En cuanto a la estrategia 2, se nota que en algunos aumentos la precisión y el F1 son incluso más bajos, posiblemente debido al ruido adicional que impide una clasificación clara de las instancias de tartamudez.

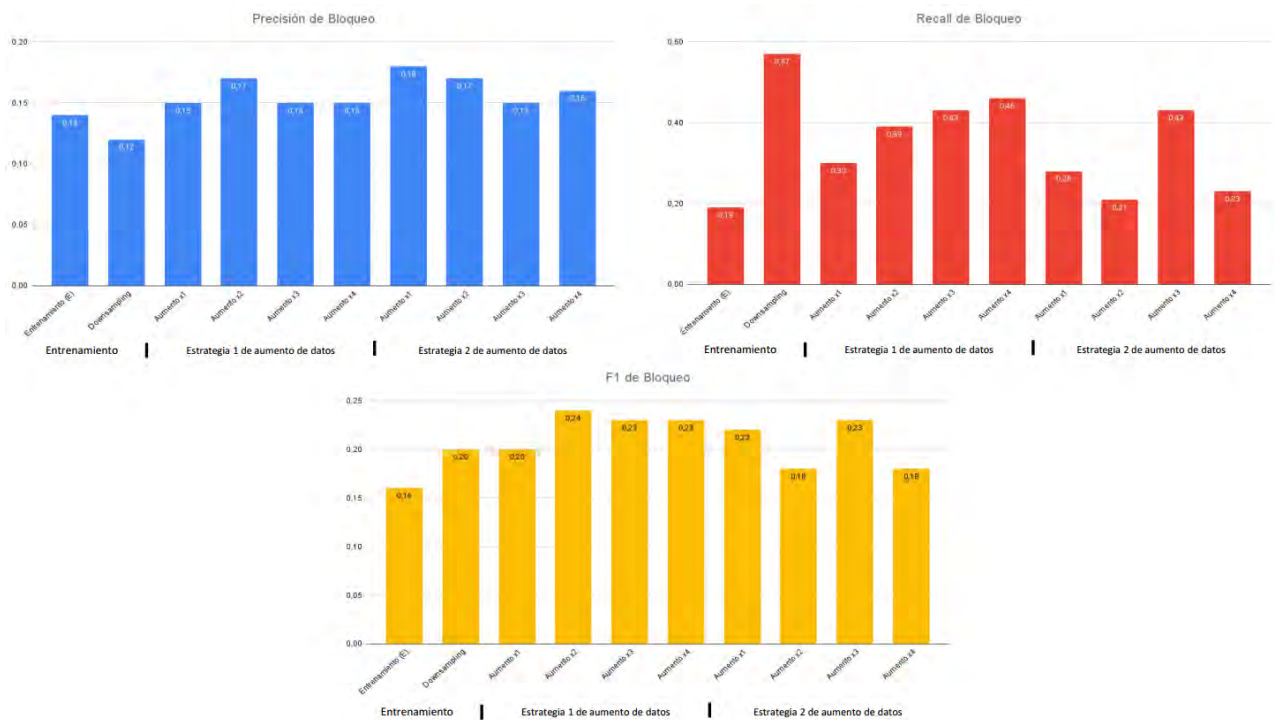


Figura 10: Resultados de prueba de las estrategias de aumento de datos por separado para *Bloqueo*.

Como se observa en la Figura 11, en el caso de los datos aumentados mediante la primera estrategia para "Repetición de sonido", las instancias con disfluencias muestran un incremento en el recall a lo largo de las rondas, alcanzando un máximo de 0.50 en la cuarta ronda en la primera estrategia.

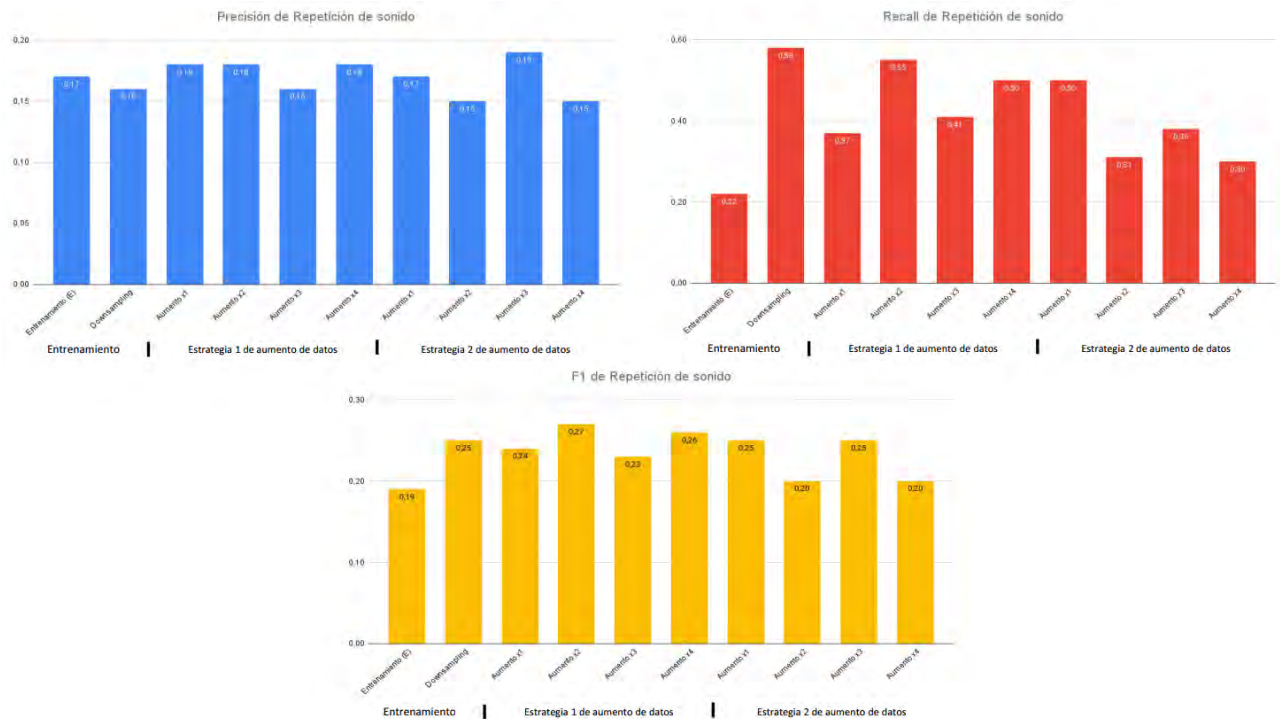


Figura 11: Resultados de prueba de las estrategias de aumento de datos por separado para Repetición de sonido.

Los resultados de las estrategias de aumento de datos para "Repetición de palabra" se presentan en la Figura 12. En el caso de la primera estrategia, el recall mejora levemente en el segundo aumento. En la segunda estrategia sucede lo mismo, pero muestra caídas significativas en los siguientes dos aumentos. Estos resultados podrían deberse a la poca cantidad de datos y al posible efecto confuso del ruido adicional sobre la clasificación de la disfluencia.

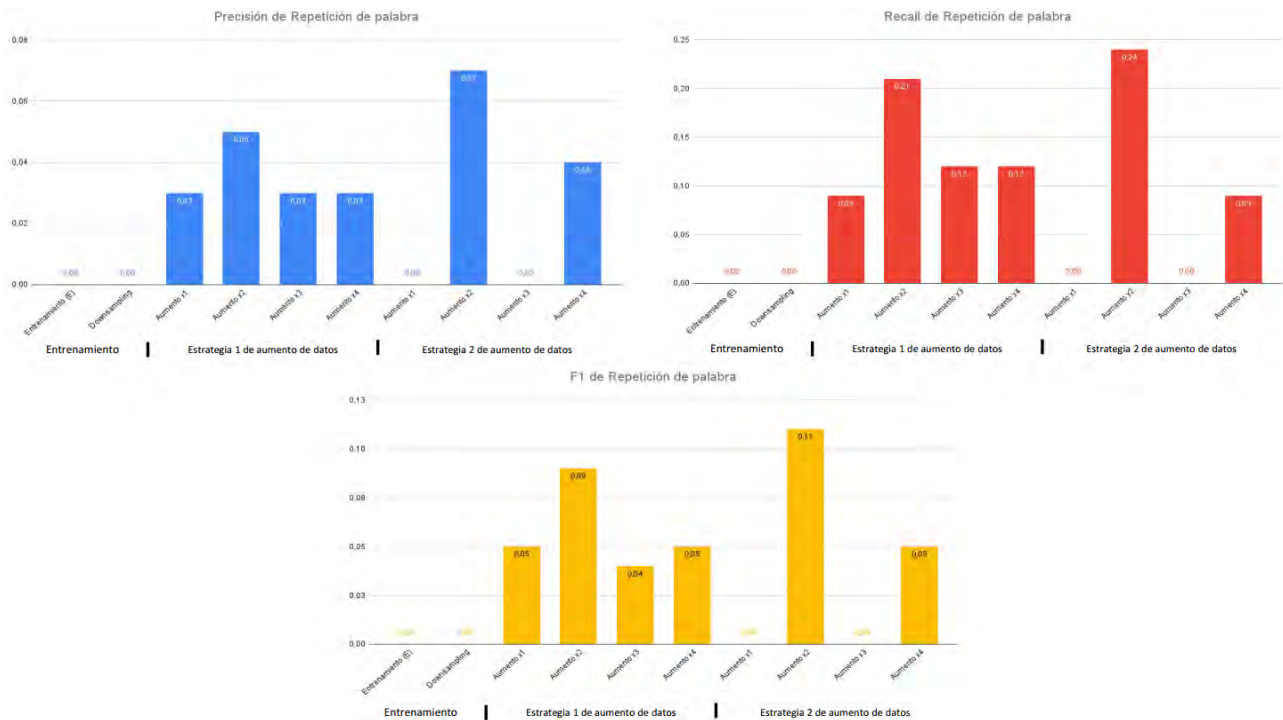


Figura 12: Resultados de prueba de las estrategias de aumento de datos por separado para *Repetición de palabra*.

Luego de la primera experimentación, se realizó una segunda experimentación, esta vez con las estrategias 1 y 2 de aumento de datos por modificación de audio aplicadas al conjunto de datos de entrenamiento original de manera aleatoria.

Al igual que en evaluaciones anteriores para datos de entrenamiento original y downsampling, el modelo muestra una alta eficiencia al clasificar instancias sin disfluencias para los datos aumentados de entrenamiento, con métricas de precisión, recall y F1 relativamente altas. Sin embargo, para las instancias con disfluencias, en todos tipos, los valores siguen siendo bajos, reflejando dificultades continuas en la detección de estas condiciones específicas.

En la Figura 13 se detallan estos resultados para la clasificación de "Prolongación". Para los datos de entrenamiento aumentados con ambas estrategias, no se observan mejoras, por el contrario, los valores tienden a reducir o a mantenerse.

A pesar de la implementación de varias rondas de aumento de datos, estos no parecen haber influido significativamente en la mejora de las métricas relacionadas con los bloqueos, como se visualiza en la Figura 14. Esto sugiere que los métodos de aumento de datos empleados podrían no estar capturando las características necesarias para mejorar la detección de esta específica disfluencia.

Por otro lado, en la Figura 15, se presentan los resultados de las métricas para "Repetición de sonido". Las variaciones entre cada aumento son generalmente mínimas en la mayoría de las métricas. No obstante, se nota un incremento notable en el recall de prueba entre el tercer y cuarto aumento, indicando una posible mejora en la detección de esta disfluencia.

Finalmente, en la Figura 16 presenta las métricas de evaluación para las instancias con disfluencia de "Repetición de palabra". Se observa que los valores de precisión, recall y F1 para las instancias con disfluencia mejoran ligeramente con cada aumento de datos, pero siguen siendo bajos. Por ejemplo, en el último aumento, el recall alcanza un máximo de 0.36 y el F1 de 0.14, indicando una limitada eficacia en la clasificación de estas disfluencias. Es notable que el aumento de datos eleve las métricas, dado que antes del aumento de datos eran 0, un logro que no se observó con el downsampling. Esto muestra un avance inicial en la identificación de instancias con disfluencia de "Repetición de palabra" que antes no eran detectadas por el modelo.

