



**Universidad Nacional de Lomas de
Zamora**
Facultad De Ingeniería



PROYECTO DE INGENIERÍA MECATRÓNICA

INFORME **TÉCNICO** **“SignGlasses”**

2025

DOCENTES:

- BLANCA, EZEQUIEL
- LUCASZEWICZ, CRISTIAN
- SZOMBACH, JUAN

INTEGRANTES	DNI
GAUNA, IGNACIO EZEQUIEL	42.874.840
SARACINO, JUAN PABLO	42.113.281

ÍNDICE

1. INTRODUCCION	4
2. FUNCIONALIDAD DEL DISPOSITIVO	5
2.1. Estructura General del Proyecto	5
2.2. Descripción del Flujo de Funciones.....	5
2.3. Sincronización y Operación Simultánea.....	6
3. EVOLUCIÓN DEL PROTOTIPO	7
3.1. Concepto Inicial	7
3.2. Model 3.....	7
4. DISEÑO ELECTRÓNICO	8
4.1. Esquema de Conexión (Fritzing).....	8
4.2. Cámara OMNIVISION OV5647	9
4.3. Preamplificador PAM8403.....	10
4.4. Pantalla OLED 0.96 SSD1306	11
4.5. MICRÓFONO INMP441	11
4.6. Raspberry Pi Zero 2WH	12
4.7. MOTOR SERVO SG90	12
4.8. Análisis de Consumo de Energía y Conexionado.....	13
5. SOFTWARE	14
5.1. Programa de Recolección, Entrenamiento y Administración de Señas	14
5.2. Programa de Evaluación en Tiempo Real	16
5.3. Programa Implementado en la Raspberry.....	17
6. MODELO DE MACHINE LEARNING Y RED NEURONAL	17
6.1. Arquitectura implementada:	17
7. ENSAYOS Y TESTEOS REALIZADOS.	19
7.1. Evaluación del Impacto de la Iluminación:.....	19
7.2. Calidad de Reproducción de Audio	20
7.3. Visibilidad de la Pantalla	20
7.4. Validación de Modelos de Machine Learning	20
8 COSTOS	21
8.1. Costo real del prototipo.....	21
8.2. Costo estimado del proyecto real.....	21
9. MEJORAS A IMPLEMENTAR.....	22
9.1. Pantalla.....	23
9.2. Micrófono.....	23
9.3. Sistema de Audio	23
9.4. Procesamiento.....	23

9.4.1. Ejecución en dispositivo móvil	23
9.4.2. Despliegue en servidor web	24
9.4.3. Sistema Embebido	24
9.5. Enriquecimiento de la traducción	24
9.5.1. Precisión y Latencia	24
9.5.2. Contextualización Lingüística	24
9.6. Reducción de Tamaño y Peso.....	25
9.7. Sistema Inalámbrico	25
9.8. Incorporar reconocimiento de expresiones faciales.....	25
9.9. Diccionario LSA Completo	26
9.10. Visión evolutiva:	26
10. CONCLUSIONES.....	26

1. INTRODUCCION

El presente informe técnico expone el desarrollo e implementación de un sistema portátil de asistencia comunicacional denominado **SignGlasses**, un prototipo de lentes inteligentes orientados a la traducción de lenguaje de señas y transcripción de voz en tiempo real. La iniciativa surge como una solución tecnológica para mejorar la accesibilidad e inclusión de personas sordas o con dificultades auditivas, facilitando la comunicación bidireccional con interlocutores oyentes en entornos cotidianos.

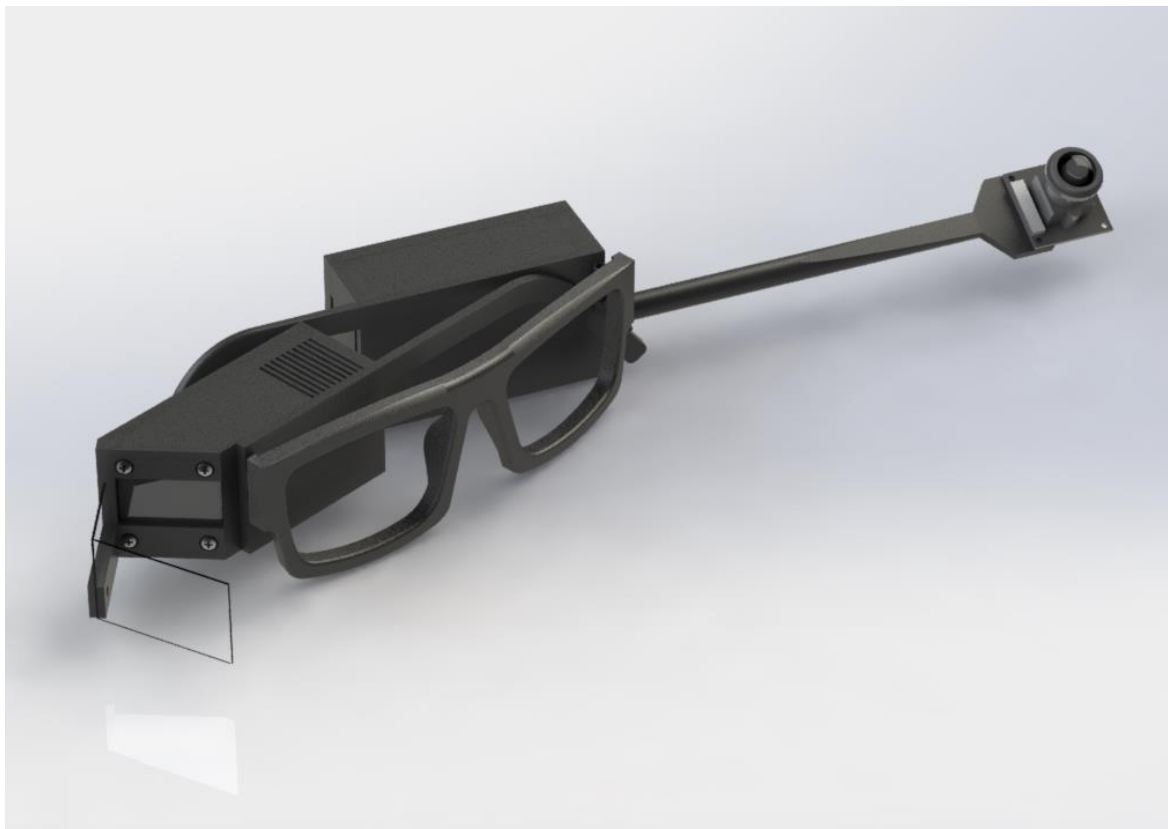


Figura 1: Vista Isométrica SignGlasses Model 3 con las patillas cerradas

El proyecto combina tecnologías de visión por computadora y procesamiento de lenguaje natural. El sistema se basa en la utilización de una placa de procesamiento Raspberry Pi, una cámara integrada para la detección y reconocimiento de señas mediante modelos de inteligencia artificial, y una pantalla HUD (Head-Up Display) para la visualización de textos transcritos. Además, el sistema incorpora un módulo de reconocimiento de voz que permite la transcripción instantánea del lenguaje hablado hacia texto visible para el usuario.

Este documento detalla el diseño, los componentes electrónicos, así como la arquitectura de software desarrollada para lograr un funcionamiento autónomo y en tiempo real. El objetivo principal es validar la factibilidad técnica de un dispositivo compacto, accesible y de bajo costo, capaz de asistir en la interpretación de lenguaje de señas argentino (LSA) y mejorar la experiencia comunicativa de sus usuarios.

2. FUNCIONALIDAD DEL DISPOSITIVO

2.1. Estructura General del Proyecto

El prototipo desarrollado bajo el sistema SignGlasses presenta una doble funcionalidad principal:

- **Traducción en tiempo real de LSA a Español oral**, permitiendo que la persona sorda adopte el rol de locutor en la comunicación.
- **Transcripción en tiempo real de Español oral a subtítulos visibles**, facilitando que la persona sorda actúe como receptor durante una conversación con oyentes.

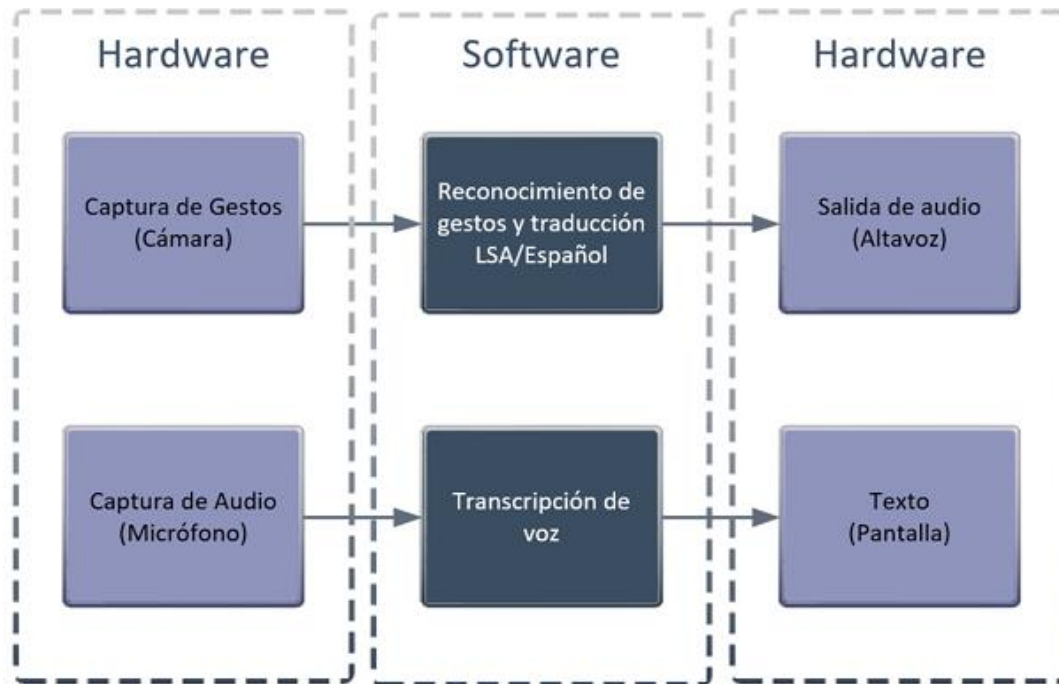


Figura 2: Diagrama de bloques a nivel macro del funcionamiento de las "SignGlasses"

