

EchoLens: Abriendo Barreras de Comunicación

Tomas Acosta, Agustin Curilef, Olivetti Angel Ezequiel,
Juan Cruz Martinez Luquez, Mauro Cambarieri

Universidad Nacional de Río Negro (UNRN). Laboratorio de Informática Aplicada (LIA). Sede
Atlántica, Viedma, Río Negro.

{tomsacosta19,agustin.curilef,angelolivetti78}@gmail.com
{mcambarieri, jcmluquez}@unrn.edu.ar

Abstract.

Este informe presenta el desarrollo de *EchoLens*, una solución tecnológica orientada a personas con discapacidades auditivas, una problemática creciente en Latinoamérica con fuerte impacto social, académico y laboral. El proyecto propone unos lentes inteligentes que proyectan subtítulos en tiempo real mediante tecnologías de Inteligencia Artificial y Cloud Computing. *EchoLens* fue presentado en la Huawei ICT Innovation Competition 2024/25, obteniendo el tercer puesto a nivel nacional y fue declarado de interés por la Legislatura de Río Negro (declaración N° 334/2025). La metodología empleada contempla el diseño físico de los lentes y la arquitectura del software, con dos enfoques principales que permitieron analizar mejoras futuras y los desafíos técnicos a largo plazo.

Palabras claves. Lentes inteligentes, subtítulos, inteligencia artificial, procesamiento de voz, diarización, realidad aumentada, cloud computing.

1. Introducción:

1.1. Contexto y motivación.

El desarrollo tecnológico es uno de los principales impulsores del desarrollo humano. Bajo esta premisa, Huawei presenta una competencia donde estudiantes de todo el mundo comparten sus ideas y proyectos que busquen innovar en pos de una sociedad más igualitaria. En este contexto, nace el proyecto *EchoLens*.

EchoLens es una propuesta tecnológica presentada en la *Huawei ICT Innovation Competition 2024/25*, con el acompañamiento académico de Mauro Cambarieri y Juan Cruz Martínez Luquez. La solución busca mejorar la comunicación de personas con discapacidades auditivas a través de la utilización de lentes inteligentes los cuales proyectan subtítulos en tiempo real.

La iniciativa responde a una problemática creciente: las discapacidades auditivas están en aumento y se proyecta un crecimiento significativo en los próximos 25 años [1]. En América Latina, se determinó que dichas problemáticas no son una prioridad para los gobiernos [2], y quienes las padecen suelen encontrarse en situaciones de vulnerabilidad social: bajos ingresos, menor nivel educativo, desempleo o pobreza [3].

Frente a esta situación, *EchoLens* promueve la inclusión social al facilitar la interacción cotidiana en entornos familiares, educativos y laborales. La solución utiliza inteligencia artificial en tres módulos principales: diarización (identificación de hablantes), transcripción automática y traducción multilingüe. Esta arquitectura permite interpretar conversaciones con precisión, adaptándose a contextos diversos como la educación, el turismo o espacios institucionales.

2. Metodología

2.1. Diseño de los lentes

El proyecto contempla el desarrollo de dos prototipos de lentes inteligentes, ambos basados en el mismo principio óptico: la refracción de luz mediante espejos y vidrios. La principal

diferencia entre ellos radica en la ubicación de la pantalla OLED y en aspectos de ergonomía que influyen en la comodidad de uso.

Principio de funcionamiento

Ambos prototipos utilizan un sistema de proyección que combina una pantalla OLED, un espejo y un vidrio espejado, permitiendo superponer información visual directamente en el campo de visión del usuario sin estorbar la percepción del entorno. Esta técnica ofrece una experiencia similar a la de un *head-up display* (HUD).

Componentes utilizados

- Pantalla OLED 128x32 (0.91"): utilizada para proyectar la información visual.
- Microcontrolador ESP32-C3 Supermini: encargado del procesamiento y comunicación inalámbrica.
- Micrófono digital INMP-441: captura de audio para aplicaciones de reconocimiento de voz o interacción por comandos.
- Módulo táctil capacitivo TTP223: permite interacción sencilla con el sistema mediante toques.
- Espejo y vidrio con film espejado: elementos ópticos para redirigir y superponer la imagen proyectada.
- Lente amplificador: mejora la visibilidad de la imagen proyectada y amplía el ángulo de visión.

Fabricación de la estructura

Los marcos de los lentes serán diseñados mediante software de modelado 3D, lo que permite optimizar la ergonomía y adaptación al hardware. Para su fabricación, se utilizará tecnología de impresión 3D con filamento PETG, seleccionado por su resistencia mecánica, flexibilidad y durabilidad, lo cual lo hace adecuado para aplicaciones de uso diario.

2.2. Elección del software.

Utilización de la Inteligencia Artificial.

