

Estudios Preliminares para un Traductor de Lenguaje Subvocal

F. J. Luis Juan Barragán^{#1}, V. G. Villanueva^{#2}, J. C. Chávez Novoa^{#3}, M. A. Delgado López^{#4}

[#]Tecnológico Nacional de México: Departamento de Ingeniería en Sistemas Computacionales, Instituto Tecnológico José Mario Molina Pasquel y Henríquez, Unidad Académica Chapala, Jalisco, México

¹francisco.luisjuan@chapala.tecmm.edu.mx, ²victor.gonzalez@chapala.tecmm.edu.mx, ³julio.chavez@chapala.tecmm.edu.mx, ⁴miguel.delgado@chapala.tecmm.edu.mx

Resumen— Este trabajo presenta avance del proyecto con procesamiento digital de señales, el cual genera las bases de construcción de un dispositivo para adquirir una señal proveniente de las cuerdas vocales del cuerpo humano, porque a través de ellas se puede obtener una gran fuente de información para la generación de una interfaz para el traductor del lenguaje subvocal, teniendo una gran relevancia en el ámbito del idioma español. La adquisición se da a través de un micrófono conectado a un amplificador de muy alta ganancia para detectar las vibraciones de las cuerdas vocales y así empezar a generar el algoritmo de caracterización en base al dominio de la frecuencia utilizando la transformada rápida de Fourier y con esto comparar los patrones predefinidos para determinar la palabra expresada.

Palabras clave—algoritmo, amplitud, cuerdas vocales, fonema, traductor

I. INTRODUCCIÓN

La sociedad debe prestar especial atención a la población que padece de discapacidades físicas o mentales. En este contexto el Instituto Nacional de Estadística y Geografía ha incorporado en los censos generales de población preguntas acerca de este segmento con la finalidad de generar información que coadyuve a generar conocimiento que permita su tratamiento, rehabilitación, e incorporación a la vida social del país. Así, en México al año 2020 hay 6, 179, 890 personas con algún tipo de discapacidad física o mental, o con un problema de salud considerable, lo que representa 4.9% de la población total. [4]

En el estado de Jalisco viven 595,000 personas discapacitadas. De la población con discapacidad, el 48% presenta limitaciones relacionadas con sus piernas o brazos, 15% habla o comunicación. [4]

Las personas que padecen la discapacidad de la falta del habla se puede dar por diferentes causas, como lo indica el siguiente investigador, “*el cerebro manda la señal a la laringe, puede estar dañada o no la tiene. El habla se genera a través del cerebro que es el centro de control del habla cuando nos comunicamos, el cerebro le transmite a la boca lo que tiene que decir, en definitiva, el habla es codificación de los pensamientos que llegan a nuestras cuerdas vocales a través de las señales nerviosas que se convierten así en palabras*”. [1]

Desde el año 1999 la NASA (por sus siglas en inglés, National Aeronautics and Space Administration) dio inicio con una serie de investigaciones que permiten trabajar en un

programa para desarrollar el sistema de habla subvocal, o habla interna, esto enfocado en el diseño de una forma de comunicación que pudiera ser utilizada durante las caminatas y construcción de las estaciones en el espacio.

La NASA ha desarrollado una tecnología que es capaz de transformar en palabras el habla subvocal que una persona emplea cuando lee en silencio o cuando habla consigo misma [2]. También se ha estado trabajando con señales electromiográficas para prótesis de músculos que se encontrarían en la mano o brazo de una persona.

Actualmente, en el Instituto Tecnológico José Mario Molina Pasquel y Henríquez Unidad Académica Chapala, se trabaja en el desarrollo de sistemas para personas discapacitadas con problemas de audición y de lenguaje que actualmente se comunican mediante el uso del Lenguaje Mexicano de Señas. Las líneas de investigación seguidas son el procesamiento del lenguaje natural y el reconocimiento de señales.