2.2. Descripción del Flujo de Funciones

a) Traducción de LSA a Español (Emisión):

Para la función de traducción LSA → Español, el sistema opera mediante el siguiente flujo técnico:

1. La captura de señas se realiza mediante una cámara ubicada sobre un soporte telescópico en los lentes, acompañada de un sistema de seguimiento de la mano en tiempo real, gestionado por un servomotor encargado de orientar el campo visual.
2. La imagen capturada es procesada inicialmente por la Raspberry Pi, la cual fragmenta el flujo de vídeo y lo transmite a una unidad de procesamiento (PC) mediante una comunicación basada en UDP sobre USB.
3. En el PC, se ejecuta un modelo de inteligencia artificial previamente entrenado, utilizando MediaPipe para la detección de landmarks de la mano y clasificadores de aprendizaje profundo para la identificación de señas.

4. La etiqueta (label) resultante es asociada a un archivo de audio pregrabado correspondiente, el cual es retornado a la Raspberry Pi mediante el mismo canal de comunicación.
5. Finalmente, el audio es reproducido por el sistema de parlantes integrado en los lentes, permitiendo al usuario sordo expresarse de forma oral ante oyentes.

b) Transcripción de Español Oral a Subtítulos (Recepción):

De manera simultánea, para la función de transcripción Español oral → subtítulos, el flujo es el siguiente:

1. El micrófono INMP441 integrado en los lentes capta fragmentos de audio en intervalos periódicos de 5 segundos.
2. La Raspberry Pi procesa el audio entrante y lo transmite a la PC mediante el mismo sistema de transporte UDP.
3. El audio es procesado por un modelo de reconocimiento de voz (speech-to-text), que convierte las señales sonoras en texto escrito.
4. El texto transcrito es reenviado a la Raspberry Pi, que lo direcciona al display OLED integrado en la montura, permitiendo al usuario visualizar subtítulos de la conversación en tiempo real.

2.3. Sincronización y Operación Simultánea

Ambos procesos (traducción de señas y transcripción de voz) se ejecutan de forma paralela, gestionados por la Raspberry Pi como núcleo coordinador, brindando al usuario sordo la capacidad de participar activamente en cualquier interacción comunicativa, tanto emitiendo mensajes orales como recibiendo información visualmente a través de subtítulos.

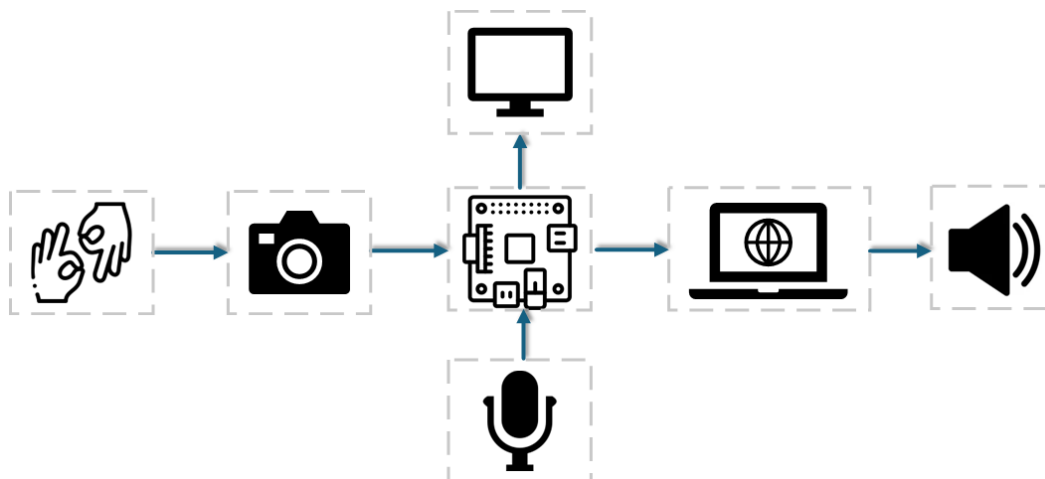


Figura 3: Esquema de operación simplificado del hardware

3. EVOLUCIÓN DEL PROTOTIPO

3.1. Concepto Inicial

El proyecto SignGlasses tuvo su origen en el marco de la cátedra de *Vigilancia Tecnológica e Inteligencia Competitiva*, donde se elaboró un primer concepto orientado a la identificación de una necesidad concreta en materia de accesibilidad comunicacional. En esta etapa inicial, el enfoque fue exclusivamente esquemático, centrado en la conceptualización de los componentes clave del sistema, su lógica de funcionamiento y las posibles aplicaciones. No se contempló la implementación física del prototipo, pero esta fase resultó fundamental para definir los lineamientos técnicos y funcionales que guiaron el desarrollo posterior del dispositivo.



Figura 4: Diseño conceptual para SignGlasses

3.2. Model 3

El Model 3 representa la iteración más reciente y refinada tras un proceso de mejora continua basado en las versiones preliminares, Model 1 y Model 2. A partir de las pruebas realizadas con los prototipos anteriores, se identificaron y resolvieron diversas limitaciones ergonómicas y estructurales. Las principales mejoras implementadas incluyen:

- Optimización de la ergonomía y del confort de uso, con un diseño que se adapta de manera más estable al cráneo y al tabique nasal.
- Incorporación de ranuras específicas para una mejor gestión del cableado interno y de las conexiones eléctricas.
- Inclusión de fichas de alimentación accesibles y estratégicamente ubicadas para mejorar la portabilidad y la autonomía del sistema.

