TecnoLógicas

Procesamiento de Señales Provenientes del Habla Subvocal usando Wavelet Packet y Redes Neuronales

Speech Subvocal Signal Processing using Packet Wavelet and Neuronal Network

Luis E. Mendoza[1]
Jesús Peña[2]
Luis A. Muñoz-Bedoya[3]
Hernando J. Velandia-Villamizar[4]

1 Ing. Telecomunicaciones, DIEEST, Universidad de Pamplona, Pamplona-Colombia
luis.mendoza@unipamplona.edu.co
2 Ing. Electrónica, DIEEST, Universidad de Pamplona, Pamplona-Colombia
jesusteel@gmail.com
3 Ing. Telecomunicaciones, DIEEST, Universidad de Pamplona, Pamplona-Colombia
luislamb99@unipamplona.edu.co
4 Ing. Telecomunicaciones, DIEEST, Universidad de Pamplona, Pamplona-Colombia
hernando.velandia@unipamplona.edu.co

Tecno. Lógicas., ISSN 0123-7799, Edición Especial, octubre de 2013, pp. 655-667

Resumen

Este artículo presenta los resultados obtenidos del registro, procesamiento, reconocimiento y clasificación de palabras del lenguaje español mediante el análisis de las señales de voz de habla subvocal. El trabajo en conjunto será en un futuro enfocado en aplicaciones de telecomunicaciones como: chat para sordo mudos. La base de datos procesada está conformada por seis palabras (adelante, atrás, derecha, izquierda, iniciar y parar). Las señales fueron sensadas con electrodos superficiales dispuestos en la superficie de la garganta y adquiridas con una frecuencia de muestreo de 50 Khz. El acondicionamiento de las señales consistió en: la ubicación de la zona de interés mediante análisis de energía, y el filtrado usando Transformada Wavelet Discreta. Finalmente, la extracción de características se hizo en el dominio del tiempo-frecuencia empleando Wavelet Packet y técnicas estadísticas por ventaneo. La clasificación se llevó a cabo con una Red Neuronal por Retropropagación cuyo entrenamiento se realizó con el 70% de la base de datos obtenida. El porcentaje de acierto encontrado fue de 75%±2.

Palabras clave

Electromiografía; habla subvocal; wavelet packet; redes neuronales.

Abstract

This paper presents the results obtained from the recording, processing and classification of words in the Spanish language by means of the analysis of subvocal speech signals. The processed database has six words (forward, backward, right, left, start and stop). In this work, the signals were sensed with surface electrodes placed on the surface of the throat and acquired with a sampling frequency of 50 kHz. The signal conditioning consisted in: the location of area of interest using energy analysis, and filtering using Discrete Wavelet Transform. Finally, the feature extraction was made in the time-frequency domain using Wavelet Packet and statistical techniques for windowing. The classification was carried out with a backpropagation neural network whose training was performed with 70% of the database obtained. The correct classification rate was 75%±2.

Keywords

Electromyography; subvocal speech; wavelet packet; neuronal networks.

1. INTRODUCCIÓN

El *habla subvocal* es el registro e interpretación de las señales bioeléctricas que controlan las cuerdas vocales y la lengua durante el proceso de comunicación verbal, estas señales proveen información y se relacionan con palabras que percibimos auditivamente. Además las señales de habla subvocal se presentan sin ser necesaria la producción física de sonido por parte del individuo (Seniam, 2012; Chuck, et al 2003; Chuck y Kim, 2005). En la actualidad, los entornos ruidosos y las patologías fisiológicas son dos de los grandes problemas que ocasionan que la comunicación oral entre individuos sea afectada de manera tal que la información se distorsione o simplemente nunca se genere. Una comunicación que no dependa de la producción física de señales audibles plantearía una solución eficaz a los problemas que aumentan la vulnerabilidad de la comunicación verbal.

El habla subvocal expone una respuesta tentativa a estas limitaciones, ya que no depende de la emisión de sonido, sino que hace uso de las señales eléctricas que el cerebro transmite a las cuerdas vocales y a la lengua (Chuck et al, 2003; Chucky y Kin, 2005). Surgen diferentes aplicaciones del concepto de "comunicación silenciosa" como: control de interfaces virtuales, comunicación interpersonal, control de robots, comunicación en entornos industriales, comunicación submarina, intercambio de información militar de alta confidencialidad, comunicación móvil, control de sistemas transporte para personas con problemas motrices, ayuda a personas con patologías de pronunciación y operaciones de rescate. Diferentes estudios sobre las aplicaciones del habla subvocal se han llevado a cabo, entre los más importantes están: (Chuck et al, 2003; Chuck Kim, 2005; Bradley et al, 2005) realizados por el Centro de Investigación Ames de la NASA. Sin embargo, la existencia del habla subvocal ha sido estudiada desde décadas atrás, en 1969 *Curtis & Lewis* (Curtis et al, 1969) analizaron diversos patrones de subvocalización, registrando las señales mediante EMG de superficie usando electrodos ubicados en lugares opuestos sobre el cartílago tiroides.

