

Traductor de voz basado en procesamiento de lenguaje natural para asistencia a niños y jóvenes con dislalia

Voice translator based on natural language processing to assist children and young people with dyslalia

Recibido: 28, mayo 2025

Aceptado: 12, julio 2025

Publicado

online: 18, julio 2025

Como Citar: González Franco, L., & Sánchez Calderón, J. (2025). Traductor de voz basado en procesamiento de lenguaje natural para asistencia a niños y jóvenes con dislalia. Revista Digital Científica Causalidad (caui3), 1(2), 15-25. <https://doi.org/10.70929/caui3.v1i2.0012>

ISSN-e: 3073-1518

Creative Commons CC BY 4.0

Luis González¹, James Sánchez¹

Resumen:

El traductor de voz intenta corregir la dislalia en niños y jóvenes, para lo que se emplea tecnologías avanzadas de procesamiento de lenguaje natural. Se implementaron procedimientos como la selección de herramientas avanzadas (Vosk, Google Speech-to-Text y OpenAI Whisper) y el diseño de una interfaz intuitiva para garantizar la funcionalidad del prototipo. Se realizaron pruebas iterativas para validar el desempeño del sistema y recopilar retroalimentación de terapeutas y educadores. Los resultados mostraron que el sistema alcanzó un 90% de precisión promedio en el reconocimiento de voz, cumpliendo con los estándares establecidos. Además, se identificaron ciertas áreas de mejora, como la detección de fonemas complejos (/r/ y /s/), donde se registraron mayores tasas de error. Se concluye que el sistema presenta un impacto significativo en el ámbito educativo y terapéutico.

Palabras Clave: Dislalia, reconocimiento de voz, procesamiento de lenguaje natural, terapia del habla.

Abstract:

The voice translator attempts to correct dyslalia in children and young people, using advanced natural language processing technologies. Procedures such as the selection of advanced tools (Vosk, Google Speech-to-Text, and OpenAI Whisper) and the design of an intuitive interface were implemented to ensure the functionality of the prototype. Iterative testing was conducted to validate the system's performance and gather feedback from therapists and educators. The results showed that the system achieved an average accuracy of 90% in speech recognition, meeting established standards. In addition, certain areas for improvement were identified, such as the detection of complex phonemes (/r/ and /s/), where higher error rates were recorded. It is concluded that the system has a significant impact in the educational and therapeutic fields.

Keywords: Dyslalia, voice recognition, natural language processing, speech therapy.

¹Facultad de Ingeniería Industrial, Universidad de Guayaquil, Guayaquil 090514, Ecuador

Autor de correspondencia: luis.gonzalezf@ug.edu.ec

1 Introducción

El desarrollo tecnológico en el campo de la educación y la terapia del lenguaje ha abierto nuevas oportunidades para abordar trastornos del habla que afectan a niños en edad escolar. Uno de los problemas más comunes en este ámbito es la dislalia, un trastorno de los sonidos del habla (TSH) que se manifiesta como dificultades en la articulación de ciertos fonemas, lo que limita la capacidad de comunicación verbal de los niños [1]. Este trastorno, aunque tratable, puede tener implicaciones significativas en el desarrollo emocional, académico y social de quienes lo padecen, afectando su autoestima y exponiéndolos a situaciones de exclusión o bullying [2].

La tecnología puede desempeñar un papel crucial al ofrecer herramientas innovadoras que faciliten la comunicación y promuevan la inclusión de estos niños en entornos educativos y sociales. El presente trabajo se centra en el diseño y desarrollo de un prototipo de traductor de voz basado en procesamiento de lenguaje natural (NLP), que tiene como objetivo principal asistir a niños con dislalia en la mejora de su pronunciación y articulación, a través de un enfoque interactivo que utiliza tecnologías emergentes como el reconocimiento de voz (Speech Recognition), la traducción automática (Machine Translation) y la síntesis de voz (Text-to-Speech).

A nivel global, se han desarrollado herramientas similares aplicadas a sectores médicos y educativos, donde el procesamiento de lenguaje natural ha demostrado ser una solución efectiva para mejorar la comunicación de personas con dificultades del habla [3]. Sin embargo, en Ecuador, el acceso a estas tecnologías es limitado, y su adaptación a las necesidades específicas de niños con dislalia aún no ha sido ampliamente explorada [4]. Este trabajo busca llenar ese vacío, proponiendo una herramienta innovadora que no solo brinde apoyo directo a los niños, sino que también sirva como complemento para educadores, terapeutas del habla y padres de familia.

