Congr. Int. en Ing. Electrónica. Mem. ELECTRO, Vol. 39, pp. 282-287, Oct 2017, Chihuahua, Chih. México
http://electro.itchihuahua.edu.mx/memorias_electro/MemoriaElectro2017.zip
ISSN 1405-2172

SISTEMA GESTUAL DE GENERACION DE VOZ PARA PERSONAS CON DISCAPACIDAD FONOLÓGICA

Ríos Ortiz, Brian Alejandro; Márquez Gutiérrez, Pedro Rafael
Instituto Tecnológico de Chihuahua
División de estudios de Posgrado e Investigación del Instituto Tecnológico de Chihuahua
Ave. Tecnológico #2909, Chihuahua, Chihuahua. México, CP 31310
+52 (614) 2-01-2000, Fax +52 (614) 4-13 5187
e-mail: {barios, pmarquez}@itchihuahua.edu.mx

RESUMEN.

Se presenta el diseño e implementación de un sistema para personas con trastornos del habla y lenguaje que les permite comunicarse con las personas mediante un lenguaje de señas especiales. El sistema está basado en el código Morse sin las restricciones estrictas de tiempo entre símbolos. El sistema incorpora una interfaz de usuario mediante la cual es posible modificar algunas propiedades del generador de voz, más algunas ayudas visuales.

Palabras Clave: Discapacidad de habla, Código Morse, control gestual.

ABSTRACT.

We present the design and implementation of a system for people with speech and language disorders that allows them to communicate with people through a special sign language. This system is based on Morse code without strict time restrictions between symbols. The system incorporates a user interface through which it is possible to modify some properties of the voice generator, plus some visual aids.

Keywords: Speech impairment, Morse code, Signal control

1. INTRODUCCIÓN

La comunicación del humano con su entorno se inicia con la propia vida. Desde antes de nacer comenzamos a captar mensajes mediante sonidos acerca de nosotros mismos y de lo que nos rodea. Existen estudios acerca del beneficio que proporciona la música al bebé durante el embarazo. De hecho, se cree que la audición es el primer sentido en despertar. Así, la comunicación auditiva es básica en la vida del hombre a lo largo de todas sus etapas [1]. La comunicación no verbal integra un conjunto de códigos que intervienen en nuestra vida diaria, emitiendo de forma permanente signos que inciden en la percepción de la gente.

Ahora, como nunca antes, la comunicación es fundamental para el desarrollo de las personas y es algo que se les dificulta enormemente a las personas con alguna discapacidad para hablar, ya que la sociedad no está adaptada ni preparada para establecer la comunicación fluida con estas personas.

Si bien es cierto que las personas con discapacidad del habla pueden comunicarse con su entorno directo o similares por medio del lenguaje de señas, el resto de la sociedad desconoce completamente este lenguaje.

Existen algunas aplicaciones diseñadas para convertir texto escrito en voz, pero no se han encontrado aplicaciones que acepten otro método de entrada que no sea un teclado, lo cual dificulta su uso extendido. Si, además, la persona tuviera alguna discapacidad motora, como la gente mayor, el uso de un teclado queda, aún más, descartado.

2. MARCO TEORICO

2.1. DISCAPACIDAD FONOLOGICA O MUTISMO

El concepto de mutismo [4] tiene su origen en el latín *mutus*, término que hace referencia a la mudez, es decir a la suspensión del habla por parte de un individuo. Las personas que sufren de mutismo se mantienen en silencio ante determinadas situaciones; las causas de este silencio pueden ser varias y activarse de forma voluntaria o como consecuencia de una condición que escapa a su control.

Las principales causas del mutismo son físicas y suelen estar vinculadas a las cuerdas vocales, la lengua, la boca, la garganta o los pulmones. En algunos casos, el mutismo está vinculado con la sordera: quien ha nacido sordo, nunca oyó hablar y, por lo tanto, no aprende a desarrollar esta capacidad. Una persona, de todas maneras, puede contraer el mutismo en cualquier momento de su vida ante una enfermedad o un accidente.