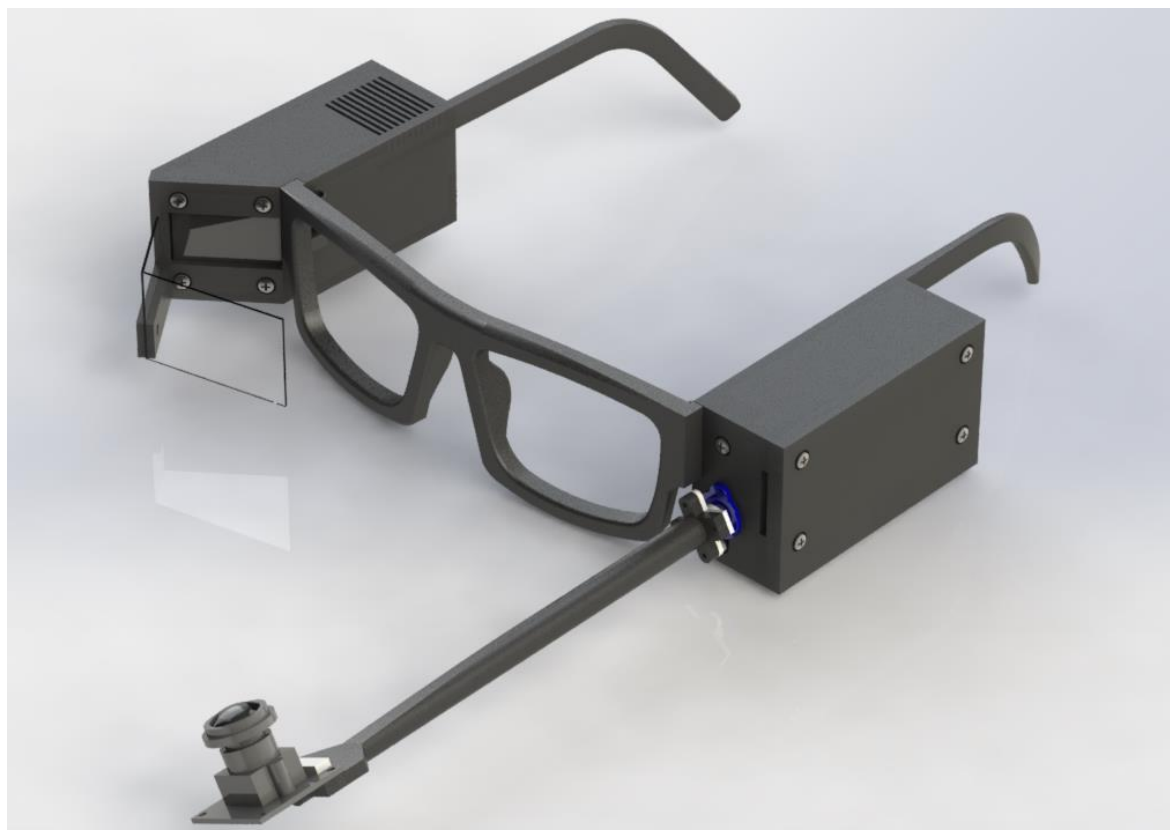


Figura 5: Vista Isométrica de las SignGlasses Model 3

4. DISEÑO ELECTRÓNICO

4.1. Esquema de Conexión (Fritzing)

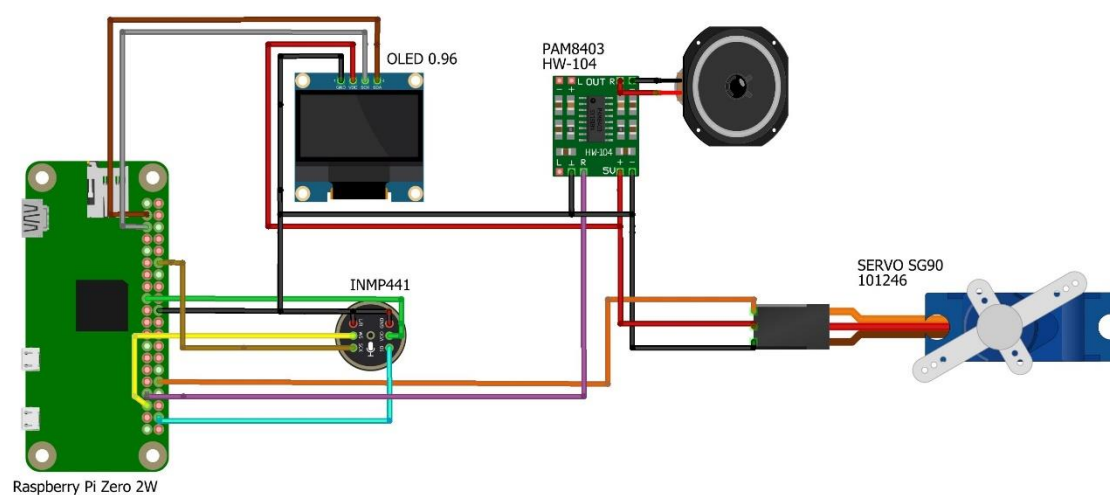


Figura 6: Conexión de los pines de la Raspberry Pi Zero 2w a los diferentes componentes del proyecto

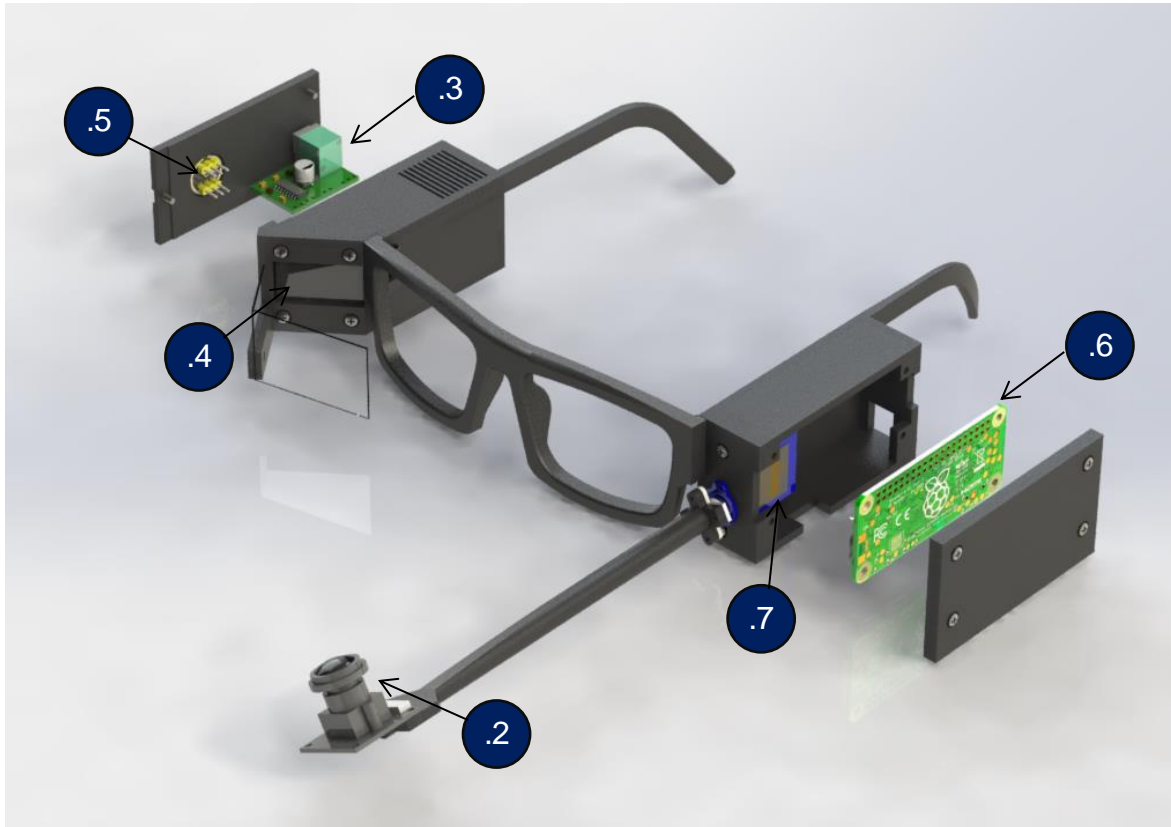


Figura 7: Vista Isométrica explosionada con numeración de los componentes.

4.2. Cámara OMNIVISION OV5647

Para la captura de video de los gestos en LSA, se utilizó el sensor OMNIVISION OV5647 cuenta con CMOS de 5 megapíxeles, ampliamente compatible con plataformas de procesamiento como Raspberry Pi a través de interfaz CSI. La cámara cuenta con un lente tipo *ojo de pez* con una apertura de **130°**, lo que permite un ángulo de visión ultra amplio, asegura la captación de gestos y movimientos corporales sin necesidad de reposicionar la cámara.



Figura 8: OV5647 imagen de catalogo

4.3. Preamplificador PAM8403

Para la reproducción de audio traducido (LSA - español) se empleó el módulo amplificador PAM8403, un amplificador de potencia clase D con una capacidad de salida de hasta 3W por canal en estéreo a 5V. El módulo integra un potenciómetro para el ajuste de volumen, permitiendo al usuario regular la intensidad del audio de salida según la intensidad a la cual quiere ser escuchado. Su elección se basó en su alta eficiencia energética, bajo nivel de distorsión armónica (THD), tamaño compacto y facilidad de integración directa con la Raspberry Pi.

PAM8403 Mini Power Amplifier Board

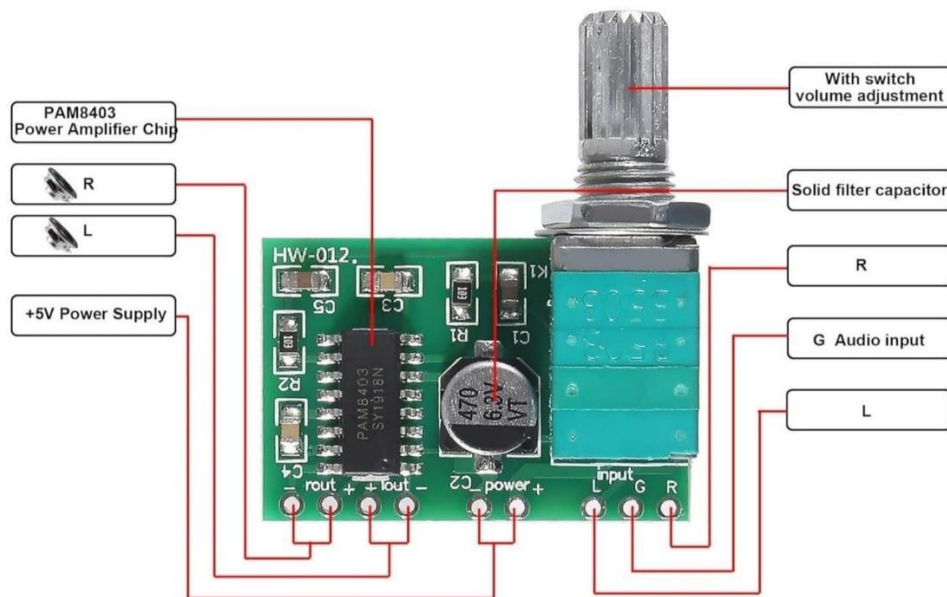


Figura 9: Descripción de los pines y componentes del PAM8403

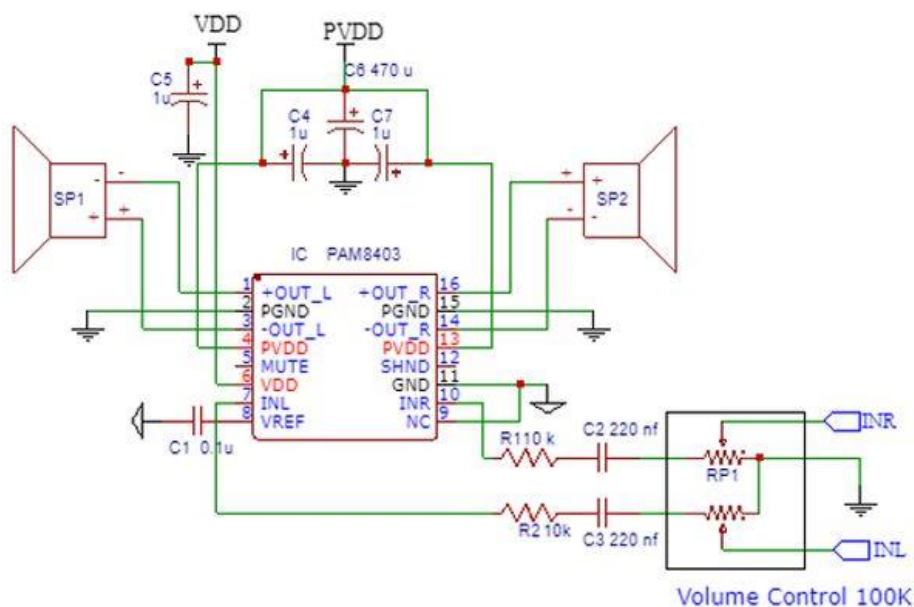


Figura 10: Circuito esquemático electrónico para el preamplificador PAM8403

4.4. Pantalla OLED 0.96 SSD1306

La visualización del texto transcripto (Speech to Text) se realiza mediante la proyección de la pantalla OLED SSD1306 sobre un plástico reflectivo, cuya orientación es directa al ojo del usuario. Su tamaño es de 0.96 pulgadas, con una resolución de 128x64 píxeles y comunicación mediante protocolo I2C. La tecnología OLED permite un alto contraste con bajo consumo energético, ideal para visores compactos. Su pequeño tamaño contribuye a un diseño ergonómico y liviano, manteniendo una buena legibilidad incluso en entornos con alta luminosidad gracias al fondo negro característico de este tipo de pantallas. La facilidad de implementación mediante librerías ampliamente soportadas en Python fue otro factor determinante para su elección. Para versiones futuras se evalúa la incorporación de sistemas ópticos de proyección más avanzados, como microproyectores o pantallas transparentes.

