

Sistema embebido como herramienta de apoyo a la comunicación para pacientes con afecciones del habla

Trabajo Terminal No. ---- - ----

*Alumnos: *Valle Martínez Luis Eduardo*

Directores: Rodolfo Romero Herrera, Jesús Yaljá Montiel Pérez

**e-mail: lvallem1400@alumno.ipn.mx*

Resumen - Las afectaciones del habla en adultos mayores suelen originarse después de una lesión o accidente comunmente con daño cerebral en el lóbulo izquierdo, limitando o completamente impidiendo la correcta comunicación verbal de las personas. Este proyecto propone un prototipo de *wearable* y sistema embebido como herramienta de apoyo para pacientes con afectación del habla mediante la traducción de un código motriz de una mano a texto, usando un modelo de IA y algoritmo DTW para su clasificación, realizando finalmente la reproducción de palabras y frases mediante servicios *Text-to-Speech*.

Palabras clave - Afectaciones del habla, Modelo de Inteligencia Artificial, Sistemas Embebidos, *Text-to-Speech*

1. Introducción

Apectos como los gestos, muecas, señas, etc. son factores influyentes en la comunicación cuando son interpretados, sin embargo el recurso predilecto para la comunicación humana no es otro que el elemento simbólico fundamental que llamamos lenguaje. El lenguaje humano es prueba del desarrollo intelectual de nuestra especie, al punto que logró convertirse en el contexto de la historia de la evolución humana, en la piedra angular de la que nuestra especie se ha valido como herramienta para generar conocimiento y perpetuarlo, comunicarnos entre nosotros, e incluso expresar nuestros sentimientos y pensamientos

A nivel oral, expresamos el lenguaje mediante el habla, alcanzada en algún punto de la evolución del sistema canal vocal-auditivo en el humano por el descenso de la laringe, otorgándonos la posibilidad de crear sonidos[1]. A través de la especialización como especie en la actividad del habla, se fueron otorgando semántica a los sonidos generados, permitiendo a la vez la asociación de un significado a estos[1].

Tratándose de la herramienta cotidiana que permite la comunicación e interacción social con los demás, es natural la relevancia que se le otorga al lenguaje en la vida del humano, resaltando específicamente su derivación oral como tema de interés en el estudio de este trabajo. La vivencia ordinaria diaria se dislumbra compleja cuando se restringe el uso del lenguaje verbal, y sin embargo la realidad presenta la existencia de sectores de la sociedad que a raíz de trastornos y afecciones, afrontan dificultades que parcial o completamente, les impiden expresarse oralmente.

A continuación se mencionan las afectaciones más habituales que tienen la posibilidad de beneficiarse de la implementación de la propuesta en este trabajo. Las primeras son las **afasias**[2-4], las cuales son problemas médicos originados por una lesión cerebral y que resulta en la pérdida o alteración del lenguaje. Las **apraxias**[2,5,6] que son trastornos neurológicos caracterizado por la pérdida de la capacidad de llevar a cabo movimientos diestros y gestos, aún cuando se tenga el deseo y la habilidad física para hacerlo, teniendo diferentes afectaciones en función de la parte lesionada en el cerebro. Finalmente existe la **disartria**[2,7,8] que es un trastorno de la ejecución motora del habla debido a un problema neurológico por la presencia de un accidente u lesiones cerebrales. Afecta gravemente la motricidad de los músculos para el habla. Importante mencionar que en algunos casos los profesionales médicos recomiendan el uso de algún dispositivo electrónico o tecnológico de apoyo para la comunicación para las disartrias.[8]

Otro tipo de trastorno que imposibilita de forma indirecta la capacidad para la expresión hablada, aunque como consecuencia de la afección principal, es la sordera de percepción total y es clasificada como un trastorno de la

audición. Este proyecto no considera a este sector de la población como el público objetivo principal por razones como: El amplio desarrollo del Lenguaje de Señas como principal recurso de comunicación, y las limitantes propias del trabajo donde no se provee una solución para una comunicación bidireccional con usuarios con este tipo de afección.