Muchas veces el mutismo se define a través del concepto de afonía, que es el concepto médico que se emplea respecto a la falta de capacidad para el habla. Un motivo frecuente de afonía es que se produzca, ya sea por una cirugía, un tumor o un accidente, un daño en el nervio laríngeo recurrente que se encarga de controlar la mayoría de los músculos que se hallan en la laringe.

2.2. EMG

Las señales electromiográficas (EMG) [5,7], pueden ser medidas utilizando elementos conductivos o electrodos sobre las superficie de la piel, o de manera invasiva sobre el músculo utilizando agujas [11].

La Electromiografía de superficie [8,13] es el método más común de medida, puesto que es no invasiva y puede ser

realizada con un mínimo de riesgo sobre la persona que la va a usar [9,10].
La amplitud de las señales EMG varía desde los µV hasta un bajo rango de mV (menor de 10mV). La amplitud, y las propiedades de las señales EMG tanto en el dominio del tiempo como en la frecuencia dependen de factores tales como:

- El tiempo y la intensidad de la contracción muscular.
- La distancia entre el electrodo y la zona de actividad muscular.
- Las propiedades de la piel (por ejemplo el espesor de la piel y tejido adiposo).
- Las propiedades del electrodo y el amplificador.
- La calidad del contacto entre la piel y el electrodo.[10]

3. ARQUITECTURA BÁSICA DEL SISTEMA

El Sistema de Control Gestual [14,15] para Generación de Voz se compone de los siguientes módulos:

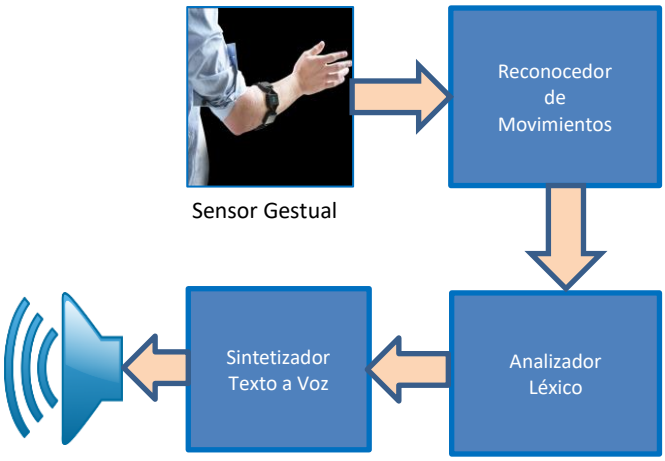


Figura 1. Arquitectura básica del sistema

3.1. Sensor Gestual

Para facilitar el uso del sistema a una población objetivo mayor, se utiliza un brazalete de control gestual, como el mostrado en la figura 2. Esto permite una forma de entrada de movimientos gruesos más fáciles de realizar y detectar. Por la misma razón, se eligió el Código Morse como base del lenguaje gestual, ya que es binario y suficientemente simple de ejecutar y reconocer.

Figura 2. Brazalete de control gestual (Myo Armband)

3.2. Reconocedor de Movimientos

Este módulo es parte integral del sensor gestual y permite reconocer un conjunto básico de movimiento del brazo y la mano [12, 14]. Dado que el código base es binario, se experimentó con diferentes movimientos de mano para seleccionar aquellos que fueran altamente diferenciables.