En los 80's con la llegada de la electrónica digital y con los dispositivos electrónicos cada vez más pequeños. Diferentes grupos de

investigación se focalizaron en proyectos de reconocimiento de habla mediante electromiografía facial y sublingual. La mayoría de estudios sobre el comportamiento de los músculos se han realizado con el fin de controlar prótesis electrónicas (Szu et al, 2008) dependiendo de características como la amplitud y frecuencia de las señales electromiográficas. El análisis de las señales subvocales no solo procesa características como la amplitud y la frecuencia sino que hace un uso más detallado de la información de la señal.

Un punto importante en este trabajo es el registro de las señales electromiográficas (EMG) las cuales son adquiridas usando electrodos superficiales dispuestos en la garganta en modo diferencial (Szu et al, 2008). En la etapa de implementación, la disposición de los electrodos se basó en la normativa propuesta por SENIAM (Electromiografía de Superficie para la Valoración no Invasiva de Músculos) (Seniam 2010). Debido a que las señales EMG son de carácter no estacionario (la frecuencia varia en el tiempo), la extracción de características se realizó usando representaciones basadas en el plano tiempo frecuencia mediante herramientas como la Transformada Wavelet Discreta y la Transformada Wavelet Packet (Englehart et al, 1999), obteniendo una efectividad de clasificación alrededor de 75%.

2. METODOLOGÍA

2.1 Adquisición de Datos

Las señales EMG fueron adquiridas usando electrodos superficiales de Cloruro de Plata (AgCl) en configuración bipolar (Chuck et al, 2003; Chuck y Kim, 2005), estos fueron ubicados en la parte inferior derecha y superior izquierda bajo la garganta, después de haber limpiado la piel con alcohol isopropílico al 70% para reducir la impedancia de la piel; debido a la baja amplitud que presentan las señales subvocales, las señales EMG fueron amplificadas en un factor de 18700. En el sistema de adquisición y con el fin de disminuir el ruido y atenuar las frecuencias que no hacen parte de la señal EMG, se usó un filtro Butterworth pasa banda de orden 8 conformado por un filtro pasa alta de 30 Hz y un filtro pasa baja

de 450 Hz; la información fue grabada con una frecuencia de muestreo de 50 kHz, en vectores de 2 segundos y 50 señales por palabra. La frecuencia de muestreo se usa con el fin de tener una resolución suficiente y conseguir sensibilidad a alta frecuencia del sistema. El hardware implementado fue diseñado específicamente para esta aplicación en el GIBUP (Ingeniería Biomédica de la Universidad de Pamplona). La implementación final se muestra en la Fig. 1.

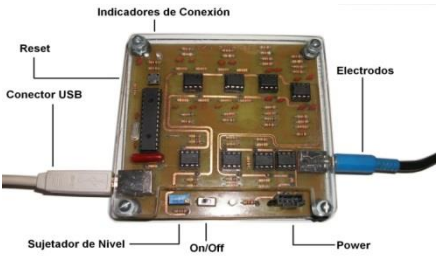

Fig. 1. Tarjeta EMG implementada para la adquisición y acondicionamiento de las señales subvocales

2.2 Acondicionamiento de la Señal

Una vez las señales fueron registradas y almacenadas, se procedió al acondicionamiento y extracción de características relevantes o patrones. Este proceso se realizó con el fin de conseguir parámetros importantes en cada grupo de señales analizadas y finalmente obtener un proceso de clasificación efectivo. El proceso de acondicionamiento se describe a continuación: Inicialmente se realizó la ubicación de la zona activa. Este proceso se hizo mediante el análisis de la energía de la señal usando ventanas sensibles a cambios repentinos en la amplitud de la señal EMG, en este caso ventanas de 400µS de ancho. La energía de la señal por ventaneo se define como:

$$E_n = \sum_{i=1}^{W} (x_{(n-1)W+i})^2 \quad (1)$$

Donde E_n es el vector de energía, x es la señal EMG y W es el tamaño de ventana. Luego se establece un umbral que indica el cambio entre el comportamiento normal de la señal EMG y el inicio de la actividad subvocal. El umbral de detección se define como:

$$U = (0.15)\, E_{max} \tag{2}$$

Donde U es el umbral y E_{max} es el pico máximo del vector energía de la señal EMG. La Fig. 2 muestra el vector energía de la señal y el umbral de detección. Seguidamente, se aplica una ventana de 0.8 S de longitud a partir del inicio de la zona activa; en esta parte de la señal se encuentra la totalidad de la información de la actividad subvocal. La Fig. 3 muestra en color verde la zona de activa de la señal de habla subvocal.