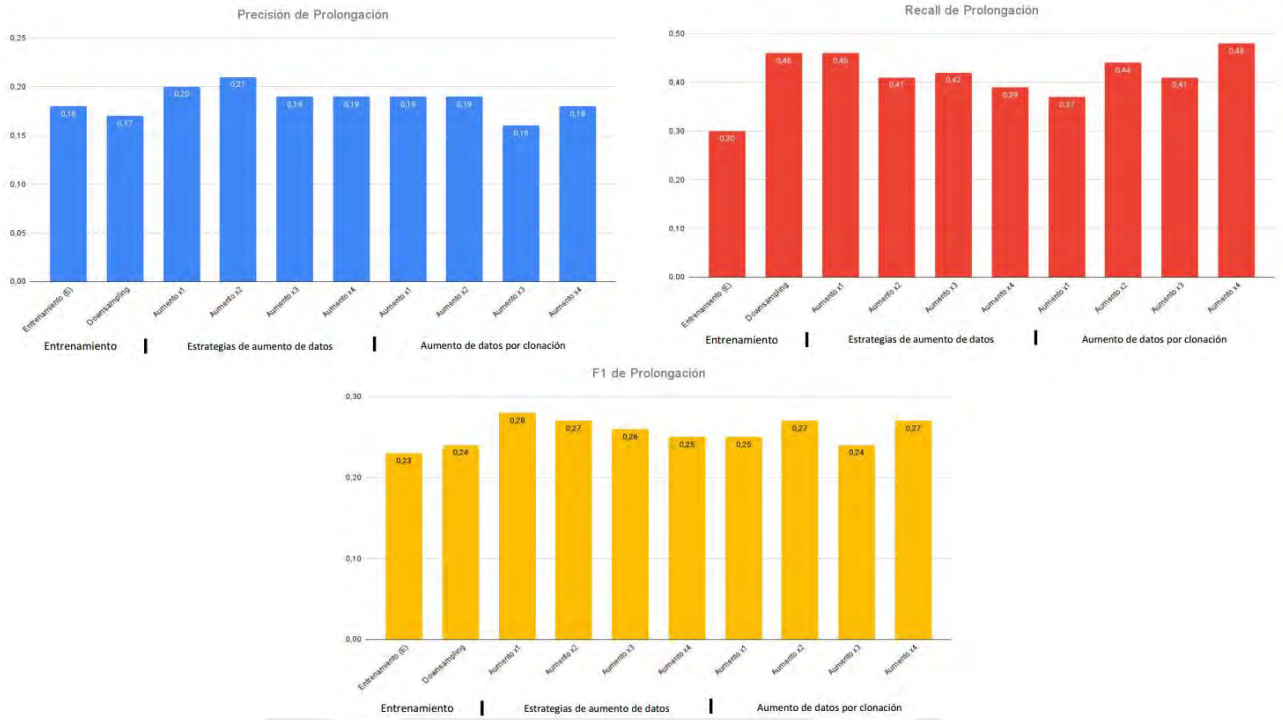


Figura 13: Resultados de prueba de todos los conjuntos de datos para Prolongación

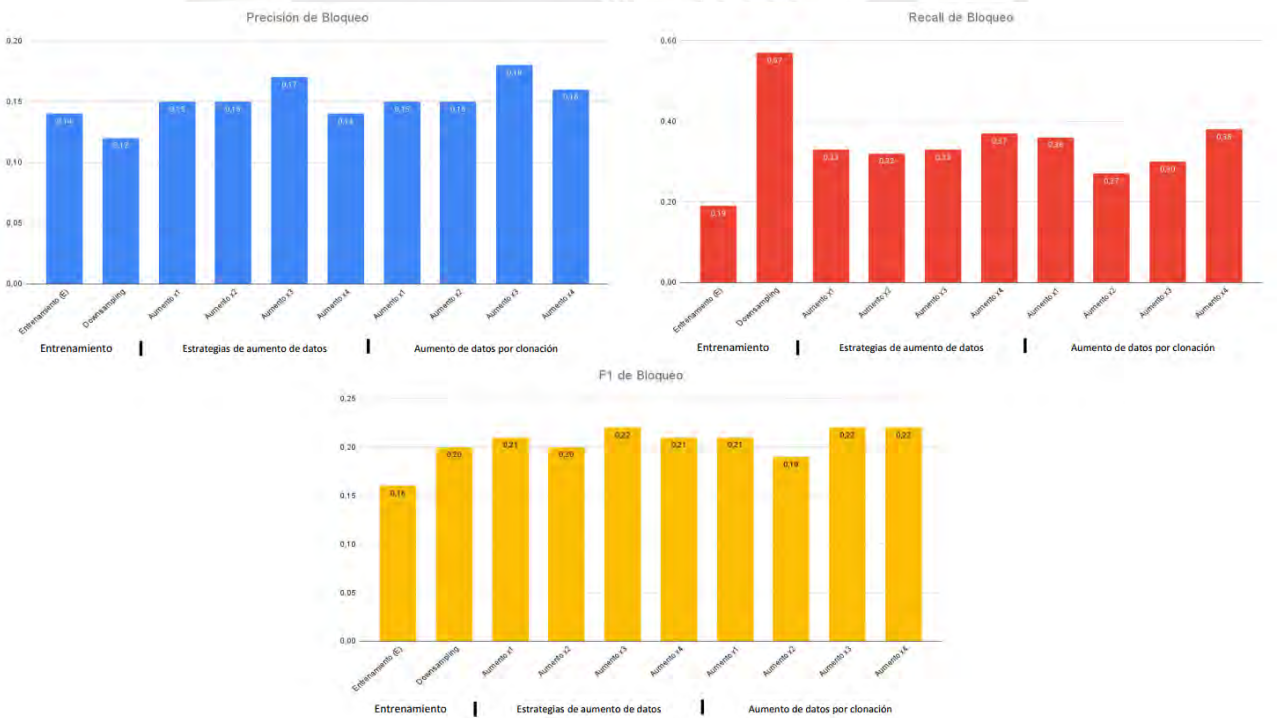


Figura 14: Resultados de prueba de todos los conjuntos de datos para Bloqueo

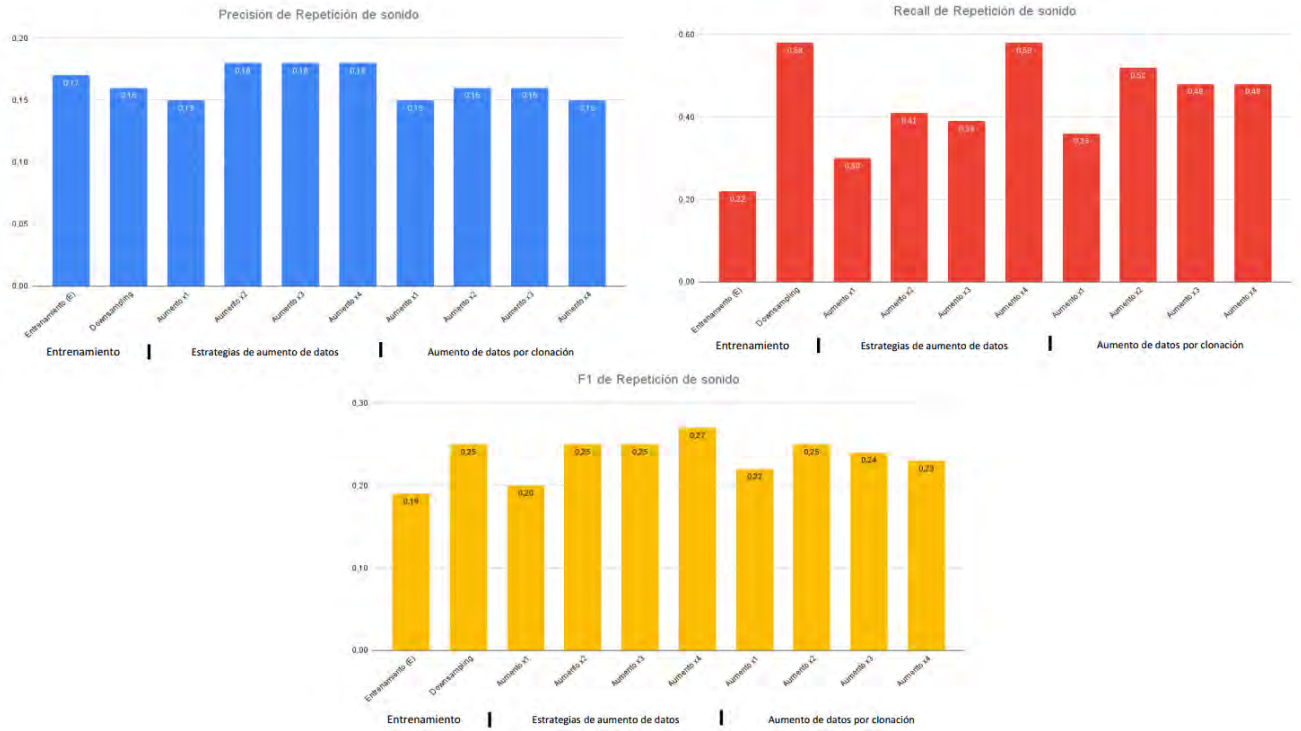


Figura 15: Resultados de prueba de todos los conjuntos de datos para Repetición de sonido

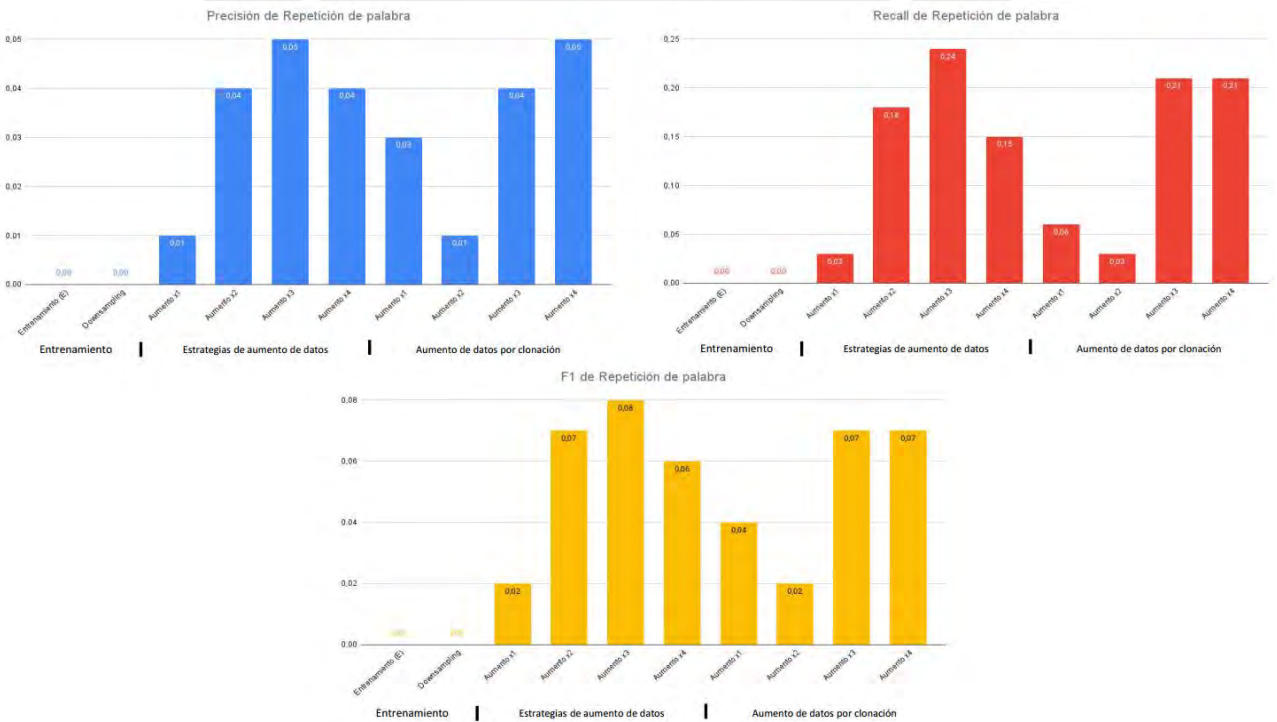


Figura 16: Resultados de prueba de todos los conjuntos de datos para Repetición de palabra

### **5.5.3.2 Resultados de la estrategia basada en clonación de voz**

En este punto se analizan los resultados de la estrategia de aumento de datos basada en la clonación de voz. En las tablas 23, 24, 25 y 26 se observan los resultados de las métricas analizadas por cada aumento realizado con la clonación de voz.

En la Figura 13 se detallan los resultados de los datos de clonación para "Prolongación", que muestran una mejora en el recall con cada aumento que se genera, superando los resultados de las métricas de los datos de entrenamiento aumentados. En el caso de los datos de clonación, se muestran mejoras para el recall, llegando a 0.48.

En la Figura 14, no obstante, es relevante destacar que los resultados de las métricas para el conjunto de datos de "Bloqueo" de clonación aumentados también muestra una similitud con las métricas de los datos de entrenamiento aumentados.

Para la "Repetición de sonido", se observa que en la Figura 15 es relevante destacar que la precisión en la clasificación de las instancias sin disfluencias en los datos de clonación aumentados muestra una disminución en comparación con los otros conjuntos de entrenamiento.

Finalmente, en la Figura 16 muestra los resultados de las métricas para "Repeticiones de palabras". En el conjunto de datos de clonación aumentados, las métricas exhiben un comportamiento distinto al observado en los datos de entrenamiento aumentados. A diferencia de los datos de entrenamiento aumentados donde el recall tiende a aumentar, en los datos de clonación aumentados el recall varía, mostrando fluctuaciones en lugar de una tendencia clara de mejora.

## Capítulo VI - Conclusiones

El propósito principal de esta tesis de maestría fue desarrollar recursos computacionales para apoyar la detección y clasificación automática de disfluencias producidas por tartamudez en el idioma español, con objetivos específicos que abarcan desde la recolección y anotación de audios hasta la implementación de un modelo efectivo y estrategias de aumento de datos. A continuación, se procederá a describir las conclusiones obtenidas a partir de la investigación realizada:

**Recolección y Preprocesamiento de Datos (OE1):** Audios de cinco participantes con tartamudez fueron recolectados y grabados, cumpliendo con los estándares del dataset SEP-28K y con la colaboración de dos anotadores especialistas en tartamudez. Este proceso aseguró una base sólida para el desarrollo del modelo, aunque la naturaleza controlada de las grabaciones limitó la diversidad de las disfluencias capturadas.

**Desarrollo del Modelo de Detección y Clasificación (OE2):** Se implementó con éxito el modelo DisfluencyNet, que demostró su capacidad para identificar y clasificar las disfluencias en español. Sin embargo, siendo el Español diferente al inglés, el rendimiento diferenciado del modelo es esperado, lo que resalta la necesidad de más datos y posiblemente una mayor adaptación cultural y lingüística de las herramientas de aprendizaje automático.