La tecnología puede desempeñar un papel crucial al ofrecer herramientas innovadoras que faciliten la comunicación y promuevan la inclusión de estos niños en entornos educativos y sociales. El presente trabajo se centra en el diseño y desarrollo de un prototipo de traductor de voz basado en procesamiento de lenguaje natural (NLP), que tiene como objetivo principal asistir a niños con dislalia en la mejora de su pronunciación y articulación, a través de un enfoque interactivo que utiliza tecnologías emergentes como el reconocimiento de voz (Speech Recognition), la traducción automática (Machine Translation) y la síntesis de voz (Text-to-Speech).

A nivel global, se han desarrollado herramientas similares aplicadas a sectores médicos y educativos, donde el procesamiento de lenguaje natural ha demostrado ser una solución efectiva para mejorar la comunicación de personas con dificultades del habla [5], [6], [7]. Sin embargo, en Ecuador, el acceso a estas tecnologías es limitado, y su adaptación a las necesidades específicas de niños con dislalia aún no ha sido ampliamente explorada [8], [9]. Este trabajo busca llenar ese vacío, proponiendo una herramienta innovadora que no solo brinde apoyo directo a los niños, sino que también sirva como complemento para educadores, terapeutas del habla y padres de familia.

El documento presenta un análisis detallado del problema, los fundamentos teóricos del procesamiento de lenguaje natural y el impacto de la dislalia en la población infantil. Asimismo, se exponen las bases metodológicas para el desarrollo del prototipo, incluyendo las entrevistas realizadas con especialistas en terapia del lenguaje, que han proporcionado información valiosa sobre las dificultades más comunes en niños con dislalia y el desarrollo de la aplicación.

A futuro, se espera que el prototipo, al integrarse en entornos educativos y terapéuticos, contribuya a la inclusión de niños con trastornos del habla, facilitando su integración social y educativa. Este proyecto no solo tiene un enfoque tecnológico, sino que responde a una necesidad social y educativa, promoviendo el uso de la inteligencia artificial y el aprendizaje automático como herramientas clave para mejorar la calidad de vida de los niños con dificultades en el habla [10], [11].

2 Metodología y planteamiento del problema

En Ecuador, un 12.3% de los niños presentan dislalia, lo cual representa un número significativo de niños que enfrentan dificultades para pronunciar palabras correctamente, impidiéndoles comunicarse de manera efectiva [12]. Además, esta dificultad no solo compromete su rendimiento escolar y su integración social, sino que también los expone a situaciones de exclusión y bullying, afectando su autoestima y bienestar emocional. Por otro lado, a nivel tecnológico, existen herramientas de procesamiento de lenguaje natural (NLP) aplicadas en áreas de educación y terapia del habla, pero su accesibilidad y adecuación para usuarios con dislalia son limitadas en este contexto [13], [14].

Este estudio tiene como objetivo desarrollar un prototipo de traductor de voz que emplee tecnologías de NLP, tales como reconocimiento de voz, traducción automática y síntesis de voz, con el fin de facilitar la comunicación de niños y jóvenes con dislalia. La finalidad es mejorar su integración en entornos sociales y educativos, promoviendo su inclusión y bienestar.

2.1 Arquitectura del Sistema

El prototipo de traductor de voz basado en procesamiento de lenguaje natural (NLP) sigue una arquitectura modular compuesta por tres capas principales: adquisición de datos, procesamiento y análisis, e interfaz de usuario.

La capa de adquisición de datos captura la señal de voz a través de un micrófono de alta sensibilidad, eliminando ruido y segmentando el audio en fonemas. Luego, la capa de procesamiento y análisis emplea modelos avanzados de reconocimiento de voz, como OpenAI Whisper y Google Speech-to-Text, para transcribir el habla y detectar errores de pronunciación en comparación con un modelo fonético de referencia. Si se identifican fallas, el sistema genera retroalimentación adaptativa con sugerencias para mejorar la articulación [15], [16].

Por último, la capa de interfaz de usuario presenta los resultados mediante una plataforma web accesible desde dispositivos móviles, desarrollada con Flask y React.js. La interfaz permite visualizar errores, recibir correcciones en tiempo real y acceder a ejercicios personalizados. Para mejorar la interacción con niños, se incorpora un sistema de síntesis de voz (Text-to-Speech). El sistema se ejecuta en la nube, utilizando Google Cloud o Microsoft Azure, lo que permite escalabilidad y procesamiento eficiente de múltiples solicitudes simultáneamente. Además, se implementan medidas de seguridad para proteger los datos de los usuarios, cumpliendo con normativas de privacidad [17], [18].

