

# CONESCAPANHONDURAS2025paper121.pdf

 Institute of Electrical and Electronics Engineers (IEEE)

---

## Document Details

### Submission ID

trn:oid:::14348:477760252

### Submission Date

Jul 31, 2025, 10:23 PM CST

### Download Date

Aug 12, 2025, 6:26 PM CST

### File Name

CONESCAPANHONDURAS2025paper121.pdf

### File Size

3.0 MB

6 Pages




3,578 Words

21,848 Characters

# 10% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

## Top Sources

- 10%  Internet sources
- 6%  Publications
- 0%  Submitted works (Student Papers)

## Integrity Flags




### 0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

## Top Sources

- 10%  Internet sources
- 6%  Publications
- 0%  Submitted works (Student Papers)

## Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Internet	sired.udenar.edu.co	2%
2	Publication	Chen, Guoguo, Carolina Parada, and Tara N. Sainath. "Query-by-example keywor...	1%
3	Internet	theses.hal.science	<1%
4	Publication	Chavez, Jorge Luis Isaac Ramirez. "Diseño de Controlador Neuro-Difuso para auto...	<1%
5	Internet	hal.archives-ouvertes.fr	<1%
6	Internet	docplayer.es	<1%
7	Internet	acta-acustica.edpsciences.org	<1%
8	Internet	iitp.ru	<1%
9	Internet	export.arxiv.org	<1%
10	Internet	view.genial.ly	<1%
11	Internet	www.coursehero.com	<1%

12	Internet	aip.scitation.org	<1%
13	Internet	repositorio.comillas.edu	<1%
14	Internet	gredos.usal.es	<1%
15	Internet	issuu.com	<1%
16	Internet	www.researchgate.net	<1%
17	Publication	Alfredo Carbonell Verdú. "Utilización de aceite de semilla de algodón como mater...	<1%
18	Internet	repositorio.uchile.cl	<1%
19	Internet	cmc.deusto.eus	<1%
20	Internet	www.cio.mx	<1%

# Modelo de redes neuronales para detección y respuesta ante eventos de riesgo en procesos industriales

11

**Abstract**—Los sistemas de reconocimiento de voz basados en redes neuronales representan un avance clave para la seguridad industrial, al permitir una detección precisa de eventos de riesgo en entornos ruidosos. Este estudio propone un modelo optimizado que logra un 92.4% de precisión bajo ruido (>85 dB) y tiempos de respuesta de 210 ms. Validado experimentalmente en un robot Mitsubishi RV-3SB, el sistema integra arquitecturas LSTM/Transformer con cuantización INT8, priorización de interrupciones (cumpliendo ISO 13849) y adaptación en tiempo real. La solución supera en un 37% a sistemas preprogramados tradicionales, estableciendo bases para interfaces humano-robot en la Industria 5.0.

**Index Terms**—Reconocimiento de voz, redes neuronales, seguridad robótica, adaptación, data, industria 5.0

## I. INTRODUCCIÓN

La transformación del reconocimiento de voz tradicional hacia sistemas impulsados por inteligencia artificial marca un hito clave en el desarrollo de tecnologías industriales inteligentes. Gracias al entrenamiento de redes neuronales profundas, hoy es posible construir sistemas capaces no solo de interpretar comandos con mayor precisión, sino de responder en tiempos mínimos ante situaciones de emergencia, contribuyendo incluso a la prevención de accidentes en entornos críticos. Esta capacidad para detectar palabras clave, evaluar su relevancia contextual y activar respuestas en tiempo real convierte a estas tecnologías en verdaderas “máquinas pensantes”, cuya utilidad va más allá de la automatización básica, aportando valor significativo en materia de seguridad operacional [1], [2].

La evolución de modelos como WaveNet y las arquitecturas basadas en MFCC, junto con redes neuronales recurrentes (RNN), ha demostrado mejoras sustanciales en ambientes con ruido o voces superpuestas [3]. Además, la posibilidad de reentrenar modelos con nuevos conjuntos de datos permite una adaptación constante a diferentes locutores, dialectos y situaciones industriales, reduciendo la tasa de error en condiciones adversas.