**Estrategias de Aumento de Datos (OE3):** Se implementaron las técnicas de aumento de datos, como la variación de pitch, la adición de reverberación y el ruido blanco para enriquecer el dataset. Aunque se duplicó la cantidad de datos disponibles, los resultados fueron variados, y se observó un incremento en el recall en ciertos aumentos. Los datos obtenidos tras aplicar downsampling para balancear la clase minoritaria frente a los datos sin tartamudez mostraron mejoras frente al entrenamiento con los datos originales. En el caso de la clonación de voz, aunque la calidad del audio generado por la clonación presentó limitaciones significativas y similares a las estrategias de aumento de datos, mostró mejoras en algunas métricas.

**Limitaciones:** La limitada diversidad en los datos debido a las condiciones controladas de recolección y las restricciones inherentes a las técnicas de aumento de datos subrayan la necesidad de expandir y diversificar el corpus de entrenamiento. Esto es crucial para optimizar el rendimiento del modelo en entornos de habla natural y espontánea y para abordar la variabilidad de disfluencias que se encuentran en diferentes dialectos y acentos del español.

Estos resultados y desafíos destacan la importancia de continuar mejorando las técnicas de aumento de datos y de explorar nuevas metodologías para recopilar y procesar datos en idiomas distintos al inglés, con el objetivo de crear modelos más robustos y adaptativos que reflejen la complejidad del habla real y sus variaciones.



## Capítulo VII - Trabajos Futuros

De cara a futuras investigaciones, se planea ampliar la anotación de disfluencias en contextos de habla espontánea, realizando entrevistas con personas que presentan tartamudez para capturar una mayor variedad y número de instancias de disfluencias naturales. Esto ayudará a mejorar la representatividad y la diversidad del corpus.

Adicionalmente, se abordarán los 23 300 clips de audio restantes de los 28 005 clips inicialmente generados, de los cuales solo se han tratado y anotado 4265 hasta el momento. Se investigará si el trabajo con estos audios restantes puede conducir a mejores resultados en la detección y clasificación de disfluencias. Por otra parte, también se contempla la necesidad de aumentar el número de participantes para mejorar los resultados.

También se explorarán métodos avanzados de clonación de voz con el objetivo de mejorar la calidad del audio, asegurando que estos sean adecuados para su uso en estrategias de aumento de datos destinadas a enriquecer los datasets de personas con tartamudez.

Finalmente, se evaluarán otras técnicas de modificación de audios, más allá de los ajustes actuales de pitch, reverberación y ruido blanco. Estas nuevas estrategias (como *Time Stretching* o *Random Cropping*) serán analizadas para determinar su utilidad en el mejoramiento de modelos de detección y clasificación de disfluencias. Estas iniciativas serán clave para avanzar en la precisión y eficiencia de los modelos computacionales aplicados a la tartamudez en el idioma español.

## Referencias

Alharbi, Sadeen (2018) Automatic Framework to Aid Therapists to Diagnose Children who Stutter. PhD thesis, University of Sheffield.

Alharbi, S., Hasan, M., Simons, A. J., Brumfitt, S., & Green, P. (2020). Sequence labeling to detect stuttering events in read speech. *Computer Speech & Language*, 62, 101052. <https://doi.org/10.1016/j.csl.2019.101052>

Ambrose, N. G., & Yairi, E. (1999). Normative Disfluency Data for Early Childhood Stuttering. *Journal of Speech Language and Hearing Research*

Arbajian, P., Hajja, A., Raś, Z. W., & Wieczorkowska, A. A. (2019). Effect of speech segment samples selection in stutter Bloqueo detection and remediation. *Journal of Intelligent Information Systems*, 53(2), 241–264. <https://doi.org/10.1007/s10844-019-00546-z>

Bayerl, S. P., Gerczuk, M., Batliner, A., Bergler, C., Amiriparian, S., Schuller, B., Nöth, E., & Riedhammer, K. (2023). Classification of stuttering – the compare challenge and beyond. *Computer Speech & Language*, 81, 101519. <https://doi.org/10.1016/j.csl.2023.101519>

Casanova, E., Davis, K., Gölge, E., Gökner, G., Gulea, I., Hart, L., Aljafari, A., Meyer, J., Morais, R., Olayemi, S., & Weber, J. (2024). XTTS: A Massively Multilingual Zero-Shot Text-to-Speech Model. *INTERSPEECH 2024*. <https://doi.org/10.48550/arXiv.2406.04904>

Conneau, A., Baevski, A., Collobert, R., Mohamed, A., & Auli, M. (2021). Unsupervised cross-lingual representation learning for speech recognition. *Interspeech 2021*. <https://doi.org/10.21437/interspeech.2021-329>

Craig, A. The Developmental Nature and Effective Treatment of Stuttering in Children and Adolescents. *Journal of Developmental and Physical Disabilities* 12, 173–186 (2000). <https://doi.org/10.1023/A:1009463703647>

Eren, G., & The Coqui TTS Team. (2021). Coqui TTS (Version 1.4) [Computer software]. <https://doi.org/10.5281/zenodo.6334862>

Falotico, R., & Quatto, P. (2014). Fleiss' Kappa Statistic Without paradoxes. *Quality & Quantity*, 49(2), 463–470. <https://doi.org/10.1007/s11135-014-0003-1>

Feng, S. Y., Gangal, V., Lee, J., Chandar, S., Vosoughi, S., Mitamura, T., & Hovy, E. (2021). A survey of Data Augmentation Approaches for NLP. arXiv (Cornell University). <https://doi.org/10.48550/arxiv.2105.03075>

Gupta, S., Shukla, R.S., Shukla, R.K., & Verma, R. (2020). Deep Learning Bidirectional LSTM based Detection of Prolongación and Repetition in Stuttered Speech using Weighted MFCC. *International Journal of Advanced Computer Science and Applications*, 11.

Hosseini, R., Walsh, B., Tian, F., & Wang, S. (2018). An fNIRS-based feature learning and classification framework to distinguish hemodynamic patterns in children who stutter. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 26(6), 1254–1263. <https://doi.org/10.1109/tnsre.2018.2829083>

Jouaiti, M., & Dautenhahn, K. (2022). Dysfluency classification in speech using a biological sound perception model. 2022 9th International Conference on Soft Computing & Machine Intelligence (ISCM). <https://doi.org/10.1109/iscmi56532.2022.10068490>

Kourkounakis, T., Hajavi, A., & Etemad, A. (2020). Detecting multiple speech disfluencies using a deep residual network with bidirectional long short-term memory. ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). <https://doi.org/10.1109/icassp40776.2020.9053893>

Kourkounakis, T., Hajavi, A., & Etemad, A. (2021). FluentNet: End-to-end detection of stuttered speech disfluencies with Deep Learning. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29, 2986–2999. <https://doi.org/10.1109/taslp.2021.3110146>

Lea, C. H., Mitra, V., Joshi, A., Kajarekar, S., & Bigam, J. P. (2021). SEP-28K: a dataset for stuttering event detection from podcasts with people who stutter. arXiv (Cornell University). <https://doi.org/10.48550/arxiv.2102.12394>.

Levis, J. M., & Suvorov, R. (2012). Automatic speech recognition. The Encyclopedia of Applied Linguistics. <https://doi.org/10.1002/9781405198431.wbeal0066>

Mahesh, B. (2020). Machine learning algorithms-a review. International Journal of Science and Research (IJSR).[Internet], 9(1), 381-386.

Manaut-Gil, E. (2005). Tartamudez del Desarrollo Y tartamudez adquirida. Semejanzas y Diferencias. Revista de Neurología, 40(10), 587. <https://doi.org/10.33588/rn.4010.2004487>

Mishra, N., Gupta, A. & Vathana, D. Optimization of stammering in speech recognition applications. Int J Speech Technol 24, 679–685 (2021). <https://doi.org/10.1007/s10772-021-09828-w>

Mohapatra, P., Pandey, A., Islam, B., & Zhu, Q. (2022). Speech disfluency detection with contextual representation and data distillation. Proceedings of the 1st ACM International Workshop on Intelligent Acoustic Systems and Applications. <https://doi.org/10.1145/3539490.3539601>

Mohapatra, P., Islam, B., Islam, M. T., Jiao, R., & Zhu, Q. (2023). Efficient stuttering event detection using Siamese networks. ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). <https://doi.org/10.1109/icassp49357.2023.10094692>

Murugan, K., Cherukuri, N. K., & Donthu, S. S. (2022). Efficient recognition and classification of stuttered word from speech signal using Deep Learning Technique. 2022 IEEE World Conference on Applied Intelligence and Computing (AIC). <https://doi.org/10.1109/aic55036.2022.9848868>

Phani Bhushan S, Vani H Y, D K Shivkumar, Sreeraksha M R, 2021, Stuttered Speech Recognition using Convolutional Neural Networks, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) NCCDS – 2021 (Volume 09 – Issue 12)

Raj, A. (2019). A review on machine learning algorithms. International Journal for Research in Applied Science and Engineering Technology, 7(6), 792–796. <https://doi.org/10.22214/ijraset.2019.6138>

Riad, R., Bachoud-Lévi, A.-C., Rudzicz, F., & Dupoux, E. (2020). Identification of primary and collateral tracks in stuttered speech. arXiv. <https://doi.org/10.48550/arXiv.2003.01018>

Rodríguez Morejón, A. (2000). Intervención Temprana en Tartamudez: Criterios para Tomar Decisiones. *Logopedia, Foniatria y Audiología*, 136-150.

Seyfeddinipur, M. (2006). *Disfluency: interrupting speech and gesture*. Wageningen: Max Planck Institute for psycholinguistics.

Shah, S. K., & Nautiyal, A. (2022). Retracted: Signal Quality Assessment for speech recognition using deep convolutional neural networks. 2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon). <https://doi.org/10.1109/mysurucon55714.2022.9972718>

Sheikh, S. A., Sahidullah, M., Hirsch, F., & Ouni, S. (2021). StutterNet: Stuttering Detection Using Time Delay Neural Network. 2021 29th European Signal Processing Conference (EUSIPCO). <https://doi.org/10.23919/eusipco54536.2021.9616063>

Sheikh, S. A., Sahidullah, M., Hirsch, F., & Ouni, S. (2023). Advancing stuttering detection via data augmentation, class-balanced loss and multi-contextual deep learning. *IEEE Journal of Biomedical and Health Informatics*, 27(5), 2553–2564. <https://doi.org/10.1109/jbhi.2023.3248281>

Simmons, E.S. (2017). Stuttering. In: Volkmar, F. (eds) *Encyclopedia of Autism Spectrum Disorders*. Springer, New York, NY. [https://doi.org/10.1007/978-1-4614-6435-8\\_990-3](https://doi.org/10.1007/978-1-4614-6435-8_990-3)

Tichenor, S. E., & Yaruss, J. S. (2019). Stuttering as defined by adults who stutter. *Journal of Speech, Language, and Hearing Research*, 62(12), 4356–4369. [https://doi.org/10.1044/2019\\_jslhr-19-00137](https://doi.org/10.1044/2019_jslhr-19-00137)

Uhle, C., & Bäckström, T. (2017). Voice Activity Detection. , 185-203. [https://doi.org/10.1007/978-3-319-50204-5\\_13](https://doi.org/10.1007/978-3-319-50204-5_13).

Valente, A. R., Jesus, L. M., Hall, A., & Leahy, M. (2015). Event- and interval-based measurement of Stuttering: A Review. *International Journal of Language*

& Communication Disorders, 50(1), 14–30. <https://doi.org/10.1111/1460-6984.12113>.