En este sentido y debido a la aparente utilidad e innovación en el uso de guantes de traducción para la comunicación entre personas con afectaciones en el habla y audición con la sociedad en general, se han desarrollado cantidad de trabajos a escala internacional, nacional e incluso interinstitucional basándose principalmente en el lenguaje de señas respectivo del país donde se investigó.

Anteriormente se han desarrollado trabajos de titulación enfocados en sistemas de apoyo para la comunicación de personas con afecciones en el habla y que se encargan de realizar una traducción del LSM(Lenguaje de Señas Mexicano) al español mediante la reproducción sonora de las letras o palabras. La solución propuesta en el trabajo *Guante traductor de señas para sordomudos*[9], se basa precisamente en un prototipo de guante equipado con sensores de flexión en los dedos y que son procesados y controlados mediante un μC que enlazado a un sintetizador de voz y una pequeña pantalla LCD se reproduce el mensaje identificado; Este guante se encuentra limitado a 26 letras del abecedario y algunas abreviaturas.

Similar al trabajo anterior, pero empleando técnicas de Visión Artificial o Visión por Computadora, el trabajo "*Sistema de comunicación auditiva para personas con problemas del habla*"[10], utiliza la tecnología infrarroja del dispositivo Kinect desarrollado por Microsoft para la obtención de imágenes con las que al aplicarse un algoritmo de clasificación de características en las familias de *Random Forest* y *Mean Shift*, son implementadas en complemento con un modelo de Redes Neuronales para obtener los resultados de las señales realizadas por el usuario.

En el ámbito internacional se han desarrollado trabajos enfocados en miniaturizar y disminuir el hardware necesario en la creación de un *wearable* sensor del movimiento de las manos y los dedos en forma de una pulsera[11]. En este trabajo desarrollado por un equipo de la Universidad de Amrita en la India, utilizan en conjunto un sensor IMU(*Inertial Measurement Unit*) integra un par de sensores: acelerómetro y giroscopio, junto con un arreglo de electrodos EMG(*Electromyography*) que permiten reconocer la contracción de los músculos para el movimiento respectivo de cada dedo.

En otro trabajo realizado por un equipo de 3 investigadores en el *National institute of Technology Puducherry Karaikal India*[12], se propone un prototipo de guante que implementa sensores de flexión, acelerómetro y giroscopio compilando sus mediciones con un Arduino Nano y que permite el envío de la recopilación vía Bluetooth a una PC que corre un algoritmo de ML para la clasificación de los gestos, SVM(*Support Vector Machine*). Este prototipo además permite la identificación de gestos correspondientes al Lenguaje de Señas Americano(ASL) y el Lenguaje de Señas Indio(ISL).

Las mayoría de las propuestas de solución en los trabajos consultados tienen por tendencia el uso extendido de prototipos guantes, o incluso *wearables*, con los que se realiza la detección para la interpretación, sin embargo esta clase de solución cuenta con inconvenientes principalmente físicos, como la cantidad de hardware requerida, la portabilidad del prototipo y la falta de estética para un uso cotidiano generalizado. En otra rama tecnológica pero con soluciones existentes en cantidades casi iguales, se tienen los proyectos enfocados en soluciones basadas en Visión Artificial, las cuales suelen ser mucho más cómodas en términos de portabilidad para el usuario, pero en contraste requieren de un ambiente de iluminación y contraste controlados para adquirir resultados aceptables con las técnicas de análisis de imágenes.

Tomando en cuenta estas consideraciones, la solución tecnológica propuesta en este trabajo se decanta por un prototipo físico *wearable* tipo guante, con el que se busca aminorar el impacto de portabilidad y estética al utilizarse un único dispositivo sensor, el SoC micro:bit, que integra un IMU del que principalmente se utiliza para el muestreo de los patrones de movimiento el acelerómetro. Estos patrones manuales serán elementos pertenecientes a un código motriz especialmente propuesto para este sistema, y con los que se busca clasificar cada patrón en una de las clases textuales con las que se logra conformar el texto, a través de la implementación conjunta del

algoritmo DTW y un modelo de Inteligencia Artificial. Una última etapa, ejecutada después de la conformación de las palabras o frases, se logrará mediante la consulta de un servicio de *Text-to-Speech* para la reproducción sonora del texto armado. El proyecto *TTS de Mozilla*, es una solución que se ejecuta como servicio en red local, aunque también se considera el uso de servicios especializados en la nube, como por ejemplo *Microsoft Azure Text-to-speech*.