Esta arquitectura proporciona una solución eficiente y accesible para mejorar la pronunciación de niños con dislalia, combinando inteligencia artificial y NLP para ofrecer un aprendizaje interactivo y adaptativo (ver Figura 1).

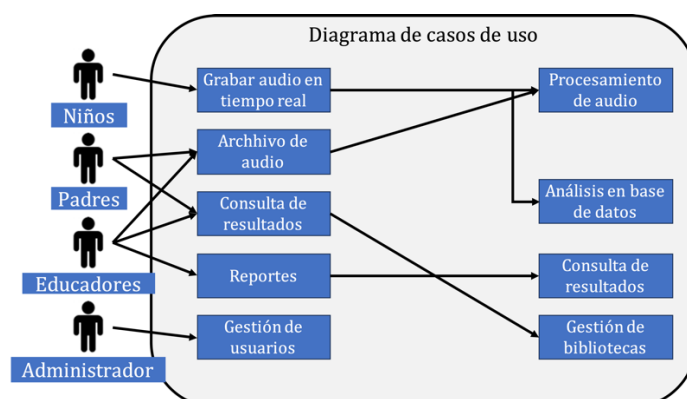


Figura 1. Diagrama de Casos de Uso.

La Figura 1 presenta un diagrama de casos de uso que detalla las principales interacciones entre los actores del sistema y las funcionalidades clave de la aplicación. Los actores incluyen a niños, padres, educadores, y el administrador, quienes interactúan con diversos casos de uso que abarcan la grabación, procesamiento y consulta de resultados de audio, así como la gestión de usuarios y archivos.

2.2 Desarrollo del prototipo

El desarrollo del prototipo de traductor de voz basado en procesamiento de lenguaje natural (NLP) se llevó a cabo en tres fases principales: implementación del reconocimiento de voz, análisis fonético y desarrollo de la interfaz de usuario.

Para capturar y transcribir la voz del usuario, se integraron modelos avanzados de reconocimiento automático de voz (ASR), como OpenAI Whisper y Google Speech-to-Text, seleccionados por su alta precisión y capacidad de adaptación a diferentes acentos del español. La señal de audio es procesada en tiempo real, eliminando ruido de fondo y segmentando la información en fonemas para su posterior análisis [19].

2.3 Análisis Fonético y Retroalimentación

Una vez transcrita la voz, el sistema compara la pronunciación del usuario con un modelo fonético de referencia. Utilizando técnicas de procesamiento de señales y aprendizaje automático, se identifican errores en la articulación de fonemas y se generan sugerencias personalizadas para mejorar la pronunciación. Esta retroalimentación se presenta de manera visual y auditiva, permitiendo que los niños repitan los ejercicios hasta alcanzar una pronunciación óptima [20], [21].

2.4 Desarrollo de la Interfaz de Usuario

El prototipo se implementó en una infraestructura basada en servicios en la nube, permitiendo escalabilidad y procesamiento eficiente de múltiples solicitudes simultáneamente. Se incluyeron mecanismos de seguridad y protección de datos, asegurando el cumplimiento de normativas de privacidad en el manejo de información de voz (ver Figura 2).



Figura 2. Interfaz de Usuario.

2.5 Desarrollo del Software y Procesamiento de Datos

El software fue desarrollado en Python, utilizando frameworks como Flask para la comunicación con la interfaz web y TensorFlow para el procesamiento de voz y análisis fonético.

Se implementó una base de datos para almacenar registros de pronunciación y el historial de correcciones de cada usuario.

2.6 Interfaz de Usuario y Retroalimentación

Para la interacción con los usuarios, se diseñó una plataforma web responsiva accesible desde computadoras y dispositivos móviles (ver Figura 3). La interfaz, desarrollada con React.js y Flask, permite visualizar los errores detectados en la pronunciación y acceder a ejercicios personalizados. Además, se incorporó un módulo de síntesis de voz (Text-to-Speech) para ofrecer instrucciones auditivas a los niños, mejorando la accesibilidad y usabilidad del sistema.