3.3. Analizador Léxico

El analizador léxico [15], recibe como entrada la secuencia de caracteres suminsitrada por el reconocedor gestual y produce como salida una sucesión de componentes léxicos o *tokens*. Estos *tokens* sirven como base de la etapa posterior del proceso de generación de voz, siendo la entrada para el analizador sintáctico, el cual se implementará en una etapa posterior.
La especificación de un lenguaje requiere de un conjunto de reglas que definen el léxico. Estas reglas consisten comúnmente en expresiones regulares que indican el conjunto de posibles secuencias de caracteres que definen un token o lexema en el caso de los lenguajes de programación. En la etapa actual de implemengtación, el análisis léxico no hace uso de una gramática, pero en el afán de aumentar la expresibilidad del lenguaje gestual, se prevee la definición de una gramática más rica y expresiva.
Este módulo se basa en una máquina de estados finitos, la cual contiene la información de las posibles secuencias de símbolos que conforman los *tokens* que forman parte del subconjunto del lenguaje generado por el sistema.

3.4. Analizador Sintáctico

Los algoritmos de análisis de lenguajes naturales no se pueden basar en gramáticas que tengan las buenas características como las gramáticas diseñadas para los lenguajes de programación. Algunos formalismos gramaticales son muy difíciles de analizar computacionalmente, por lo que, en general se usa una aproximación libre de contexto incluso si la estructura en sí no es libre de contexto con el fin de obtener una primera simplificación.
Los algoritmos que usan gramáticas libres de contexto [15] se suelen basar en alguna variante del algoritmo Cocke-Younger-Kasami (CYK) y heurísticas para la poda de análisis infructuosos. En todo caso algunos enfoques sacrifican la velocidad por la precisión usando, por ejemplo, versiones

lineales del algoritmo «shift-reduce». Enfoques recientemente desarrollados utilizan un algoritmo que genera a partir de múltiples análisis y otro que escoge la mejor opción.

3.5. Sintetizador Texto a Voz

La conversión texto-voz [24] es la generación por medios automáticos de una voz artificial que genera un sonido similar al producido por una persona al leer un texto cualquiera en voz alta. Es decir, permite la conversión de textos en voz sintética. Los conversores de texto-voz son conocidos también con las siglas CTV o por las siglas en inglés TTS (Text To Speech).

Los conversores CTV/TTS deben producir una voz sintética que resulte natural y sea inteligible. La síntesis del habla ha de ser completamente automática, sin tener que introducir algún tipo de reajuste manual en alguna parte del proceso. Además, el texto introducido al sistema debe poder ser arbitrario.

En la primera fase del sintetizador se realiza una representación lingüística simbólica como se describe a continuación. Primero se realiza una normalización del texto que convierte el texto en una forma textual canónica. Esto afecta principalmente a las cifras, abreviaturas, etc. Segundo, se realiza una conversión fonética. Al texto normalizado se le asignan transcripciones fonéticas a cada palabra. Tercero, se realiza la división prosódica en unidades sintagmáticas, proposiciones y frases. Finalmente, se toma la representación fonética y se transforma en voz sintética.

3.6. Interfaz de Configuración

El sistema posee una interfaz de usuario intuitiva, Figura 7, que permite la modificación y ajuste de diferentes opciones de operación del sistema, como, por ejemplo, el volumen de la pronunciación y la entonación de palabras y frases. También presta diversas ayudas sobre cómo utilizar correctamente el sistema y proporciona información acerca del estado del dispositivo, como el nivel de la batería o el estado de la conexión inalámbrica.

4. ASPECTOS DE IMPLEMENTACIÓN

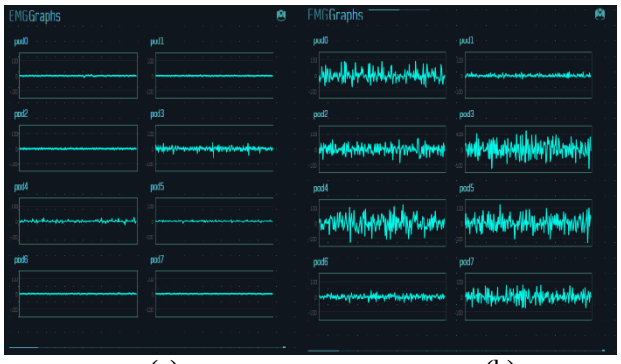
3.1 Descripción Funcional del Sistema

La plataforma de hardware [23], sobre la cual se ha implementado el primer prototipo del sistema es el kit Myo, mostrado en la figura 3. Myo es un brazalete capaz de recoger los impulsos eléctricos generados por nuestro brazo al realizar movimientos del antebrazo y la mano. Además, registra algunos otros datos como lo son orientación y la aceleración.

