

Prototipo embebido de la interpretación de patrones de movimiento en una mano para la reproducción de fonemas y palabras completas en el idioma Español de México

Trabajo Terminal No. ____ - ____

*Alumnos: *Valle Martínez Luis Eduardo*

Directores: Rodolfo Romero Herrera, Dr. Jesús Yaljá Montiél Pérez

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

Resumen -

Palabras clave - Afectaciones del habla, *Machine Learning*, Sistemas Embebidos, *Text-to-Speech*

1. Introducción

Aspectos como los gestos, muecas, sonidos, etc. son factores influyentes en la comunicación del estado consciente o inconsciente de una persona que lo hace visible para ser interpretable por los demás. Sin embargo en la comunicación humana, aunque relevantes estos aspectos, es inevitable reconocer al lenguaje como el elemento simbólico fundamental en la interacción entre individuos y grupos.

El lenguaje humano como la prueba de especímenes que han desarrollado un intelecto sofisticado, ha sido en la historia de la evolución humana la piedra angular de la que nuestra especie se ha valido como herramienta para generar conocimiento y perpetuarlo para ser compartido y adquirido por generaciones posteriores, que ha permitido la comunicación para la organización de grupos con un objetivo en común, incluso como el medio para la expresión de sentimientos y pensamientos en forma de literatura y poemas.

A nivel oral, expresamos el lenguaje mediante el habla, alcanzado en algún punto de la evolución del sistema canal vocal-auditivo en el humano, derivado del descenso de la laringe y lo que nos dio la posibilidad de crear sonidos[5]. A través de la especialización como especie en la actividad del habla, se fueron otorgando semántica a los sonidos generados y nos permitió la asociación de un significado a estos[5].

Es entonces sencillo visualizar la importancia del lenguaje, resaltando específicamente el lenguaje oral como tema de estudio en este trabajo, pues es la herramienta cotidiana utilizada por la gran mayoría de las personas para la comunicación e interacción social. Contrastado por un porcentaje muy mayor de personas que no cuentan con ninguna circunstancia que le impida el habla, un sector de la población en la sociedad se ve limitado en el aprovechamiento de este recurso humano (Siendo 2,234,303 personas, un aproximado del 1.76% de la población total en México para el Censo 2020 para personas con dificultades o discapacidad para hablar o comunicarse[6]), originado por trastornos y afecciones que dificultan, o incluso impiden e imposibilitan, que una persona tenga la capacidad de expresarse de manera oral mediante el habla. Entre las afecciones más habituales se encuentran 3 de ellas que podrían verse beneficiadas por el uso de la solución propuesta en este trabajo.

Las primeras son las **afasias**[7-9], las cuales son problemas médico originado por una lesión cerebral y que resulta en la pérdida o alteración del lenguaje. Las **apraxias**[7,11,12] 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**[7,13,14] 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.[14]

Otro tipo de trastorno que también afecta de forma indirecta a la capacidad para la expresión hablada del lenguaje como consecuencia de la afección principal, es la sordera de percepción total y es clasificado como un trastorno de la audición. Aún cuando este proyecto propuesto podría ser útil en algunas situaciones para sujetos con este trastorno, sus condiciones limitan el uso y por lo tanto no es el principal sector de la población considerada para beneficiarse del trabajo. Este tipo de afecciones requiere de consideraciones adicionales que garanticen la comunicación bidireccional entre un individuo con sordera y otro sin trastornos auditivos, esto con el objetivo de proponer una tecnología atractiva y útil para su uso por el público objetivo.

Frente a esta problemática, se propone en este documento el desarrollo de una solución tecnológica basada en el SoC micro:bit, que permita el sensado del movimiento de una mano utilizando el acelerometro integrado, y que a partir de unos patrones de movimientos

predefinidos con un sencillo lenguaje propuesto, se puedan identificar letras del alfabeto en español como interfaz para la reproducción sonora de los fonemas en el idioma español de México y su conformación en sílabas y finalmente palabras completas. Para conseguir esto a partir de los patrones de movimiento del lenguaje propuesto, se identifican y clasifican utilizando modelos de *Machine Learning*, los movimientos que son traducidos a una cadena de texto y que finalmente para su reproducción sonora se utiliza un servicio de *Text-to-Speech*, como el proyecto *TTS de Mozilla* y servicios especializados en la nube como *Microsoft Azure Text-to-speech*.