2. Objetivo

Objetivo general:

Crear un sistema embebido como herramienta de apoyo a la comunicación de personas con afecciones del habla a través de un prototipo *wearable* sensor del movimiento en una mano, un modelo de Inteligencia Artificial en conjunto con el algoritmo DTW para la clasificación de los patrones, y un servicio de *Text-to-Speech* para la reproducción sonora de palabras y frases.

Objetivos específicos:

1. Proponer un conjunto de patrones de movimiento como código motriz que permita conformar al usuario, palabras y frases en texto.
2. Ensamblar el prototipo de guante *wearable* sensor implementando el SoC micro:bit para el muestreo del cambio en las aceleraciones y un sensor de pulso para indicar el inicio y fin del muestreo de un movimiento.
3. Implementar de bibliotecas especializadas y en conjunto, el algoritmo DTW y un modelo de Inteligencia Artificial, para la clasificación eficiente de los patrones de movimiento en las respectivas clases textuales.
4. Configurar el servicio local y en la nube de *Text-to-Speech*, para la reproducción de las palabras y frases.

3. Justificación

Las afectaciones en el habla en personas que han superado la etapa de niñez(adolescentes, adultos jóvenes, adultos y adultos mayores) suelen originarse principalmente por accidentes o lesiones que dañaron alguna zona de la masa encefálica, más comunmente afectando las funciones del lenguaje cuando se localiza en el hemisferio izquierdo del cerebro. Estos daños pueden darse por la falta de circulación del torrente sanguíneo, daño directo en las conexiones entre hemisferios, etc[2]. Se han estudiado para estas condiciones sus diferentes ramificaciones[2], mencionándose aquellas reconocidas como las que encontrarían mayor beneficio en el trabajo propuesto:

■ Afasias:

- **Afasia de Broca**[4,13]: El área de Broca es una región localizada en el lóbulo cerebral izquierdo y está relacionada con el uso del lenguaje. Específicamente la afasia en la que se sufre un daño en esta área, tienen dificultad para la expresión fluida, la pronunciación y modulación del tono de voz. La producción de los sonidos correctos y encontrar las palabras correctas suelen ser trabajos laboriosos, sin embargo la comprensión del habla de otras personas es relativamente buena, por lo que entender textos o lenguaje oral en comparación con su capacidad de hablar y escribir se encuentra mejor conservada.
- **Afasia motora transcortical**[4]: Parecida a la afasia de Broca en la dificultad del paciente para la emisión de un lenguaje fluido y coherente conservando una relativa buena comprensión de lenguaje, esta afectación difiere en el hecho de que los pacientes si son capaces de repetir lo que se les dice, mientras que aquellos que sufren de afasia de Broca son incapaces de repetir frases.

■ Apraxia:

- **Apraxia bucofacial u orofacial**[5]: Incapacidad de realizar movimientos faciales a voluntad, como pasar la lengua por los labios, silbar, toser o guiñar el ojo.
- **Apraxia Verbal**[11]: Dificultad para coordinar los movimientos de la boca y del habla

- **Disartria:** Trastorno de la ejecución motora del habla, derivado de un problema neurológico debido a la presencia de un accidente cerebrovascular, traumas craneoencefálicos u otras lesiones cerebrales.[2]
Entre los síntomas se tienen el habla entrecortada jadeante, irregular, imprecisa o monótona, lenta, o rápida y .^{entre} dientes”, entonación anormal, cambios del timbre del voz, ronquera, babeo o escases del control de la saliva y la movilidad limitada de la lengua, los labios y la mandíbula.[7,8]