Figura 3. Contenido del kit Myo

Cuenta, además, con un SDK[1] que el fabricante Thalmic Labs [2] pone a disposición de manera gratuita para el desarrollo de aplicaciones. Estas aplicaciones pueden ser programadas en cualquier entorno integrado de desarrollo (IDE), por lo que se eligió el Microsoft Visual Studio. El SDK permite tomar lectura de la señal eléctrica generada por el brazo y censada por el brazalete cada vez que los diferentes músculos del brazo se alteran al hacer una pose con nuestra mano. Estas posiciones representan la codificación gestual de la clave Morse.

El sensor gestual utiliza una técnica conocida como electromiografía, la cual es el registro gráfico de la actividad eléctrica producida por los músculos esqueléticos [5]. Esta actividad eléctrica es conocida como el electromiograma o EMG y puede ser monitoreado a través de electrodos insertados dentro de los músculos (electrodos intramusculares) o a través de electrodos en la superficie de la piel sobre el músculo (electrodos superficiales) [6][7], como es el caso del Myo.



(a) (b)

Figura 4. a) Actividad del sensor cuando el brazo está en reposo. b) Actividad del sensor cuando se realiza algún movimiento con la mano.

[1] Software Development Kit

El rango típico de repetición de una unidad motora muscular es de alrededor 7–20 Hz dependiendo del tamaño del músculo [7, 8]. El daño a las unidades esperadas puede ser entre rangos de 450 y 780 mV [13].

El brazalete lee las señales eléctricas de los músculos del antebrazo, y los asocia a los gestos realizados por la mano. La figura 4 muestra la señal (a) cuando el brazo esta en reposo, y (b) cuando se realiza alguna actividad.

El código Morse básico, mostrado en la figura 5, está conformado por signos temporales de distinta duración, denominados puntos y rayas [3]. Generalmente, una raya equivale a tres veces el tiempo de un punto. La separación entre palabras (espacio en blanco) es aproximadamente 3 veces la duración de una raya. Sin embargo, con el fin de optimizar el tiempo de respuesta del sistema, se optó por seguir únicamente la regla de duración de los puntos y las rayas, y separar caracteres y palabras empleando otros gestos preestablecidos que el dispositivo reconoce.

Alfabeto				Números	
A	.-	N	-.	1	.----
B	-...	O	---	2	..---
C	-.-.	P	.--.	3	...--
CH	----	Q	--.-	4	....-
D	-..	R	.-.	5	.....
E	.	S	...	6	-....
F	..-.	T	-	7	--...
G	--.	U	..-	8	---..
H	....	V	...-	9	----.
I	..	W	.--	0	-----
J	.---	X	-..-		
K	-.-	Y	-.--		
L	.-..	Z	--..		
M	--				

Figura 5. Alfabeto morse.

El dispositivo es capaz de reconocer 5 gestos básicos preestablecidos (Figura 6) que se utilizan para generar distintos comandos del sistema. El gesto de *puño cerrado* se emplea como el gesto emisor de puntos y rayas correspondiente a la clave Morse. La *palma adentro* es el gesto empleado para separar las letras de cada palabra La *palma afuera* es el gesto que introduce un espacio en blanco entre cada palabra y, por último, la *mano abierta* convierte el texto en voz y lo reproduce en los altavoces del sistema.

