Memorias del Congreso Científico Tecnológico de las carreras de
Ingeniería Mecánica Eléctrica, Industrial y Telecomunicaciones, sistemas y electrónica

AÑO 8. No. 8. ISSN-2448-7236. SEPTIEMBRE 2023 – AGOSTO 2024. IC-01, pág.: 1 a la 5.



# Manipulación de objetos domóticos mediante comando de voz

***Tania Abigail Lira Baca***[1] , **Fernando Gudiño Peñaloza**[2]*

## RESUMEN

Para hablar de discapacidad debemos de abarcar un tema muy amplio que implica abordar varios aspectos alrededor de este término. La discapacidad se refiere a las limitaciones o dificultades que tiene una persona para llevar a cabo acciones o tareas en situaciones cotidianas y vitales debido a una condición física o mental. En México, alrededor del 6% de la población vive con alguna discapacidad; ya sea física, mental y sensorial, y puede ser a causa o agravada por el entorno económico y social. En las casas habitación las diferentes discapacidades limitan el desarrollo de las actividades diarias, acciones sencillas como, prender y apagar una estufa, trasladarse de un lugar a otro, utilizar un electrodoméstico o cerrar una persiana, se convierten en actividades complejas difíciles de realizar para ciertas personas. Es por ello que la automatización de dichos procesos rutinarios es una necesidad inmediata que se debe solucionar. Actualmente, con la ayuda de los asistentes de voz la automatización de unos procesos se ha vuelto más sencilla. El habla ha emergido como uno de los más increíbles medios de comunicación humana, como escritura, lenguaje corporal y el lenguaje gesticular, el habla se considera como la forma más natural y directa de comunicación. Es por ello que naturalmente, se crean interfaces hombre-maquina, para la interacción con diferentes dispositivos. Este articulo habla sobre la utilización de reconocimiento de voz para la automatización de procesos repetitivos dentro de las casas habitación, con el fin de mejorar la calidad de vida de las personas con discapacidad motriz.

## ABSTRACT

To talk about disability we need to include a very broad topic that involves addressing various aspects around this term. Disability refers to the limitations or difficulties that a person has to carry out actions or tasks in daily and vital situations due to a physical or mental condition. In Mexico, around 6% of the population lives with some kind of disability; whether physical, mental and sensory, and may be caused by or aggravated by the economic and social environment. In dwelling houses, the different disabilities limit the development of daily activities, simple actions such as turning a stove on and off, moving from one place to another, using an electrical appliance or closing a blind, become complex activities that are difficult for certain people to carry out. people. That is why the automation of these routine processes is an immediate need that must be solved. Currently, with the help of voice assistants, the automation of some processes has become easier. Speech has emerged as one of the most incredible means of human communication, like writing, body language, and gestural language, speech is considered to be the most natural and direct form of communication. That is why naturally, man-machine interfaces are created for interaction with different devices. This article talks about the use of voice recognition for the automation of repetitive processes within the houses, in order to improve the quality of life of people with motor disabilities.

**Palabras claves**: Casa inteligente, Automatización, Inclusión Social, Interfaz Hombre-Máquina

## INTRODUCCIÓN

De acuerdo con el Consejo Nacional para el Desarrollo y la Inclusión de las Personas con Discapacidad (CONADIS), la prevalencia en varones es de 3.3 millones y de 3.8 millones en mujeres [1,4].

Existen tres tipos de discapacidad: la sensorial y de la comunicación (discapacidad para ver, oír y hablar), motriz (problemas para caminar, manipular objetos y de coordinación para realizar actividades) y mental (personas que tienen dificultades para aprender y relacionarse con otras personas).

La discapacidad más frecuente es la motriz, ya que de acuerdo con el Instituto Nacional de Estadística y Geografía (INEGI), las dificultades para ver y caminar son las más frecuentes, mientras que las menos reportadas son las de habla o comunicación. La discapacidad motriz se presenta con más frecuencia en poblaciones productivas y económicamente activas, mientras que la sensorial se presenta más en niños y en adultos mayores.