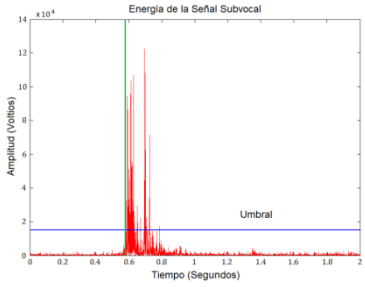

Fig. 2. Ubicación del inicio de la zona de interés mediante el umbralizado del vector de la energía de la señal EMG

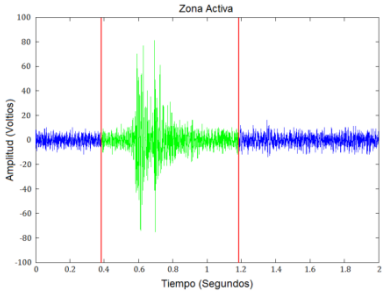

Fig. 3. Zona activa de la señal subvocal de la palabra *(Adelante)*

Seguidamente al proceso de ubicación de la zona activa se procedió a filtrar la señal usando la Transformada Wavelet Discreta (TWD) que matemáticamente se expresa, (Marcos et al, 2005):

$$\psi_{jk}[n] = 2^{\frac{j}{2}}\psi(2^j n - k) \qquad j,k \ \in \ Z \tag{3}$$

Donde j es el parámetro de escalamiento, k es el parámetro de traslación y ψ es la wavelet madre; en este trabajo la señal subvocal fue descompuesta en cuatro niveles usando la wavelet madre db 5. El umbral de filtrado se define:

$$U_f = \sqrt{2Log(n)} \tag{4}$$

Donde n es el número de muestras de la señal y U_f el umbral. Una vez hallado el umbral U_f, este se aplica mediante el método *Hard* a todos los coeficientes de descomposición de la señal. El método de umbralización *Hard* se define, (Dora et el, 2004):

$$f(x) = \begin{cases} x, & |x| > U_f \\ 0, & |x| \leq U_f \end{cases} \tag{5}$$

Donde x es cada uno de los coeficientes de descomposición, U_f es el valor umbral y $f(x)$ es la señal umbralizada. Luego la señal subvocal es reconstruida a partir de los coeficientes umbralizados, usando la Transformada Wavelet Inversa, la cual se expresa, (Marcos et al, 2005):

$$x[n] = \sum_{j\in Z}\sum_{k\in Z} C[j,k].\psi_{j,k}[n] \tag{6}$$

Donde $x[n]$ es la señal reconstruida, C son los coeficientes umbralizados y ψ es la base wavelet. La reconstrucción de la señal subvocal se muestra en la Fig. 4, donde la señal en color azul es la señal original y la señal en color rojo es la señal filtrada.

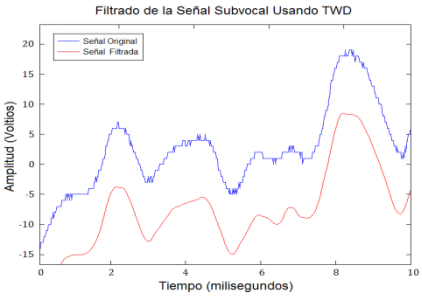

Fig. 4. Comparación entre la señal subvocal registrada y la señal subvocal filtrada con TWD

2.3 Extracción de Patrones

La extracción de características relevantes se realizó mediante la Transformada Wavelet Packet (TWP) , (Sepúlveda, 2004). Esta transformada se utilizó, ya que permite un analisis multiresolucional de las señales y por ende permite realizar una extracción de patrones más generalizada que por ejemplo Fourier o discreta del coseno. La TWP descompone los coeficientes de aproximación y los de detalle formando una estructura de árbol. En este caso la señal subvocal fue descompuesta en tres niveles generando 14 coeficientes tal y como se muestra en la Fig. 5.