1) 2) 3) 4) 5)

Figura 6. Gestos Reconocidos por el brazalete. 1) Doble toque 2) Puño 3) Mano abierta 4) Palma adentro 5) Palma afuera

El TTS empleado en este prototipo es el Microsoft Speech Platform y su SDK, para convertir el texto generado por el sistema a voz.

Por último, se ha diseñado una interfaz de usuario que mantiene al usuario al tanto de los errores y el estado actual del sistema. Asimismo, contiene un pequeño tutorial que enseña cómo se debe colocar correctamente el brazalete para evitar errores de funcionamiento. Asimismo, posee un pequeño sistema de ayuda que muestra el alfabeto Morse y como medio de entrenamiento y aprendizaje. La interfaz muestra el estado del dispositivo en todo momento (bloqueado/desbloqueado, nivel de batería etc.).

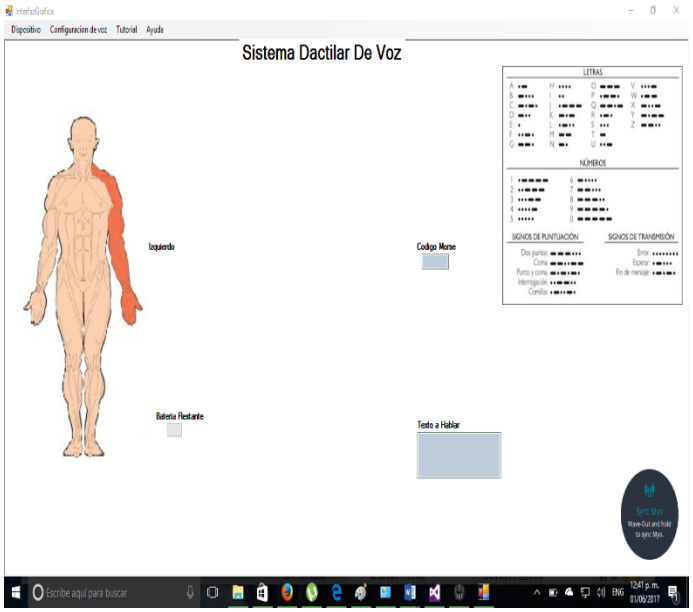


Figura 7. Interfaz de usuario del sistema.

Este monitoreo se puede habilitar y deshabilitar según el usuario lo desee.

La interfaz muestra las rayas y los puntos que se han detectado, así como las letras y palabras formadas. Esto retroalimenta al usuario, y le permite saber si está creando correctamente las palabras y oraciones.

4.2. Descripción Básica de la Aplicación

El sistema se ha desarrollado empleando Windows Forms, el cual es el nombre dado a la interfaz de programación de aplicación gráfica (API) que se incluye como parte de Microsoft .NET Framework [18, 19, 20, 21, 22], el cual proporciona acceso a los elementos de la interfaz de Microsoft Windows nativas envolviendo la API de Windows existente y algunos controles sólo envuelven componentes subyacentes de Windows.

1. Iniciar la aplicación de Windows forms (necesario para que la aplicación funcione correctamente [16, 17])
2. Encontrar un Myo y establecer conexión.

3. Si no hubo conexión, ir al paso 15.
4. Crear una nueva instancia del colector de datos del dispositivo y ejecutarlo.
5. Escuchar al sensor para detectar gestos.
6. Si la pose es un *puño* seguido de reposo, calculamos el tiempo que mantuvo la posición, para decidir si se trata de un punto o una raya y regresar al paso 5.
7. Si la pose es *palma adentro*, continuar al paso 9.
8. Si la pose detectada es *mano abierta*, ir al paso 11.
9. Decodificar la señal para convertir de Morse a alfabeto.
10. Regresar al paso 5.
11. Crear nueva instancia del sintetizador de voz y ejecutar.
12. Seleccionar la salida de audio
13. Seleccionar el timbre de voz
14. Reproducir el texto en voz, y regresar al paso 5
15. ¿Se desea intentar una conexión de nuevo?
16. Si es así, regresar al paso 2.
17. Cerrar la aplicación de Windows forms.