La domótica es el conjunto de tecnologías aplicadas al control y la automatización inteligente de la vivienda, que permite una gestión eficiente del uso de la energía, que aporta seguridad y confort, además de comunicación entre el usuario y el sistema. Un sistema domótico es capaz de recoger información proveniente de unos sensores o entradas, procesarla y emitir órdenes a unos actuadores o salidas, asimismo, puede acceder a redes exteriores de comunicación o información. [3,5]

* FES-C, Departamento de Ingeniería, itse.lira@gmail.com. **FES-C, Departamento de Ingeniería,** fernando.cuautitlan@comunidad.unam.mx

AÑO 8. No. 8. ISSN-2448-7236. SEPTIEMBRE 2023 – AGOSTO 2024. IC-01, pág.: 1 a la 5.

## ANTECEDENTES

### La discapacidad en México

Según la Organización Mundial de la Salud al 2020, más de 1,000 millones de personas viven en todo el mundo con algún tipo de discapacidad, aproximadamente el 15 % de la población mundial; de ellas, casi 190 millones tienen dificultades en su funcionamiento y requieren con frecuencia servicios de asistencia. El número de personas con discapacidad va en aumento debido al envejecimiento de la población y al incremento de enfermedades crónicas.

De acuerdo con el Censo de Población y Vivienda 2020, en México hay 6,179,890 personas con algún tipo de discapacidad, lo que representa 4.9 % de la población total del país. De ellas 53 % son mujeres y 47 % son hombres. [1]


**Figura 1. Distribución de personas con discapacidad por genero**

El INEGI identifica a las personas con discapacidad como aquellas que tienen dificultad para llevar a cabo actividades consideradas básicas, como: ver, escuchar, caminar, recordar o concentrarse, realizar su cuidado personal y comunicarse, tal como se ve en la figura 2.

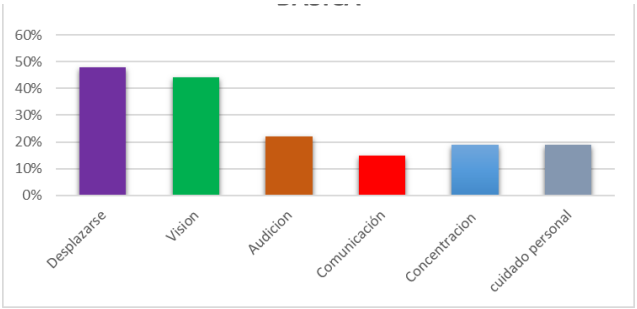



**Figura 2. Porcentaje por dificultad actividad Básica. Nota: La suma de porcentajes es mayor a 100 por la población que presenta más de una dificultad. Fuente: INEGI. Censo de Población y Vivienda 2020.[2]**

## Desarrollo

El reconocimiento automático de voz es la conversión de ondas de voz o de audio en una representación textual de palabras. Por lo general, es el primer paso en las interfaces de usuario de voz como Apple Siri y Amazon Alexa.

Hay muchas propiedades del lenguaje que lo hacen diferente para realizar un reconocimiento automático de la voz con precisión. Algunos de los cuales son el ruido, la elocución, los límites ambiguos de las palabras y la falta de contexto[6].

Para nuestra propuesta seguiremos la siguiente metodología. Descrita en las secciones siguientes

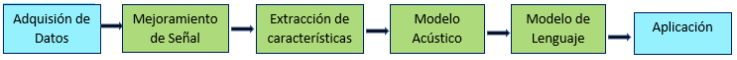



**Figura 3. Canalización de reconocimiento de voz[6]**

### Adquisición de datos.

En la primera etapa se encuentra la adquisición de señales. Mediante un medio de adquisición y digitalización (ADC) se recopilan muestras de audio. Para ello se utilizó el módulo de reconocimiento de voz V.3 de Arduino, cuyas características electrónicas son las siguientes:

- Voltaje de Operación: 4.5V - 5V DC
- Consumo de corriente: <40mA
- Interface digital: 5V TTL UART y GPIO
- Interface analógica: Conector 3.5mm para micrófono
- Precisión de detección: 99% (bajo condiciones controladas

Además se utilizó de un micrófono omnidireccional de solapa, con las siguientes características:

- Transductor: condensador electret
- Rango de frecuencia: 20 Hz - 16,000 XNUMX Hz
- Sensibilidad: -38 dB ± 3 dB
- Impedancia de salida: ≤ 100 K
- Relación señal-ruido: 58 dB SPL

### Mejoramiento de la señal.

AÑO 8. No. 8. ISSN-2448-7236. SEPTIEMBRE 2023 – AGOSTO 2024. IC-01, pág.: 1 a la 5.