## Anexos

### Anexo A. Relación de lecturas.

Día	Lectura	URL
Día 1	Lectura 1	<a href="https://elcomercio.pe/economia/coronavirus-peru-midis-creara-nuevo-bono-para-quienes-se-registren-en-padron-de-reniec-bono-familiar-universal-nndc-noticia/">https://elcomercio.pe/economia/coronavirus-peru-midis-creara-nuevo-bono-para-quienes-se-registren-en-padron-de-reniec-bono-familiar-universal-nndc-noticia/</a>
Día 1	Lectura 2	<a href="https://larepublica.pe/politica/2019/07/26/daniel-salaverry-maritzagarcia-no-integrara-unidos-por-la-republica-ni-apoyara-a-presidente-legislativo-para-la-mesa-directiva/">https://larepublica.pe/politica/2019/07/26/daniel-salaverry-maritzagarcia-no-integrara-unidos-por-la-republica-ni-apoyara-a-presidente-legislativo-para-la-mesa-directiva/</a>
Día 1	Lectura 3	<a href="https://peru21.pe/peru/sismo-de-magnitud-4-se-reporto-en-piura-senala-igp-nndc-noticia/">https://peru21.pe/peru/sismo-de-magnitud-4-se-reporto-en-piura-senala-igp-nndc-noticia/</a>
Día 1	Lectura 4	<a href="https://peru21.pe/peru/dia-de-la-educacion-inicial-cinco-recomendaciones-para-mejorar-la-relacion-padre-docente-en-el-aprendizaje-infantil-en-linea-noticia/">https://peru21.pe/peru/dia-de-la-educacion-inicial-cinco-recomendaciones-para-mejorar-la-relacion-padre-docente-en-el-aprendizaje-infantil-en-linea-noticia/</a>
Día 1	Lectura 5	<a href="https://peru21.pe/lima/lima-motociclistas-anuncian-marcha-nacional-488782-noticia/?ref=p21r">https://peru21.pe/lima/lima-motociclistas-anuncian-marcha-nacional-488782-noticia/?ref=p21r</a>
Día 1	Lectura 6	<a href="https://peru21.pe/politica/mercedes-araoz-he-decidido-renunciar-irrevocablemente-al-cargo-de-vicepresidenta-noticia/">https://peru21.pe/politica/mercedes-araoz-he-decidido-renunciar-irrevocablemente-al-cargo-de-vicepresidenta-noticia/</a>
Día 1	Lectura 7	<a href="https://peru21.pe/lima/coronavirus-peru-conozca-los-nuevos-horarios-del-metropolitano-los-corredores-complementarios-y-el-tren-electrico-coronavirus-pandemia-estado-de-emergencia-nacional-emergencia-sanitaria-noticia/">https://peru21.pe/lima/coronavirus-peru-conozca-los-nuevos-horarios-del-metropolitano-los-corredores-complementarios-y-el-tren-electrico-coronavirus-pandemia-estado-de-emergencia-nacional-emergencia-sanitaria-noticia/</a>
Día 1	Lectura 8	<a href="https://peru21.pe/lima/coronavirus-coronavirus-peru-coronavirus-en-peru-lima-registro-800-denuncias-por-robo-durante-la-cuarentena-en-mayo-coronavirus-covid-19-emergencia-sanitaria-estado-de-emergencia-nacional-noticia/">https://peru21.pe/lima/coronavirus-coronavirus-peru-coronavirus-en-peru-lima-registro-800-denuncias-por-robo-durante-la-cuarentena-en-mayo-coronavirus-covid-19-emergencia-sanitaria-estado-de-emergencia-nacional-noticia/</a>
Día 1	Lectura 9	<a href="https://elcomercio.pe/peru/andahuaylas-declaran-duelo-provincial-levantaran-bandera-media-asta-asesinato-ninas-noticia-nndc-634807-noticia/">https://elcomercio.pe/peru/andahuaylas-declaran-duelo-provincial-levantaran-bandera-media-asta-asesinato-ninas-noticia-nndc-634807-noticia/</a>
Día 1	Lectura 10	<a href="https://peru21.pe/lima/alertan-nueva-forma-de-robo-electronico-noticia/">https://peru21.pe/lima/alertan-nueva-forma-de-robo-electronico-noticia/</a>
Día 2	Lectura 11	<a href="https://elcomercio.pe/peru/apurimac-contara-con-laboratorio-para-analizar-pruebas-moleculares-de-covid-19-nnpp-noticia/">https://elcomercio.pe/peru/apurimac-contara-con-laboratorio-para-analizar-pruebas-moleculares-de-covid-19-nnpp-noticia/</a>
Día 2	Lectura 12	<a href="https://peru21.pe/lima/puente-piedra-mujer-queda-herida-tras-recibir-un-balazo-en-frustrado-asalto-a-bodega-noticia/">https://peru21.pe/lima/puente-piedra-mujer-queda-herida-tras-recibir-un-balazo-en-frustrado-asalto-a-bodega-noticia/</a>
Día 2	Lectura 13	<a href="https://peru21.pe/lima/sjm-joven-16-anos-estaria-involucrada-asesinato-mujer-bodega-pnp-nndc-485382-noticia/">https://peru21.pe/lima/sjm-joven-16-anos-estaria-involucrada-asesinato-mujer-bodega-pnp-nndc-485382-noticia/</a>
Día 2	Lectura 14	<a href="https://elcomercio.pe/peru/ancash/ancash-alcalde-de-chimbote-fue-intervenido-en-reunion-social-durante-toque-de-queda-noticia/">https://elcomercio.pe/peru/ancash/ancash-alcalde-de-chimbote-fue-intervenido-en-reunion-social-durante-toque-de-queda-noticia/</a>

Día 2	Lectura 15	<a href="https://peru21.pe/lima/policiales/villa-salvador-balean-prestamista-rostro-bodega-nndc-461972-noticia/">https://peru21.pe/lima/policiales/villa-salvador-balean-prestamista-rostro-bodega-nndc-461972-noticia/</a>
Día 2	Lectura 16	<a href="https://larepublica.pe/sociedad/2019/10/13/delincuencia-en-puente-piedra-disparan-a-madre-para-robar-su-bodega-inseguridad/">https://larepublica.pe/sociedad/2019/10/13/delincuencia-en-puente-piedra-disparan-a-madre-para-robar-su-bodega-inseguridad/</a>
Día 2	Lectura 17	<a href="https://larepublica.pe/sociedad/2019/07/23/pnp-capturan-a-banda-de-delincuentes-extranjeros-en-el-agustino-raqueteros-video/">https://larepublica.pe/sociedad/2019/07/23/pnp-capturan-a-banda-de-delincuentes-extranjeros-en-el-agustino-raqueteros-video/</a>
Día 2	Lectura 18	<a href="https://gestion.pe/peru/ciro-maguina-afirma-que-ola-del-covid-19-debe-salir-del-pais-en-dos-o-tres-semanas-noticia/">https://gestion.pe/peru/ciro-maguina-afirma-que-ola-del-covid-19-debe-salir-del-pais-en-dos-o-tres-semanas-noticia/</a>
Día 2	Lectura 19	<a href="https://gestion.pe/peru/sunedu-nuevo-reglamento-de-licenciamiento-no-representa-una-segunda-oportunidad-para-universidades-con-cese-nndc-noticia/">https://gestion.pe/peru/sunedu-nuevo-reglamento-de-licenciamiento-no-representa-una-segunda-oportunidad-para-universidades-con-cese-nndc-noticia/</a>
Día 2	Lectura 20	<a href="https://gestion.pe/economia/coronavirus-peru-produce-ni-rappi-ni-farennet-tomaron-previsiones-para-evitar-el-contagio-del-covid-19-nndc-noticia/?ref=gesr">https://gestion.pe/economia/coronavirus-peru-produce-ni-rappi-ni-farennet-tomaron-previsiones-para-evitar-el-contagio-del-covid-19-nndc-noticia/?ref=gesr</a>
Día 3	Lectura 21	<a href="https://peru21.pe/politica/martin-vizcarra-anuncia-el-cierre-del-congreso-noticia/">https://peru21.pe/politica/martin-vizcarra-anuncia-el-cierre-del-congreso-noticia/</a>
Día 3	Lectura 22	<a href="https://elcomercio.pe/peru/puno-bolivianos-varados-en-el-peru-regresan-a-su-pais-por-desaguadero-noticia/?ref=ecr">https://elcomercio.pe/peru/puno-bolivianos-varados-en-el-peru-regresan-a-su-pais-por-desaguadero-noticia/?ref=ecr</a>
Día 3	Lectura 23	<a href="https://elcomercio.pe/peru/coronavirus-en-peru-70-reclusos-del-penal-varones-trujillo-vencieron-al-covid-19-nnpp-la-libertad-recuperados-noticia/?ref=ecr">https://elcomercio.pe/peru/coronavirus-en-peru-70-reclusos-del-penal-varones-trujillo-vencieron-al-covid-19-nnpp-la-libertad-recuperados-noticia/?ref=ecr</a>
Día 3	Lectura 24	<a href="https://peru21.pe/peru/roque-benavides-tia-maria-proyecto-minero-perfectamente-factible-nndc-490500-noticia/">https://peru21.pe/peru/roque-benavides-tia-maria-proyecto-minero-perfectamente-factible-nndc-490500-noticia/</a>
Día 3	Lectura 25	<a href="https://peru21.pe/peru/15-junio-venezolanos-entraran-peru-pasaporte-visa-anuncio-martin-vizcarra-expulsados-migraciones-482899-noticia/">https://peru21.pe/peru/15-junio-venezolanos-entraran-peru-pasaporte-visa-anuncio-martin-vizcarra-expulsados-migraciones-482899-noticia/</a>
Día 3	Lectura 26	<a href="https://peru21.pe/politica/edwin-donayre-fue-capturado-en-puente-piedra-noticia/">https://peru21.pe/politica/edwin-donayre-fue-capturado-en-puente-piedra-noticia/</a>
Día 3	Lectura 27	<a href="https://peru21.pe/peru/piura-interceptan-narcosubmarino-con-dos-toneladas-de-droga-noticia/">https://peru21.pe/peru/piura-interceptan-narcosubmarino-con-dos-toneladas-de-droga-noticia/</a>
Día 3	Lectura 28	<a href="https://peru21.pe/lima/coronavirus-peru-la-victoria-fiscalizadores-decomisaron-mercaderia-a-ambulantes-en-alrededores-de-av-grau-nndc-noticia/">https://peru21.pe/lima/coronavirus-peru-la-victoria-fiscalizadores-decomisaron-mercaderia-a-ambulantes-en-alrededores-de-av-grau-nndc-noticia/</a>
Día 3	Lectura 29	<a href="https://peru21.pe/peru/arequipa-identifican-sospechosos-extranjeros-crimen-esposos-nndc-481394-noticia/">https://peru21.pe/peru/arequipa-identifican-sospechosos-extranjeros-crimen-esposos-nndc-481394-noticia/</a>
Día 3	Lectura 30	<a href="https://peru21.pe/politica/vladimir-cerron-cumplira-su-condena-en-libertad-noticia/">https://peru21.pe/politica/vladimir-cerron-cumplira-su-condena-en-libertad-noticia/</a>
Día 4	Lectura 31	<a href="https://peru21.pe/vida/viajes/lima-y-junin-los-destinos-favoritos-para-el-proximo-feriado-noticia/">https://peru21.pe/vida/viajes/lima-y-junin-los-destinos-favoritos-para-el-proximo-feriado-noticia/</a>
Día 4	Lectura 32	<a href="https://peru21.pe/lima/pueblo-libre-bus-se-queda-sin-techo-por-intentar-pasar-debajo-de-un-puente-nndc-noticia/">https://peru21.pe/lima/pueblo-libre-bus-se-queda-sin-techo-por-intentar-pasar-debajo-de-un-puente-nndc-noticia/</a>

Día 4	Lectura 33	<a href="https://peru21.pe/lima/mininter-ejecuto-segunda-expulsion-venezolanos-antecedentes-policiales-video-nndc-477483-noticia/">https://peru21.pe/lima/mininter-ejecuto-segunda-expulsion-venezolanos-antecedentes-policiales-video-nndc-477483-noticia/</a>
Día 4	Lectura 34	<a href="https://peru21.pe/lima/viaducto-armendariz-aliviara-congestion-costa-verde-486967-noticia/">https://peru21.pe/lima/viaducto-armendariz-aliviara-congestion-costa-verde-486967-noticia/</a>
Día 4	Lectura 35	<a href="https://peru21.pe/lima/plan-de-restriccion-vehicular-pico-y-placa-continuara-noticia/?ref=p21r">https://peru21.pe/lima/plan-de-restriccion-vehicular-pico-y-placa-continuara-noticia/?ref=p21r</a>
Día 4	Lectura 36	<a href="https://peru21.pe/economia/tercera-convocatoria-del-bono-renta-joven-saldra-en-diciembre-noticia/">https://peru21.pe/economia/tercera-convocatoria-del-bono-renta-joven-saldra-en-diciembre-noticia/</a>
Día 4	Lectura 37	<a href="https://peru21.pe/lima/atu-ajustara-las-rutas-de-transporte-publico-noticia/">https://peru21.pe/lima/atu-ajustara-las-rutas-de-transporte-publico-noticia/</a>
Día 4	Lectura 38	<a href="https://elcomercio.pe/lima/sucesos/coronavirus-peru-podemos-peru-plantea-servicio-militar-para-quienes-incumplan-inmovilizacion-social-por-covid-19-nndc-noticia/?ref=ecr">https://elcomercio.pe/lima/sucesos/coronavirus-peru-podemos-peru-plantea-servicio-militar-para-quienes-incumplan-inmovilizacion-social-por-covid-19-nndc-noticia/?ref=ecr</a>
Día 4	Lectura 39	<a href="https://elcomercio.pe/lima/sucesos/coronavirus-peru-migraciones-anuncio-el-reinicio-de-tramites-desde-este-3-de-junio-de-manera-virtual-covid-19-estado-de-emergencia-nndc-noticia/">https://elcomercio.pe/lima/sucesos/coronavirus-peru-migraciones-anuncio-el-reinicio-de-tramites-desde-este-3-de-junio-de-manera-virtual-covid-19-estado-de-emergencia-nndc-noticia/</a>
Día 4	Lectura 40	<a href="https://elcomercio.pe/lima/sucesos/sunedu-anuncia-que-se-ampliara-plazo-de-cierre-de-universidades-denegadas-de-2-a-3-anos-por-emergencia-nacional-coronavirus-en-peru-noticia/">https://elcomercio.pe/lima/sucesos/sunedu-anuncia-que-se-ampliara-plazo-de-cierre-de-universidades-denegadas-de-2-a-3-anos-por-emergencia-nacional-coronavirus-en-peru-noticia/</a>
Día 5	Lectura 41	<a href="https://peru21.pe/opinion/san-marcos-abandono-y-estigmatizacion-noticia/">https://peru21.pe/opinion/san-marcos-abandono-y-estigmatizacion-noticia/</a>
Día 5	Lectura 42	<a href="https://peru21.pe/lima/costa-verde-se-empezaran-a-aplicar-papeletas-por-exceso-de-velocidad-noticia/">https://peru21.pe/lima/costa-verde-se-empezaran-a-aplicar-papeletas-por-exceso-de-velocidad-noticia/</a>
Día 5	Lectura 43	<a href="https://peru21.pe/lima/centro-lima-clausuran-centro-comercial-hueco-inspeccion-zona-3-mesa-redonda-fotos-478805-noticia/">https://peru21.pe/lima/centro-lima-clausuran-centro-comercial-hueco-inspeccion-zona-3-mesa-redonda-fotos-478805-noticia/</a>
Día 5	Lectura 44	<a href="https://peru21.pe/economia/adecuado-aumento-3-1-lima-marzo-mayo-484626-noticia/">https://peru21.pe/economia/adecuado-aumento-3-1-lima-marzo-mayo-484626-noticia/</a>
Día 5	Lectura 45	<a href="https://peru21.pe/lima/paro-de-colectivos-colectiveros-ilegales-acataran-huelga-hoy-en-lima-y-callao-noticia/">https://peru21.pe/lima/paro-de-colectivos-colectiveros-ilegales-acataran-huelga-hoy-en-lima-y-callao-noticia/</a>
Día 5	Lectura 46	<a href="https://elcomercio.pe/lima/coronavirus-peru-realizan-pruebas-rapidas-de-covid-19-a-transportistas-de-carga-pesada-por-segundo-dia-consecutivo-fotos-nndc-noticia/?foto=3">https://elcomercio.pe/lima/coronavirus-peru-realizan-pruebas-rapidas-de-covid-19-a-transportistas-de-carga-pesada-por-segundo-dia-consecutivo-fotos-nndc-noticia/?foto=3</a>
Día 5	Lectura 47	<a href="https://elcomercio.pe/lima/sucesos/coronavirus-peru-subio-a-4230-la-cifra-de-fallecidos-por-covid-19-estado-de-emergencia-nndc-noticia/?ref=ecr">https://elcomercio.pe/lima/sucesos/coronavirus-peru-subio-a-4230-la-cifra-de-fallecidos-por-covid-19-estado-de-emergencia-nndc-noticia/?ref=ecr</a>
Día 5	Lectura 48	<a href="https://elcomercio.pe/lima/sucesos/coronavirus-peru-martin-vizcarra-esta-pandemia-del-covid-19-no-se-supera-en-15-dias-o-un-mes-toma-6-u-8-meses-estado-de-emergencia-nndc-noticia/?ref=ecr">https://elcomercio.pe/lima/sucesos/coronavirus-peru-martin-vizcarra-esta-pandemia-del-covid-19-no-se-supera-en-15-dias-o-un-mes-toma-6-u-8-meses-estado-de-emergencia-nndc-noticia/?ref=ecr</a>