5. RESULTADOS

La principal parte de los resultados obtenidos se presenta en la figura 7. En la cual se observa la interfaz de usuario diseñada. La cual contiene diversas ayudas visuales que pueden ayudar al usuario a saber si el dispositivo está funcionando correctamente.

También como se observa en la figura 5 se proporciona el alfabeto morse para así garantizar que cualquier persona, aunque no sepa nada del código morse, se familiarice con este código y así pueda generar las secuencias de caracteres tras un breve periodo de familiarización con este.

En la secuencia de las figuras 6 se puede observar que la interfaz cuenta con unos cuadros de texto los cuales sirven para ir viendo si la secuencia del código morse es correcta y una vez que esta es correcta y el usuario a realizado el gesto de terminación de secuencia, el cuadro de texto del texto a hablar se escribirán la secuencia de letras que el usuario quiera leer en voz alta.

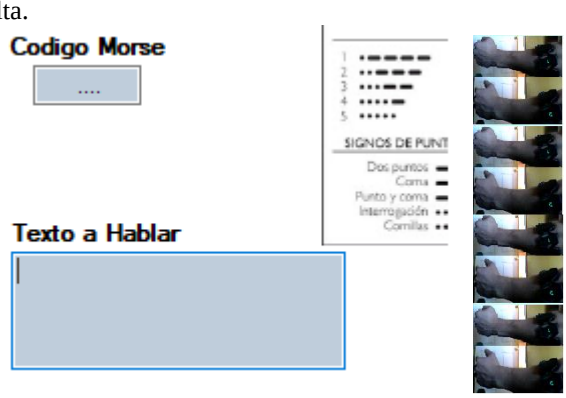


Figura 8 . Secuencia de la letra “H”

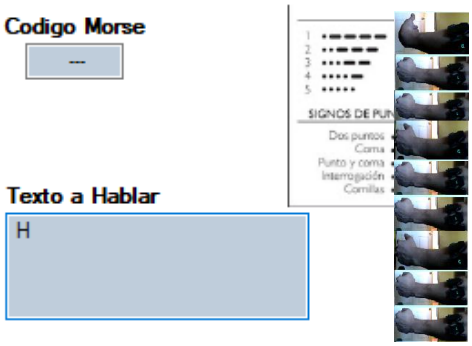


Figura 9. Secuencia de la letra “O”, con la letra “H” ya agregada al texto a hablar

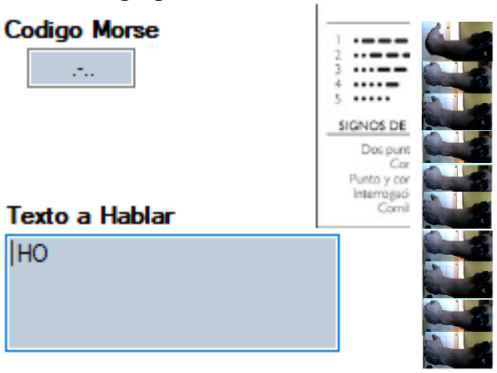


Figura 10. Secuencia de la letra “L”, con la letra “O” ya agregada al texto a hablar

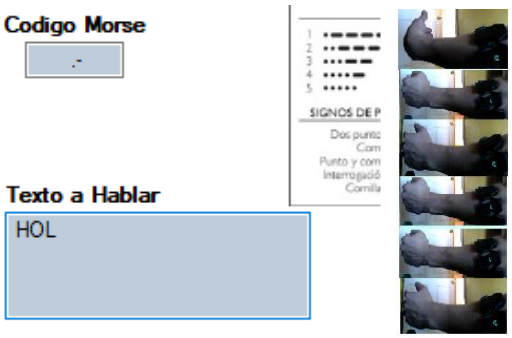


Figura 11. Secuencia de la letra “A”, con la letra “L” ya agregada al texto a hablar