En la segunda etapa se realiza el mejoramiento de la señal, en esta etapa las técnicas adecuadas para el mejoramiento de la calidad de las muestras adquiridas son utilizadas: el filtrado, la amplificación etc. Para ello se realizaron los siguientes procesos:

- Filtrado pasa bajas y pasa bandas. El filtrado pasa bandas, también conocido como filtro de banda, es un tipo de filtro electrónico que permite el paso de señales dentro de una determinada banda de frecuencia mientras atenúa o bloquea las señales fuera de esa banda. Esencialmente, este tipo de filtro "filtra" una banda específica de frecuencias de una señal más amplia. Un filtro pasa bandas tiene una respuesta de frecuencia que muestra una ganancia alta o una transmisión eficiente dentro de la banda de frecuencia deseada, y una atenuación significativa fuera de esa banda. La banda de frecuencia se define por una frecuencia de corte inferior y una frecuencia de corte superior.
- Cambio de nivel (offset). el "offset" se refiere a un desplazamiento o corrimiento de un nivel de voltaje o corriente de una señal respecto a una referencia establecida. Es la diferencia entre el valor real de la señal y el valor de referencia.
-
- El offset puede ser tanto positivo como negativo, dependiendo de si la señal se desplaza hacia arriba o hacia abajo en relación con la referencia. Por lo general, se expresa en voltios (V) o en porcentaje de la amplitud de la señal
- Filtrado de línea. Un filtro de línea es un dispositivo utilizado en electrónica para reducir o eliminar las interferencias y ruidos presentes en la corriente eléctrica de la línea de alimentación. Estas interferencias pueden ser generadas por equipos electrónicos cercanos, fluctuaciones en la red eléctrica o fuentes de radiofrecuencia externas.

**Extracción de características.**

En la tercera etapa se trata de la extracción de características.
Este proceso busca extraer, patrones determinantes en la señal de audio caracterizándolos, permitiendo enfocarse en parámetros distintivos y relevantes en el proceso de reconocimiento de voz.

En el reconocimiento de voz, las características se extraen a partir de las señales de audio que contienen el habla. Algunas de las técnicas comunes utilizadas para la extracción de características en el reconocimiento de voz incluyen:

1. Mel Frequency Cepstral Coefficients (MFCC): Los MFCC son una representación ampliamente utilizada en el reconocimiento de voz. Se calculan a través de una serie de pasos, que incluyen el cálculo del espectro de potencia de la señal de audio, la aplicación de una escala de frecuencia no lineal llamada escala mel y la aplicación de la transformada de coseno discreta.
2. Filtro bancos de energía: Se utilizan filtros de paso de banda en un rango de frecuencias para medir la energía de la señal de audio en diferentes bandas. Estos filtros capturan las características de frecuencia del habla.
3. Delta y Delta-Delta: Además de las características estáticas, se pueden calcular características dinámicas, como las velocidades y las aceleraciones de los coeficientes MFCC. Estas características dinámicas, conocidas como Delta y Delta-Delta, proporcionan información temporal sobre la señal de audio.
4. Perceptual Linear Prediction (PLP): Esta técnica modela la percepción humana del sonido y extrae características perceptualmente relevantes, utilizando una combinación de análisis espectral y técnicas de predicción lineal.

Estas son solo algunas de las técnicas comunes utilizadas para la extracción de características en el reconocimiento de voz. El objetivo final es obtener un conjunto de características que capturen de manera efectiva las propiedades distintivas del habla y que permitan una correcta clasificación y reconocimiento de los patrones de voz.

En nuestro caso, antes de que se proporcione voz o audio a los modelos, es necesario convertirlo en formas apropiadas para el modelo. Los cuales consisten en espectrogramas

**Modelo Acústico.**

En la cuarta etapa es el modelo acústico en la cual se intenta asignar la señal de audio a las unidades básicas del habla.