Buscando proporcionar una herramienta de apoyo para la comunicación a pacientes que han recientemente sufrido alguna accidente u lesión como las mencionadas anteriormente, no se enfoca el proyecto a un grupo de la sociedad con un amplio desarrollo del lenguaje de señas tal como suele ser común con las personas sordomudas. Esta propuesta consiste en un prototipo simplificado que se apoyado de la idea de un sujetador de los dispositivos hardware con orificios para los dedos tipo *wearable*, no un guante completo. Aprovechando la anatomía del *wearable*, el componente SoC micro:bit, se encontraría en el dorso de la mano, posición en la que se evita cualquier interferencia motriz o de sensibilidad en dedos y palma. Tanto la implementación de los LEDs del SoC para indicar ciertas acciones acompañadas de animaciones buscando aportar una estética moderna y de interacción sencilla para el usuario; Como la sencillez de este con el uso casi exclusivo del sensor acelerómetro en un intento de evitar sensores extras como los de flexión, atienden a la filosofía de disminuir los puntos negativos identificados en proyectos posteriores proponiendo estrategias que los eviten, o incluso, puedan aportar nuevos beneficios al prototipo.

4. Productos o Resultados esperados

La solución que se propone en este trabajo consiste en un prototipo *wearable* sensor, que utiliza el SoC micro:bit y el sensor acelerómetro integrado para el muestreo de los patrones de movimientos realizados por el usuario, con el objetivo de realizar su clasificación mediante un sistema embebido que permita identificar la clase textual del patrón, usando en conjunto el algoritmo DTW y un modelo de Inteligencia Artificial. A partir de esta clasificación, se realiza la conformación de palabras y frases en texto, que finalmente son reproducibles sonoramente en una tarjeta de sonido cuando se hace uso de los servicios que ofrecen el proyecto *TTS de Mozilla* en la red local, o los servicios especializados de *Text-to-speech* ofrecidos por las nubes de compañías privadas.

Desglose de los productos particulares requeridos para cumplir la descripción de solución previa:

1. Como primer propuesta al sistema embebido y prototipo mencionado, se muestra el diagrama de la solución en la Figura 1
2. Código motriz como conjunto de patrones de movimiento, que en el contexto de la solución, tienen respectivamente una clasificación en una clase textual utilizada para la conformación de palabras y frases.
3. El prototipo *wearable* tipo guante, se compone de un par de elementos hardware esencialmente:
 - El SoC micro:bit como elemento sensor principal, que implementa un IMU integrado y del cuál principalmente se utiliza el sensor acelerómetro para el muestreo de los patrones motrices.
 - Un sensor de pulso, que a través de un gesto que puede interpretar el sensor como un pico en el ritmo cardiaco, permita al usuario indicar el inicio del muestreo de un patrón motriz y el fin de este

El prototipo *wearable* al convertirse en el componente con el que ocurrirá la mayor parte de la interacción del usuario, funge como la interfaz principal de la solución, ofreciendo posibilidades para la configuraciones de unos pocos parámetros con el uso de los botones integrados en el SoC, y la visualización de resultados de acciones mediante animaciones en su arreglo de LEDs.

4. El sistema embebido permitirá la comunicación entre el prototipo sensor, la clasificación de los patrones y la consulta de los servicios de *Text-to-Speech*. Su conformación en la primer propuesta de solución considera al menos los siguientes componentes:
 - Prototipo sensor *wearable*
 - Mini PC Raspberry Pi 4

- Tarjeta de Sonido WM8960 de Waveshare
5. La identificación de los patrones de movimiento a sus respectivas clases textuales, se podrá lograr implementando de bibliotecas especializadas, el algoritmo DTW y un modelo de Inteligencia Artificial, que usados en conjunto permitan la clasificación eficiente de los patrones.
 6. La configuración de los servicios de *Text-to-Speech* en red local con el proyecto *TTS de Mozilla*, y el servicio especializado en la nube, permitirá el uso de al menos 2 voces femeninas y 2 voces masculinas en diferentes rangos de tono de voz y edad, al afinarse la combinación de los valores de los parámetros modificables en estos servicios.
 7. Artículo de publicación con el desarrollo de la solución propuesta