Día 5	Lectura 49	<a href="https://elcomercio.pe/lima/sucesos/coronavirus-en-peru-bajas-temperaturas-pueden-ocasionar-incremento-de-infecciones-respiratorias-segun-essalud-covid-19-estado-de-emergencia-nndc-noticia/">https://elcomercio.pe/lima/sucesos/coronavirus-en-peru-bajas-temperaturas-pueden-ocasionar-incremento-de-infecciones-respiratorias-segun-essalud-covid-19-estado-de-emergencia-nndc-noticia/</a>
Día 5	Lectura 50	<a href="https://elcomercio.pe/peru/por-pandemia-se-extiende-el-plazo-de-cierre-de-universidades-sin-licenciamiento-noticia/">https://elcomercio.pe/peru/por-pandemia-se-extiende-el-plazo-de-cierre-de-universidades-sin-licenciamiento-noticia/</a>
Día 6	Lectura 51	<a href="https://peru21.pe/lima/cercado-lima-extranjero-roba-60-mil-soles-tienda-mesa-redonda-nndc-491541-noticia/">https://peru21.pe/lima/cercado-lima-extranjero-roba-60-mil-soles-tienda-mesa-redonda-nndc-491541-noticia/</a>
Día 6	Lectura 52	<a href="https://peru21.pe/peru/silencio-sismico-igp-sismo-terremoto-peru-nnda-nnlt-496835-noticia/">https://peru21.pe/peru/silencio-sismico-igp-sismo-terremoto-peru-nnda-nnlt-496835-noticia/</a>
Día 6	Lectura 53	<a href="https://peru21.pe/lima/destinan-policias-patrulleros-surco-479035-noticia/?ref=p21r">https://peru21.pe/lima/destinan-policias-patrulleros-surco-479035-noticia/?ref=p21r</a>
Día 6	Lectura 54	<a href="https://peru21.pe/economia/empleo-femenino-se-incremento-en-49-noticia/">https://peru21.pe/economia/empleo-femenino-se-incremento-en-49-noticia/</a>
Día 6	Lectura 55	<a href="https://peru21.pe/lima/parque-leyendas-extranjeros-causan-terror-gritar-leon-habia-escapado-video-nndc-497723-noticia/">https://peru21.pe/lima/parque-leyendas-extranjeros-causan-terror-gritar-leon-habia-escapado-video-nndc-497723-noticia/</a>
Día 6	Lectura 56	<a href="https://elcomercio.pe/lima/coronavirus-peru-a-164476-se-elevo-el-numero-de-casos-confirmados-de-covid-19-en-el-pais-nndc-noticia/">https://elcomercio.pe/lima/coronavirus-peru-a-164476-se-elevo-el-numero-de-casos-confirmados-de-covid-19-en-el-pais-nndc-noticia/</a>
Día 6	Lectura 57	<a href="https://elcomercio.pe/economia/peru/gremio-mundial-de-transporte-maritimo-advierte-intrusion-del-gobierno-peruano-en-comercio-exterior-nndc-noticia/?ref=ecr">https://elcomercio.pe/economia/peru/gremio-mundial-de-transporte-maritimo-advierte-intrusion-del-gobierno-peruano-en-comercio-exterior-nndc-noticia/?ref=ecr</a>
Día 6	Lectura 58	<a href="https://elcomercio.pe/lima/sucesos/coronavirus-peru-reportan-242-bomberos-contagiados-y-8-fallecidos-a-nivel-nacional-nndc-noticia/?ref=ecr">https://elcomercio.pe/lima/sucesos/coronavirus-peru-reportan-242-bomberos-contagiados-y-8-fallecidos-a-nivel-nacional-nndc-noticia/?ref=ecr</a>
Día 6	Lectura 59	<a href="https://elcomercio.pe/economia/peru/coronavirus-en-peru-se-aprobo-norma-para-reanudar-actividades-economicas-que-sectores-estan-listos-para-operar-economia-peruana-coronavirus-pesca-centros-comerciales-delivery-pymes-restaurantes-martin-vizcarra-decreto-supremo-construccion-turismo-noticia/?ref=ecr">https://elcomercio.pe/economia/peru/coronavirus-en-peru-se-aprobo-norma-para-reanudar-actividades-economicas-que-sectores-estan-listos-para-operar-economia-peruana-coronavirus-pesca-centros-comerciales-delivery-pymes-restaurantes-martin-vizcarra-decreto-supremo-construccion-turismo-noticia/?ref=ecr</a>
Día 6	Lectura 60	<a href="https://elcomercio.pe/peru/coronavirus-en-peru-reportan-16-gestantes-y-14-obstetras-con-covid-19-en-tumbes-nnpp-noticia/">https://elcomercio.pe/peru/coronavirus-en-peru-reportan-16-gestantes-y-14-obstetras-con-covid-19-en-tumbes-nnpp-noticia/</a>
Día 7	Lectura 61	<a href="https://peru21.pe/politica/separan-a-funcionaria-de-la-corte-de-lima-por-contratar-a-familiares-noticia/?ref=p21r">https://peru21.pe/politica/separan-a-funcionaria-de-la-corte-de-lima-por-contratar-a-familiares-noticia/?ref=p21r</a>
Día 7	Lectura 62	<a href="https://peru21.pe/politica/presidente-martin-vizcarra-plantea-construccion-aeropuerto-oxapampa-472237-noticia/">https://peru21.pe/politica/presidente-martin-vizcarra-plantea-construccion-aeropuerto-oxapampa-472237-noticia/</a>
Día 7	Lectura 63	<a href="https://peru21.pe/lima/diferencia-feriado-dias-laborables-nnda-nnlt-490211-noticia/">https://peru21.pe/lima/diferencia-feriado-dias-laborables-nnda-nnlt-490211-noticia/</a>
Día 7	Lectura 64	<a href="https://peru21.pe/politica/cuellos-blancos-puerto-tuvieron-hermanitos-pais-497660-noticia/">https://peru21.pe/politica/cuellos-blancos-puerto-tuvieron-hermanitos-pais-497660-noticia/</a>
Día 7	Lectura 65	<a href="https://peru21.pe/lima/lima-restauracion-cien-esculturas-centro-historico-498635-noticia/">https://peru21.pe/lima/lima-restauracion-cien-esculturas-centro-historico-498635-noticia/</a>

Día 7	Lectura 66	<a href="https://peru21.pe/peru/feriado-agosto-dias-laborables-mes-santa-rosa-lima-nnda-nnlt-497537-noticia/">https://peru21.pe/peru/feriado-agosto-dias-laborables-mes-santa-rosa-lima-nnda-nnlt-497537-noticia/</a>
Día 7	Lectura 67	<a href="https://peru21.pe/lima/migraciones-numero-venezolanos-peru-aumento-708-mil-nndc-463298-noticia/">https://peru21.pe/lima/migraciones-numero-venezolanos-peru-aumento-708-mil-nndc-463298-noticia/</a>
Día 7	Lectura 68	<a href="https://peru21.pe/lima/policiales/interpol-detiene-a-exabogado-de-fujimori-noticia/">https://peru21.pe/lima/policiales/interpol-detiene-a-exabogado-de-fujimori-noticia/</a>
Día 7	Lectura 69	<a href="https://gestion.pe/peru/peru21-lima-municipalidad-lima-recibe-facultades-ministerio-vivienda-iniciar-construccion-telefericos-nndc-275415-noticia/">https://gestion.pe/peru/peru21-lima-municipalidad-lima-recibe-facultades-ministerio-vivienda-iniciar-construccion-telefericos-nndc-275415-noticia/</a>
Día 7	Lectura 70	<a href="https://elcomercio.pe/lima/sucesos/coronavirus-en-peru-sis-y-clinicas-privadas-llegan-a-un-acuerdo-para-atender-a-pacientes-con-covid-19-referidos-de-hospitales-publicos-nndc-noticia/">https://elcomercio.pe/lima/sucesos/coronavirus-en-peru-sis-y-clinicas-privadas-llegan-a-un-acuerdo-para-atender-a-pacientes-con-covid-19-referidos-de-hospitales-publicos-nndc-noticia/</a>
Día 8	Lectura 71	<a href="https://elcomercio.pe/lima/sucesos/coronavirus-en-peru-subio-a-4767-la-cifra-de-fallecidos-por-covid-19-cuarentena-nndc-noticia/?ref=ecr">https://elcomercio.pe/lima/sucesos/coronavirus-en-peru-subio-a-4767-la-cifra-de-fallecidos-por-covid-19-cuarentena-nndc-noticia/?ref=ecr</a>
Día 8	Lectura 72	<a href="https://elcomercio.pe/lima/sucesos/coronavirus-peru-avenidas-colonial-brasil-y-universitaria-presentan-el-mayor-indice-de-contagios-de-covid-19-cuarentena-estado-de-emergencia-nndc-noticia/?ref=ecr">https://elcomercio.pe/lima/sucesos/coronavirus-peru-avenidas-colonial-brasil-y-universitaria-presentan-el-mayor-indice-de-contagios-de-covid-19-cuarentena-estado-de-emergencia-nndc-noticia/?ref=ecr</a>
Día 8	Lectura 73	<a href="https://elcomercio.pe/lima/sucesos/coronavirus-peru-regidores-denuncian-al-alcalde-del-callao-por-presunta-venta-de-pollos-que-habrian-sido-donados-estado-de-emergencia-covid-19-nndc-noticia/?ref=ecr">https://elcomercio.pe/lima/sucesos/coronavirus-peru-regidores-denuncian-al-alcalde-del-callao-por-presunta-venta-de-pollos-que-habrian-sido-donados-estado-de-emergencia-covid-19-nndc-noticia/?ref=ecr</a>
Día 8	Lectura 74	<a href="https://elcomercio.pe/lima/sucesos/coronavirus-peru-denuncian-que-el-ins-proceso-gratis-pruebas-moleculares-de-covid-19-a-clinicas-privadas-que-si-cobraron-a-pacientes-noticia/">https://elcomercio.pe/lima/sucesos/coronavirus-peru-denuncian-que-el-ins-proceso-gratis-pruebas-moleculares-de-covid-19-a-clinicas-privadas-que-si-cobraron-a-pacientes-noticia/</a>
Día 8	Lectura 75	<a href="https://elcomercio.pe/lima/sucesos/coronavirus-peru-estas-son-las-disposiciones-del-ministerio-de-salud-para-la-realizacion-de-actividades-fisicas-en-la-cuarentena-nndc-noticia/?ref=ecr">https://elcomercio.pe/lima/sucesos/coronavirus-peru-estas-son-las-disposiciones-del-ministerio-de-salud-para-la-realizacion-de-actividades-fisicas-en-la-cuarentena-nndc-noticia/?ref=ecr</a>
Día 8	Lectura 76	<a href="https://elcomercio.pe/economia/coronavirus-peru-vuelos-comerciales-al-interior-del-pais-se-permitirian-desde-julio-preve-el-mtc-nndc-noticia/?ref=ecr">https://elcomercio.pe/economia/coronavirus-peru-vuelos-comerciales-al-interior-del-pais-se-permitirian-desde-julio-preve-el-mtc-nndc-noticia/?ref=ecr</a>
Día 8	Lectura 77	<a href="https://elcomercio.pe/lima/sucesos/coronavirus-peru-colegio-medico-del-peru-rindio-homenaje-a-los-45-doctores-fallecidos-a-causa-del-covid-19-estado-de-emergencia-nndc-noticia/">https://elcomercio.pe/lima/sucesos/coronavirus-peru-colegio-medico-del-peru-rindio-homenaje-a-los-45-doctores-fallecidos-a-causa-del-covid-19-estado-de-emergencia-nndc-noticia/</a>
Día 8	Lectura 78	<a href="https://elcomercio.pe/lima/callao/coronavirus-peru-gobierno-regional-del-callao-instalara-dos-plantas-de-oxigeno-para-atender-a-pacientes-con-covid-19-nndc-noticia/?ref=ecr">https://elcomercio.pe/lima/callao/coronavirus-peru-gobierno-regional-del-callao-instalara-dos-plantas-de-oxigeno-para-atender-a-pacientes-con-covid-19-nndc-noticia/?ref=ecr</a>
Día 8	Lectura 79	<a href="https://elcomercio.pe/lima/sucesos/coronavirus-peru-ambulantes-se-resisten-abandonar-la-avenida-grau-y-calles-aledanas-video-covid-19-estado-de-emergencia-nndc-noticia/?ref=ecr">https://elcomercio.pe/lima/sucesos/coronavirus-peru-ambulantes-se-resisten-abandonar-la-avenida-grau-y-calles-aledanas-video-covid-19-estado-de-emergencia-nndc-noticia/?ref=ecr</a>