En el reconocimiento de voz, el modelo acústico es una parte fundamental del sistema. Es un componente que se encarga de convertir la señal de audio de entrada en una secuencia de unidades de sonido discretas, como fonemas o subfonemas, que representan las unidades básicas del lenguaje hablado.

El modelo acústico se construye mediante técnicas de aprendizaje automático, como los modelos ocultos de Márkov (HMM, por sus siglas en inglés) o las redes neuronales, que se entrenan utilizando grandes cantidades de datos de voz etiquetados. Estos datos de entrenamiento consisten en grabaciones de voz junto con las transcripciones correspondientes, lo que permite al modelo aprender a asociar las características acústicas de la señal de audio con las unidades de lenguaje correspondientes.

El modelo acústico se basa en la extracción de características de la señal de audio, como los coeficientes MFCC o las características PLP, que se mencionaron anteriormente. Estas características se

AÑO 8. No. 8. ISSN-2448-7236. SEPTIEMBRE 2023 – AGOSTO 2024. IC-01, pág.: 1 a la 5.

utilizan como entradas al modelo acústico para realizar la clasificación y el reconocimiento del habla.

Durante la etapa de reconocimiento, el modelo acústico compara las características extraídas de la señal de audio de entrada con las características aprendidas durante el entrenamiento. Utilizando algoritmos de clasificación, el modelo acústico asigna probabilidades a cada unidad de sonido posible (fonemas o subfonemas) para cada tramo de la señal de audio. De esta manera, el modelo acústico busca determinar qué unidades de sonido son más probables de estar presentes en la señal de entrada.

El modelo acústico es solo una parte de un sistema de reconocimiento de voz completo, que también incluye otros componentes como el modelo de lenguaje y el decodificador. El modelo acústico y estos componentes trabajan en conjunto para convertir la señal de audio en texto reconocido, generando la salida final del sistema de reconocimiento de voz.

**Modelo de Lenguaje**

En la última etapa es el modelo del lenguaje, en donde el modelo acústico se encarga de asignar correctamente el audio a las palabras, este modelo es el cual permitirá diferenciarlos entre sí. Para lograr esta clasificación adecuada se decidió utilizar redes neuronales artificiales (RNA), Las cuales son una técnica conocida para el reconocimiento de patrones .

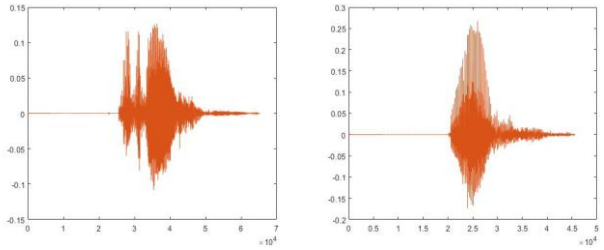
Las redes neuronales funcionan de la siguiente manera:

- **Entrada de datos**: Los datos se introducen en la red neuronal a través de una capa de entrada. Estos datos pueden ser imágenes, texto, señales de audio u otros tipos de información, en nuestro caso son de audio.
- **Propagación hacia adelante (Forward propagation)**: Los datos se propagan a través de la red neuronal desde la capa de entrada hacia las capas ocultas y, finalmente, hacia la capa de salida. Cada nodo en una capa oculta toma las entradas de los nodos de la capa anterior, realiza un cálculo utilizando una función de activación y envía la salida a los nodos de la capa siguiente.
- **Pesos y sesgos (Weights and biases)**: Cada conexión entre los nodos en la red neuronal tiene un peso asociado. Los pesos determinan la influencia que una entrada tiene en la salida de un nodo. Además, cada nodo puede tener un sesgo (bias) asociado, que es un valor constante que se suma a la entrada ponderada antes de aplicar la función de activación.
- **Función de activación**: Cada nodo utiliza una función de activación para introducir no linealidad en la red neuronal. Ejemplos de funciones de activación comunes son la función sigmoide, la función ReLU (Rectified Linear Unit) y la función tangente hiperbólica.
- **Capa de salida y función de pérdida (Loss function)**: La capa de salida de la red neuronal produce los resultados finales. Dependiendo del tipo de problema, se utiliza una función de pérdida apropiada para calcular la diferencia entre las salidas de la red y los valores esperados. El objetivo es minimizar esta función de pérdida durante el entrenamiento.
- **Retropropagación (Backpropagation)**: La retropropagación es el proceso de ajuste de los pesos y sesgos de la red neuronal para reducir el error de salida. Utiliza el algoritmo de descenso de gradiente para calcular las derivadas parciales de la función de pérdida con respecto a los pesos y sesgos de la red. Estas derivadas se utilizan para actualizar los pesos y sesgos en la dirección que reduce el error.
- **Iteración y entrenamiento**: El proceso de propagación hacia adelante, retropropagación y ajuste de pesos se repite durante múltiples iteraciones o épocas hasta que la red neuronal logre un nivel de precisión o rendimiento deseado en los datos de entrenamiento.