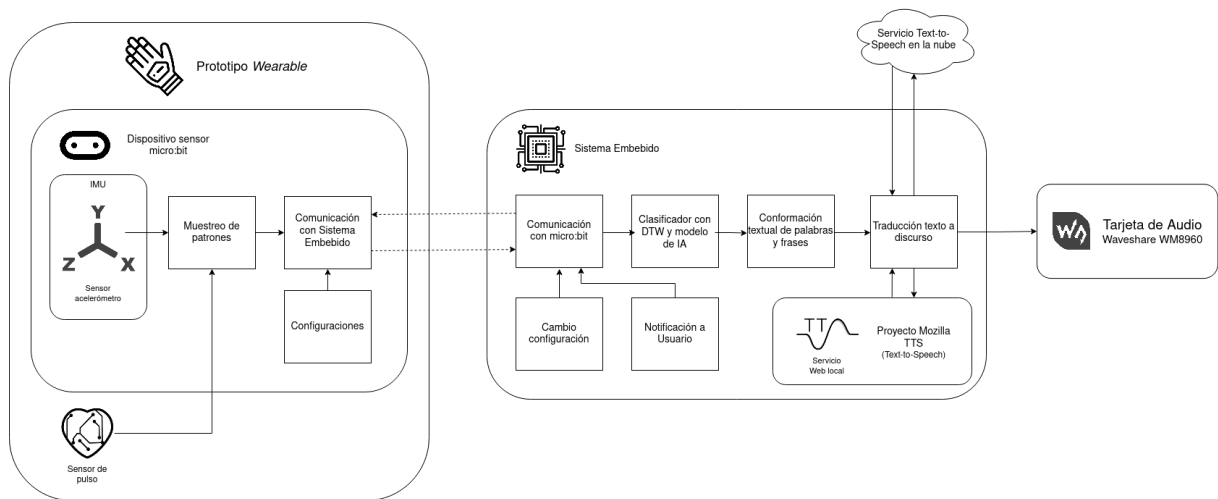


Figura 1: Diagrama como primer propuesta de solución

5. Metodología

El Método en V define un procedimiento para el desarrollo de productos para las TIC, siendo importante mencionar que suele ser un estándar para los proyectos de la Administración Federal alemana y defensa. Se trata de un modelo rígido y con una gran cantidad de iteraciones siendo similar al modelo de cascada.

El nombre lo toma de su estructura gráfica que suele asemejarse a una letra V, dividiéndose por su parte izquierda en las actividades que descomponen las necesidades y la creación de las especificaciones del sistema. En su parte derecha, se representa la integración de las piezas y su verificación.

La implementación de este tipo de modelo aporta una serie de ventajas que son importantes como objetivos a alcanzar en el proyecto. Este modelo aporta una mejora y garantía de la calidad gracias a las medidas de control de calidad firmemente integradas. También minimiza los riesgos del proyecto, permitiendo una detección temprana de las desviaciones y riesgos para la mejora de la gestión de los procesos. Otra gran ventaja es la reducción de los gastos totales durante el proyecto y sistema de ciclo de vida, consiguiéndose gracias al procesamiento transparente a lo largo de todo el ciclo del producto.

Aun con todas estas ventajas este modelo puede ser poco flexible a cambios durante el desarrollo, promoviendo un curso lineal del proyecto, sin embargo si el modelo es entendido y se utiliza correctamente es posible utilizar el modelo V para el desarrollo ágil.

El modelo V define el curso del proyecto en fases individuales cada vez más detalladas(Definición del proyecto):

- **Fase de especificaciones:** Prevé un análisis de las especificaciones del sistema a grandes rasgos
- **Fase funcional:** Se complementa el proyecto con requisitos funcionales y no funcionales para la arquitectura del sistema
- **Fase de diseño:** Se planifican los componentes y las interfaces con un diseño detallado
- **Codificación:** Completadas estas fases inicial el desarrollo de la arquitectura en software

En la corriente de pruebas del lado derecho(Prueba e integración del proyecto) consiste de:

- **Pruebas unitarias:** Permiten declarar que un módulo esta listo y terminado: Lógica de módulos(caja blanca), Funciones(Caja negra)
- **Pruebas de integración:** Implican una progresión ordenada de pruebas que van desde los componentes o módulos y que culmina en el sistema completo
- **Pruebas de Sistema:** Se verifica el cumplimiento de los objetivos y se validan los desajustes entre el software y los requisitos planteados
- **Pruebas de Aceptación:** El usuario final comprueba que el sistema hace lo especificado en el contrato