En el área de procesamiento de señales se está trabajando en el reconocimiento de señales electromiográficas (EMG) y a través de detección de vibraciones, aplicadas al discurso subvocal (esto es, cuando el hablante no emite sonido, por ejemplo, al leer en voz baja). En este momento se cuenta con un sistema de adquisición de señales EMG, sin embargo, en la actualidad existen aplicaciones dirigidas a proyectos de investigación específicos no para la población en general sino para cierto grupo y actividades en particular, como por ejemplo el ejército, proyectos aeroespaciales y todo se encuentra enfocado en el idioma del inglés, además de que aún no se tiene una aplicación práctica debido a la falta de etapas de procesamiento.

Por lo cual realizando un dispositivo capaz de interpretar las vibraciones que generan las cuerdas vocales de una persona con discapacidad del habla podrá comunicarse en el entorno en que interactúa habitualmente.

El objetivo del proyecto de investigación es desarrollar un dispositivo que detecte las señales emitidas por las cuerdas vocales para la identificación de fonemas en individuos silentes a través de un algoritmo de caracterización que genere como resultados los descriptores de palabras.

II. MATERIALES Y MÉTODOS

El trabajo de investigación que actualmente se está realizando en las instalaciones del tecnológico es del tipo aplicada, porque se está buscando generar conocimiento con

una aplicación directamente de un problema actual que afecta a un porcentaje de la población de nuestro país, esto basado fundamentalmente en hallazgos tecnológicos de una investigación básica que genera un enlace entre la teoría y el producto al que se quiere llegar. Con un apoyo de la investigación o metodología cuantitativa usando datos numéricos que se podrán tratar mediante herramientas estadísticas en la identificación de patrones.

En la investigación se usa el método en primera instancia de la observación para una percepción directa que nos permita conocer la realidad de las frecuencias adquiridas y su comportamiento para generar el algoritmo de identificación de patrones, en el transcurso puede convertirse en un procedimiento propio del método utilizado en la comprobación de la hipótesis, ya que se hará una recolección de información de cada una de las variables.

La metodología propuesta para la adquisición de las señales emitidas por las cuerdas vocales es desarrollar un dispositivo que se conectará en la parte media del cuello de una persona con discapacidad del habla a la altura de la laringe, para generar una secuencia de recolección de vibraciones a través de un sensor de movimientos, posteriormente este paso entrará a un proceso de análisis de la misma, la cual lleva la implementación de amplificación y filtros para la limpieza de los datos obtenidos, finalmente entrará a un proceso de detección de patrones a través de un algoritmo de identificación de las ondas adquiridas de fonemas generados a través del cerebro y enviadas a las cuerdas vocales para poder identificarlas y dar como resultado un dispositivo que traduzca los pensamientos en sonidos. En la Figura 1 se muestra el diagrama del proceso a tomar en cuenta en el proyecto.

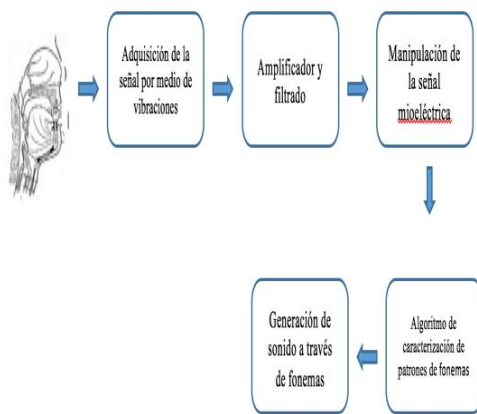


Fig. 1.- Diagrama a bloques Traductor Lenguaje Subvocal.

La metodología propuesta para la adquisición de las señales electromiográficas superficiales sigue el gráfico de la Figura 2. La primera etapa consiste en la adquisición de las señales EMG mediante el uso de un micrófono con un sensor integrado que identificará las señales y almacenará en una base de datos para poder analizarlas.

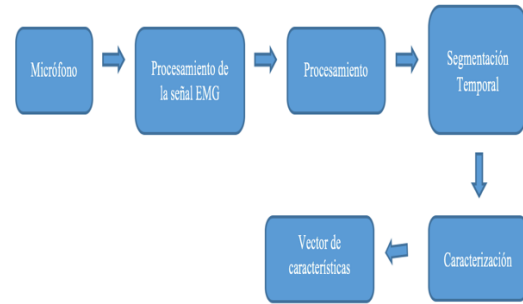


Fig. 2. Metodología para la caracterización de señales electromiográficas.