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*[1], 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"[2], 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 característica de la familia de *Random Forest* y *Mean Shift*, son aplicadas 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[3]. En este trabajo desarrollado por un equipo de la Universidad de Amrita en la India, utilizan en conjunto un sensor IMU(*Inertial Measurement Unit*) que implementa un sensor acelerómetro y un 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*[4], 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 recompilación 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).

De los trabajos realizados y que se encuentran relacionados a la solución que se ofrece en este trabajo, se aprecia que las principales aplicaciones suelen desarrollarse bajo el área de la Visión por Computadora; Lo que acarrea sus complicaciones debido al específico ambiente de contraste e iluminación que requieren las técnicas de análisis de imágenes, así como su poca portabilidad e intuitivo uso. Como así también los prototipos de guantes o *wearables*, que principalmente encuentran sus inconvenientes en la cantidad de hardware y estética, para un extendido uso como herramientas tecnológicas de apoyo en la comunicación.

La solución propuesta es un prototipo decantado por un prototipo físico tipo *wearable*, aunque más sencillo que los mencionados en el sentido del uso de menos hardware, y por lo tanto menos sensores utilizados, aunque aportando mayor portabilidad y sencillez de uso. A pesar de esto, se opta utilizar un lenguaje específico para este trabajo y no utilizando, como comunmente se implementa en estos proyectos, el lenguaje de señas del país. Inevitable realizar la comparación del LSM(Lenguaje de Señas Mexicano) con el lenguaje por proponerse, sin embargo, y aún cuando este trabajo tiene el objetivo común de permitir la comunicación principalmente de una comunidad con impedimento del habla, difieren primordialmente en algunos aspectos:

- El LSM es un lenguaje que requiere para su expresión el involucramiento de gestos con movimiento desde encima de la cabeza y hasta debajo de la cadera, incluyendo movimiento de manos, expresiones faciales y mirada intencional. Por su parte el lenguaje motriz que se propondrá se limita únicamente a 1 mano, donde se colocará el SoC para el sensado.
- La interpretación de ambos lenguajes se realiza en dominios sensoriales humanos distintos. Completamente visual para el LSM, y parcialmente visual para el lenguaje propuesto, pero principalmente auditivo.
- Otra diferencia importante se encuentra en la complejidad y relación con el idioma español. Mientras el LSM cuenta con una rica y compleja gramática, y vocabulario abundante, también existe desacoplado de la estructura, gramática y normas propias del español. El enfoque principal del lenguaje que se propondrá, es la expresión de palabras y letras del español, sobrepasando sus capacidades la formulación de normas gramaticales o vocabulario propio, fungiendo únicamente como un simplificado codificado motriz para la expresión sonora del español.

- Finalmente su implementación, aún cuando es común para las personas con impedimento del habla u afecciones relacionadas a la expresión oral, el LSM brinda mayores posibilidades para la población sordo-muda. Mientras tanto el prototipo de este trabajo podrá ser poco beneficioso para la población sordo-muda frente al LSM, pero propone una alternativa para la comunicación(dentro de los límites de este trabajo, aún unilateral) entre personas con impedimento del habla hacia la población con afecciones relacionadas a la percepción visual.

2. Objetivo

Objetivo general:

Creación de un sistema embebido que opere en conjunto a un prototipo de guante sensor utilizado para el muestreo de los patrones de movimiento en una mano, comunicando los patrones a un microcomputador Raspberry que será el encargado del procesamiento y clasificación de los patrones con técnicas de *Machine Learning*, y alojando o solicitando los servicios para la reproducción sonora del texto que identifique al fonema o palabra completa en español asociados al movimiento ejecutado.

Objetivos específicos:

1. Ensamblar el prototipo de guante sensor implementando el SoC micro:bit como dispositivo microcontrolador y sensor del cambio de aceleraciones con el acelerómetro integrado. Añadiendo 1 sensor táctil usado para indicar el inicio y fin del muestreo de un movimiento.
2. Muestrear y preprocesar los patrones tridimensionales utilizando el SoC micro:bit.
3. Comunicar y transmitir los patrones muestreados a través de Bluetooth entre el SoC micro:bit y el mini PC RaspberryPi 4.
4. Clasificar los patrones de movimiento utilizando un modelo de *Machine Learning*.
5. Concatenar las asociaciones de los movimientos para convertirlos en una cadena textual que pueda ser interpretada por el servicio en la nube de *Text-to-speech*
6. Reproducir el audio resultante del servicio en la nube mediante la tarjeta de sonido WM8960 de Waveshare.