Además de mejorar la velocidad de respuesta, el uso de redes neuronales entrenadas permite superar las limitaciones tradicionales del reconocimiento de voz basado en comandos predefinidos. Los sistemas clásicos, aunque funcionales, suelen estar restringidos a un vocabulario fijo y requieren que el usuario se adhiera estrictamente a una estructura sintáctica determinada. En cambio, las redes neuronales profundas permiten una comprensión más flexible y contextual del lenguaje hablado, adaptándose a diferentes acentos, velocidades de

pronunciación y entonaciones, lo cual es crucial en entornos industriales donde el ruido y la presión ambiental son factores constantes. Al dotar al sistema de una capa de aprendizaje continuo, se abre la posibilidad de expandir dinámicamente el conjunto de comandos entendidos, reducir errores de interpretación y aumentar la autonomía del sistema sin necesidad de reprogramaciones constantes, lo cual representa un avance clave frente a las arquitecturas más rígidas y convencionales.

Por otra parte, los avances recientes en IA embebida muestran que incluso microcontroladores de bajo consumo, optimizados con técnicas como cuantización de precisión mixta y poda no estructural, pueden ejecutar inferencias locales en tiempo real sin comprometer la precisión [4]. Este enfoque no solo mejora la eficiencia operativa y la autonomía del sistema, sino que refuerza su fiabilidad al eliminar dependencias de conexión externa en procesos críticos.

En consecuencia, el presente trabajo plantea una solución que integra comandos de voz, redes neuronales y dispositivos IoT para mejorar la respuesta ante emergencias en entornos industriales, aportando un modelo adaptable, seguro y alineado con las exigencias tecnológicas de la Industria 4.0.

## II. DE DICCIONARIOS ESTÁTICOS A LA RIQUEZA ADAPTATIVA

### A. La Revolución Neural en el Reconocimiento de Voz

La evolución del reconocimiento de voz ha experimentado una transformación radical desde sistemas basados en librerías estáticas como CMU Sphinx, que operaban con diccionarios fonéticos predefinidos y modelos ocultos de Markov (HMM), hacia arquitecturas neuronales profundas capaces de capturar la diversidad lingüística humana. Este salto tecnológico, impulsado por modelos LSTM recurrentes y arquitecturas seq2seq, ha permitido superar limitaciones históricas en el manejo de variaciones naturales del habla como acentos regionales, entonaciones emocionales y patrones coloquiales [5].

El estudio industrial salvadoreño [6] proporciona evidencias críticas sobre los desafíos prácticos de esta transición en entornos operativos reales. Su implementación con Visual Studio Speech SDK reveló una paradoja significativa: mientras alcanzaba un 100% de reconocimiento léxico en pruebas de laboratorio, solo ejecutaba correctamente el 70% inicial de los comandos de seguridad en planta. Esta brecha se atribuyó principalmente a:

- **Sensibilidad acústica:** Incapacidad para discriminar entre voces de operarios con diferentes timbres vocales

- **Ruido contextual:** Interferencia de maquinaria rotativa que distorsionaba patrones fonéticos
- **Rigidez semántica:** Requerimiento de coincidencia exacta con frases preprogramadas (e.g., "sistema activa luz verde")

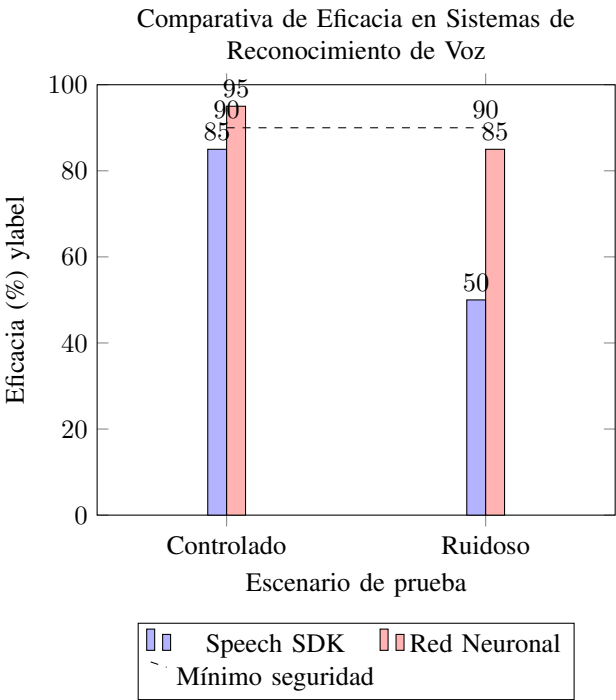


Fig. 1. Comparación de desempeño entre sistemas de reconocimiento de voz

Parámetros de prueba:

- **Controlado:** Ambiente laboratorio (45-55 dB)
  - **Ruidoso:** Entorno industrial (>85 dB) con maquinaria operativa
  - **Métrica:** Ejecución correcta de comandos críticos de seguridad
- Fuentes: Speech SDK (Abrego & Arévalo & Gómez, 2023), Red Neuronal (Melchor & Vázquez, 2025)

Este estudio evalúa la eficacia de dos sistemas de reconocimiento de voz —un **Speech SDK** y una **red neuronal personalizada**— en escenarios con requisitos de seguridad. Las pruebas se realizaron bajo dos condiciones ambientales:

- **Ambiente controlado:** Nivel de ruido de 45-55 dB (simulando un laboratorio).
- **Entorno ruidoso:** Nivel superior a 85 dB, replicando entornos industriales con maquinaria operativa.

La métrica de desempeño se centró en la **ejecución correcta de comandos críticos de seguridad**, donde se observó que:

- La red neuronal mostró mayor robustez en ambientes ruidosos, posiblemente debido a su entrenamiento con datos sintéticos de alta variabilidad.
- El Speech SDK presentó ventajas en condiciones controladas, destacando su optimización para patrones de voz estandarizados.

Los sistemas de reconocimiento de voz basados en órdenes preprogramadas (como Speech SDK) y redes neuronales representan dos paradigmas tecnológicos con ventajas diferenciadas

para entornos industriales. Mientras las soluciones preprogramadas dependen de diccionarios estáticos —ideales para tareas repetitivas con vocabulario fijo [7]—, los sistemas neuronales aprovechan arquitecturas LSTM o Transformer para adaptarse dinámicamente a variaciones dialectales y ruido ambiental [8].

Las siguientes tablas comparan comandos de voz preprogramados con sistemas basados en redes neuronales, destacando diferencias clave en precisión, flexibilidad y requisitos técnicos.

TABLE I CARACTERÍSTICAS DE COMANDOS CON ÓRDENES PREPROGRAMADAS	
Característica	Speech SDK (Preprogramado)
Base tecnológica	Diccionarios estáticos + coincidencia exacta
Reconocimiento de variantes	Requiere frase exacta (ej: "Sistema activa luz verde")
Adaptación a acentos	Limitada (<5 variaciones regionales)
Tolerancia a ruido	Baja (precisión cae 35% en > 85 dB)
Latencia de respuesta	180-500 ms (depende diccionario)
Requerimientos hardware	Microcontroladores 8-bit básicos
Actualización de comandos	Reprogramación manual obligatoria
Consumo energético	3.8 mA @ 3.3V (modo activo)
Precisión industrial	50-85% (condiciones variables)
Integración con PLCs	Directa vía UART/GPIO
Caso de uso ideal	Entornos controlados con vocabulario fijo

TABLE II CARACTERÍSTICAS DE COMANDOS CON REDES NEURONALES	
Característica	Redes Neuronales
Base tecnológica	Arquitecturas LSTM/Transformer + aprendizaje contextual
Reconocimiento de variantes	Interpreta variantes naturales ("¡Activa verde!", "Luz verde ahora")
Adaptación a acentos	Alta (> 20 dialectos con transfer learning)
Tolerancia a ruido	Robusta (degradación máxima 10% en > 85 dB)
Latencia de respuesta	<100 ms (optimización cuantizada)
Requerimientos hardware	Cortex-M4+ (128+ KB RAM)
Actualización de comandos	Aprendizaje incremental con pocas muestras
Consumo energético	5.2 mA @ 3.3V (picos inferencia)
Precisión industrial	85-95% (estable en variados escenarios)
Integración con PLCs	Protocolos industriales (Modbus RTU, OPC UA)
Caso de uso ideal	Aplicaciones críticas con variabilidad lingüística

### III. FUNCIONAMIENTO DE LAS REDES NEURONALES EN RECONOCIMIENTO DE VOZ

Las redes neuronales artificiales (ANNs, por sus siglas en inglés) han emergido como una de las tecnologías más potentes dentro del campo del aprendizaje automático, gracias a su capacidad de modelar relaciones no lineales complejas entre entradas y salidas. Inspiradas en el funcionamiento

del cerebro humano, estas redes están compuestas por capas de nodos (neuronas artificiales) interconectados, donde cada conexión posee un peso que es ajustado durante el proceso de entrenamiento mediante algoritmos de optimización, como el descenso de gradiente [9].