Las señales adquiridas tendrán como origen individuos de género masculino y femenino, abarcando un rango de edades de 17 a 45 años, la adquisición se llevará a cabo en un ambiente controlado en diversos instantes del día (mañana, medio día y en la tarde), para la adquisición se pedirá a los individuos que pronuncien de manera subvocal cada una de las vocales: a, e, i, o y u, además de fonemas como naranja, hola, uno, en una primera etapa experimental de identificación de los primeros patrones en base al algoritmo (Figura 3).

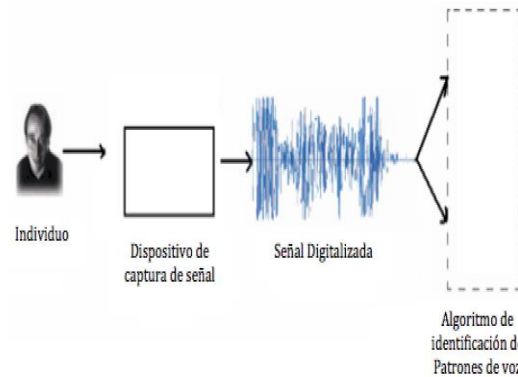


Fig. 3. Proceso de adquisición y procesamiento de la señal adquirida.

El siguiente paso es el preprocesamiento de la señal en el cual se aplicará un conjunto de filtros y transformaciones con el fin de acondicionar la señal que posteriormente será procesada. Una vez que se tenga la señal acondicionada se aplicará una segmentación temporal donde se normalizan y se recortan para tener registros de longitud estándar en el tiempo. Este procedimiento se realizará buscando el canal donde se presenta mayor energía y ubicando el punto inicial del movimiento como inicio de la venta. Una vez ubicado este punto se recortarán simultáneamente todos los canales.

Posteriormente se tiene que realizar el proceso de caracterización, utilizando su propio algoritmo y un conjunto de descriptores, por ejemplo, la transformada wavelet, la transformada Hilbert-Huang, y la transformada rápida de Fourier. El resultado será un conjunto de vectores característicos para cada vocal, lo cual permitirá construir un

plano multidimensional de patrones identificados para cada fonema.

Para la realización de estos procedimientos se llevará a cabo una investigación de campo y documental sobre los temas que se requieren para el desarrollo idóneo del proyecto.

A las personas que apoyen como voluntarios del proyecto se aplicará la técnica de prueba, es decir, pruebas del dispositivo y software para identificar si el sensor está trabajando de manera correcta y detectando las señales correspondientes que son necesarias para el proyecto, para dar paso al instrumento de medición de los datos a procesar y haciendo la evaluación adecuada que identifica el rango de valores de los fonemas detectados en los individuos.

III. RESULTADOS

El resultado que se ha logrado es la creación del dispositivo para detectar las señales electromiográficas a partir de las vibraciones que producen las cuerdas vocales, esta señal se va adquiriendo a través de un micrófono conectado a un amplificador de muy alta ganancia para detectar el tipo de señal que se requiere y así poder trabajar con su análisis y el proceso del algoritmo.

La primera versión del prototipo generado por la investigación y propuesto para este proyecto se trabajo con un sensor electromiográfico y electrodos para capturar la señal EMG generadas en la zona de la laringe del cuerpo humano, diseñando el software que recibía la señal para poder realizar el análisis de filtrado y eliminación del ruido con esto se pretendía una señal limpia que se ubicará entre los valores aproximados del rango entre la frecuencia y amplitud de una cierta palabra y así poder generar un patrón del fonema a través del algoritmo de caracterización, en la Figura 4 se muestra el diagrama del primer prototipo.