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, ya sea por la falta de circulación del torrente sanguíneo, daño directo en las conexiones entre hemisferios, etc[7]. Se han estudiado para estas condiciones sus diferentes ramificaciones[7], mencionándose aquellas reconocidas como las que encontrarían mayor beneficio en el trabajo propuesto:

• Afasias:

- **Afasia de Broca**[9,10]: 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**[9]: Parecida a la afasia de Broca en la dificultad del paciente para la emisión de un lenguaje fluido y coherente mientras una relativa comprensión de lenguaje se mantiene, con diferencia en que los sujetos con esta afección 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**[11]: 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, debido a un problema neurológico debido a la presencia de un accidente cerebrovascular, traumas craneoencefálicos u otras lesiones cerebrales.[7]

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.[13,14]

Buscando proporcionar una herramienta de apoyo para la comunicación a pacientes que han recientemente sufrido alguna accidente u lesión como las mencionadas anteriormente, no está precisamente enfocado este proyecto a un grupo de la sociedad con un amplio desarrollo del lenguaje de señas tal como suele ser común con las personas sordomudas. Se propone un prototipo simplificado apoyado en la idea de un sujetador de los dispositivos hardware con orificios para los dedos y no un guante completo. De esta forma el componente principal, el SoC micro:bit, se encuentra en el dorso de la mano y no interfiere de otra forma en la sensibilidad, o limitación motriz de los dedos. Esto hace al prototipo sencillo de usar, estético con la implementación de leds directamente en el SoC y sencillo para la interacción.

Precisamente y debido a la sencillez del prototipo, se utilizan a los más el par de sensores: acelerómetro y magnetómetro, descartando el agregado de sensores de flexión u otros componentes sensores adicionales.

La creación del sencillo lenguaje codificado en un contexto de patrones de movimiento, le servirá al usuario para indicar el sonido que desea sea reproducido, y debido a que la forma más intuitiva para la mayor parte de la población (cerca del 95% de la población que es alfabeta en México[15]) la asociación de la pronunciación de las letras en las palabras es natural mediante la relación en su representación escrita, se descarta un desglose directo en fonémas de las letras o palabras, siendo prácticamente desconocido un marco referencial como el Alfabeto Fonético Internacional y las asignaciones de los sonidos propios del español.

4. Productos o Resultados esperados

1. Prototipo de *wearable* tipo guante para la portabilidad del SoC micro:bit como elemento sensor y preprocesador de los patrones de movimiento, así como principal interfaz de interacción con el usuario
2. Sistema embebido con comunicación entre el prototipo sensor y la tarjeta de sonido:
 - Prototipo sensor *wearable*
 - Mini PC Raspberry Pi 4
 - Tarjeta de Sonido WM8960 de Waveshare
3. Modelo basado en *Machine Learning* para el procesamiento y clasificación de los patrones de movimiento a las respectivas clases textuales (letras, sílabas o palabras)
4. Configuración de al menos 2 voces femeninas y 2 voces masculinas en diferentes rangos de tono de voz y edad
5. Documento de publicación con el desarrollo de la solución propuesta