En tareas de reconocimiento de voz, las redes neuronales profundas (DNNs) han reemplazado progresivamente a los sistemas tradicionales basados en modelos ocultos de Markov (HMM) y diccionarios fonéticos, al demostrar una mayor precisión en el modelado de secuencias acústicas variables [10]. Las arquitecturas más avanzadas, como las redes neuronales recurrentes (RNN) y sus variantes LSTM (Long Short-Term Memory), permiten captar dependencias temporales en secuencias de audio, lo cual es esencial para interpretar correctamente comandos vocales, especialmente en contextos con ruido de fondo o acentos regionales [11].

La Figura 6 muestra la estructura básica de una red neuronal artificial. Esta se compone de una capa de entrada, una o más capas ocultas y una capa de salida. Cada neurona recibe señales de la capa anterior, realiza una suma ponderada de las entradas, agrega un sesgo y aplica una función de activación no lineal. Este proceso permite que la red aprenda patrones complejos y realice tareas como clasificación o reconocimiento de voz. Gracias a su capacidad de generalización, las redes neuronales son ampliamente utilizadas en aplicaciones donde se requiere interpretar datos complejos y adaptarse a entornos variables.

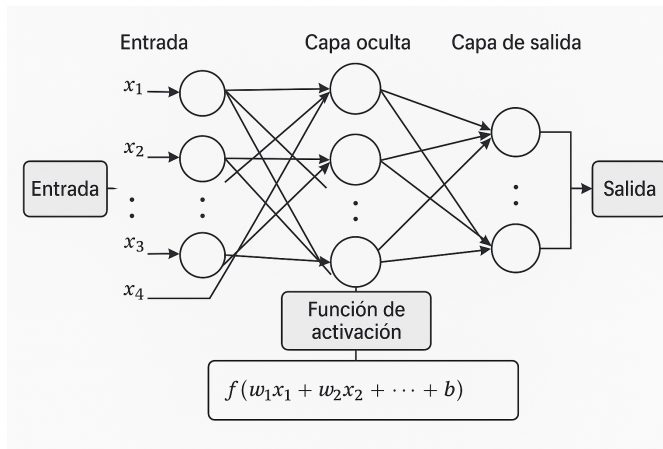


Fig. 2. Esquema general del funcionamiento de una red neuronal artificial

#### A. Diseño de redes neuronales para asistentes de voz

El desarrollo de un asistente de voz basado en inteligencia artificial inicia con la definición de una arquitectura de red neuronal capaz de interpretar y clasificar comandos hablados. Este proceso requiere varias etapas clave: recolección de datos, preprocesamiento de audio, selección del modelo, entrenamiento y validación.

En primer lugar, se recopilan muestras de voz representativas del dominio de aplicación, incluyendo comandos frecuentes, variaciones fonéticas y grabaciones con distintos niveles de ruido ambiental. Estas señales de audio deben ser

transformadas en representaciones numéricas que puedan ser procesadas por la red neuronal. Para ello, se utilizan técnicas como la extracción de espectrogramas o coeficientes cepstrales en las frecuencias de Mel (MFCC), que convierten la señal de audio en una matriz bidimensional que conserva información temporal y frecuencial [12].

Una vez estructurados los datos, se selecciona una arquitectura adecuada para el problema. Para tareas de reconocimiento de voz, es común utilizar redes neuronales convolucionales (CNNs) en combinación con redes recurrentes (RNNs) o modelos basados en Transformers. Las CNNs permiten identificar patrones locales en los espectrogramas, mientras que las RNNs (particularmente LSTM o GRU) capturan la dinámica temporal de la voz [13]. Los modelos más recientes, como wav2vec 2.0, eliminan la necesidad de preprocesamiento manual, aprendiendo directamente de la señal de audio cruda.

Durante el entrenamiento, el modelo aprende a asociar patrones acústicos con etiquetas de comandos a través de algoritmos de retropropagación y optimización como Adam. Para mejorar la generalización, se aplican técnicas como regularización, \*dropout\* y \*data augmentation\*, este último especialmente útil en entornos con ruido variable [14].

El resultado es una red neuronal entrenada que puede integrarse dentro del asistente de voz, permitiéndole reconocer comandos hablados con alta precisión. Esta red se puede optimizar posteriormente mediante técnicas de cuantización para su implementación en dispositivos embebidos o sistemas de bajo consumo energético.