Para abordar las tareas de diarización, transcripción y traducción automática, se consideraron dos enfoques: Un desarrollo local y una integración externa.

El desarrollo local implica utilizar modelos de inteligencia artificial de desarrollo propio o de código abierto. Partiendo de esta premisa, se seleccionaron modelos de inteligencia artificial de código abierto ampliamente validados, alojados en la plataforma *HuggingFace*. Esta decisión se fundamenta en la necesidad de contar con herramientas confiables, reproducibles y adaptables a distintos contextos lingüísticos. Los modelos considerados son:

- I. Pyannote.audio: utilizado para la diarización del hablante. Este modelo se destaca por su desempeño en entornos reales con múltiples interlocutores.
- II. Whisper (OpenAI): empleado para la transcripción automática del habla, reconocido por su alta precisión y robustez frente a variaciones en el acento y ruido de fondo.
- III. Helsinki-NLP (Tatoeba Translation Models): utilizado para la traducción automática multilingüe, con soporte para más de 100 idiomas.

Estos modelos se ejecutan en un entorno virtual proporcionado por Google Cloud Platform (GCP), lo cual permite externalizar el procesamiento de audio sin requerir hardware especializado embebido en los lentes. La decisión de externalizar el procesamiento

En cambio, en el segundo enfoque, se analizó la posibilidad de integrar el servicio Google Gemini, una solución basada en IA generativa que ofrece capacidades unificadas de transcripción, traducción y comprensión contextual. En las pruebas preliminares, Gemini demostró un rendimiento superior en precisión y con una mayor eficiencia operativa, al unificar las tres tareas en una única API.

No obstante, esta alternativa plantea desafíos relevantes en términos de privacidad y seguridad de los datos. Por defecto, los servicios de Gemini pueden retener temporalmente las conversaciones para fines de monitoreo [4], lo cual resulta incompatible con los estándares

éticos de privacidad requeridos por aplicaciones sensibles como *EchoLens*. Para mitigar estos riesgos, se identificaron las siguientes condiciones técnicas necesarias:

- Configuración del servicio dentro de Google Workspace, con el historial de interacciones completamente desactivado [5].
- Garantizar que los datos procesados no se utilicen para entrenamiento del modelo, deshabilitando explícitamente la opción de aprendizaje automático con datos de usuario [4], [5].
- En caso de implementación institucional, establecer acuerdos de procesamiento de datos (DPA/CDPA) [5] formales con el proveedor, que aseguren el cumplimiento de las normativas de protección de datos personales .

La decisión final entre continuar con una arquitectura modular basada en modelos open-source, o avanzar hacia una solución unificada mediante Gemini, se encuentra aún en evaluación. Dicha elección dependerá de factores como escalabilidad, robustez en entornos reales, latencia en tiempo real y garantías de privacidad y soberanía de datos.

Desarrollo de aplicación móvil.

Se definió como componente esencial el desarrollo de una aplicación móvil intermediaria entre los lentes inteligentes y la infraestructura en la nube, encargada de gestionar la conexión, captar el audio y derivarlo al procesamiento en Google Cloud Platform.

Aunque la elección tecnológica aún está en análisis, se contemplan frameworks multiplataforma como Flutter y React Native, junto con opciones nativas como Kotlin, Java y Swift.

El integral de la solución tiene como objetivo priorizar la privacidad, procesando los datos en tiempo real sin almacenamiento de audio ni transcripciones. Cumple con la Ley N.º 25.326 de Protección de Datos Personales, garantizando el tratamiento informado y seguro de la información. En caso de expansión internacional, se prevé la adaptación normativa según la legislación vigente en cada jurisdicción

3. Líneas de trabajo.

3.1. Prototipos planificados.

Como parte de la etapa inicial del proyecto, se planificó el desarrollo de dos prototipos con diferentes configuraciones en la ubicación de la pantalla OLED.

El primer prototipo posiciona la pantalla sobre la línea superior de los lentes, cerca de la frente del usuario, mientras que el segundo la ubica en uno de los laterales del marco. Ambas configuraciones presentan ventajas y desventajas relacionadas principalmente con la óptica del sistema, ya que modifican los ángulos de reflexión y la forma en que la imagen se proyecta hacia el ojo.

Actualmente, nos encontramos evaluando estos diseños a través de pruebas preliminares con un grupo seleccionado de usuarios. El feedback recopilado será clave para determinar cuál de los prototipos resulta más eficiente y cómodo, así como cuál ofrece mayor potencial de mejora en futuras iteraciones. Esta etapa representa el punto de partida para una evolución iterativa del diseño basada en la experiencia del usuario y el rendimiento óptico observado.