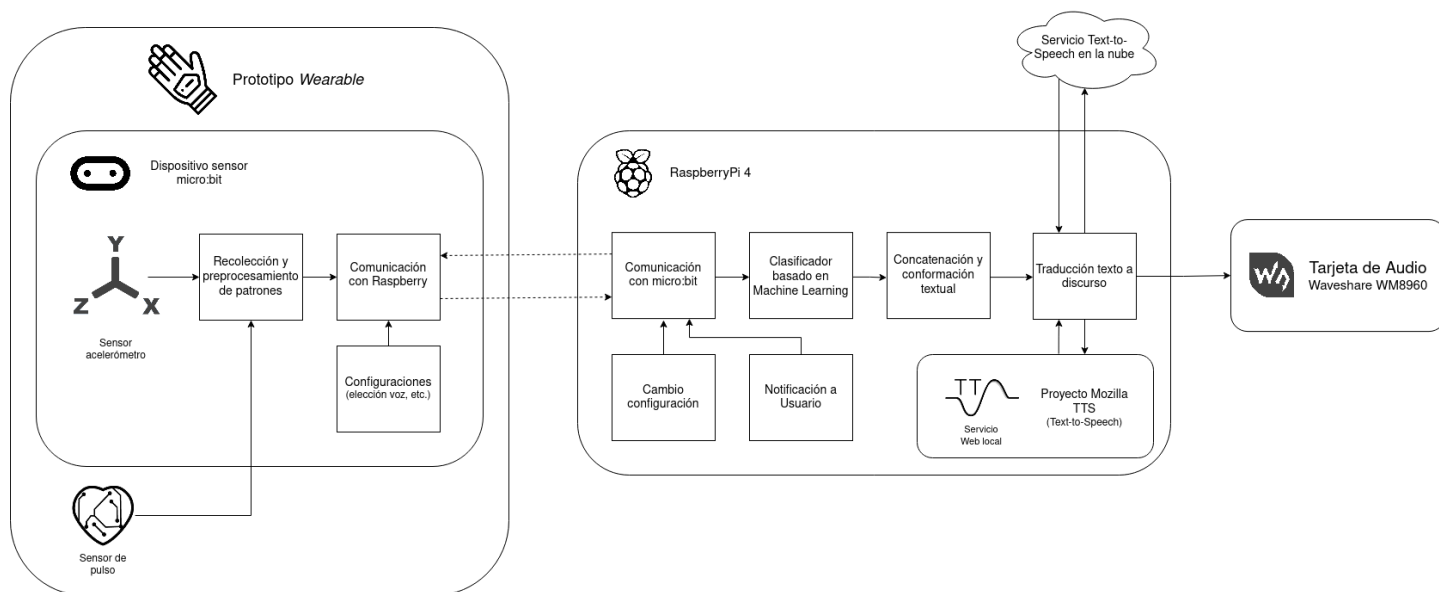


Figure 1: Diagrama de la arquitectura del sistema completo considerando al prototipo de sensado, el mini PC RaspberryPi 4 y la tarjeta de Audio.

La línea punteada indica una comunicación *wireless* tipo Bluetooth entre el micro:bit y la Raspberry Pi 4

5. Metodología

Inicialmente fue considerada una Metodología de Prototipos, que cuenta con características similares a la metodología final escogida, difiriendo fundamentalmente en la filosofía de un desarrollo acelerado que pueda ofrecer una visión previa del producto, requiriendo una costante integración con el fin de realizar pruebas con los clientes e ir realizando las modificaciones o ajustes necesarios para cumplir con los requisitos dispuestos. Este enfoque aunque práctico, puede fácilmente caer en el error de perder de vista los compromisos de calidad y mantenimiento que se acordaron con el cliente, esto debido a las continuas pruebas presentadas al cliente lo que hace que el desarrollo se riga a base de prueba y error.

Se ha decantado el desarrollo del proyecto por una Metodología basada en Componentes, por la naturaleza modular de la solución y los diferentes nodos y componentes que pueden irse mejorando independiente a la conformación completa del producto. Importante mencionar que el desarrollo independiente de cada componente requiere una definición correcta de los medios de conexión y protocolos de comunicación, facilitando y promoviendo una exitosa integración de todos los componentes.

Este paradigma considera un progreso de forma evolutiva implementando muchas bondades del modelo en espiral, siendo una de las herramientas más beneficiosas y también las más problemáticas si no se aplican de forma correcta, el análisis de riesgos. La eficiente predicción de riesgos requiere de experiencia y habilidad derivada del conocimiento profundo de la problemática y la solución por aplicarse, lo que requerirá asesoría con especialistas en las tecnologías que se utilizarán y permitiendo disminuir la incertidumbre de la detección errada de los riesgos.

Entre los beneficios de la implementación de esta metodología se encuentra la simplificación de pruebas, pues estas son ejecutadas para cada uno de los componentes antes de ser ensamblados en el conjunto. A su vez esto simplifica el mantenimiento general del sistema, así como su escala y propicia la calidad del sistema completo mejorando continuamente los componentes de forma independiente.

Finalmente resaltar la ventaja que aporta su cualidad incremental respecto a la búsqueda del modelo de aprendizaje automático, asegurando aquel con la mejor precisión de predicción y eficiencia en el desempeño durante su entrenamiento y finalmente la clasificación implementado directamente ya en el sistema embebido.