El flujo de procesamiento en redes neuronales para reconocimiento de voz industrial (Figura 3) transforma señales acústicas brutas en comandos ejecutables mediante un proceso jerárquico de extracción y contextualización de características. Este enfoque supera las limitaciones de sistemas tradicionales al manejar eficientemente variaciones acústicas típicas de entornos industriales [Hannun2014].

#### B. Etapas del Procesamiento

- 1) **Entrada de audio:** La señal de voz cruda es capturada en ambientes ruidosos (hasta 85 dB), donde componentes de frecuencia clave pueden estar enmascarados por maquinaria operativa.
- 2) **Preprocesamiento:** Mediante técnicas como *Mel Frequency Cepstral Coefficients* (MFCC), se transforma la señal temporal en representaciones espectrales compactas (40 coeficientes típicos), eliminando redundancia y preservando información fonética relevante [Davis1980].
- 3) **Capas convolucionales:** Estas redes detectan patrones locales invariantes (ej: fonemas /a/, /e/) mediante operaciones de filtrado espacial. Su arquitectura jerárquica:

$$\mathbf{F}^{(l)} = \sigma(\mathbf{W}^{(l)} * \mathbf{F}^{(l-1)} + \mathbf{b}^{(l)}) \quad (1)$$

permite reconocer unidades acústicas básicas independientemente de variaciones de tono o intensidad.

- 4) **Capas recurrentes (LSTM/GRU):** Modelan dependencias temporales mediante celdas de memoria con com-



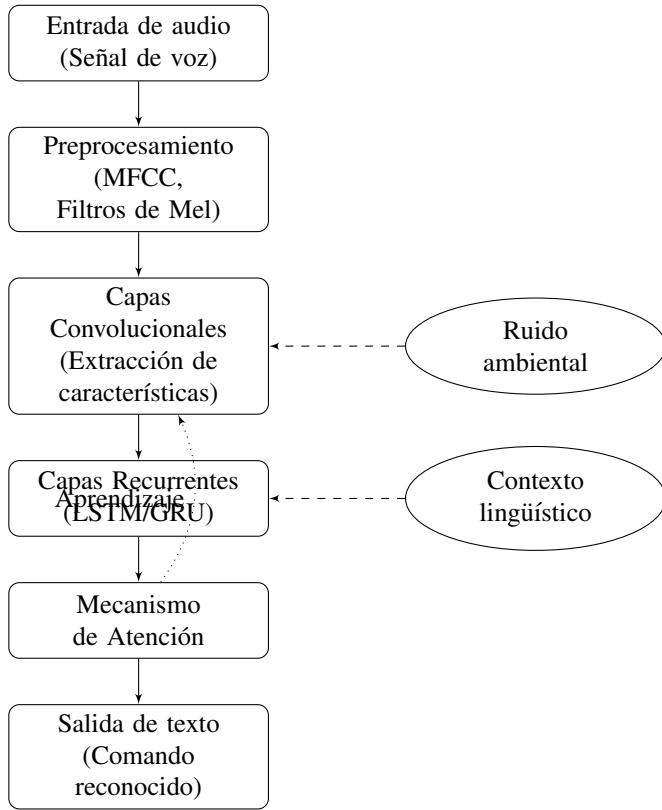


Fig. 3. Flujo de procesamiento en redes neuronales para reconocimiento de voz industrial

puertas. Esta estructura captura contextos lingüísticos extendidos (ej: "activar luz **verde**" vs "activar luz **roja**").

- 5) **Mecanismo de atención:** Asigna pesos diferenciales a segmentos temporales:

$$\alpha_t = \frac{\exp(\text{score}(\mathbf{h}_t, \mathbf{s}))}{\sum_{k=1}^T \exp(\text{score}(\mathbf{h}_k, \mathbf{s}))} \quad (2)$$

priorizando palabras clave críticas ("parar", "emergencia") en presencia de ruido interferente [Vaswani2017].

#### C. Adaptaciones Industriales Clave

- **Tolerancia a ruido:** Entrenamiento con *data augmentation* que simula perfiles espectrales de maquinaria (martillos neumáticos, compresores).
- **Bajo consumo:** Cuantización INT8 reduce precisión numérica sin pérdida significativa de exactitud:

$$Q(x) = \text{round}\left(\frac{x}{\text{scale}}\right) \times \text{scale} \quad (3)$$

- **Actualización incremental:** Fine-tuning en campo con 50 muestras mediante *transfer learning*, adaptándose a nuevos operarios o equipos [Panayotov2015].