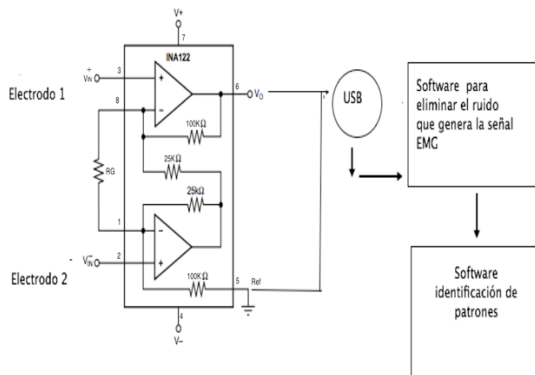


Fig. 4. Diagrama del primer prototipo del Traductor de Lenguaje Subvocal.

Este prototipo presentó el problema de no obtener de manera correcta la detección de señales por la baja intensidad de estas, lo que provocó que no llegara de manera ideal al proceso de eliminación del ruido y filtrado por la inestabilidad en los valores que se registraban en dispositivo, además de que los electrodos no eran los adecuados porque sólo se podían

estar utilizando una sola vez por ser desechables. En la Figura 5 se muestra la versión del dispositivo.

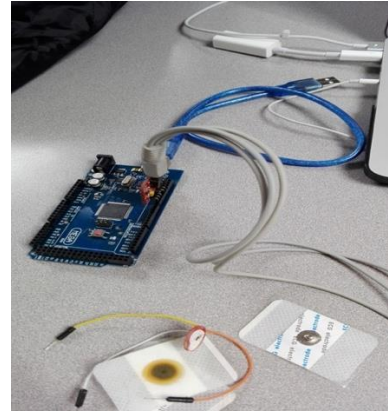


Fig. 5. Primera versión del prototipo.

Esta señal es capturada a través de un micrófono diseñado para detectar vibraciones de bajo nivel generadas por las cuerdas vocales a través de un microcontrolador Arduino Mega (Figura 6) con un puerto analógico (ADC), a una frecuencia específica, la cual convierte una señal analógica a una de tipo digital y está a su vez es enviada a la computadora a través de un medio de comunicación inalámbrico para su procesamiento. La interfaz que se diseño para recibir la señal de manera continúa procesa los datos obtenidos para así poder determinar las palabras que realizan grupos de captura en base a la amplitud de la señal, una vez obtenido un grupo, ese se procesa y se convierte al dominio de la frecuencia utilizando la Transformada Rápida de Fourier (FFT) (1) [3], teniendo una frecuencia máxima que no sobrepase la frecuencia de la voz que oscila 1KHz. La señal en el dominio de la frecuencia es analizada para determinar las frecuencias base del grupo, las cuales se comparan con los patrones predefinidos para determinar la palabra.

$$\left. \begin{aligned} X(k) &= \sum_{j=1}^N x(j)\omega_N^{(j-1)(k-1)} \\ x(j) &= \frac{1}{N} \sum_{k=1}^N x(k)\omega_N^{-(j-1)(k-1)} \end{aligned} \right\} \omega_N = \exp\left(-\frac{2\pi i}{N}\right) \quad (1)$$

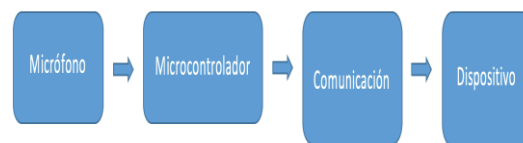


Fig. 6. Estructura del micrófono conectado al dispositivo.

Se realizó una prueba implementando el dispositivo a un joven de 18 años de edad que tiene problemas del habla, el proceso fue que se le ubicó el micrófono en la parte de la laringe donde se encuentra las cuerdas vocales para dar comienzo el proceso de adquisición de las señales y empezar

a recibir la información que es requerida, al momento de dar inicio el software que se liga al dispositivo empezó a recibir toda la información adquirida por las vibraciones que las cuerdas vocales generaban al momento que el joven trataba de pronunciar una palabra indicada, en este proceso se amplificó la señal para poder fortalecer las señales eléctricas nerviosas, a partir de ahí se procesan limpiando la señal y así obtener las partes que son realmente útiles para el proyecto y así pasar a la siguiente etapa del algoritmo de caracterización de patrones. En la Figura 7 se muestra la imagen donde se localiza el joven con el dispositivo colocado en la parte del cuello correspondiente.