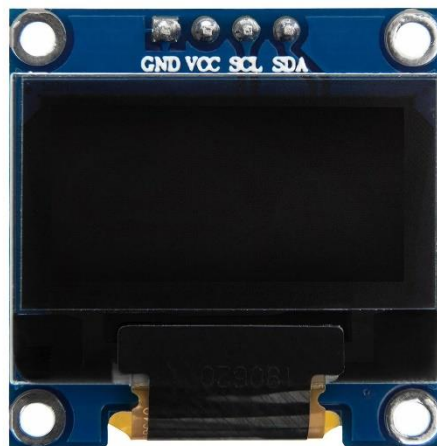


Figura 11: SSD1306 imagen de catálogo

4.5. MICRÓFONO INMP441

Para la captura de audio se incorporó el micrófono digital INMP441, un sensor de tipo MEMS (Micro-Electro-Mechanical Systems) con salida digital mediante protocolo I2S, compatible directamente con la Raspberry Pi sin necesidad de conversores analógicos externos. Este micrófono es omnidireccional, permitiendo captar sonidos provenientes de cualquier dirección. Su bajo consumo, pequeño tamaño y alta sensibilidad lo convierten en una opción adecuada para aplicaciones portátiles como este prototipo. Las principales ventajas son su pequeño tamaño y costo.

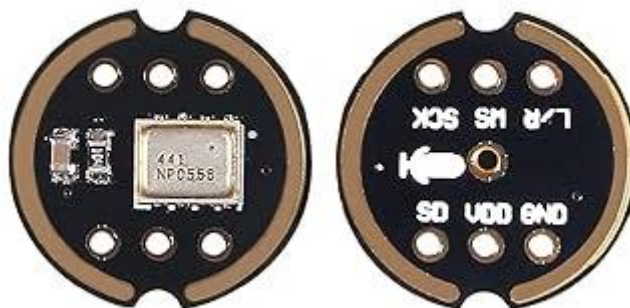


Figura 12: INMP441 imagen de catálogo

4.6. Raspberry Pi Zero 2WH

Como unidad principal de procesamiento se seleccionó la Raspberry Pi Zero 2 WH, una placa compacta que incorpora un procesador quad-core ARM Cortex-A53 a 1 GHz, 512 MB de RAM y conectividad Wi-Fi integrada. Con un precio competitivo se destaca por su excelente relación entre capacidad de procesamiento y consumo energético, soportando sistemas operativos completos basados en Linux. Su principal uso es la comunicación de los periféricos con la pc mediante la conexión USB y el transporte veloz de los datos para una respuesta en tiempo real. Sus interfaces nativas CSI (para cámara), I2C, SPI y GPIO facilitan la integración directa de los diferentes periféricos del sistema.

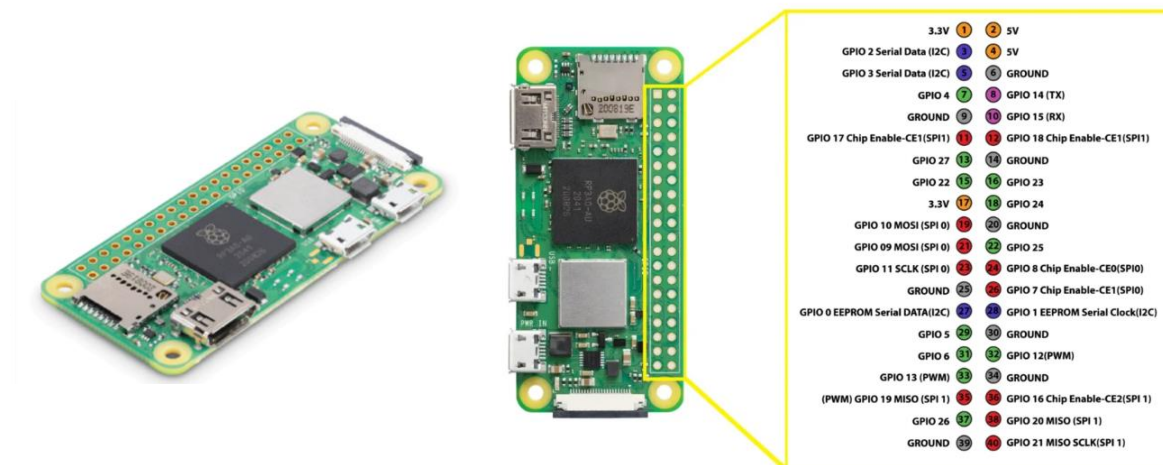


Figura 13: Distribución de pines I/O de la Raspberry Pi Zero 2W

4.7. MOTOR SERVO SG90

Para la movilidad de la cámara se utilizó el **servomotor SG90**, un microservo compacto con un ángulo de rotación de aproximadamente **180°**, controlado mediante señales PWM. Su bajo peso, tamaño reducido y bajo consumo de corriente lo hacen ideal para aplicaciones portátiles. En este proyecto, el SG90 es responsable de proporcionar un sistema básico de *tracking* (seguimiento) permitiendo ajustar la orientación de la cámara según el movimiento del usuario o de la seña, aumentando la efectividad del reconocimiento gestual.

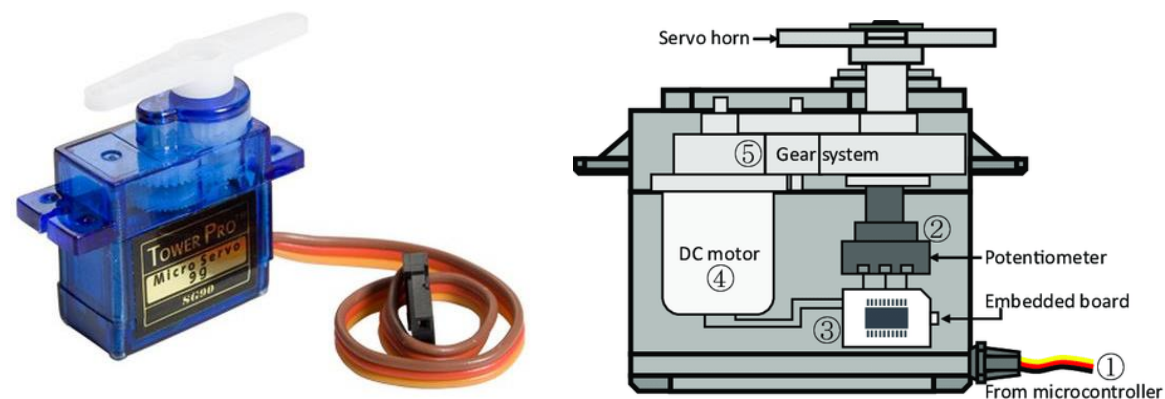


Figura 14: SG90 Imagen de catálogo y composición interna

4.8. Análisis de Consumo de Energía y Conexionado

1. Componentes y Requisitos Eléctricos:

Componentes	Tensión	Corriente Máxima	Observaciones
Servo SG90	5V DC	500 mA	Pico durante movimiento
PAM8403	5V DC	1.5 A	Máximo a volumen alto (3W + 3W)
OLED SSD1306	5V DC	25 mA	Consumo constante
Fuente externa	5V DC	2A (máx)	Alimenta los 4 componentes anteriores
Raspberry Pi Zero 2W	5V USB	1.2 A	Cámara OV5647 (alimentada por Pi)
INMP441	3.3V DC	1.5 mA	Utiliza el 1117-3.3 de la raspberry
Fuente PC	5V USB	USB 2.0 200ma / USB 3.0 1.5A	Alimenta Raspberry + Cámara + micrófono

2. Cálculos de Consumo (Fuente Externa de 5V/2A):

- Corriente Total:

$$I_{total} = I_{servo} + I_{PAM8403} + I_{OLED}$$

$$I_{total} = I_{servo} + I_{PAM8403} + I_{OLED}$$

$$I_{total} = 0.5 A + 0.8 A + 0.025 A = 1.325 A$$

$$I_{total} = 0.5A + 0.8A + 0.025A = 1.325A$$

- Margen de Seguridad:

$$Margen = \frac{2A - 1.325A}{2A} * 100\% = 33.67\%$$

Cálculos para Raspberry Pi (Alimentada por USB de PC):

$$I_{RPi} + \text{Cámara} = 1.2 A$$

(Recomendado: USB 3.0 con capacidad > 1.5A)