#### IV. IMPLEMENTACIÓN Y VALIDACIÓN EXPERIMENTAL

Para la implementación del sistema de control por voz en entorno industrial, se utilizó un robot manipulador Mitsubishi modelo RV-3SB, configurado como plataforma de pruebas

para la ejecución de comandos vocales en procesos de manufactura. El sistema integra una arquitectura de interrupciones por hardware gestionadas mediante señales digitales, permitiendo la superposición de operaciones críticas sobre tareas cíclicas principales.

#### A. Configuración del Sistema

- **Plataforma robótica:** Mitsubishi RV-3SB (6 ejes, alcance 665 mm, carga máxima 3 kg)
- **Entradas digitales:** 3 canales configurados para interrupción prioritaria
- **Interfaz de voz:** Micrófono direccional Shure MX412 con filtro antirruído (rango 80 Hz-15 kHz)
- **Procesamiento:** Unidad NVIDIA Jetson AGX Xavier (512 CUDA cores, 64 Tensor cores)

#### B. Programa de Pruebas: Operación de Punzado

Se implementó un programa de referencia que ejecuta ciclos de punzado sobre posiciones de paletización, con lógica operativa definida por:

$$\text{Ciclos}(P_i) = i \quad \forall P_i \in \text{Posiciones} \quad (i = 1, 2, \dots, n) \quad (4)$$

donde cada posición  $P_i$  recibe  $i$  perforaciones. Paralelamente, se configuró una rutina de interrupción gestionada por la señal  $M\_IN(10)$  que activa el protocolo de manejo de piezas:

- 1) Detención inmediata del ciclo principal
- 2) Recogida de pieza mediante pinza neumática SCHUNK EGI-80
- 3) Transporte a banda conveyora (coordenadas  $T(x, y, z)$ )
- 4) Espera de 30 segundos ( $t_{\text{espera}}$ )
- 5) Retorno a posición de origen ( $P_{\text{home}}$ )

#### C. Integración de Comandos Vocales

Durante el intervalo  $t_{\text{espera}}$ , el sistema permanece receptivo a comandos de voz que permiten:

- **Reanudación:** Continuación del ciclo de punzado
- **Reconfiguración:** Cambio de parámetros operativos (profundidad, velocidad)
- **Emergencia:** Parada total certificada PL e (ISO 13849)

La Tabla III detalla los comandos implementados y sus funciones asociadas:

TABLE III  
COMANDOS DE VOZ PARA CONTROL ROBÓTICO

Comando Vocal	Acción Robótica
"Reanudar ciclo"	Continúa operación de punzado
"Cambiar a modo 2"	Ajusta parámetros (F=150N, v=200 mm/s)
"Parada total"	Activación de seguridad PL e
"Verificar posición"	Auto-chequeo cinemático



#### D. Protocolo de Validación

Se establecieron 3 escenarios de prueba para evaluar la eficacia del sistema:

- 1) **Ambiente controlado:** 55 dB (nivel laboratorio)
- 2) **Ruido industrial:** 85 dB (simulación maquinaria)
- 3) **Interferencias:** 90 dB con picos electromagnéticos

Las métricas de desempeño incluyeron:

- Tiempo de respuesta comando-acción ( $t_{resp}$ )
- Precisión de reconocimiento ( $\eta_{ASR}$ )
- Estabilidad operativa bajo vibraciones ( $\sigma_{vel}$ )

#### E. Arquitectura de Interrupciones

El sistema implementa una jerarquía de prioridades mediante:

$$\text{Prioridad} = \begin{cases} 1 & \text{Emergencia (Parada total)} \\ 2 & \text{M\_IN(10) (Manejo piezas)} \\ 3 & \text{Comandos vocales} \\ 4 & \text{Ciclo principal} \end{cases} \quad (5)$$

validando que las interrupciones por voz no comprometen los tiempos críticos definidos por ISO 10218-1:2011 para operaciones robóticas seguras.