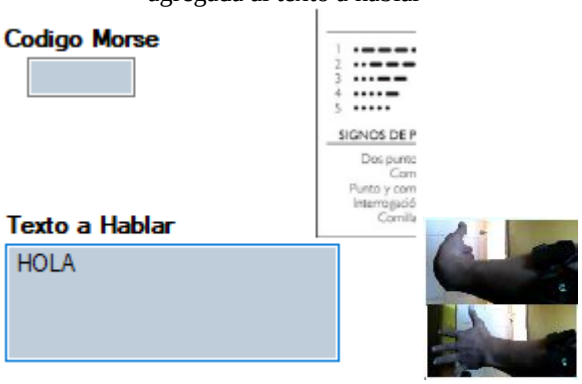


Figura 12. Palabra generada

Para la figura 8 se observa que es un patrón de poses, en la que la pose *puño* dura menos de 250 mS seguida de la pose de *descanso* indicándole al sistema que se ha generado un punto(“.”) para asi volver a meter la pose *puño* para hacer la sucesión de puntos para conformar la letra “H” mostrada en la figura 9.
Para la figura 9 se comienza usando la pose *palma adentro* para indicarle al sistema que la sucesión de puntos y rallas ha terminado y la letra se ponga en la casilla de texto a hablar.
En la sucesión de figuras anterior se usa una doble imagen de *puño* para indicar que el puño duro mas de 250 mS es decir que se ha conformado una ralla(“-“).
En la figura 12 se termina con la pose *palma abierta* para indicarle al sistema que puede reproducir la palabra generada.

6. CONCLUSIONES Y TRABAJO FUTURO

El sistema está orientado a facilitar la vida diaria de personas que se les dificulta hablar legiblemente, ya que de una manera sencilla es posible generar texto mediante gestos manuales que luego se convertirá en voz, para así poder lograr un mayor grado de comunicación e integración para la persona discapacitada y el resto de la sociedad, por lo que posee un alto valor social.

Al usar el código Morse para generar palabras facilita mucho a que cualquier persona pueda usar el sistema ya que es fácil de aprender y es un alfabeto universal, si bien no es un código eficiente. Aun así, es relativamente sencillo diseñar entrenarse en él al ser un alfabeto binario. Para ello, en la interfaz de usuario se muestran las ayudas necesarias para que cualquiera pueda escribir las palabras que desee.

El módulo de ayuda permite que cualquier persona pueda usar rápidamente el sistema ya lo guía desde el momento en el que el usuario se coloca el sensor en el brazo y verifica en todo momento el estado de comunicación y uso del sistema.

La finalidad de este sistema es proporcionar un sistema portable, económico y confiable, para que la gente con dificultades del habla pueda comunicarse e integrarse mejor a la sociedad. Para alcanzar esta meta se continúa trabajando en una versión móvil para dispositivos Android, que elimine el uso de una computadora portátil o tableta. Esta versión móvil (ya bajo desarrollo) contará con todas las características del sistema basado en computadora, e incluso podrá detectar el idioma que el dispositivo está usando para así poder elegir la voz a usar y otras facilidades no disponibles en computadora.