Fig. 7. Prueba del dispositivo con una persona que tiene problemas del habla.

A partir del procesamiento que se realiza para limpiar la señal se enviaron los datos a un dispositivo móvil esto permitió observar los valores que estaban generando debido al proceso de adquisición, esto ayudará a identificar un patrón de cierto fonema indicado, como se muestra en la Figura 8.

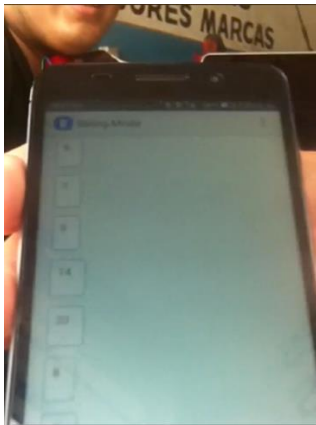


Fig. 8. Valores proporcionados por el dispositivo.

En el proceso de identificación de patrones se implementa en una fase del algoritmo de caracterización, en el cual utiliza procesos inteligentes para poder asignar los valores correspondientes seguido de la presentación del sonido claro.

En la Figura 9 y 10 se muestra como el software de adquisición de señales desarrollado en MATLAB va

recibiendo los datos analógicos y haciendo la conversión a digital, en la primera parte se muestra las ondas originales con rangos en amplitud y tiempo de cada onda generada por los sonidos capturados de las vibraciones de la persona, en la segunda gráfica se muestra la onda ya aplicando la FFT, donde ya se alcanza a percibir información útil para el proceso de identificación de patrones de fonemas, así mismo se observa el comportamiento de la señal en amplitud y frecuencia estos valores se guardan en arreglos para llevar un registro de cada onda y así con esta información obtenida se implementa el algoritmo para poder detectar la palabra que se le indicó al joven que trataba de pronunciar.

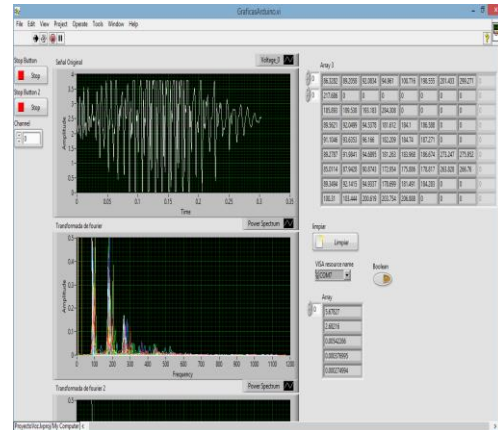


Fig. 9. Interfaz de procesamiento de la señal EMG.

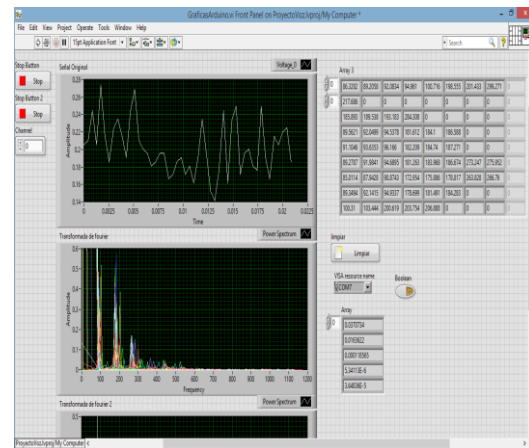


Fig. 10. Interfaz de procesamiento de la señal EMG.

En las figuras anteriores se muestran filas del arreglo que corresponden a las frecuencias de amplitud máxima de un grupo de datos obtenidos que forman una palabra, donde cada palabra nueva se va agregando en una fila diferente del arreglo, para poder identificar los patrones de frecuencia en base a un rango generado por las señales. En la Figura 9 se muestra las ondas de la palabra de prueba “hola” y en la Figura 10 se observan las ondas generadas de la palabra “uno”, en esta parte se elimina el ruido que no se necesita en las distintas etapas del proceso con esto se pretende dejar la señal requerida para su análisis e interpretación de los datos.