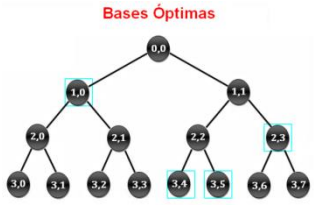

Fig. 5. Selección de las Bases Óptimas de la TWP

Para escoger las bases óptimas o coeficientes que mejor representan la señal subvocal se usa una *función de costo* que mide la concentración de la información. En este caso se uso como función de costo la entropía de *Shannon* que se define, (Sepúlveda, 2004):

$$\mathcal{H}(q) = \sum_{k} |q(k)| Log \frac{1}{|q(k)|} \tag{7}$$

Donde $\mathcal{H}$ es la entropía de Shannon y $q(k)$ es la energia normalizada de los coeficientes wavelet. Una vez obtenida la función de costo de cada coeficiente, se procede a escoger las bases óptimas siguiendo los parámetros propuestos en [Sepúlveda, 2004]. Como resultado se obtuvo que la señal EMG de habla subvocal esta mejor representada por los coeficientes (1,0), (2,3), (3,4) y (3,5) indicados en la Fig. 5. Teniendo los resultados de la TWP, se procedió a extraer los patrones de las bases óptimas empleando métodos estadísticos como la Raíz Media Cuadrada (RMS), Valor Medio Absoluto (VMA), Valor Medio Absoluto Diferencial ($VMAD$) y la Varianza (σ^2) en ventanas de 16 ms, con el fin de reducir el tamaño del vector de caracterización de las señales subvocales. La definición de estos métodos esta representada en (8), (9), (10) y (11) respectivamente:

$$RMS = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(x_k)^2} \tag{8}$$

$$VMA = \frac{1}{N}\sum_{k=1}^{N}|x_k| \tag{9}$$

$$VMAD = \frac{1}{N}\sum_{k=1}^{N}|x_{k+1} - x_k| \tag{10}$$

$$\sigma^2 = \frac{\sum_{k=1}^{N}(x_i - \bar{x})^2}{N-1} \tag{11}$$

Donde N es la longitud de la ventana y x es el coeficiente óptimo. Una vez se obtuvieron los vectores de características usando los métodos estadísticos, se procedió a aplicar Análisis de Componentes Principales (PCA) para la reducción de dimensionalidad de los datos, (Lindsay, 2011). En este se encontró

que haciendo uso de PCA el 98% de la información importante se conserva en las cuatro primeras componentes principales.

2.4 Clasificación de Datos

Finalmente y ya con los procesos de acondicionamiento y extracción de patrones, se procedió a realizar la clasificación. La clasificación fue realizada con una red neuronal (RN) perceptrón multicapa con aprendizaje supervisado por retropropagación, (Bonifacio, 2006). La RN consta de 6 neuronas de entrada P, 25 neuronas ocultas b y una neuronas de salida y. La Fig. 6 muestra la estructura de la RN que se utilizó. El entrenamiento de la red neuronal consta de 300 ciclos y se realizó con el 70% de las bases de datos registrada.

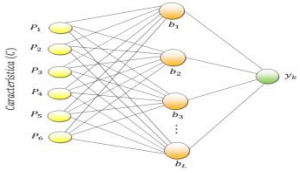

Fig. 6. Arquitectura de la RN implementada en la clasificación de las señales subvocales

3 RESULTADOS

3.1 Efectividad de Clasificación

El test de efectividad del sistema de clasificación de habla subvocal se realizó usando el 30% de la base de datos obtenida. La base de datos fue tomada de una persona de sexo masculino y 22 años de edad en varias sesiones de registro con condiciones de contaminación auditiva normal.

En la Tabla 1 son mostrados los resultados de la fase de clasificación, los cuales muestran como la efectividad del algoritmo varía dependiendo de la señal subvocal clasificada, teniendo un promedio de acierto del 75%. La tasa de efectividad puede variar en relación a la correcta ubicación de los electrodos, a la sensibili-

dad del sistema de adquisición, al acondicionamiento de la señal y a la extracción de las características.

Tabla 1. Efectividad del proceso de clasificación

	Clasificación					
Palabra	Derecha	Izquierda	Adelante	Atrás	Inicio	Parada
Efectividad (%)	75.5	78	74.4	77.5	82.8	61.22

3.2 Resultados en Tiempo Real

Los algoritmos de acondicionamiento, extracción de características y clasificación fueron implementados en un sistema de reconocimiento en tiempo real diseñado con el software MatLab, este sistema consta de una interfaz de monitoreo la cual indica el tiempo de la registro, el estado de conexión de la tarjeta EMG y el resultado de la clasificación. La Fig. 7 muestra el registro de una señal subvocal de 2 segundos de duración y el reconocimiento de la palabra "Adelante".