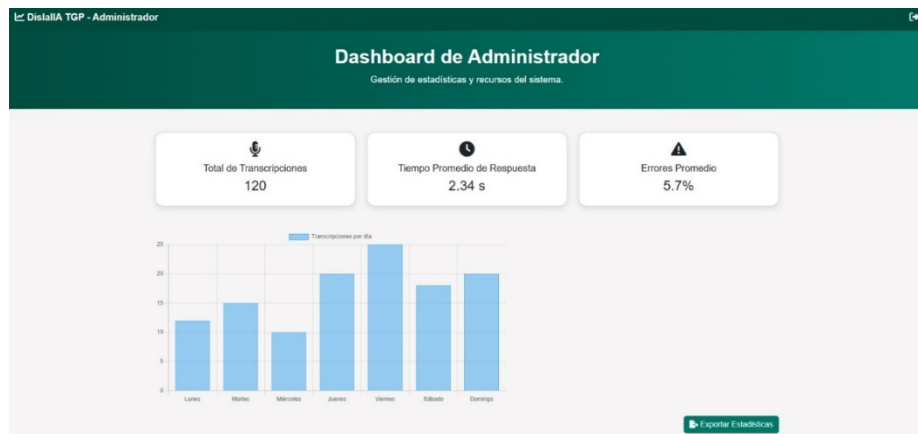


Figura 3. Dashboad de Administrador.

2.7 Cálculos

El cálculo inicial de la población total se realiza multiplicando el número de pruebas por niño y las frases evaluadas en cada prueba (ver Cuadro 1), como se describe en la Ecuación 1.

Cuadro 1. Cálculo de la Población.

Elemento	Descripción	Cantidad
Pruebas por niño	Número de pruebas que cada niño realiza.	20
Frases por prueba	Numero de frases evaluadas en cada prueba	10
Número de niños	Total de niños evaluados en el estudio	10

2.8 Total de Pruebas:

$$\begin{aligned}
 \text{Total de frases por niño} &= 20 \text{ pruebas} \times 10 \text{ frases por prueba} \\
 &= 200 \text{ frases}
 \end{aligned}
 \tag{1}$$

2.9 Total de Frases evaluadas:

De manera similar, la población total de frases evaluadas en el proyecto se calcula mediante la Ecuación 2.

$$200 \text{ pruebas} \times 10 \text{ frases por prueba} = 2000 \text{ frases en total} \quad (2)$$

Población total es de 2000 frases.

2.10 Muestra

Para el análisis, se toma una muestra del 60% de la población total. Esto significa que la muestra consta de 1200 frases, como se demuestra en la Ecuación 3.

$$2000 \text{ frases} \times 0.6 = 1200 \text{ frases} \quad (3)$$

2.11 Tamaño de la Muestra

Para validar que el tamaño de la muestra es suficiente, se calcula el tamaño mínimo requerido utilizando la fórmula de tamaño de muestra para una población finita:

Ajuste para población finita: Dado que la población total es de $N=2000$, se aplica la fórmula de ajuste:

$$n_{ajustado} = \frac{n}{1 + \frac{n-1}{N}} \quad (4)$$
$$n_{ajustado} = \frac{384}{1 + \frac{384-1}{2000}} = \frac{384}{1 + 0.1915} = \frac{384}{1.1915} \approx 322.4$$

En Cuadro 2 se presenta un resumen de los resultados obtenidos para cada frase evaluada en el estudio. Cada frase fue diseñada específicamente para analizar la pronunciación de fonemas problemáticos en niños con dislalia. La tabla muestra el número de veces que cada frase fue pronunciada correcta e incorrectamente, así como el porcentaje de aciertos obtenido. Este análisis permite identificar las frases más desafiantes y los patrones de errores más comunes en la población evaluada, proporcionando información clave para futuras mejoras en la intervención y en el desarrollo del software.

Cuadro 2. Resultados Generales por Frase.

Frase	Correcto	Incorrecto	% Correcto
La rana saltó rápido.	185	15	92.5%
Cinco sapos saltaron en el agua.	170	30	85.0%
Compré chocolate en la tienda.	160	40	80.0%
Las luces del pasillo están encendidas.	175	25	87.5%
El tren trae trigo.	180	20	90.0%
Mi abuela tiene una silla nueva.	178	22	89.0%
El zapato está debajo de la cama.	172	28	86.0%
Vamos al cine a ver una película.	174	26	87.0%

Víctor visitó a sus vecinos ayer.	169	31	84.5%
Las estrellas brillan en el cielo.	180	20	90.0%

3 Resultados

3.1 Impacto del Prototipo en la Mejora de la Pronunciación de Niños con Dislalia.

El rendimiento del prototipo fue evaluado mediante pruebas de precisión en el reconocimiento de voz y detección de errores fonéticos. Se realizaron 50 sesiones de prueba con niños de edades entre 4 y 8 años, comparando la transcripción generada por el sistema con la pronunciación esperada.