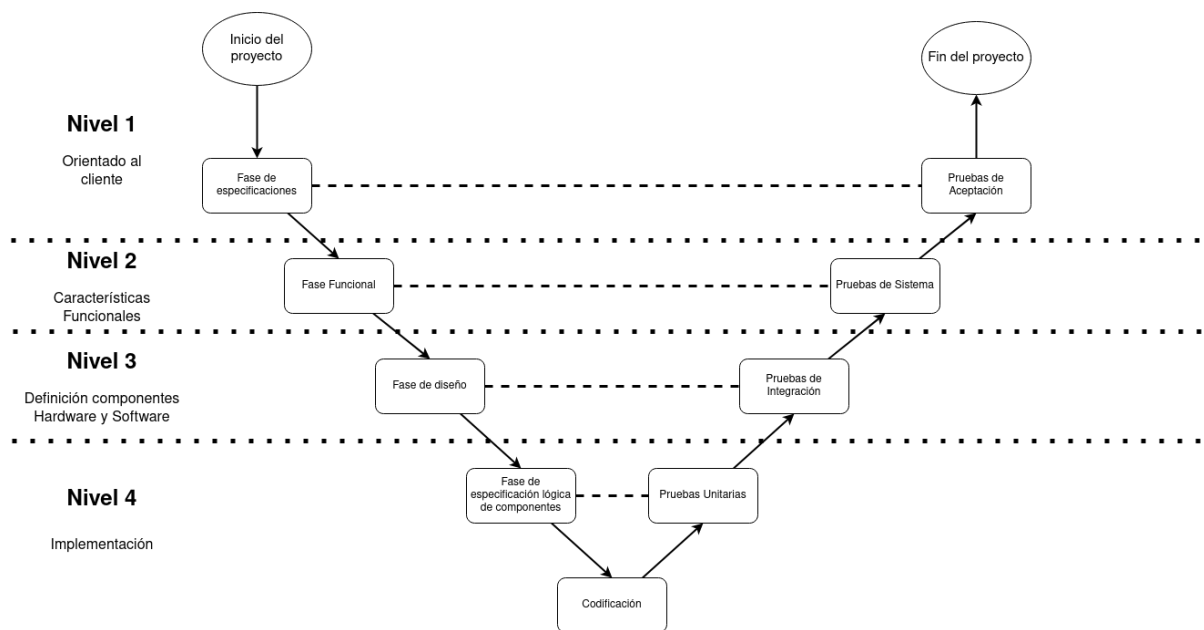


Figura 2: Diagrama mostrando las fases del Modelo V

6. Cronograma

Actividades	Ago	Sep	Oct	Nov	Dic	Ene	Feb	Mar	Abr	May	Jun	Jul
Investigación en dispositivo sensor y acelerómetro integrado	X											
Investigación en algoritmo DTW y optimización de este	X											
Investigación en modelos clasificadores de IA	X	X										
Análisis del prototipo <i>wearable</i>		X										
Análisis del modelo clasificador		X	X									
Diseño alto nivel del prototipo <i>wearable</i>			X									
Diseño alto nivel del modelo clasificador			X									
Diseño del código motriz propuesto			X	X								
Diseño detallado del prototipo <i>wearable</i>				X								
Diseño detallado del modelo clasificador <i>wearable</i>				X	X							
Especificación componentes lógicos del sistema					X							
Evaluación TT I					X	X						
Ensamblado del prototipo de guante <i>wearable</i>						X	X					
Pruebas del funcionamiento del prototipo <i>wearable</i>							X					
Desarrollo de los componentes lógicos en el nodo sensor							X	X				

[illegible]