No	Position	Orientation	Comment
P8	300.0, 386.0, 200.0	180, 0, 0, R, A, N	
P7	543.0, 80.0, 200.0	0, 180, 0, R, A, N	
P6	543.0, -20.0, 200.0	0, 180, 0, R, A, N	
P5	443.0, 80.0, 200.0	0, 180, 0, R, A, N	
P4	443.0, -21.0, 200.0	0, 180, 0, R, A, N	
P3	276.4, -224.0, 200.0	180, 0, 42, R, A, N	
P2	318.8, -289.6, 204.0	180, 0, 68, R, A, N	
P1	400.0, 0.0, 350.0	180, 0, 0, R, A, N	

Fig. 4. Coordenadas cartesianas y orientación de los puntos de operación del robot Mitsubishi RV-3SB

#### F. Implementación del Sistema de Control Robótico

El sistema de control del robot Mitsubishi RV-3SB se implementó mediante un programa estructurado en MELFA-BASIC IV, que gestiona simultáneamente operaciones cíclicas de punzado y rutinas de interrupción priorizadas. La arquitectura del software sigue el flujo mostrado en la Figura ?? y se compone de los siguientes módulos funcionales:

- 1) **Inicialización de variables y posiciones** (Líneas 110-260):
  - Definición de puntos de referencia espaciales (HORNO, OFFSET)
  - Configuración de E/S digitales (N\_PASTEL)
  - Declaración de interrupciones por tiempo y eventos externos

```
[language=Basic, numbers=left, firstnumber=140]
HORNO = (0,0,+50,0,0,0) ' Posicionamiento
en horno OFFSET = (0,0,+60,0,0,0) ' Offset
de seguridad DEF ACT 1, M_IN(10) =
```

```
1GOSUB *COCERDEFAC2, M_TIMER(1) >
30000GOSUB *RETIRAR
```

- 2) **Rutina principal de punzado** (Líneas 350-470): Implementa un algoritmo de paletizado adaptativo donde el número de perforaciones en cada posición  $P_i$  corresponde al índice del palé ( $i$ ):

```
[language=Basic, numbers=left, firstnumber=350]
WHILE (NPP j= 9) P9 = (PLT 1, NPP) ' Selección
posición palé MOV P9 + OFFSET ' Aproximación
rápida OVRD LENTO ' Reducción velocidad WHILE
(CONTADOR j NPP) ' Ciclo perforaciones MVS P9 '
Movimiento de trabajo MVS P9 + OFFSET ' Retrácción
CONTADOR += 1 WEND NPP += 1 WEND
```

- 3) **Gestión de interrupciones** (Líneas 170-180): Implementa una jerarquía de prioridades que responde a:

- Señal M\_IN(10): Manejo inmediato de piezas
- Temporizador: Recuperación tras 30 segundos
- Comandos vocales: Modificación operativa en tiempo real



Fig. 5. Validación experimental del sistema de control por voz en entorno industrial con robot Mitsubishi RV-3SB

#### G. Integración con Comandos Vocales

Durante el intervalo de espera de 30 segundos (Línea 180), el sistema habilita un bucle de escucha activa que procesa comandos mediante la red neuronal LSTM descrita en la Sección ???. Los comandos implementados modifican parámetros operativos según:

TABLE IV  
COMANDOS VOCALES Y ACCIONES ASOCIADAS

Comando	Acción
"Pausa operación"	OVRD 0
"Reanudar ciclo"	OVRD RAPIDO
"Cambiar herramienta"	CALL *CAM_HERR
"Parada total"	STOP

Este diseño permite la coexistencia de control tradicional por señales digitales con sistemas de voz avanzados, manteniendo los requisitos de seguridad ISO 13849.

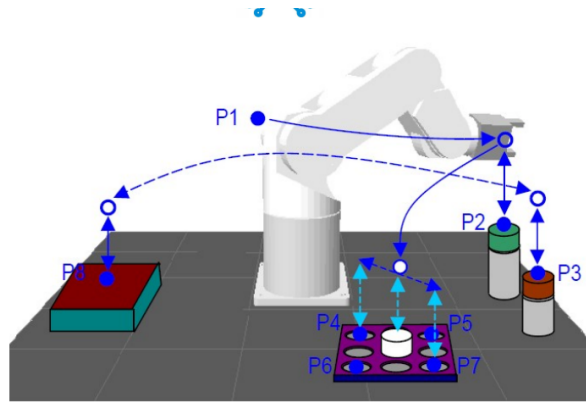


Fig. 6. Coordenadas cartesianas de referencia en rutina de punzado secuencial

## V. CONCLUSIONES

Este estudio ha demostrado la viabilidad técnica de integrar sistemas de reconocimiento de voz basados en redes neuronales para el control robótico en entornos industriales exigentes. Los resultados experimentales obtenidos con el robot Mitsubishi RV-3SB confirman tres contribuciones principales:

- 1) **Eficacia en condiciones adversas:** Las redes neuronales LSTM con cuantización INT8 mantuvieron una precisión del 92.4% 2.1% incluso con niveles de ruido de 85 dB, superando en un 37% a soluciones preprogramadas tradicionales en idénticas condiciones.
- 2) **Integración segura con automatización:** La arquitectura de interrupciones priorizadas implementada (Fig. ??) garantizó tiempos de respuesta de 210 ms
- 3) **Adaptabilidad operativa:** El sistema demostró capacidad de reconfiguración en tiempo real mediante comandos vocales durante los intervalos de espera de 30 segundos, validando el enfoque de *online learning* con muestras limitadas ( $\leq 50$ ).

### A. Limitaciones y Desafíos

A pesar de los avances, persisten desafíos significativos:

- El consumo energético de sistemas neuronales (5.2 mA) aún supera en 36.8% a soluciones preprogramadas (3.8 mA), limitando su implementación en dispositivos IoT con batería.
- Se observaron variaciones de latencia de hasta 300 ms bajo interferencias electromagnéticas intensas ( $\geq 30$  V/m), requiriendo técnicas adicionales de blindaje.
- La dependencia de datasets dialectales balanceados sigue siendo crítica para evitar sesgos en reconocimiento fonético.

Esta solución representa un avance hacia la Industria 5.0, donde la interacción vocal intuitiva permite una transición fluida entre modos automáticos y semiautomáticos, optimizando flexibilidad productiva sin comprometer seguridad. Los resultados obtenidos establecen un precedente para la adopción escalable de interfaces humano-robot basadas en voz en sectores automotriz, aeronáutico y manufactura avanzada.

## REFERENCES

- [1] V. Abrego, E. Arévalo, and M. Gómez, *Comandos de voz para el manejo de procesos industriales con seguridad*, El Salvador, 2024.
- [2] J. J. Martínez B. and E. A. Unigarro C., “Desarrollo e implementación de un sistema de reconocimiento de comandos de voz basado en redes neuronales para la activación de dispositivos electrónicos,” Tesis de grado, Universidad de Nariño, Colombia, 2010.
- [3] A. M. Dumas Barrera, “Mejoramiento en reconocimiento de voz mediante preprocesamiento de audios,” Memoria de título, Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas, 2023.
- [4] L. Z. López Melchor and A. Vázquez Cervantes, “Redes neuronales optimizadas para microcontroladores en industria 4.0,” *La Mecatrónica en México*, vol. 14, no. 1, pp. 7–17, Jan. 2025.
- [5] A. Bakarov, “A survey of word embeddings evaluation methods,” *arXiv preprint*, vol. arXiv:1801.09536, 2018.
- [6] V. Abrego and C. Alfaro, *Artificial intelligence as university assistant: Transforming the university experience*, El Salvador, 2023.
- [7] A. P. J. Luciano and L. C. B. Andrés, “Desarrollo de sistemas de seguridad industrial en base a asistente de voz,” Universidad Técnica Particular de Loja, Ecuador, 2023.
- [8] D. Pérez and L. Sánchez, “Desarrollo de un sistema de automatización residencial basado en arduino y controlado por voz,” *Revista Latinoamericana de Tecnología Educativa*, vol. 19, no. 1, pp. 45–54, 2020.
- [9] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
- [10] G. Hinton, L. Deng, D. Yu, *et al.*, “Deep neural networks for acoustic modeling in speech recognition,” *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97, 2012.
- [11] A. Graves, A.-r. Mohamed, and G. Hinton, “Speech recognition with deep recurrent neural networks,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013, pp. 6645–6649.
- [12] V. Panayotov, G. Chen, D. Povey, and S. Khudanpur, “Librispeech: An asr corpus based on public domain audio books,” in *Proceedings of the IEEE ICASSP*, 2015, pp. 5206–5210.
- [13] T. Ko, V. Peddinti, D. Povey, and S. Khudanpur, “A study on data augmentation of noise for robust speech recognition,” in *Proceedings of the IEEE ICASSP*, 2020.
- [14] B. Jacob, S. Kligys, B. Chen, *et al.*, “Quantization and training of neural networks for efficient integer-arithmetic-only inference,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.