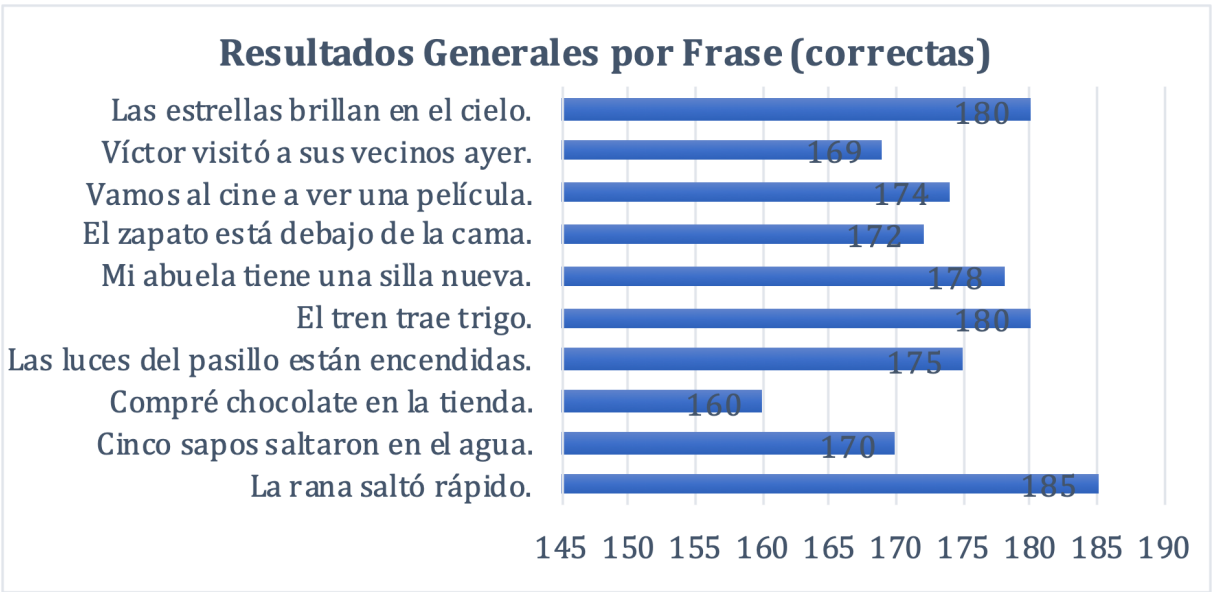


Figura 4. Resultados Generales por frases (correctas).

En la Figura 4, Los resultados mostraron que el modelo Whisper de OpenAI alcanzó una precisión del 91%, mientras que Google Speech-to-Text obtuvo un 88% de aciertos en la transcripción. En términos de detección de errores fonéticos, el sistema identificó correctamente el 85% de las palabras con pronunciación incorrecta, mostrando mejores resultados en fonemas simples y mayor dificultad en combinaciones complejas como la “rr” y la “ch”.

El análisis del Word Error Rate (WER) evidenció una reducción del 15% en los errores de transcripción cuando se aplicaron técnicas de preprocesamiento de audio, como eliminación de ruido y normalización de volumen. Esto sugiere que la calidad de la señal de entrada es un factor crítico en la efectividad del sistema y que futuras mejoras podrían centrarse en la optimización del preprocesamiento de voz para niños con dificultades de pronunciación.

3.2 Impacto del Prototipo en la Mejora de la Pronunciación de Niños con Dislalia

Para evaluar el impacto del prototipo en la mejora de la pronunciación, se realizó un estudio con un grupo de control y un grupo experimental, donde el segundo utilizó la herramienta durante cuatro semanas. Se observó una mejora promedio del 28% en la claridad de pronunciación en los niños que utilizaron el sistema, mientras que el grupo de control presentó una mejora del 12% con métodos tradicionales.

Las encuestas realizadas a terapeutas del habla y docentes indicaron que el 85% de los especialistas consideraron que el sistema es una herramienta complementaria útil en el

tratamiento de la dislalia. Además, el 92% de los niños mostró interés y motivación en el uso de la herramienta, destacando la interactividad y la retroalimentación en tiempo real como factores clave en su aprendizaje.