3. Diagrama de Conexionado Simplificado:

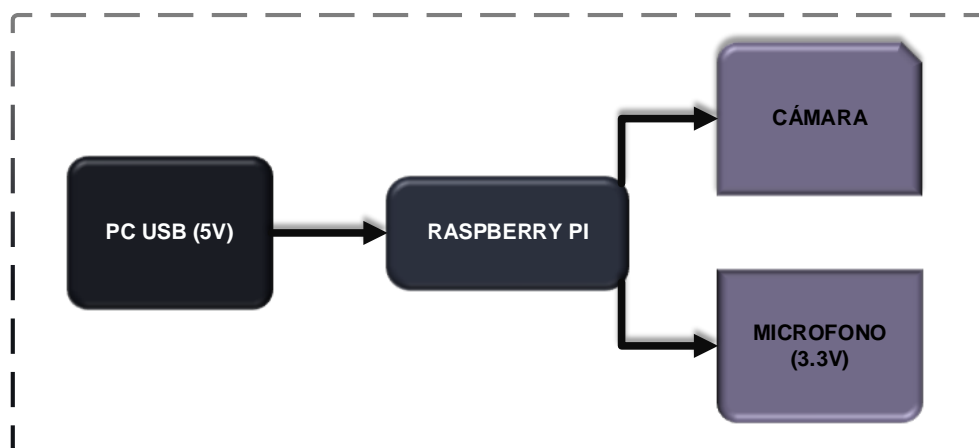


Figura 15: Diagrama de alimentación mediante Raspberry

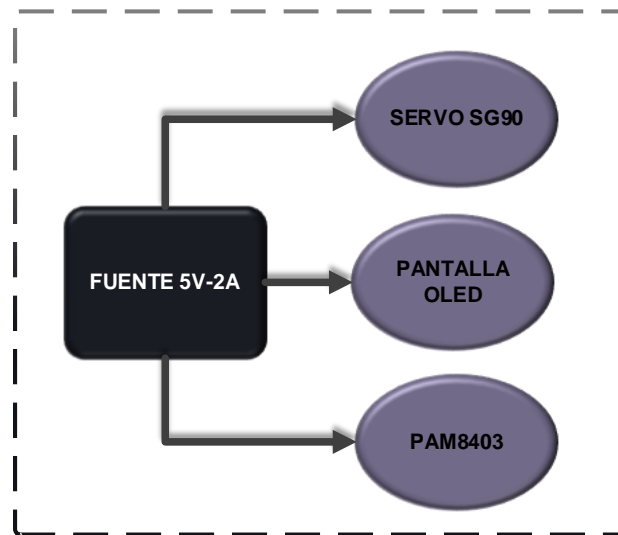


Figura 16: Diagrama de alimentación mediante fuente externa

4. Recomendaciones a tener en cuenta:

- **Fuente Externa 5V/2A:**

- Suficiente para los componentes (consumo máximo: $1.33A < 2A$).

- **Raspberry Pi:**

- Utilizar puerto USB 3.0 (soporte $> 1.5A$).
- Evitar extensiones USB largas para prevenir caída de tensión.

- **Advertencias:**

- El servo y el PAM8403 pueden generar picos de corriente: tener en cuenta de incluir un filtro RC pasivo ósea un capacitor electrolítico ($100-470\mu F$) en la entrada de 5V.
- Separar físicamente cables de audio ($I_{INMP441}$) y potencia para reducir interferencia.

5. SOFTWARE

El sistema **SignGlasses** se soporta en un entorno software distribuido que contempla tres módulos principales, cada uno con una funcionalidad específica.

5.1. Programa de Recolección, Entrenamiento y Administración de Señas

Este módulo está diseñado para la gestión y entrenamiento del modelo de aprendizaje automático, ejecutado en una PC. Su objetivo es facilitar la administración del dataset y la actualización del modelo neuronal. Las funcionalidades principales son:

1. Recolección de Nuevas Señas

- Permite registrar nuevas clases de señas. El operador asigna un nombre identificativo (label) y el sistema inicia la captura.

- Recibe vídeos por UDP desde la Raspberry Pi, procesados en tiempo real mediante MediaPipe, extrayendo 21 landmarks por mano (42 puntos en total).
- Se generan vectores de 126 características (3 coordenadas por landmark), muestreados cada 50 ms, alcanzando un total de 5000 frames (aproximadamente 10 minutos de grabación).
- El dataset se almacena en formato “.pkl”, incluyendo los vectores y sus respectivas etiquetas, facilitando la carga posterior sin necesidad de regrabar.

2. Entrenamiento del Modelo

- Se encarga de la normalización del dataset mediante StandardScaler, codificación de etiquetas con LabelEncoder y entrenamiento del modelo.
- El modelo es un clasificador secuencial denso con capas de 64 y 32 neuronas (activación ReLU), Batch Normalization, Dropout y capa de salida Softmax.
- Optimización mediante el algoritmo Adam, guardado en “.h5” y posterior conversión a formato TensorFlow Lite (.tflite) para reducir la carga computacional en inferencia.
- Genera los gráficos de matriz de confusión del modelo y las métricas de entrenamiento que permiten validar el modelo entrenado.

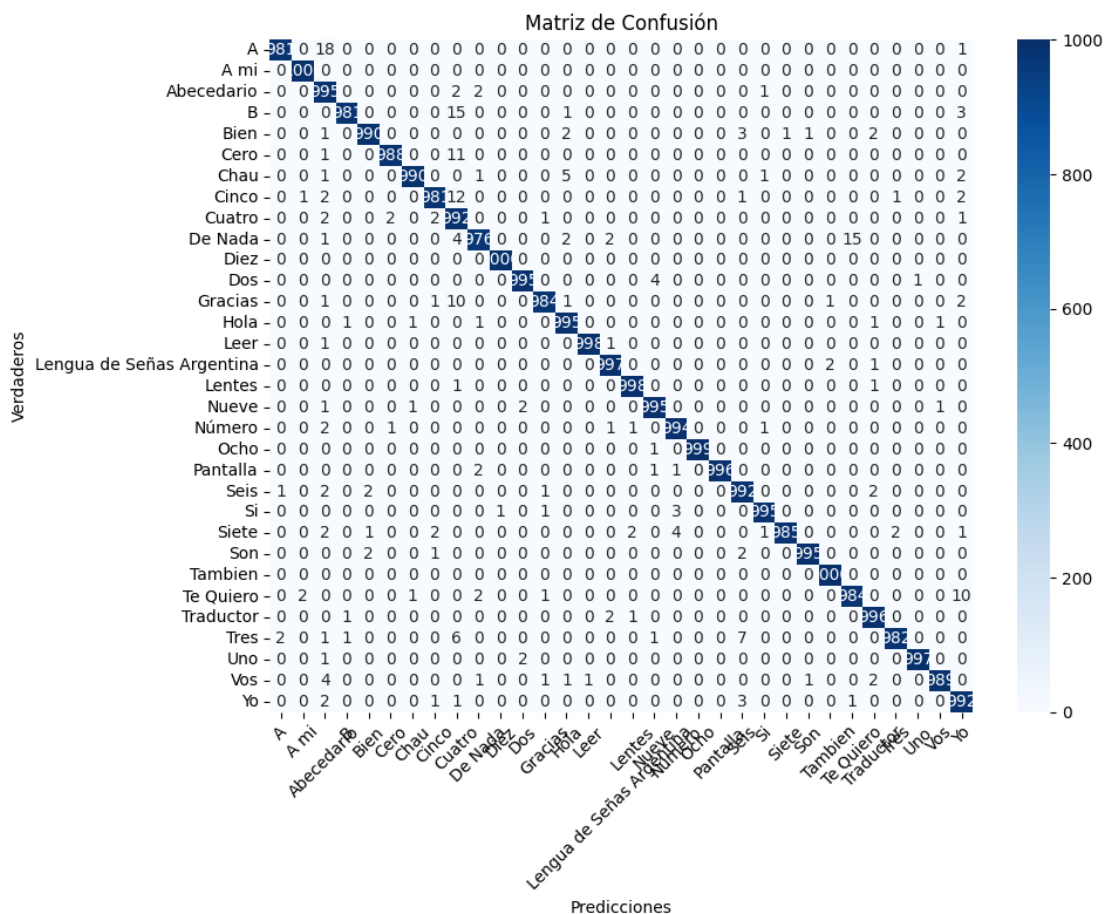


Figura 17: Matriz de Confusion para el modelo actual de las SignGlasses

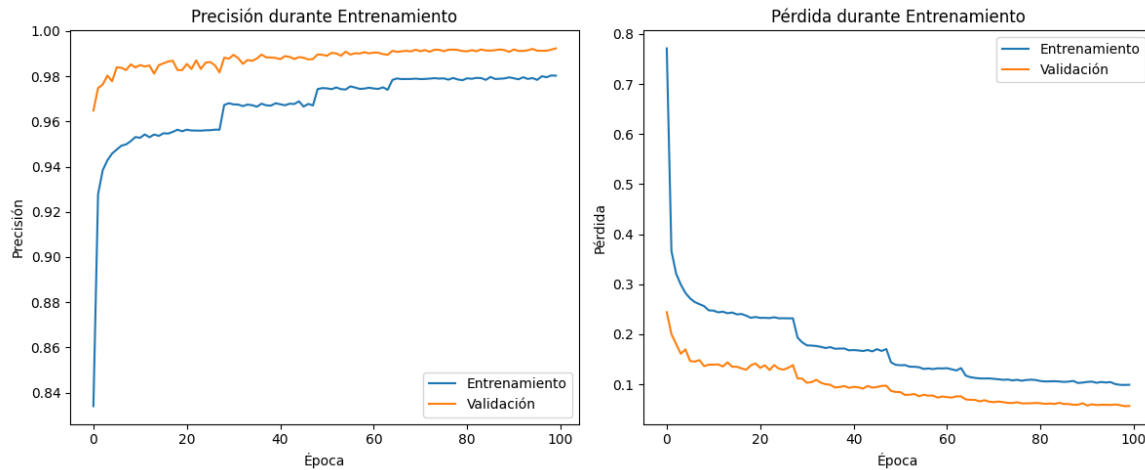


Figura 18: Métricas de Entrenamiento para la versión actual de las SignGlasses

3. Listado de Señas

- Visualiza el inventario actual de señas registradas en el dataset junto a sus etiquetas correspondientes.