En las pruebas que se presentan intervienen dos palabras en particular, con el objetivo de analizar su comportamiento en cuestión de los valores y ver cómo se comporta el algoritmo de caracterización con los datos que se están introduciendo y así poder identificar el patrón de comportamiento de la palabra indicada, este proceso se contempla como fundamental en el proyecto, en esta etapa del proyecto se trabaja realizando las modificaciones correspondientes por la complejidad del mismo.

Las señales obtenidas de las regiones musculares del cuello en dirección de la laringe se identificó diferentes unidades motrices laríngeas las cuales son bastantes pequeñas y oscilan entre los 215 a 340 microvoltios con una duración media de 3.1 a 4.7 milisegundos. Esto está dado por una causa natural del cuerpo humano, porque este genera una resistencia eléctrica natural de la piel, lo que indicó que en las pruebas necesitábamos usar un filtraje y amplificación de los datos más eficiente.

IV. CONCLUSIÓN

Los resultados que se llevan hasta el momento han demostrado ser confiables y satisfactorios dentro de las primeras pruebas seleccionadas para las necesidades actuales del proyecto y especialmente a los requerimientos del joven que interactuó con el dispositivo, donde se obtuvieron datos a procesar para la construcción de fonemas a través de la identificación de patrones. Se llevaron a cabo 50 muestras con el joven con problemas del habla y otras 50 con una persona que no tiene el problema, de estos se logró identificar un 60% de aciertos comparando las dos muestras, dando efectividad en el algoritmo, pero se tiene que mejorar más para que el porcentaje de aciertos pueda ser más fiable con respecto a las pruebas que se llevaron a cabo.

La interfaz visual mostró ser capaz de leer y almacenar todos los valores que llegan por parte del dispositivo con datos libre de ruido y con calidad. El tiempo de llegada de los datos son en milisegundos frente al número de señales adquiridas

durante la generación de las vibraciones hechas por las cuerdas vocales, esto dando como resultado que se esté almacenando la información en los arreglos correspondientes del software para poder generar las graficas y así identificar los patrones del fonema a buscar en base a las ondas mostradas en pantalla.

El alcance hasta el momento de la investigación realizada ha sido satisfactorio por los resultados obtenidos hasta a la fecha, se requiere seguir trabajando en la mejora de los algoritmos y así identificar los patrones de los fonemas para llegar a un resultado eficiente para el proyecto.

AGRADECIMIENTOS

Este artículo fue desarrollado gracias al trabajo de los profesores investigadores y a la cooperación de personas e instituciones que ayudaron a las investigaciones de esta fase del proyecto.

REFERENCIAS

- [1] M. Callahan. (2014) Proyecto Aurea. [En línea]. Disponible en: <http://iipdigital.usembassy.gov/st/spanish/article/2008/01/20080102183242f10.8968622.html#axzz4S4R9Rz4p>
- [2] C. Martin. (2006) Novedades científicas. [En línea]. Disponible en: <https://www.novaciencia.com/2006/10/29/>.
- [3] A. F. Garcia. (2016) MATLAB Análisis de los datos. [En línea]. Disponible en: http://www.sc.ehu.es/sbweb/fisica3/datos/fourier/fourier_1.html.
- [4] INEGI. (2020) Cuéntame. [En línea]. Disponible en: <https://cuentame.inegi.org.mx/poblacion/discapacidad.aspx>.
- [5] J. A. Gutiérrez Calderón, E. N. Gama Melo, D. Amaya Hurtado and O. F. Avilés Sánchez, "Desarrollo de interfaces para la detección del habla sub-vocal", *Tecnura*, 17(37), pp.138-152, 2013.
- [6] L.J. Hargrove, K. Englehart, and B. Hudgins, "A comparison of surface and intramuscular myoelectric signal classification", *IEEE transactions on biomedical engineering*, 54(5), pp.847-853, 2007.