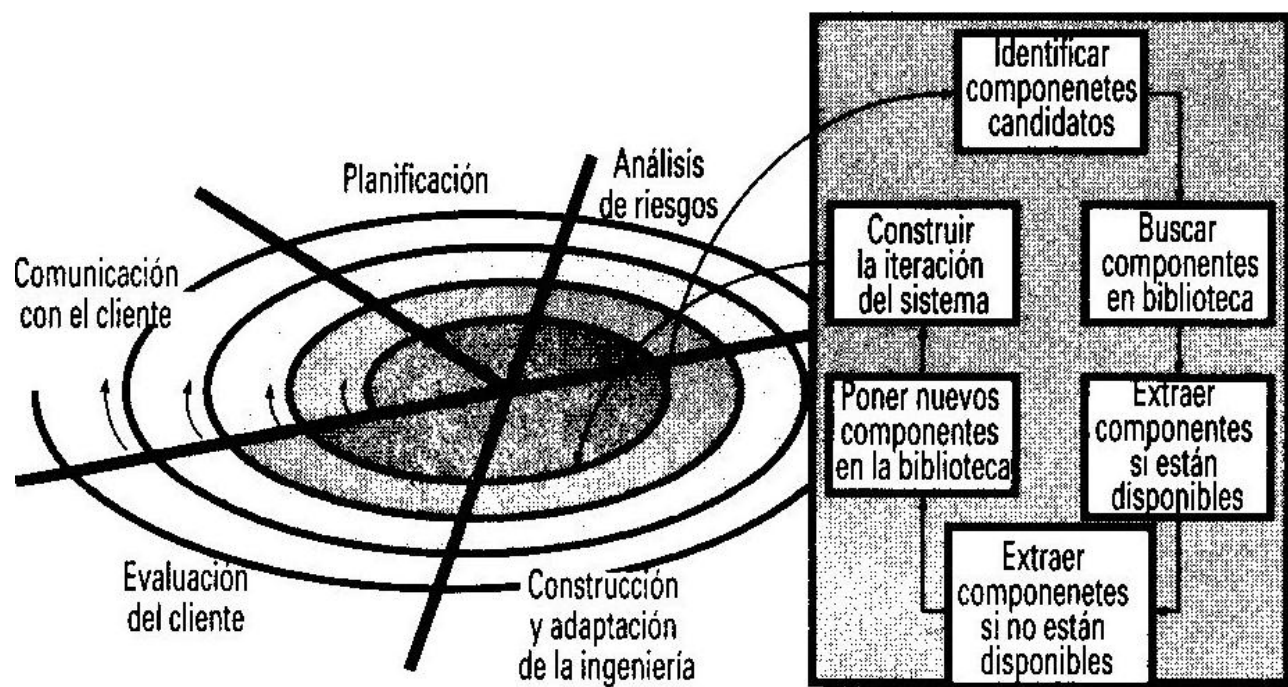


Figure 2: Diagrama mostrando el flujo iterativo y evolutivo de las etapas consideradas por la Metodología basada en el desarrollo por Componentes

6. Cronograma

7. Referencias

- [1] C. J. G. Ayala Aburto, "Guante traductor de señas para sordomudos," Tesis título licenciatura, ESIME, unidad Azcapotzalco, Ciudad de México, México, 2018.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] N. Marques, "¿Qué es el lenguaje?," *Babbel*, 02, 2018 [En línea]. Disponible en <https://es.babbel.com/es/magazine/que-es-lenguaje>
- [6] "Presentación de Resultados," INEGI Censo 2020., México, 2020. [En línea]. Disponible en https://www.inegi.org.mx/contenidos/programas/ccpv/2020/doc/Censo2020_Principales_resultados_EUM.pdf
- [7] O. Castellero Mimenza, "Los 8 tipos de trastornos del habla," *Psicología y Mente*. [En línea]. Disponible en <https://psicologiaymente.com/clinica/tipos-trastornos-habla>
- [8] 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>
- [9] A. Triglia, "Afalias:los principales trastornos del lenguaje," *Psicología y Mente*. [En línea]. Disponible en <https://psicologiaymente.com/clinica/afalias-trastornos-lenguaje>
- [10] "Afasia de Broca," National Aphasia Association.[En línea]. Disponible en <https://www.aphasia.org/es/afasia-de-broca/>
- [11] "Apraxia", *Instituto Nacional de Trastornos Neurológicos y Accidentes Cerebrovasculares*. 03, 2022.[En línea]. Disponible en <https://espanol.ninds.nih.gov/es/trastornos/apraxia>
- [12] 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>
- [13] "La Disartria," *American Speech Language Hearing Association*. [En línea]. Disponible en <https://www.asha.org/public/speech/Spanish/La-Disartria/>
- [14] 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>
- [15] INEGI (2020).[En línea]. Disponible en <https://www.inegi.org.mx/temas/educacion/>

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: _____

Dr. 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