4. Eliminación de Señas

- Permite eliminar registros específicos del dataset, ya sea por errores de grabación o por bajo rendimiento durante la inferencia.

5. Generación de Audios

- Mediante pytsx3 y pydub se generan archivos de audio .WAV asociados a cada etiqueta registrada, los cuales serán utilizados para la síntesis de voz en tiempo real.

5.2. Programa de Evaluación en Tiempo Real

Este programa se ejecuta en la PC y corresponde al entorno final de usuario, enfocado en la inferencia y operación en tiempo real, sin permitir la modificación del modelo ni del dataset.

1. Listado de Señas Cargadas

- Muestra la lista de señas habilitadas para inferencia con sus respectivas etiquetas.

2. Evaluación en Tiempo Real

- Recibe flujos de imágenes en fragmentos JPEG por UDP desde la Raspberry Pi, reconstruyendo los frames con OpenCV.
- Procesa los landmarks con MediaPipe, calcula la posición normalizada de la muñeca derecha y envía comandos al servo de seguimiento.

- Normaliza el vector de 126 características, ejecuta inferencia con TensorFlow Lite, aplicando StandardScaler y LabelEncoder previamente guardados.
- Si la predicción supera un umbral de confianza del 90%, muestra en pantalla la etiqueta correspondiente junto con los landmarks y barras de progreso.
- Envía la etiqueta detectada por UDP a la Raspberry Pi para síntesis de audio, además de gestionar señales de control como "HANDS_DETECTED" o "NO_HANDS" mediante heartbeat para asegurar la robustez del sistema.

5.3. Programa Implementado en la Raspberry

El script embebido en la Raspberry Pi cumple la función de **punto de comunicación y procesamiento ligero**, realizando tareas de captura y gestión periférica. Debe trabajar en simultaneo con uno de los dos programas de la PC para cumplir con su función:

- Captura vídeo desde la cámara y lo fragmenta en paquetes UDP de **1460 bytes** para transmisión a la PC.
- Adquiere audio desde el micrófono **INMP441** a **48 kHz**, realiza remuestreo a **16 kHz** y lo transmite cada **5 segundos**.
- Reproduce archivos de audio recibidos vía UDP por el altavoz incorporado.
- Recibe ángulos de movimiento para el **servo SG90** y realiza control en tiempo real del seguimiento.
- Muestra la transcripción de voz recibida en la **pantalla OLED** para el usuario sordo.
- Mantiene un sistema de control de estado mediante **heartbeat bidireccional**, asegurando comunicación activa y detección de caídas de conexión.

6. MODELO DE MACHINE LEARNING Y RED NEURONAL

En este proyecto empleamos aprendizaje automático para traducir gestos de Lengua de Señas Argentina en palabras audibles. El modelo principal es una red neuronal que recibe vectores de características extraídos de la mano, mediante el uso del modelo ya creado por Google conocido como mediapipe, el cual discretiza la mano en puntos y líneas, los cuales denominamos "landmarks". Nuestro modelo produce una probabilidad para cada clase de seña. Esta "red" funciona como un conjunto de capas de procesamiento lineal, donde cada capa transforma su entrada y la pasa a la siguiente, aprendiendo durante el entrenamiento a distinguir patrones en los datos.

6.1. Arquitectura implementada:

Capa de Entrada: La red se inicia con una capa de entrada que recibe un vector de 126 valores (los 21 landmarks de cada mano, con coordenadas x, y, z).

Capas Ocultas: A continuación, una capa densa de 64 neuronas aplica multiplicación por pesos, suma de sesgos y la función de activación ReLU, generando una representación intermedia, detecta patrones simples. Una segunda capa densa de 32 neuronas repite este proceso, filtrando la extracción de características o features y detecta patrones complejos.

Capa de Salida: Finalmente, una capa de salida utiliza la activación Softmax para convertir las salidas lineales en un vector de probabilidades que suman 1, de donde se selecciona la clase de señal con mayor probabilidad. Su función es la de clasificar el texto. Cada neurona realiza cálculos como: $salida = \max(0, entrada * peso + sesgos)$

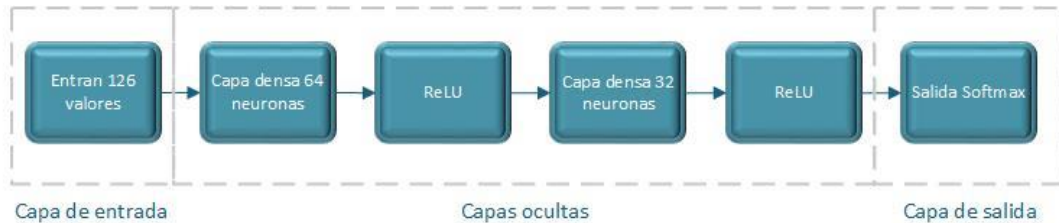


Figura 19: Arquitectura del modelo

6.2. La red neuronal

En conjunto, el sistema transforma una secuencia de imágenes en tiempo real en un vector numérico, estandariza sus valores, los pasa por una red denominada feed-forward entrenada para clasificación multiclase, y devuelve el nombre de la señal que tenga la mayor confianza o que es lo mismo, el mayor valor de probabilidad. Esa etiqueta se utiliza tanto para generar audio como para integrarse en la interfaz de los lentes, cerrando el ciclo y entregando la traducción en tiempo real.

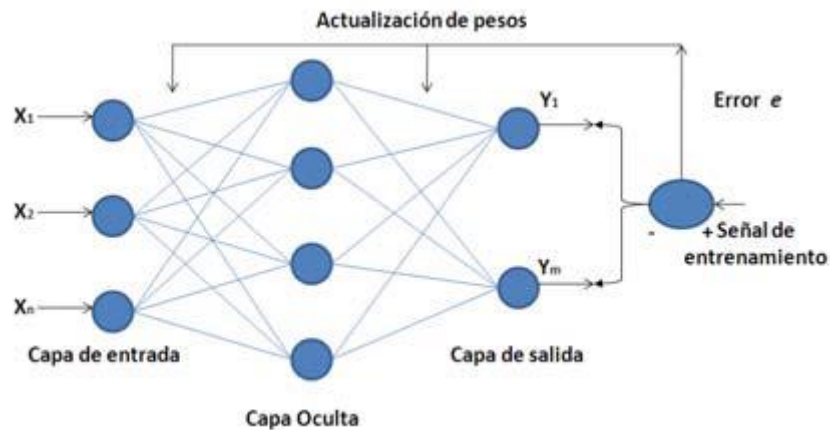


Figura 20: Representación de una red neuronal tipo feed-forward

7. ENSAYOS Y TESTEOS REALIZADOS.

Con el objetivo de validar el correcto funcionamiento del sistema SignGlasses en condiciones reales de uso, se llevaron a cabo una serie de ensayos experimentales. Los ensayos se realizaron en tres entornos diferenciados: dos de ellos en domicilios particulares, y un tercero en la Facultad de Ingeniería de la Universidad Nacional de Lomas de Zamora, con el fin de obtener resultados representativos bajo distintas condiciones ambientales.

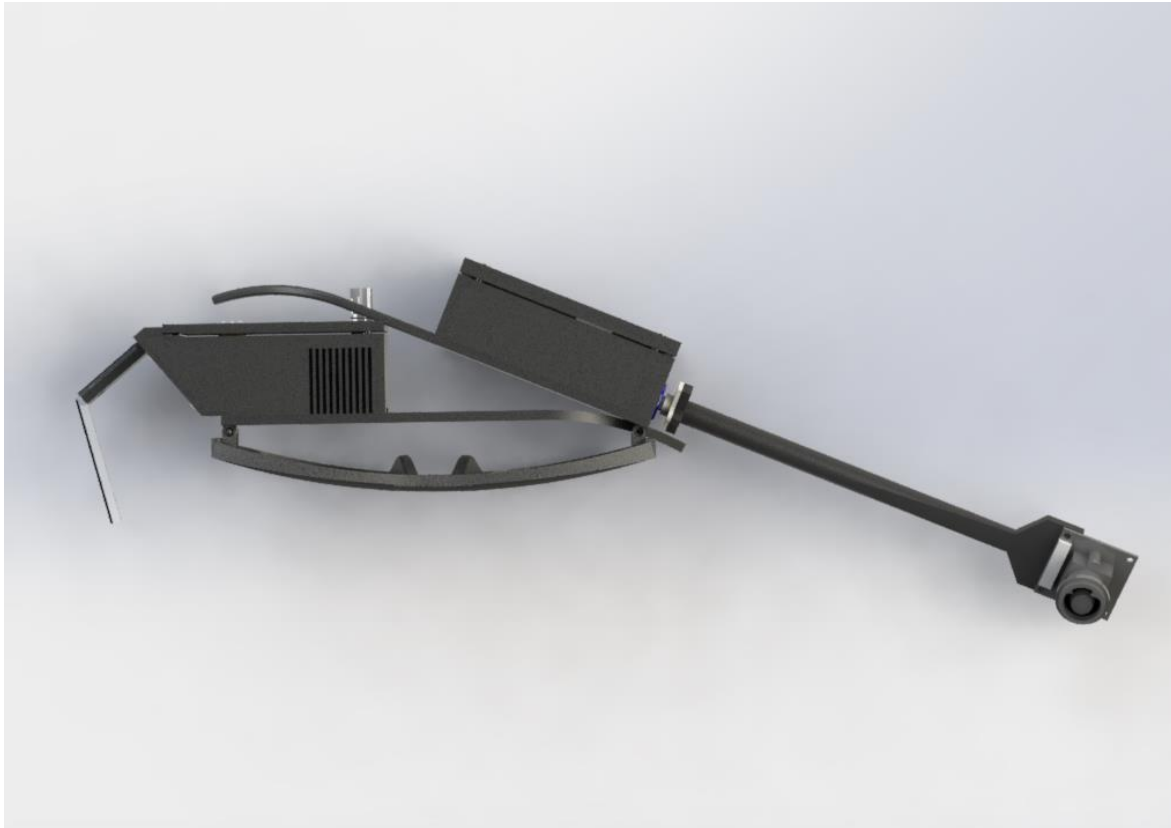


Figura 21: Vista superior y con las patillas cerradas de las SignGlasses Model 3

7.1. Evaluación del Impacto de la Iluminación:

Las pruebas de funcionamiento evidenciaron que las variaciones en las condiciones de iluminación afectan de manera directa la **precisión en la detección de señas** mediante visión artificial. La intensidad y dirección de la luz modifican la calidad de la captación de imágenes, lo que repercute en la eficiencia del reconocimiento en tiempo real.