Día 8	Lectura 80	<a href="https://elcomercio.pe/peru/piura/coronavirus-en-peru-minsa-otorga-autorizaciones-excepcionales-a-plantas-de-oxigeno-en-piura-nnpp-noticia/?ref=ecr">https://elcomercio.pe/peru/piura/coronavirus-en-peru-minsa-otorga-autorizaciones-excepcionales-a-plantas-de-oxigeno-en-piura-nnpp-noticia/?ref=ecr</a>
Día 9	Lectura 81	<a href="https://peru21.pe/politica/daniel-urresti-no-registro-en-hoja-de-vida-condena-por-difamacion-elecciones-2020-noticia/">https://peru21.pe/politica/daniel-urresti-no-registro-en-hoja-de-vida-condena-por-difamacion-elecciones-2020-noticia/</a>
Día 9	Lectura 82	<a href="https://peru21.pe/lima/sea-precavido-manana-habra-8-eventos-masivos-en-lima-noticia/">https://peru21.pe/lima/sea-precavido-manana-habra-8-eventos-masivos-en-lima-noticia/</a>
Día 9	Lectura 83	<a href="https://peru21.pe/lima/alertan-del-caos-que-persiste-en-mercado-central-y-mesa-redonda-noticia/">https://peru21.pe/lima/alertan-del-caos-que-persiste-en-mercado-central-y-mesa-redonda-noticia/</a>
Día 9	Lectura 84	<a href="https://peru21.pe/lima/promueven-reciclaje-traves-aplicativo-463591-noticia/">https://peru21.pe/lima/promueven-reciclaje-traves-aplicativo-463591-noticia/</a>
Día 9	Lectura 85	<a href="https://peru21.pe/lima/agrava-congestion-vehicular-molina-495943-noticia/">https://peru21.pe/lima/agrava-congestion-vehicular-molina-495943-noticia/</a>
Día 9	Lectura 86	<a href="https://elcomercio.pe/lima/sucesos/coronavirus-en-peru-76228-pacientes-se-recuperaron-y-fueron-dados-de-alta-informo-martin-vizcarra-covid-19-estado-de-emergencia-nndc-noticia/?ref=ecr">https://elcomercio.pe/lima/sucesos/coronavirus-en-peru-76228-pacientes-se-recuperaron-y-fueron-dados-de-alta-informo-martin-vizcarra-covid-19-estado-de-emergencia-nndc-noticia/?ref=ecr</a>
Día 9	Lectura 87	<a href="https://elcomercio.pe/economia/peru/coronavirus-peru-bono-familiar-universal-la-proxima-semana-iniciaria-entrega-a-los-hogares-que-solicitaron-el-subsidio-760-soles-nndc-noticia/?ref=ecr">https://elcomercio.pe/economia/peru/coronavirus-peru-bono-familiar-universal-la-proxima-semana-iniciaria-entrega-a-los-hogares-que-solicitaron-el-subsidio-760-soles-nndc-noticia/?ref=ecr</a>
Día 9	Lectura 88	<a href="https://elcomercio.pe/lima/sucesos/coronavirus-peru-intervienen-municipalidad-de-brena-por-presuntas-irregularidades-en-entregas-de-canastas-cuarentena-estado-de-emergencia-covid-19-nndc-noticia/?ref=ecr">https://elcomercio.pe/lima/sucesos/coronavirus-peru-intervienen-municipalidad-de-brena-por-presuntas-irregularidades-en-entregas-de-canastas-cuarentena-estado-de-emergencia-covid-19-nndc-noticia/?ref=ecr</a>
Día 9	Lectura 89	<a href="https://elcomercio.pe/lima/sucesos/coronavirus-peru-la-victoria-con-azucar-impalpable-y-colorante-falsificaban-paracetamol-y-naproxeno-video-covid-19-nndc-noticia/?ref=ecr">https://elcomercio.pe/lima/sucesos/coronavirus-peru-la-victoria-con-azucar-impalpable-y-colorante-falsificaban-paracetamol-y-naproxeno-video-covid-19-nndc-noticia/?ref=ecr</a>
Día 9	Lectura 90	<a href="https://elcomercio.pe/economia/peru/fraccionamiento-de-deuda-recibo-claro-entel-movistar-bitel-osiptel-aplaza-prohibicion-para-que-las-operadoras-corten-los-servicios-por-falta-de-pago-hasta-el-15-de-junio-coronavirus-peru-covid-19-nndc-noticia/?ref=ecr">https://elcomercio.pe/economia/peru/fraccionamiento-de-deuda-recibo-claro-entel-movistar-bitel-osiptel-aplaza-prohibicion-para-que-las-operadoras-corten-los-servicios-por-falta-de-pago-hasta-el-15-de-junio-coronavirus-peru-covid-19-nndc-noticia/?ref=ecr</a>
Día 10	Lectura 91	<a href="https://peru21.pe/politica/luz-salgado-martin-vizcarra-poner-gente-nuestra-nndc-497303-noticia/">https://peru21.pe/politica/luz-salgado-martin-vizcarra-poner-gente-nuestra-nndc-497303-noticia/</a>
Día 10	Lectura 92	<a href="https://peru21.pe/politica/salvador-del-solar-no-somos-la-resistencia-vamos-a-luchar-por-el-nuevo-peru-noticia/">https://peru21.pe/politica/salvador-del-solar-no-somos-la-resistencia-vamos-a-luchar-por-el-nuevo-peru-noticia/</a>
Día 10	Lectura 93	<a href="https://peru21.pe/lima/sujeto-salva-linchado-vecinos-callao-479684-noticia/">https://peru21.pe/lima/sujeto-salva-linchado-vecinos-callao-479684-noticia/</a>
Día 10	Lectura 94	<a href="https://peru21.pe/lima/lima-anuncia-cierre-mesa-redonda-tres-semanas-473399-noticia/">https://peru21.pe/lima/lima-anuncia-cierre-mesa-redonda-tres-semanas-473399-noticia/</a>
Día 10	Lectura 95	<a href="https://peru21.pe/politica/ana-teresa-revilla-ola-de-criticas-a-la-ministra-por-declararse-en-navidad-noticia/">https://peru21.pe/politica/ana-teresa-revilla-ola-de-criticas-a-la-ministra-por-declararse-en-navidad-noticia/</a>
Día 10	Lectura 96	<a href="https://elcomercio.pe/tecnologia/ciencias/coronavirus-queremos-fabricar-5000-oxigenadores-no-invasivos-y-donarlos-noticia/">https://elcomercio.pe/tecnologia/ciencias/coronavirus-queremos-fabricar-5000-oxigenadores-no-invasivos-y-donarlos-noticia/</a>

Día 10	Lectura 97	<a href="https://elcomercio.pe/lima/sucesos/coronavirus-peru-municipalidad-de-lima-proyectaran-peliculas-y-clases-de-baile-en-jesus-maria-para-que-vecinos-lo-vean-desde-su-balcon-covid-19-nndc-noticia/">https://elcomercio.pe/lima/sucesos/coronavirus-peru-municipalidad-de-lima-proyectaran-peliculas-y-clases-de-baile-en-jesus-maria-para-que-vecinos-lo-vean-desde-su-balcon-covid-19-nndc-noticia/</a>
Día 10	Lectura 98	<a href="https://elcomercio.pe/lima/sucesos/coronavirus-en-peru-tres-personas-que-saben-que-estan-contagiadas-de-covid-19-fueron-detenido-transitando-en-la-calle-nndc-noticia/">https://elcomercio.pe/lima/sucesos/coronavirus-en-peru-tres-personas-que-saben-que-estan-contagiadas-de-covid-19-fueron-detenido-transitando-en-la-calle-nndc-noticia/</a>
Día 10	Lectura 99	<a href="https://elcomercio.pe/lima/sucesos/coronavirus-peru-capturan-a-sujeto-que-vendia-oxigeno-industrial-y-lo-hacia-pasar-como-medicinal-en-cercado-de-lima-covid-19-estado-de-emergencia-nndc-noticia/">https://elcomercio.pe/lima/sucesos/coronavirus-peru-capturan-a-sujeto-que-vendia-oxigeno-industrial-y-lo-hacia-pasar-como-medicinal-en-cercado-de-lima-covid-19-estado-de-emergencia-nndc-noticia/</a>
Día 10	Lectura 100	<a href="https://elcomercio.pe/lima/sucesos/coronavirus-en-peru-comerciantes-estan-a-la-espera-de-ser-trasladados-a-la-tierra-prometida-noticia/">https://elcomercio.pe/lima/sucesos/coronavirus-en-peru-comerciantes-estan-a-la-espera-de-ser-trasladados-a-la-tierra-prometida-noticia/</a>

#### Anexo B. Distribución de audios de cada participante

Etiqueta	Cantidad	Porcentaje de Distribución (%)
Prolongación	51	5,79%
Bloqueo	84	9,53%
Repetición de sonido	73	8,29%
Repetición de palabra	27	3,06%
Sin palabras tartamudeadas	646	73,33%
Interjección	0	0,00%
<b>Total</b>	<b>881</b>	<b>100,00%</b>

Tabla 9. Distribución de audios del participante ATMA

Etiqueta	Cantidad	Porcentaje de Distribución (%)
Prolongación	228	26,54%
Bloqueo	89	10,36%
Repetición de sonido	92	10,71%
Repetición de palabra	18	2,10%
Sin palabras tartamudeadas	428	49,83%
Interjección	4	0,47%
<b>Total</b>	<b>859</b>	<b>100,00%</b>

Tabla 10. Distribución de audios del participante RRYR

Etiqueta	Cantidad	Porcentaje de Distribución (%)
Prolongación	153	20,03%
Bloqueo	64	8,38%
Repetición de sonido	89	11,65%
Repetición de palabra	38	4,97%
Sin palabras tartamudeadas	419	54,84%
Interjección	1	0,13%
<b>Total</b>	<b>764</b>	<b>100,00%</b>

Tabla 11. Distribución de audios del participante SSF

Etiqueta	Cantidad	Porcentaje de Distribución (%)
Prolongación	70	8,06%
Bloqueo	132	15,19%
Repetición de sonido	205	23,59%
Repetición de palabra	5	0,58%
Sin palabras tartamudeadas	453	52,13%
Interjección	4	0,46%
<b>Total</b>	<b>869</b>	<b>100,00%</b>

Tabla 12. Distribución de audios del participante PVM

Etiqueta	Cantidad	Porcentaje de Distribución (%)
Prolongación	47	5,27%
Bloqueo	55	6,17%
Repetición de sonido	25	2,80%
Repetición de palabra	15	1,68%
Sin palabras tartamudeadas	742	83,18%
Interjección	8	0,90%
<b>Total</b>	<b>892</b>	<b>100,00%</b>

Tabla 13. Distribución de audios del participante LPJM

Anexo C. Resultados de la evaluación para los audios de cada participante

Train_Data		Validation			Test		
Etiqueta	IsStuttered	Precision	Recall	F1	Precision	Recall	F1
Prolongación	No	0,95	0,73	0,82	0,93	0,73	0,82
	Sí	0,11	0,47	0,18	0,08	0,29	0,13
Bloqueo	No	0,89	0,82	0,85	0,88	0,76	0,81
	Sí	0,14	0,21	0,17	0,05	0,12	0,07
Repetición de sonido	No	0,91	0,80	0,86	0,91	0,75	0,82
	Sí	0,16	0,33	0,22	0,14	0,36	0,20
Repetición de palabra	No	0,96	1,00	0,98	0,96	1,00	0,98
	Sí	0,00	0,00	0,00	0,00	0,00	0,00

Tabla 14. Resultados de la evaluación para los audios del participante ATMA

Train_Data		Validation			Test		
Etiqueta	IsStuttered	Precision	Recall	F1	Precision	Recall	F1
Prolongación	No	0,68	0,74	0,71	0,64	0,64	0,64
	Sí	0,44	0,37	0,40	0,34	0,34	0,34
Bloqueo	No	0,83	0,83	0,83	0,82	0,80	0,81
	Sí	0,20	0,20	0,20	0,12	0,14	0,13
Repetición de sonido	No	0,82	0,78	0,80	0,86	0,78	0,82
	Sí	0,21	0,25	0,23	0,26	0,38	0,31
Repetición de palabra	No	0,97	1,00	0,98	0,95	1,00	0,98
	Sí	0,00	0,00	0,00	0,00	0,00	0,00

Tabla 15. Resultados de la evaluación para los audios del participante RRYR

Train_Data	Balance = False	Validation			Test		
------------	-----------------	------------	--	--	------	--	--