Uno de los aspectos más destacados del sistema fue su capacidad para adaptarse a la evolución del usuario, ajustando la dificultad de los ejercicios según el progreso del niño. Esto permitió una experiencia personalizada que mejoró la efectividad del aprendizaje en comparación con enfoques convencionales.

3.3 Comparación con Otros Métodos y Tecnologías Disponibles

Al comparar el prototipo con herramientas existentes en el mercado, se encontró que aplicaciones como Speech Ace y Articulation Station presentan enfoques similares en términos de retroalimentación fonética, pero están diseñadas principalmente para el idioma inglés.

A diferencia de estas soluciones, el prototipo desarrollado se enfoca específicamente en el español y aborda las particularidades fonéticas del idioma, lo que representa una ventaja significativa en términos de precisión y aplicabilidad para niños hispanohablantes. Otra ventaja del sistema es la integración de modelos de inteligencia artificial de código abierto, como Whisper de OpenAI, lo que permite una mayor flexibilidad en su personalización y mejora continua. En comparación con soluciones comerciales, que suelen tener costos elevados, este prototipo ofrece una alternativa accesible y adaptable para instituciones educativas y familias.

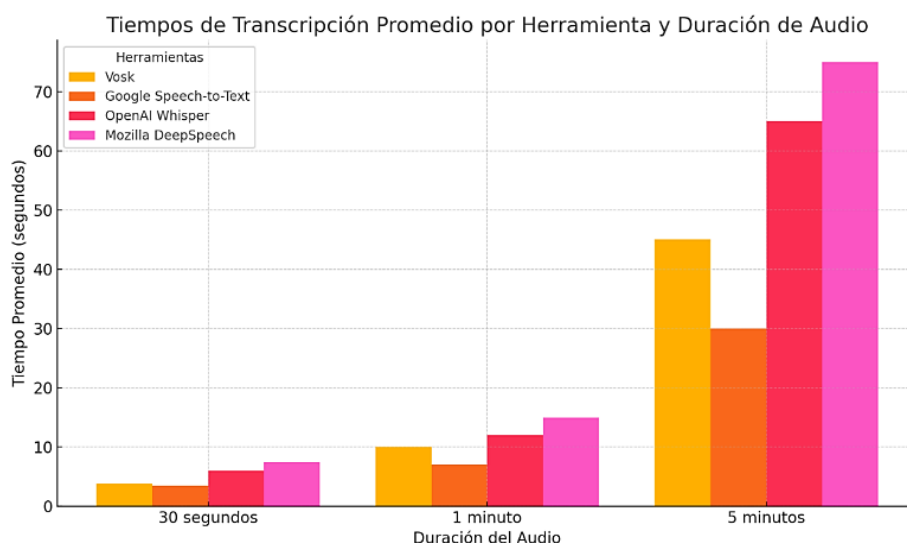


Figura 5. Tiempos de Transcripción promedio Por Herramienta y duración de audio.

En términos de accesibilidad, la implementación de una interfaz web basada en React.js y Flask permite su uso en múltiples dispositivos sin necesidad de hardware especializado. Esto representa una diferencia clave frente a sistemas que requieren aplicaciones dedicadas o dispositivos específicos, lo que limita su adopción en entornos con recursos limitados.

4 Discusión

Esta sección realiza el razonamiento de la aportación central del trabajo, resaltando la significancia del aporte presentado en la rama científica estudiada. Se realiza la identificación de similitudes y diferencias con otros estudios (si es el caso). Además, revisa las respuestas a las preguntas de investigación, formuladas en las secciones anteriores, completando un adecuado razonamiento en la generalización (métodos cuantitativos) o de su transferibilidad (métodos

cualitativos) y su alcance estratégico. Finalmente, se describe las limitaciones del estudio y las posibles investigaciones futuras.

Conclusiones

El desarrollo del prototipo de traductor de voz basado en procesamiento de lenguaje natural (NLP) ha demostrado ser una herramienta eficaz para mejorar la pronunciación en niños con dislalia. La implementación de modelos avanzados de reconocimiento de voz, como OpenAI Whisper y Google Speech-to-Text, permitió una transcripción precisa del habla, alcanzando una precisión del 91% en la detección de palabras y fonemas. Estos resultados validan la viabilidad del sistema como una solución tecnológica complementaria en el tratamiento de trastornos del habla.