Acciones implementadas:

- Se calibró el umbral de confianza del modelo de reconocimiento gestual basado en MediaPipe, adaptando la sensibilidad del sistema a distintas condiciones lumínicas.
- Se diseñó e integró un sistema auxiliar de iluminación externa, montado junto a la cámara, con el objetivo de estabilizar la captación de los puntos clave (*landmarks*) en ambientes con iluminación deficiente o irregular. Esta tan solo fue una prueba, no tiene implementación final en las SignGlasses Model 3.

Observaciones: La calidad del reconocimiento mejoró significativamente en entornos con iluminación controlada. Se recomienda considerar tanto la variabilidad de usuarios como los diferentes entornos lumínicos durante futuras calibraciones.

7.2. Calidad de Reproducción de Audio

En cuanto a la reproducción de audio a través del parlante integrado, se observaron mejoras notorias tras la optimización digital de la ganancia. Mediante ajustes por software se logró incrementar el nivel de salida sin alcanzar la saturación del amplificador.

Optimización aplicada:

- Incremento de la ganancia digital en +24 dB durante la etapa de post-procesamiento del audio.

Resultados:

- Se obtuvo un aumento del volumen percibido, mejorando la relación señal/ruido (SNR) sin generar distorsión o saturación prematura.
- La inteligibilidad del audio fue mejorada, generando una experiencia de usuario más clara y efectiva.

Conclusión: La implementación de ganancia digital controlada permitió optimizar la calidad sonora sin comprometer la integridad del sistema de amplificación.

7.3. Visibilidad de la Pantalla

Las pruebas sobre la pantalla OLED reflejaron limitaciones en situaciones de alta luminosidad ambiente. La baja capacidad de brillo generó problemas de legibilidad especialmente a contraluz, afectando la percepción del contenido proyectado.

Optimización implementada: Incorporación de un filtro polarizado sobre la superficie reflectante, lo cual permitió atenuar la incidencia de luz externa y mejorar la legibilidad del texto proyectado.

Observaciones: Si bien se mejoró parcialmente la visibilidad mediante soluciones ópticas pasivas, se recomienda para futuras iteraciones explorar pantallas transparentes o sistemas ópticos más avanzados de proyección.

7.4. Validación de Modelos de Machine Learning

Se realizaron ensayos comparativos con diferentes configuraciones de modelos de aprendizaje profundo para el reconocimiento gestual.

Pruebas realizadas:

- Modelos basados en redes neuronales tipo MLP, CNN, RNN LSTM y Transformer.
- Evaluación de distintas arquitecturas, variando el número de capas ocultas, cantidad de neuronas, tipo de características de entrada e hiperparámetros.
- Comparación del rendimiento en función del tiempo de respuesta y precisión de reconocimiento.

Resultado final: La mejor configuración se describe en el Ítem 4 del presente informe, donde se obtuvo un equilibrio óptimo entre precisión, velocidad de inferencia y consumo de recursos.

8 COSTOS

8.1. Costo real del prototipo

Componente	Precio (\$ARS)	Precio (\$USD)	Proveedor	Fecha de Compra	Fecha de Entrega
Cámara OV4547 ojo de pez 130G 5MP	\$ 10.829,00	\$ 10,18	Starware	15/1/2025	20/1/2025
Cable Cámara Raspberry Pi Zero 15cm	\$ 1.959,00	\$ 1,84	Starware	15/1/2025	20/1/2025
Modulo amplificador audio PAM8403 5V 2ch stereo 3w con volumen	\$ 2.239,00	\$ 2,10	Starware	15/1/2025	20/1/2025
Pantalla display oled SSD1306 0.96p I2C128x64 (blanco)	\$ 5.549,00	\$ 5,21	Starware	15/1/2025	20/1/2025
Raspberry Pi Zero 2WH	\$ 38.957,00	\$ 36,61	Siranet (Mercado libre)	9/1/2025	11/1/2025
Micrófono INMP441	\$ 7.676,00	\$ 7,21	Synergeek (Mercado libre)	26/3/2025	31/3/2025
TOTAL	\$ 67.209,00	\$ 63,16			

Dólar de
Referencia:

1064,15

8.2. Costo estimado del proyecto real

El costo del prototipo funcional es de aproximadamente \$67.209 ARS o 63,16 USD (con un dólar de referencia a 1064,15 ARS/USD).

Este costo contempla solo unidades sueltas de cada componente y sin optimización de producción.

En un proyecto real, deberías considerar:

- **Producción a escala** (baja, media o alta).
- **Costos adicionales de desarrollo:** diseño de PCB, manufactura de carcasa específica, empaquetado de hardware.
- **Costos de software:** licencias, mantenimiento y actualizaciones.
- **Logística:** embalaje, distribución y soporte post-venta.
- **Calidad industrial:** certificaciones, pruebas de fiabilidad.

Una regla habitual para pasar de prototipo a producto es considerar un factor multiplicador que oscila entre 3 y 10 veces el costo de prototipo, dependiendo de la escala y complejidad del producto.

Estimación Final Aproximada:

- Orden de Magnitud (aproximado):
 - Para una producción en pequeña escala o MVP comercializable (por ejemplo, 20 a 100 unidades), se podría estimar un costo unitario cercano a los 250 - 400 USD.
 - En pesos argentinos, con la misma referencia cambiaria, sería aproximadamente 266.000 - 426.000 ARS por unidad.
- Esto contempla materiales, impresión 3D o carcasas moldeadas, PCBs integradas, optimización de cableado y encapsulado, y una primera versión estable de software.
- Si el desarrollo apunta a una producción a mayor escala (>1000 unidades) y se optimiza diseño y producción, el costo podría reducirse por economía de escala, pero se sumarían costos iniciales fijos elevados (moldes, certificaciones).

9. MEJORAS A IMPLEMENTAR

A partir de los ensayos realizados y el análisis funcional del prototipo, se identificaron una serie de posibles mejoras técnicas destinadas a optimizar el rendimiento general del sistema SignGlasses. Estas mejoras abarcan aspectos de hardware, software, procesamiento, ergonomía y escalabilidad futura del dispositivo.

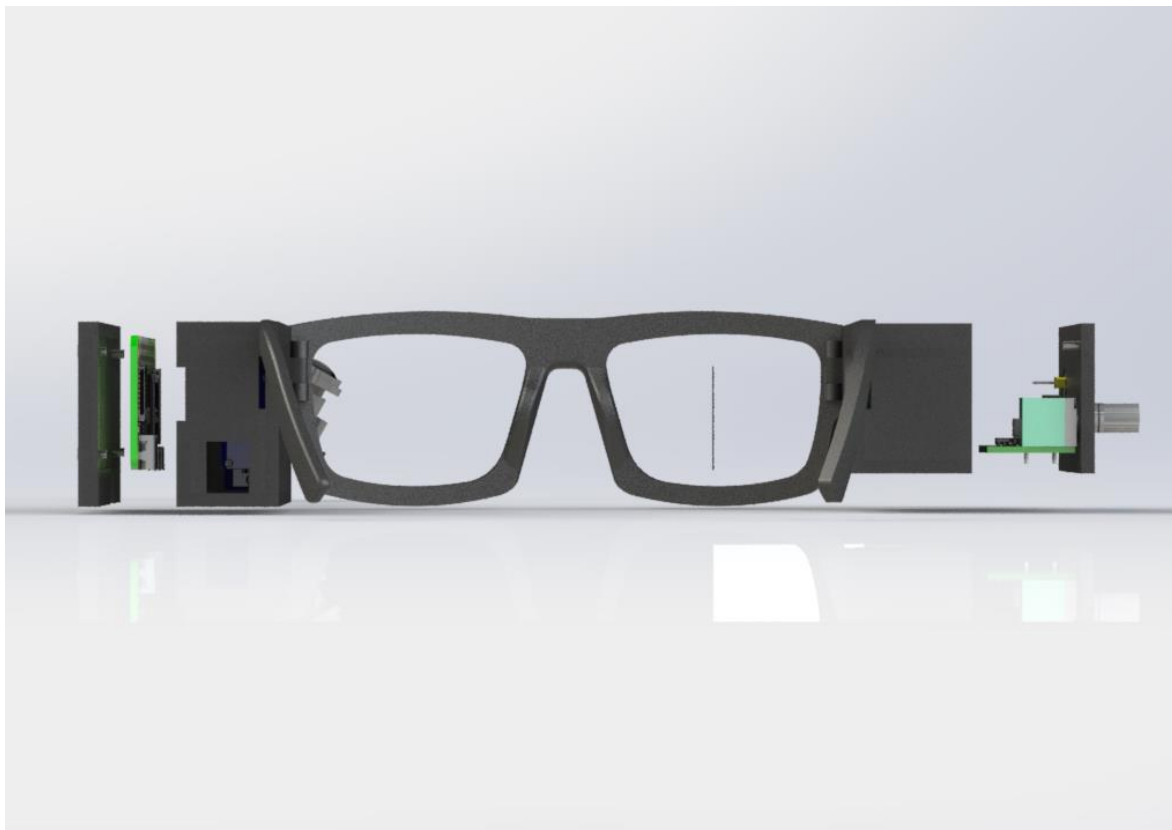


Figura 22: Vista trasera y explosionada de las SignGlasses Model 3

9.1. Pantalla

Propuestas técnicas:

- Incorporación de una pantalla de alto brillo con control dinámico de retroiluminación, para garantizar una correcta visualización en ambientes exteriores y bajo condiciones de contraluz.
- Implementación de un display transparente OLED integrado en la montura, permitiendo la proyección directa sobre el campo visual del usuario sin obstrucción.
- Evaluación de un sistema avanzado de proyección óptica, similar a los utilizados en dispositivos de realidad aumentada o realidad virtual.