<i>Etiqueta</i>	<i>IsStuttered</i>	<i>Precision</i>	<i>Recall</i>	<i>F1</i>	<i>Precision</i>	<i>Recall</i>	<i>F1</i>
<i>Prolongación</i>	No	0,75	0,75	0,75	0,71	0,72	0,71
	Sí	0,29	0,29	0,29	0,22	0,22	0,22
<i>Bloqueo</i>	No	0,86	0,78	0,82	0,87	0,79	0,83
	Sí	0,09	0,14	0,11	0,15	0,23	0,18
<i>Repetición de sonido</i>	No	0,83	0,83	0,83	0,82	0,83	0,82
	Sí	0,18	0,17	0,18	0,15	0,13	0,14
<i>Repetición de palabra</i>	No	0,91	1,00	0,96	0,92	1,00	0,96
	Sí	0,00	0,00	0,00	0,00	0,00	0,00

Tabla 16. Resultados de la evaluación para los audios del participante SSF

<i>Train_Data</i>		<i>Validation</i>			<i>Test</i>		
<i>Etiqueta</i>	<i>IsStuttered</i>	<i>Precision</i>	<i>Recall</i>	<i>F1</i>	<i>Precision</i>	<i>Recall</i>	<i>F1</i>
<i>Prolongación</i>	No	0,89	0,77	0,82	0,88	0,70	0,78
	Sí	0,19	0,35	0,24	0,17	0,39	0,23
<i>Bloqueo</i>	No	0,80	0,83	0,82	0,78	0,81	0,79
	Sí	0,32	0,27	0,30	0,28	0,24	0,26
<i>Repetición de sonido</i>	No	0,68	0,76	0,72	0,65	0,75	0,70
	Sí	0,27	0,19	0,22	0,24	0,16	0,20
<i>Repetición de palabra</i>	No	0,99	1,00	1,00	0,99	1,00	0,99
	Sí	0,00	0,00	0,00	0,00	0,00	0,00

Tabla 17. Resultados de la evaluación para los audios del participante PVM

<i>Train_Data</i>		<i>Validation</i>			<i>Test</i>		
<i>Etiqueta</i>	<i>IsStuttered</i>	<i>Precision</i>	<i>Recall</i>	<i>F1</i>	<i>Precision</i>	<i>Recall</i>	<i>F1</i>
<i>Prolongación</i>	No	0,96	0,81	0,88	0,93	0,79	0,86
	Sí	0,12	0,40	0,18	0,07	0,22	0,11
<i>Bloqueo</i>	No	0,93	0,85	0,89	0,93	0,86	0,90

	Sí	0,03	0,06	0,04	0,11	0,21	0,14
<i>Repetición de sonido</i>	No	0,96	0,81	0,88	0,97	0,84	0,90
	Sí	0,00	0,00	0,00	0,03	0,12	0,04
<i>Repetición de palabra</i>	No	0,98	1,00	0,99	0,98	1,00	0,99
	Sí	0,00	0,00	0,00	0,00	0,00	0,00

Tabla 18. Resultados de la evaluación para los audios del participante LPJM

#### Anexo D. Resultados de validación de la aplicación de las estrategias de aumento de datos por separado

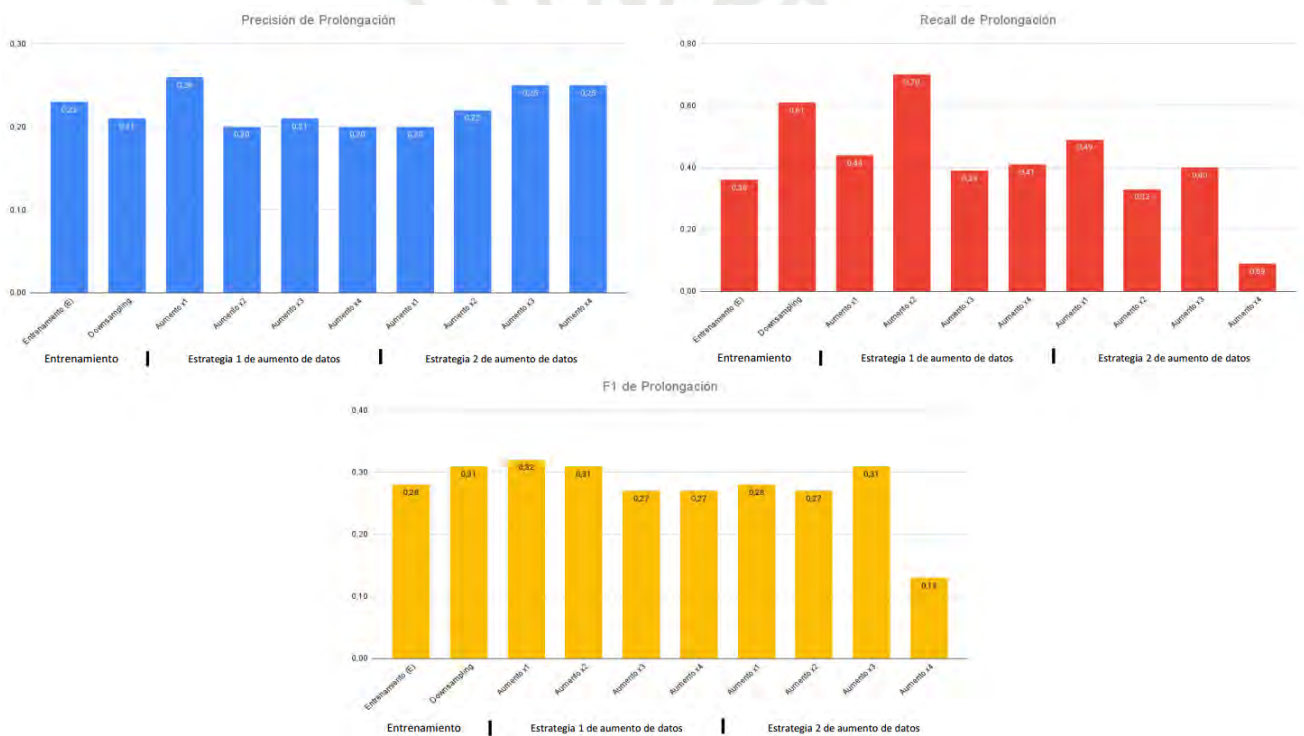


Figura 17: Resultados de validación de las estrategias de aumento de datos por separado para Prolongación

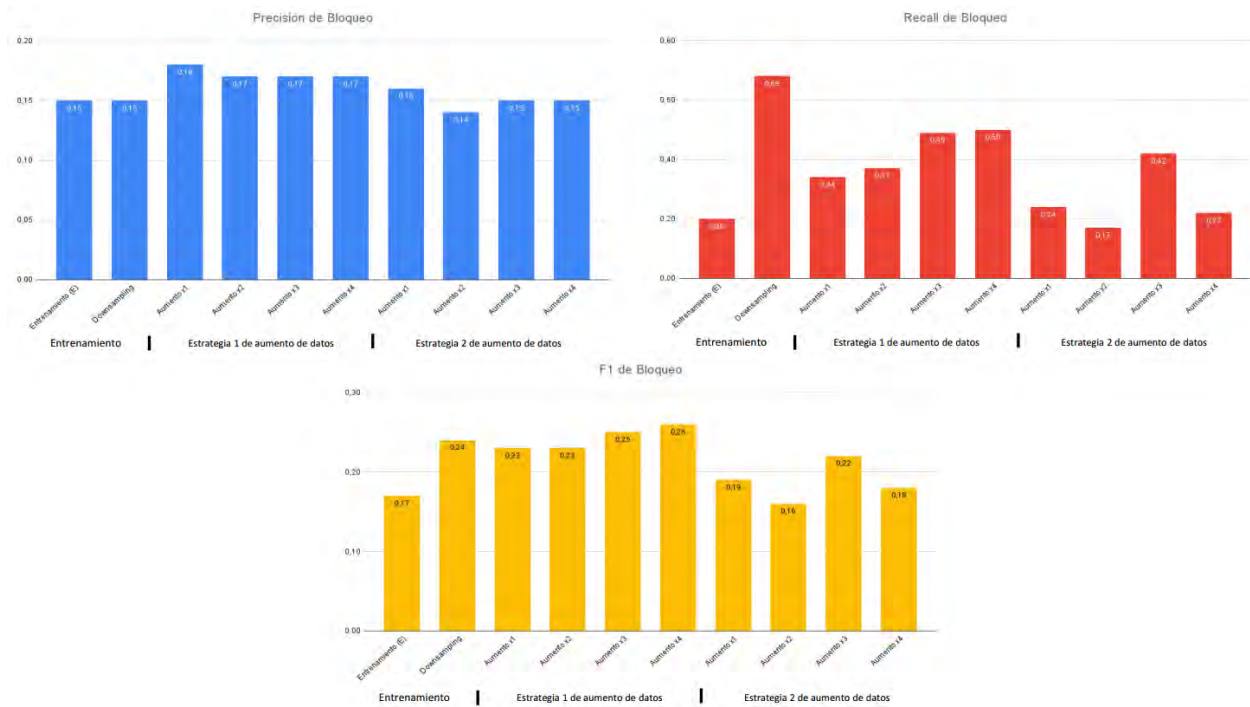


Figura 18: Resultados de validación de las estrategias de aumento de datos por separado para Bloqueo

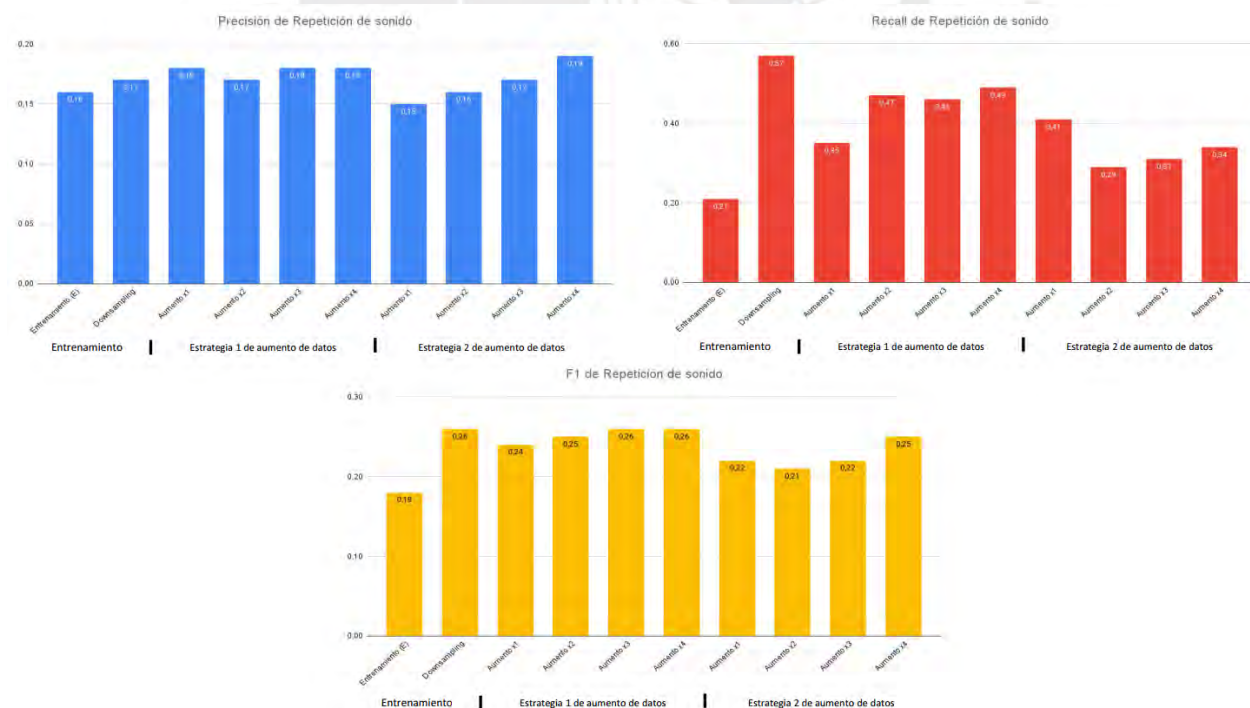


Figura 19: Resultados de validación de las estrategias de aumento de datos por separado para Repetición de sonido

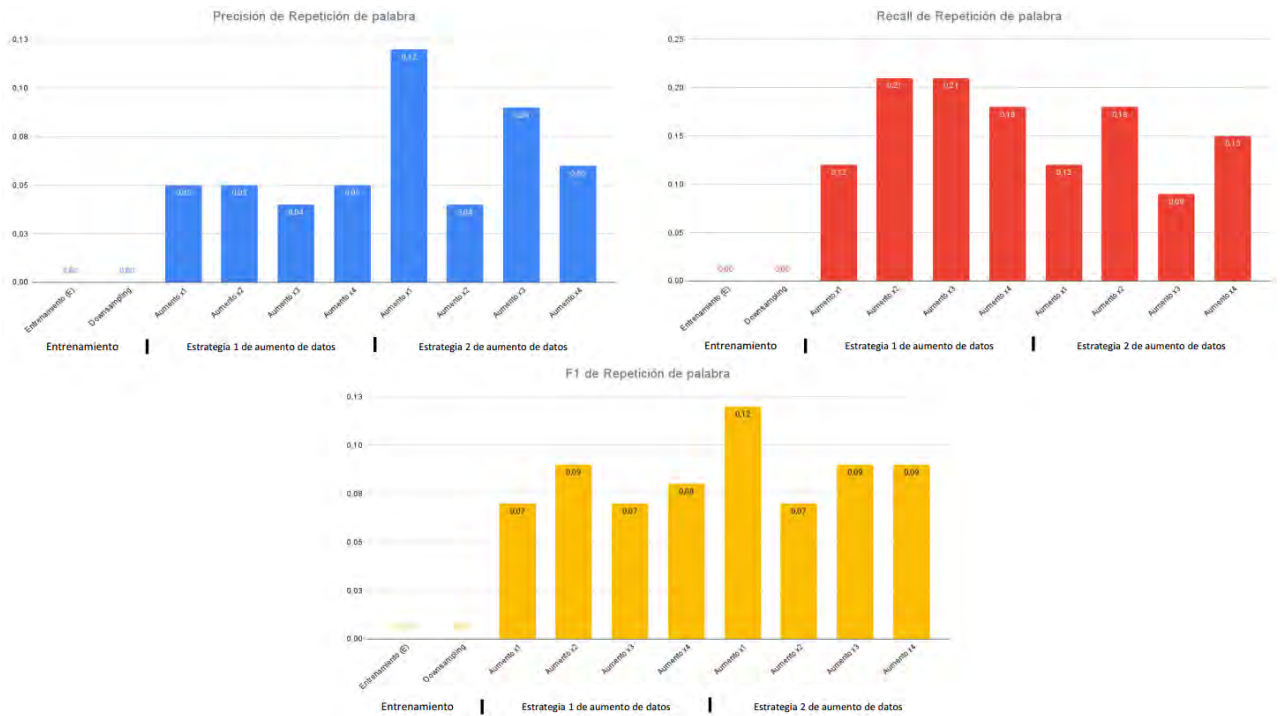


Figura 20: Resultados de validación de las estrategias de aumento de datos por separado para Repetición de palabra