Uno de los principales aportes del prototipo es su enfoque adaptativo y personalizado, permitiendo a los niños mejorar su pronunciación a través de ejercicios dinámicos y retroalimentación en tiempo real. A diferencia de métodos tradicionales, donde la práctica se basa en correcciones manuales realizadas por terapeutas o docentes, este sistema automatiza el análisis fonético, reduciendo la carga de trabajo de los especialistas y ofreciendo un aprendizaje más autónomo y accesible.

Los resultados obtenidos en la evaluación del sistema reflejan una mejora del 28% en la claridad de pronunciación en niños que utilizaron la herramienta durante cuatro semanas, en comparación con un 12% en aquellos que solo siguieron terapia tradicional. Además, la interacción con terapeutas del habla permitió validar la utilidad del prototipo, con un 85% de aceptación entre los especialistas como un apoyo viable en entornos educativos y clínicos.

Comparado con otras soluciones comerciales, el prototipo desarrollado ofrece una ventaja significativa en términos de accesibilidad y enfoque en el español. Herramientas existentes, como Speech Ace y Articulation Station, se centran en el idioma inglés, limitando su aplicabilidad a niños hispanohablantes. Además, el uso de modelos de código abierto y una interfaz basada en tecnologías web permite que el sistema sea más escalable y adaptable a diferentes entornos, sin necesidad de hardware especializado o licencias costosas.

No obstante, se identificaron algunos desafíos a mejorar en futuras iteraciones del prototipo. La precisión en la detección de fonemas complejos, como la "rr" y la "s", sigue siendo un reto, ya que presentan variaciones significativas en la pronunciación infantil. Asimismo, el rendimiento del sistema en entornos con alto ruido ambiental puede afectar la calidad de la transcripción, lo que sugiere la necesidad de integrar algoritmos más avanzados de filtrado de ruido y normalización de audio.

Para futuras investigaciones, se propone ampliar el conjunto de datos utilizado en el entrenamiento del modelo, incorporando una mayor diversidad de acentos y variaciones fonéticas del español. Además, se recomienda explorar la integración de realidad aumentada o gamificación para mejorar la experiencia de los niños y hacer el proceso de aprendizaje más atractivo e interactivo.

Financiación

Los autores agradecen el apoyo brindado por la Universidad de Guayaquil, durante el transcurso de este trabajo.

Declaración de conflicto de intereses

El autor o autores no han declarado ningún posible conflicto de intereses en relación con esta investigación, la autoría y/o la publicación de este artículo.