Resultado esperado:

Mejorar significativamente la legibilidad de la información proyectada bajo cualquier condición lumínica, además de reducir tamaño y peso del módulo óptico, acercando el prototipo a un diseño más compacto y discreto.

9.2. Micrófono

Propuestas técnicas:

- Sustitución del micrófono omnidireccional por un micrófono direccional, con mayor selectividad para captar únicamente voces humanas.
- Implementación de técnicas de preprocesamiento digital, incluyendo cancelación de eco y supresión adaptativa de ruido mediante software.
- Evaluación de micrófonos de alta sensibilidad y mayor alcance efectivo, con capacidades de diferenciación de fuentes sonoras para futuras funciones avanzadas de identificación de locutor.

Resultado esperado:

Mayor precisión en la transcripción de voz, mejor desempeño en entornos ruidosos y ampliación del rango operativo efectivo hasta 5 metros, sin degradación en la calidad del reconocimiento de audio.

9.3. Sistema de Audio

Propuestas técnicas:

- Reemplazo del actual módulo de audio por parlantes integrados de mayor potencia, posicionados lateralmente en el armazón.
- Incorporación de un sistema táctil de control de volumen o interfaz más compacta, eliminando potenciómetros físicos.
- Visualización en pantalla del nivel de volumen para una interacción más intuitiva.

Resultado esperado:

Mejorar la calidad y claridad del audio emitido, reducir distorsión y proporcionar una experiencia auditiva más personalizable y estética para el usuario.

9.4. Procesamiento.

9.4.1. Ejecución en dispositivo móvil

Propuestas técnicas:

- Portar el modelo de inferencia a TensorFlow Lite para su ejecución en Android/iOS.

- Integrar un sistema de comunicación ligera mediante WiFi, Bluetooth o MQTT con el hardware principal.

Resultado esperado:

Desvincular el procesamiento de computadoras externas, reduciendo la latencia, mejorando la portabilidad y explotando la capacidad de cómputo de dispositivos móviles modernos.

9.4.2. Despliegue en servidor web**Propuestas técnicas:**

- Servidor en la nube AWS(Amazon) o GCP(Google) que reciba imágenes y devuelva la etiqueta de señal.
- Servidores remotos, conexión WAN. El servidor que procese el programa y los modelos ML, está ubicado en un determinado lugar.

Resultado esperado

Reducir la carga de procesamiento local, permitiendo mayor escalabilidad y mantenimiento centralizado del modelo de IA.

9.4.3. Sistema Embebido**Propuestas técnicas:**

- Desarrollo de un sistema totalmente embebido utilizando placas avanzadas con procesadores integrados de mayor rendimiento.

Resultado esperado:

Eliminar completamente la dependencia de dispositivos externos, obteniendo un prototipo completamente autónomo.

9.5. Enriquecimiento de la traducción**9.5.1. Precisión y Latencia****Propuestas técnicas:**

- Mejorar el modelo de la red neuronal, en base a un servidor con mejor poder de procesamiento.
- Aumentar la cantidad de frames por secuencias, como la cantidad de secuencias por señal.
- Obtener diversas grabaciones y recolecciones de diversas fuentes. Como distintas personas, distintas locaciones.

Resultado esperado

Mejorar la robustez y exactitud de la traducción, garantizando detección efectiva sin importar la persona o locación.

9.5.2. Contextualización Lingüística

La traducción palabra a palabra en LSA carece de conectores y fluidez gramatical.

Propuesta técnica:

- Mediante el modelo de lenguaje basado en RNN/LSTM, se puede ir tomando las señas, llenar un array, y que la IA complete mediante los conectores y después que lo reproduzca todo junto. Si la capacidad de procesamiento es la adecuada no se verá reflejado en retrasos entre hacer las señas y emitir el texto. Sería en tiempo real.
- Post-procesamiento en tiempo real: rellenar conectores (“me gustaría que, el auto es rojo, y ajustar el orden sintáctico según reglas del español.

Resultado esperado:

- Generación de oraciones naturales, mejor comprensión y mayor aceptación social del sistema.
- Ejemplo antes/después:
Sin IA contextual: “Hola gustar color cielo”
Con IA contextual: “Hola, me gusta el color del cielo.”

9.6. Reducción de Tamaño y Peso**Propuesta técnica:**

- Optimización del diseño mecánico utilizando **materiales más livianos y resistentes**, como polímeros técnicos o aleaciones livianas.
- Reorganización interna de componentes para una mejor distribución por medio de un PCB con tamaño y forma de lentes completamente funcional.

Resultado esperado:

Conseguir un diseño ergonómico y liviano, visualmente similar a anteojos convencionales.

9.7. Sistema Inalámbrico**Propuestas técnicas:**

- Incorporación de baterías integradas recargables y reemplazo total de conexiones físicas mediante enlaces inalámbricos.

Resultado esperado:

Dispositivo completamente inalámbrico, con mayor autonomía y libertad de movimiento para el usuario.

9.8. Incorporar reconocimiento de expresiones faciales.**Propuestas técnicas:**

- Implementar el módulo MediaPipe Face, para el reconocimiento de expresiones faciales y gestos asociados al rostro.
- Ampliar la lógica de procesamiento para integrar la información facial con la manual.

Resultado esperado:

Expandir el diccionario de señas incluyendo expresiones faciales, fundamentales en la semántica del LSA.

9.9. Diccionario LSA Completo

Propuestas técnicas:

- Implementar el módulo MediaPipe Face, para el reconocimiento de expresiones faciales y gestos asociados al rostro.
- Ampliar la lógica de procesamiento para integrar la información facial con la manual.

Resultado esperado:

Expandir el diccionario de señas incluyendo expresiones faciales, fundamentales en la semántica del LSA.

9.10. Visión evolutiva:

La implementación de estas mejoras permitirá establecer una hoja de ruta de evolución tecnológica, orientada a transformar el actual prototipo en un producto final robusto, autónomo, estético y funcional, capaz de ofrecer asistencia comunicacional en diversos entornos de uso, promoviendo la inclusión y la accesibilidad de las personas sordas o hipoacúsicas en la vida cotidiana.

10. CONCLUSIONES

Analizando el grado de avance alcanzado en el desarrollo del prototipo, se concluye que el sistema SignGlasses ha logrado satisfacer los objetivos técnicos planteados en la etapa inicial. El dispositivo demostró ser capaz de traducir correctamente un conjunto de al menos 20 señas distintas del Lenguaje de Señas Argentino (LSA), alcanzando un nivel de precisión del 93 % en condiciones controladas.



Figura 23: Vista Superior de las SignGlasses Model 3

Sin embargo, a lo largo del presente informe se ha evidenciado que el sistema cuenta con un amplio margen de mejora, especialmente en términos de robustez, escalabilidad y capacidad de adaptación a entornos no controlados. Desde una perspectiva funcional, el prototipo cumple satisfactoriamente con su propósito fundamental: traducir en tiempo real señas a palabras audibles, permitiendo que personas sordas puedan comunicarse de forma más efectiva con interlocutores oyentes. El diseño en formato de lentes representa un acierto ergonómico, ya que libera las manos del usuario, respetando la naturalidad del lenguaje de señas y facilitando su uso prolongado.

No obstante, se identificaron limitaciones relacionadas con la formación gramatical de oraciones, dado que la versión actual realiza traducción palabra a palabra sin conectores sintácticos. Este aspecto constituye una oportunidad de mejora a nivel de software, lo cual representa una ventaja significativa al evitar la necesidad de rediseños en el hardware existente, y facilitar actualizaciones a futuro mediante simples revisiones del modelo de inteligencia artificial.

Desde el punto de vista económico, el prototipo se ha mantenido dentro de un rango de costos accesible, considerando que se trata de una fabricación artesanal y sin optimización por escala de producción. Es esperable que, mediante fabricación en serie o adquisición de componentes en volumen, se logre una reducción significativa del costo unitario. Asimismo, la futura migración parcial o total del procesamiento hacia plataformas móviles o servicios en la nube abriría nuevas oportunidades para abaratar el hardware físico y optimizar la autonomía del dispositivo.

Finalmente, desde un enfoque social y de accesibilidad, el proyecto SignGlasses demuestra un impacto potencial considerable dentro de la comunidad sorda e hipoacúsica, facilitando la integración social sin la necesidad de que el entorno domine el lenguaje de señas. Cabe destacar que este desarrollo no busca reemplazar la enseñanza del lenguaje de señas, sino brindar una herramienta complementaria que reduzca las barreras comunicativas en situaciones cotidianas, ofreciendo mayor independencia a los usuarios.

En síntesis, el prototipo desarrollado representa un primer paso sólido y funcional en la búsqueda de una solución tecnológica accesible y eficaz. Con cada iteración futura se apunta a cerrar progresivamente la brecha comunicacional, acercándose a un producto final con mayor precisión, mejor portabilidad y un diseño más amigable para el uso diario.