7. Referencias

- [1] N. Marques, "¿Qué es el lenguaje?," *Babel*, 02, 2018 [En línea]. Disponible en <https://es.babel.com/es/magazine/que-es-lenguaje>
- [2] O. Castellero Mimenza, "Los 8 tipos de trastornos del habla," *Psicología y Mente*. [En línea]. Disponible en <https://psicologiaymente.com/clinica/tipos-trastornos-habla>
- [3] National Institute on Deafness and Other Communication Disorders, (2017, 03. 06). "La afasia". [En línea]. Disponible en <https://www.nidcd.nih.gov/es/espanol/afasia>
- [4] A. Triglia, "Afasia: los principales trastornos del lenguaje," *Psicología y Mente*. [En línea]. Disponible en <https://psicologiaymente.com/clinica/afasia-trastornos-lenguaje>
- [5] "Apraxia", *Instituto Nacional de Trastornos Neurológicos y Accidentes Cerebrovasculares*. 03, 2022. [En línea]. Disponible en <https://espanol.ninds.nih.gov/es/trastornos/apraxia>
- [6] J. Huang, "Apraxia," *MSD*. 10, 2021. [En línea]. Disponible en <https://www.msmanuals.com/es-mx/hogar/enfermedades-cerebrales,-medulares-y-nerviosas/disfunci%C3%B3n-cerebral/apraxia>
- [7] "La Disartria," *American Speech Language Hearing Association*. [En línea]. Disponible en <https://www.asha.org/public/speech/Spanish/La-Disartria/>
- [8] J. Huang, "Disartria," *MSD*. 10, 2021. [En línea]. Disponible en <https://www.msmanuals.com/es-mx/hogar/enfermedades-cerebrales,-medulares-y-nerviosas/disfunci%C3%B3n-cerebral/disartria>
- [9] C. J. G. Ayala Aburto, "Guante traductor de señas para sordomudos," Tesis título licenciatura, ESIME, unidad Azcapotzalco, Ciudad de México, México, 2018.
- [10] E. D. Jiménez Carbajal, G. E. Rivera Taboada, "Sistema de comunicación auditiva para personas con problemas del habla", Tesis para título de licenciatura, ESCOM, Ciudad de México, México, 2013.
- [11] D. Vishal, H. M. Aishwarya, K. Nishkala, B. T. Royan and T. K. Ramesh, "Sign Language to Speech Conversion," (en inglés) 2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), 2017, pp. 1-4, doi: 10.1109/ICCIC.2017.8523832.
- [12] M. M. Chandra, S. Rajkumar and L. S. Kumar, "Sign Languages to Speech Conversion Prototype using the SVM Classifier," TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON), 2019, pp. 1803-1807, doi: 10.1109/TENCON.2019.8929356.
- [13] "Afasia de Broca," *National Aphasia Association*. [En línea]. Disponible en <https://www.aphasia.org/es/afasia-de-broca/>

8. Alumnos y Directores

Luis Eduardo Valle Martínez.- Alumno de la carrera de Ing. en Sistemas Computacionales en ESCOM, Especialidad Sistemas, Boleta: 2015090780, Tel. 5566143276, email: lvallem1400@alumno.ipn.mx

Firma: _____

Rodolfo Romero.- Profesor de tiempo completo Laboratorio de posgrado Sistemas computacionales móviles. Candidato a Doctor en ciencias en Comunicaciones y Electrónica. Maestría en ciencias en Ingeniería electrónica. Ingeniería en comunicaciones y Electrónica. Área de trabajo Inteligencia Artificial y Procesamiento digital de señales. Tel. 5535216128, email: rromeroh@ipn.mx.

Firma: _____

Jesús Yaljá Montiel Pérez.- Profesor de tiempo completo adscrito al Laboratorio de Robótica y Mecatrónica del Centro de Investigación en Computación del Instituto Politécnico nacional. Doctor en comunicaciones y electrónica. Maestro en Ciencias en Ingeniería electrónica e Ingeniero Físico. Sus intereses son: la Inteligencia Artificial, sensores y robótica. Tel: 5524940919, Ext. IPN: 56665, email: yalja@ipn.mx

Firma: _____

CARÁCTER: Confidencial

FUNDAMENTO LEGAL: Artículo 11 Fracc. V y Artículos 108, 113 y 117 de la Ley Federal de Transparencia y Acceso a la Información Pública.

PARTES CONFIDENCIALES: Número de boleta y teléfono