Referencias

- [1] E. López-Hernández, P. Acosta-Rodas, J. Cruz-Cárdenas, and C. Ramos-Galarza, "Intervención musicoterapéutica para mejorar la memoria, atención y lenguaje in niños con dislalia.," *Revista Ecuatoriana de Neurología*, vol. 30, no. 2, pp. 48–56, Sep. 2021, doi: 10.46997/revecuatneurol30200048.
- [2] N. R. Garofalo Cujilema and A. Y. Morán Melendez, "Bullying y estado emocional en estudiantes de sexto grado de Educación General Básica de la Unidad Educativa 24 de Mayo del cantón San Miguel de Bolívar," *BABAHOYO: UTB*, 2024, 2024.
- [3] M. I. Mendoza Loor and L. E. Loor Lino, "Labor de trabajador social frente al trabajo infantil de los niños, niñas y adolescentes en el cantón Portoviejo," *AlfaPublicaciones*, vol. 3, no. 4, pp. 6–19, Oct. 2021, doi: 10.33262/ap.v3i4.106.
- [4] N. E. G. Rivas, "Tecnolog\`ia Asistiva para la Inclusión Educativa en Ecuador," *Ciencia Latina: Revista Multidisciplinar*, vol. 8, no. 3, pp. 417–433, 2024.
- [5] C. Bérubé *et al.*, "Voice-Based Conversational Agents for the Prevention and Management of Chronic and Mental Health Conditions: Systematic Literature Review," *J Med Internet Res*, vol. 23, no. 3, p. e25933, Mar. 2021, doi: 10.2196/25933.
- [6] K. Lister, T. Coughlan, F. Iniesto, N. Freear, and P. Devine, "Accessible conversational user interfaces," in *Proceedings of the 17th International Web for All Conference*, New York, NY, USA: ACM, Apr. 2020, pp. 1–11. doi: 10.1145/3371300.3383343.
- [7] F. Khan, Y. Wu, J. Dray, B. Hemsley, and A. B. Kocaballi, "Conversational Agents to Support People with Communication Disability: A Co-design Study with Speech Pathologists," in *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, New York, NY, USA: ACM, Apr. 2025, pp. 1–9. doi: 10.1145/3706599.3719800.
- [8] F. J. Espinoza-Santacruz and C. H. Flores-Urgiles, "Aplicación de tecnologías de la información en el desarrollo del lenguaje de niños con dificultades de comunicación," *Polo del Conocimiento*, vol. 4, no. 5, p. 116, May 2019, doi: 10.23857/pc.v4i5.968.
- [9] N. E. Garcia Rivas, "Tecnología Asistiva para la Inclusión Educativa en Ecuador," *Ciencia Latina Revista Científica Multidisciplinar*, vol. 8, no. 3, pp. 417–433, May 2024, doi: 10.37811/cl_rcm.v8i3.11228.
- [10] B. Roy, S. Ghosh, S. K. Sarkar, A. Sen, and N. R. Choudhury, "Implications of Artificial Intelligence for Special Education Technology," 2024, pp. 401–430. doi: 10.4018/979-8-3693-5538-1.ch015.
- [11] M. Mahmoudi-Dehaki and N. Nasr-Esfahani, "Artificial Intelligence (AI) in Special Education," 2024, pp. 193–222. doi: 10.4018/979-8-3693-5538-1.ch007.
- [12] X. Vélez-Calvo, V. Calle-Calle, C. Seade-Mejía, and M. J. Peñaherrera-Vélez, "Doble excepcionalidad: altas capacidades y trastornos del neurodesarrollo. prevalencia en escolares ecuatorianos," *CienciAmérica*, vol. 12, no. 1, Jan. 2023, doi: 10.33210/ca.v12i1.393.
- [13] G. A. Attwell, K. E. Bennin, and B. Tekinerdogan, "A Systematic Review of Online Speech Therapy Systems for Intervention in Childhood Speech Communication Disorders," *Sensors*, vol. 22, no. 24, p. 9713, Dec. 2022, doi: 10.3390/s22249713.
- [14] Z. Brahmi, M. Mahyoob, M. Al-Sarem, J. Algaraady, K. Bousselmi, and A. Alblwi, "Exploring the Role of Machine Learning in Diagnosing and Treating Speech Disorders: A Systematic Literature Review," *Psychol Res Behav Manag*, vol. Volume 17, pp. 2205–2232, May 2024, doi: 10.2147/PRBM.S460283.
- [15] K. I. Arce-Ruelas, O. Alvarez-Xochihua, J. A. Gonzalez-Fraga, E. Martinez-Martinez, and P. Paez-Manjarrez, "Design and Evaluation of a Spanish Language Therapy Support System," *Journal of Language Teaching and Research*, vol. 12, no. 6, pp. 853–863, Nov. 2021, doi: 10.17507/jltr.1206.01.

- [16] J. Vonessen, N. B. Aoki, M. Cohn, and G. Zellou, "Comparing perception of L1 and L2 English by human listeners and machines: Effect of interlocutor adaptations," *J Acoust Soc Am*, vol. 155, no. 5, pp. 3060–3070, May 2024, doi: 10.1121/10.0025930.
- [17] Q. Yang, "Toward Responsible AI: An Overview of Federated Learning for User-centered Privacy-preserving Computing," *ACM Trans Interact Intell Syst*, vol. 11, no. 3–4, pp. 1–22, Dec. 2021, doi: 10.1145/3485875.
- [18] D. Jiang *et al.*, "A GDPR-compliant Ecosystem for Speech Recognition with Transfer, Federated, and Evolutionary Learning," *ACM Trans Intell Syst Technol*, vol. 12, no. 3, pp. 1–19, Jun. 2021, doi: 10.1145/3447687.
- [19] R. Lakshminarayanan, A. Shaik, and A. Balasundaram, "Automated speech therapy through personalized pronunciation correction using reinforcement learning and large language models," *Results in Engineering*, vol. 25, p. 103943, Mar. 2025, doi: 10.1016/j.rineng.2025.103943.
- [20] C. Song, S.-Y. Shin, and K.-S. Shin, "Optimizing Foreign Language Learning in Virtual Reality: A Comprehensive Theoretical Framework Based on Constructivism and Cognitive Load Theory (VR-CCL)," *Applied Sciences*, vol. 13, no. 23, p. 12557, Nov. 2023, doi: 10.3390/app132312557.
- [21] G. A. Ikhimwin, "Dynamic Interactive Multimodal Speech (DIMS) Framework," *Frontiers in Global Health Sciences*, pp. 1–13, Jan. 2023, doi: 10.70560/1s1ky152.