Una vez que la red neuronal ha sido entrenada, puede utilizarse para hacer predicciones o clasificar nuevos datos que no se utilizaron durante el entrenamiento.

**Resultados**

Hasta el momento el sistema de reconocimiento se encuentra en las etapas iniciales, esto incluye la adquisición de señales y su mejoramiento para la extracción de características- tal como se muestra en la figura 4.



(a) (b)

**Figura 4. Adquisición de las señales de muestra. En (a) la palabra ABRIR. En (b) la letra A.**

**Agradecimientos.**

Este proyecto fue realizado gracias al apoyo de PIAPIME PE105123 Y PAPIIT IA102323

AÑO 8. No. 8. ISSN-2448-7236. SEPTIEMBRE 2023 – AGOSTO 2024. IC-01, pág.: 1 a la 5.

**Conclusiones y trabajo futuro.**

Los asistentes de voz son herramientas que nos ayudan a la inclusión social de personas con discapacidad, su principio de funcionamiento se basa en el reconocimiento de patrones auditivos complejos que pueden representarse mediante fonogramas y un ambiente lingüístico adecuado, como trabajo a futuro se deberá de realizar la extracción de características, la separación de fonogramas y el entrenamiento del modelo para su implementación.

**Referencias**

1. INEGI, «Discapacidad en México,» [En línea]. Available: https://cuentame.inegi.org.mx/poblacion/discapacidad.aspx. [Último acceso: 30 AGOSTO 2022].
2. S.A., [En línea]. Available: https://1library.co/document/zp2066vy-procesamiento-voz-tiempo-real-empleando-procesador-digital-senales.html. [Último acceso: 14 Agosto 2022].
3. CEDOM, «Asociación Española de Domótica e Inmótica - CEDOM,» [En línea]. Available: http://www.cedom.es/sobre-domotica/que-es-domotica. [Último acceso: 21 agosto 2022].
4. GACETA UNAM, «La discapacidad en México. Una situación que nos compete a todos,» 27 SEPTIEMBRE 2021. [En línea]. Available: https://www.gaceta.unam.mx/la-discapacidad-en-mexico-una-situacion-que-nos-compete-a-todos/. [Último acceso: 21 JUNIO 2022].
5. Fuentes Euan, Jonathan. (2022). Desarrollo de aplicaciones domóticas en MATLAB App Designer, implementadas en una tarjeta de desarrollo Arduino. Universidad Autónoma de México, Facultad de Estudios Superiores Cuautitlán.
6. Towards Data Science, «Automatic Speech Recognition: Breaking Down Components of Speech,» 2021.

## INFORMACIÓN ACADÉMICA

**Tania Abigail Lira Baca**. Estudiante de la carrera de ingeniería en Telecomunicaciones, Sistemas y Electrónica, de la Facultad de Estudios Superiores Cuautitlán de la UNAM, sus intereses en sistemas de control robóticos , sistemas inteligentes.

**Fernando Gudiño Peñaloza.** Ingeniero Mecánico Electricista egresado de la Facultad de Estudios Superiores Cuautitlán de la UNAM, Obtuvo el título de Maestro en Ciencias de la computación y el Doctorado en Ciencias Computacionales por el Tecnológico de Monterrey. Actualmente profesor del departamento de ingeniería de la Fes Cuautitlán, sus intereses se