7. REFERENCIAS.

[1] I. Monsalvo Escamilla, Antología sobre comunicación: un enfoque humanista, México, Instituto Politécnico Nacional, 1998, 1.
[2] https://developer.thalmic.com/start/.
[3] G.R. Lazarovich, 100 años de radiotelegrafía naval en Chile, *Revista Marina*, 120(883)., 2004, 4.
[4] Pérez J., Gardey A. “Definición de mutismo “disponible: https://definicion.de/mutismo.
[5] M. B. I. Reaz, M. S. Hussain, F. Mohd-Yasin, “Techniques of EMG Signal Analysis: Detection, Processing, Classification and Applications”, Biological Procedures Online, vol. 8, issue 1, pp. 11–35, March 2006
[6] Nikias CL, Raghuveer MR. Bispectrum estimation: A digital signal processing framework. IEEE Proceedings on Communications and Radar. 1987;75(7):869–891.
[7] Basmajian, JV.; de Luca, CJ. Muscles Alive - The Functions Revealed by Electromyography. The Williams & Wilkins Company; Baltimore, 1985.
[8] Graupe D, Cline WK. Functional separation of EMG signals via ARMA identification methods for prosthesis control purposes. IEEE Transactions on Systems, Man and Cybernetics, 1975;5(2):252-259.
[9] Kleissen RFM, Buurke JH, Harlaar J, Zilvold G. Electromyography in the biomechanical analysis of human movement and its clinical application. Gait Posture. 1998;8(2):143–158. doi: 10.1016/S0966-6362(98)00025-3. [PubMed]
[10] Villamizar J., Padilla, R., & Cabrera G. (2012). Brazo robótico controlado por electromiografía. *Scientia et technica*, *17*(52).
[11] Cram, JR.;Kasman, GS.; Holtz, J. Introduction to Surface Electromyography. Aspen Publishers Inc.; Gaithersburg, Maryland, 1998.
[12] Ferguson, S.; Dunlop, G. Grasp Recognition From Myoelectric Signals. Procedures Australasian Conference Robotics and Automation 2002; pp. 78–83.
[13] Stanford V. Biosignals offer potential for direct interfaces and health monitoring. Pervasive Computing, IEEE. 2004;3(1):99–103.
[14] Wheeler KR, Jorgensen CC. Gestures as input: neuroelectric joysticks and keyboards. Pervasive Computing, IEEE. 2003;2(2):56–61.
[15] Aho, Alfred V.; Ravi Sethi, Jeffrey D. Ullman (2008). «Introducción a la Compilación». Compiladores: Principios, técnicas y prácticas. México: Addison Wesley..
[16] Sells, Chris (September 6, 2003). Windows Forms Programming in C# (1st ed.). Addison-Wesley Professional. p. xxxviiii.
[17] "Design and Implementation Guidelines for Web Clients by Microsoft Pattern and Practices". Microsoft. November 2003.
[18] "Introduction to Windows Forms" (Visual Studio 2003 documentation). Microsoft 2003.
[19] Griffiths, Ian; Adams, Matthew (March 2003). NET Windows Forms in a Nutshell. O'Reilly Media. p. 4.
[20] Smith, Josh (2007-09-05). "WPF vs. Windows Forms". Josh Smith on WPF. Retrieved 2011-08-25. "WPF is not intended to replace Windows Forms. [...] Windows Forms is still alive and well, and will continue to be enhanced and supported by Microsoft for years to come. WPF is simply another tool for Windows desktop application developers to use, when appropriate."
[21] "A WPF Q&A". infoq.com. 2014-04-03. Retrieved 2014-04-21. "Windows Forms is continuing to be supported, but in maintenance mode. They will fix bugs as they are discovered, but new functionality is off the table"
[22] Allen, Jonathan (2014-05-06). "High DPI Improvements for Windows Forms in .NET 4.5.2". InfoQ. Retrieved 2015-02-10.
[23] S. Benford; H. Schnadelbach; B. Koleva; B. Gaver; A. Schmidt; A. Boucher; A. Steed; R. Anastasi; C. Greenhalgh; T. Rodden; H. Gellersen. "Sensible, sensable and desirable: a framework for designing physical interfaces" (PDF). CiteSeerX 10.1.1.190.2504 Freely accessible. Archived from the original (PDF) on January 26, 2006.
[24] Allen, Jonathan; Hunnicutt, M. Sharon; Klatt, Dennis (1987). From Text to Speech: The MITalk system. Cambridge University Press. ISBN 0-521-30641-8.