Anexo E. Resultados de validación de todos los conjuntos de datos

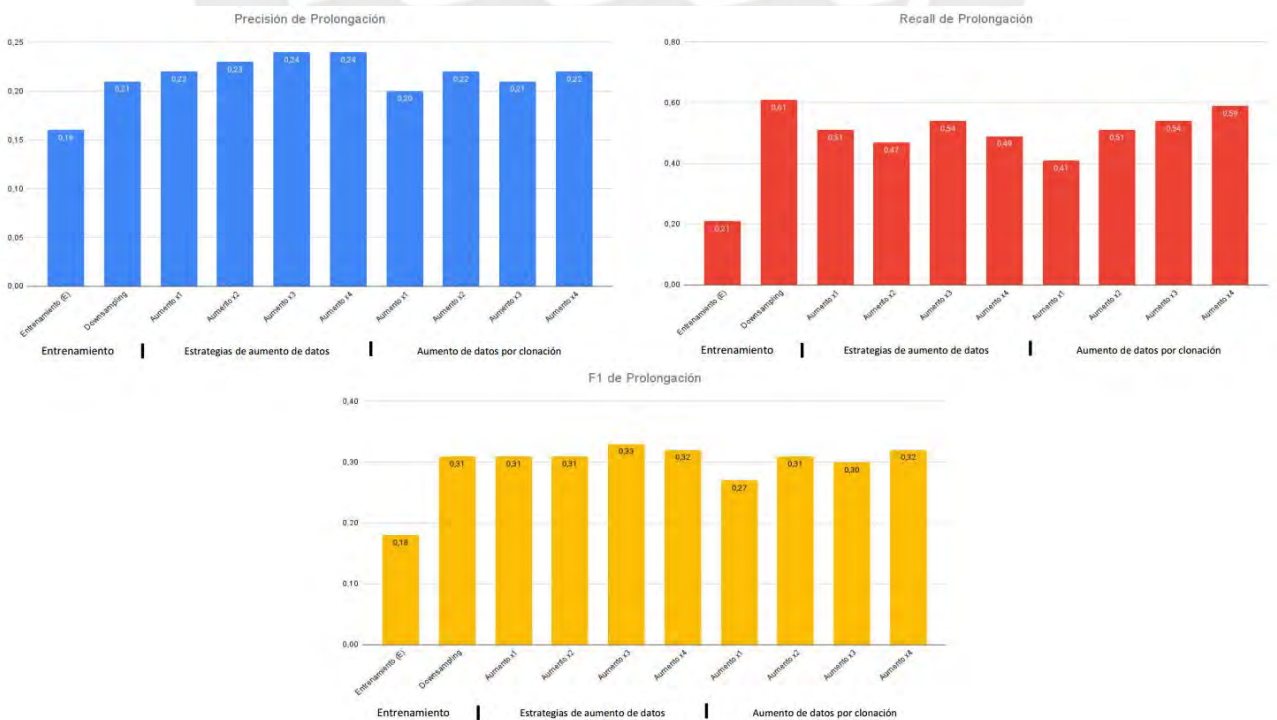


Figura 21: Resultados de validación de todos los conjuntos de datos para Prolongación

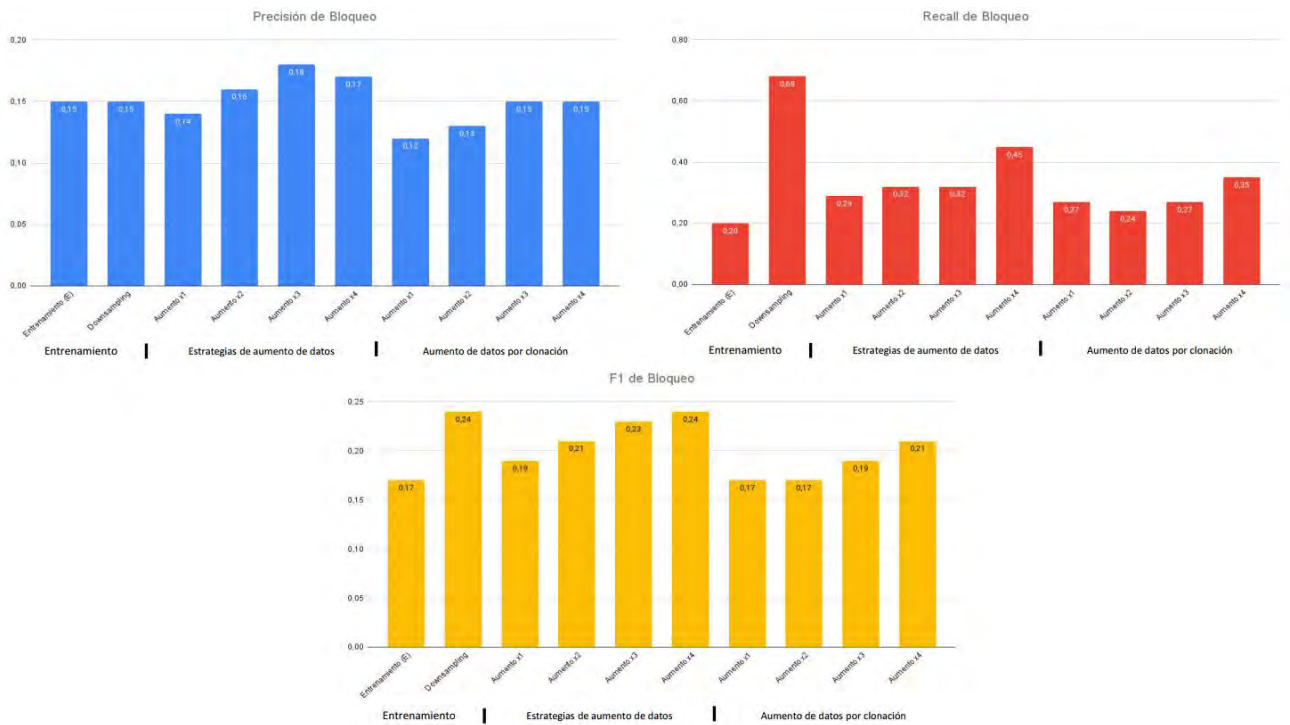


Figura 22: Resultados de validación de todos los conjuntos de datos para Bloqueo

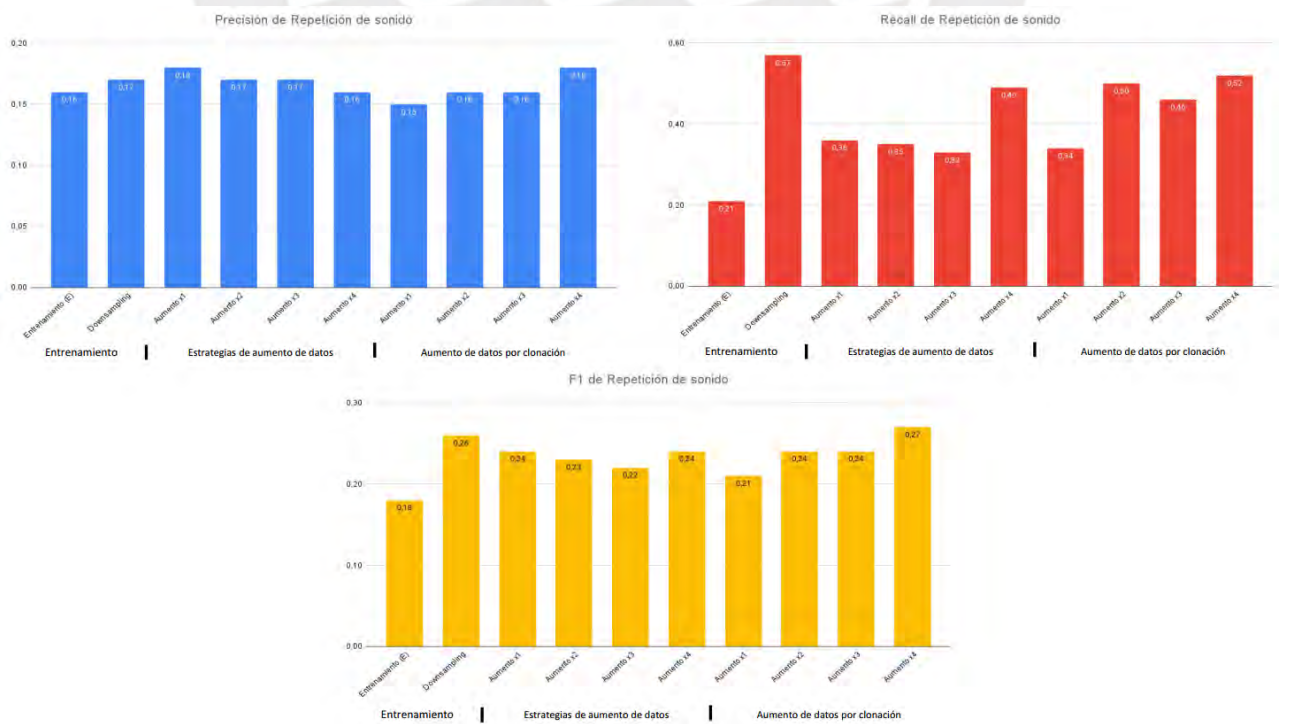


Figura 23: Resultados de validación de todos los conjuntos de datos para Repetición de sonido

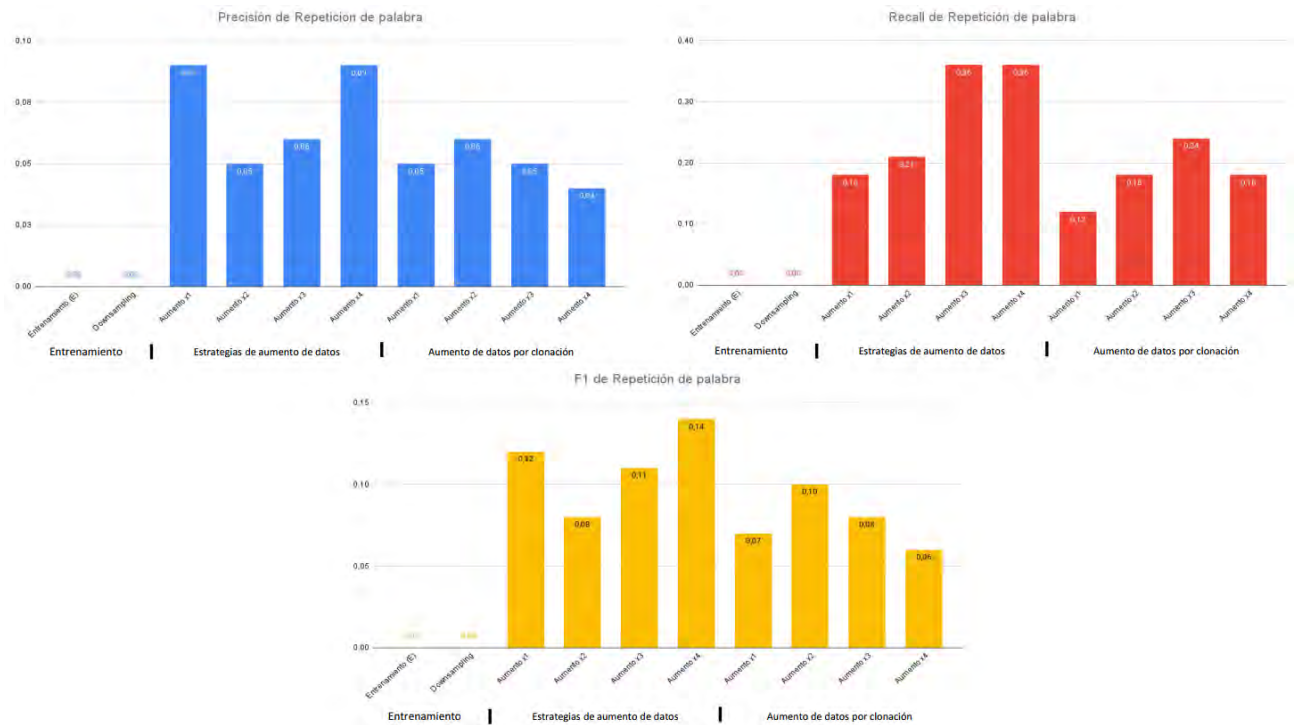


Figura 24: Resultados de validación de todos los conjuntos de datos para Repetición de palabra