3.2. Modelo ideal planificado.

El desarrollo de los prototipos apunta a un modelo ideal que integre inteligencia artificial y realidad aumentada (RA) de forma robusta. Este diseño contempla una doble diarización —mediante audio e imagen— utilizando una cámara incorporada en los lentes, lo que permitirá superponer subtítulos en tiempo real directamente debajo de cada interlocutor. Para ello, se prevé la incorporación de componentes como cámaras, pantallas específicas para subtítulos y una unidad de procesamiento mejorada, lo cual implicará un rediseño hacia una estructura de mayor volumen, que se asemeja a dispositivos de RA como *Meta Quest*. El objetivo final es optimizar la legibilidad, reducir el esfuerzo visual y ofrecer una interfaz moderna y funcional.

3.3. Limitaciones técnicas actuales.

Durante el desarrollo de los prototipos iniciales, se han identificado diversas limitaciones técnicas que condicionan tanto el rendimiento como la experiencia de usuario. Estas limitaciones constituyen desafíos clave a resolver en futuras iteraciones del proyecto.

Latencia en el procesamiento

Uno de los principales desafíos actuales es la latencia en tareas que requieren procesamiento de voz o interacción en tiempo real, debido al envío de audio a un dispositivo externo para su análisis y posterior respuesta. Una posible solución futura es el uso de procesamiento multihilo en el servicio externo, lo que permitiría realizar distintas tareas en paralelo y reducir así los tiempos de espera percibidos por el usuario.

Cuestiones ópticas

A nivel óptico, uno de los retos es lograr una proyección clara y legible sin interferir con la visión del entorno. Factores como la ubicación de los espejos, el ángulo de reflexión, la intensidad del brillo de la pantalla OLED y la calidad del film espejado impactan directamente en la legibilidad del contenido. Aún se están realizando ajustes en los prototipos para encontrar un equilibrio adecuado entre visibilidad y transparencia.

Autonomía energética

La duración de la batería es otra limitación importante. Debido al tamaño compacto del dispositivo, la capacidad de incorporar baterías de gran autonomía es reducida. El uso eficiente de energía por parte del microcontrolador y la pantalla OLED, así como la posibilidad de implementar modos de bajo consumo o sistemas de apagado automático cuando el dispositivo no está en uso, serán fundamentales para mejorar la autonomía sin comprometer la funcionalidad.

4. Conclusión

4.1. Resumen de los aportes claves.

EchoLens nació bajo el marco de una competencia que busca desarrollar tecnologías innovadoras que promuevan la igualdad, la inclusión, la integración y solucionen problemáticas globalizadas.

Desde los albores de la idea hasta el día de la publicación de este informe, el proyecto ha evolucionado profundamente. Se destaca principalmente la incorporación de tecnologías de procesamiento en la nube, cuyos resultados demostraron una mejoría significativa en tiempos de procesamiento. De la misma forma, se logró aumentar la precisión de las transcripciones y traducciones al utilizar el servicio de Gemini en lugar de modelos de desarrollo propio.

Por otro lado, la declaración de interés por parte de la Legislatura de Río Negro es otro aporte clave; al darle visibilidad al proyecto y promover su desarrollo mediante financiamiento, demostró la importancia del acompañamiento docente y la generación de espacios de interacción entre alumnos para el desarrollo profesional y científico de las universidades. Además, dicho financiamiento permitió transformar el proyecto de algo teórico a algo tangible.

Finalmente, *EchoLens* proyecta seguir creciendo, explorando nuevas aplicaciones tecnológicas y ampliando su alcance social. Este proyecto es testimonio del valor de la formación universitaria, del trabajo en equipo y del impulso brindado por espacios como el Laboratorio de Informática Aplicada (LIA) de la UNRN, sin los cuales esta idea no se habría transformado en realidad.

5. Referencias

1. World Health Organization. (2021). World report on hearing. World Health Organization. <https://www.who.int/publications/i/item/world-report-on-hearing>
2. Madriz, J. J. (2000). Hearing impairment in Latin America: An inventory of limited options and resources [Trastornos auditivos en América Latina: Un inventario de opciones y recursos limitados]. *International Journal of Audiology*, 39(4), 212–220. <https://doi.org/10.3109/00206090009073081>
3. Fuentes-López, E., & Fuente, A. (2020). Access to healthcare for deaf people: A model from a middle-income country in Latin America. *Revista de Saúde Pública*, 54, Article 13. <https://doi.org/10.11606/s1518-8787.2020054001864>

4. Google AI. (n.d.). Gemini API: Additional usage policies. Retrieved July 17, 2025, from <https://ai.google.dev/gemini-api/docs/usage-policies>
5. Google Cloud. (n.d.). Data handling and privacy for Gemini in Google Cloud. Retrieved July 17, 2025, from <https://cloud.google.com/vertex-ai/docs/generative-ai/data-governance>