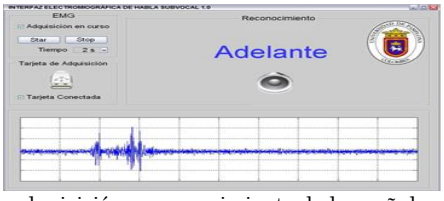

Fig. 7. Interfaz de adquisición y reconocimiento de las señales de habla subvocal

El sistema en tiempo real clasifica 3 señales subvocales por minuto teniendo una efectividad del 75%, en un entorno de prueba expuesto a 70 dB de ruido acústico. Es importante mencionar que estos resultados son preliminares y que la investigación va en continuo avance, y es importante también mencionar que el tema es muy novedoso y sus aplicaciones en el área de las telecomunicaciones son muy potenciales, como por ejemplo, control de video juegos, navegación de páginas web y otros.

4. DISCUSIÓN Y CONCLUSIONES

Los resultados de este trabajo son comparables con los estudios realizados en (Chuck et al, 2003; Chuck Kim, 2005; Bradley et al, 2005), donde obtuvieron tasas de acierto similares. La efectividad del sistema subvocal es baja en relación a los sistemas de reconocimiento de voz, pero a diferencia de estos el sistema subvocal no es afectado por factores de contaminación auditiva o de problemas de pronunciación. Otros sistemas de reconocimiento de habla usan señales cerebrales (EEG) y EMG, registrando múltiples señales a la vez, teniendo una complejidad de procesamiento mucho más alta y una tasa de reconocimiento menor. Por otro lado, el sistema de comunicación subvocal usa sólo un canal EMG lo cual reduce el costo computacional en el procesamiento y mejora el porcentaje de acierto. Se concluye que son resultados muy coherentes y que son un punto de partida para trabajo a futuro, es importante mencionar que el dominio wavelet fue de gran ayuda para poder extraer patrones que diferencian los grupos y junto con las redes neuronales fueron fundamentales para conseguir inicialmente los resultados que se presentan en este trabajo.

5. REFERENCIAS

Ballesteros Larrotta (2004). "Aplicación de la Transformada Wavelet Discreta en el Filtrado de Señales Bioeléctricas," Umbral Científico, Fundación Universitaria Manuela Beltrán, Bogotá, Colombia, pp. 92-98, Dic.

Bonifacio M. del Brío y Alfreda S. Molina. (2006) Redes Neuronales y Sistemas Borrosos, 2nd ed., Alfaomega S.A. de C.V., Ed., México D.F.

Bradley J. Betts and Charles Jorgensen (2005). "Small Vocabulary Recognition Using Surface Electromyography in an Acoustically Harsh Environment," Neuro-Engineering Laboratory, NASA Ames Research Center, Moffett Field, California, EEUU.

Chuck Jorgensen, Diana D. Lee and Shane Agabon (2003). "Sub Auditory Speech Recognition Based on EMG Signals," Proceedings of the International Joint Conference on Neural Networks (IJCNN), IEEE, vol. 4, pp. 3128–3133.

Chuck Jorgensen and Kim Binsted (2005),."Web Browser Control Using EMG Based Subvocal Speech Recognition," Proceedings of the 38th

Annual Hawaii International Conference on System Sciences (HICSS), IEEE, pp. 294c.1–294c.8.

Curtis D. Hardyck and Lewis F. Petrinovich (1969).“Treatment of Subvocal Speech During Reading”, en Journal of Reading, pp. 361-368.

Englehart, B. Hudgins, P.A. Parker, and M. Stevenson (1999). “Classification of the Myoelectric Signal using Time-Frequency Based Representations,” Special Issue Medical Engineering and Physics on Intelligent Data Analysis in Electromyography and Electroneurography, Summer.

Lindsay I. Smith (2002). “A Tutorial on Principal Components Analysis,” Feb.

Marcos C. Goñi y Alejandro P. de la Hoz (2005). “Análisis de Señales Biomédicas Mediante Transformada Wavelet”, Concurso de Trabajos Estudiantiles EST, Universidad Nacional de San Martín, Argentina.

SENIAM (2010). Website (www.seniam.org). Available: http://www.seniam.org/

Sepúlveda (2004). “Extracción de Parámetros de Señales de Voz usando Técnicas de Análisis en Tiempo-Frecuencia,” Universidad Nacional de Colombia, Manizales, Colombia.

Szu Chen S. Jou (2008). “Automatic Speech Recognition on Vibrocervigraphic and Electromyographic Signals”, Language Technologies Institute, Carnegie Mellon University, Pittsburgh PA 15213, EEUU, Oct.