Desarrollo de un sistema para adquisición y procesamiento de señales de voz para fonoaudiología

Melisa Kuzman[1], Alejandro Uriz[2], Pablo Agüero[1], Juan C. Tulli[1], Esteban González[1] y Graciela Moscardi[3]
[1]*Facultad de Ingeniería – Universidad Nacional de Mar del Plata,*
{melisakzuman,pdaguero}@fi.mdp.edu.ar, 0223-4816600, Juan B. Justo 4302 (7600) Mar del Plata.

[2]*CONICET – Facultad de Ingeniería – Universidad Nacional de Mar del Plata,*
ajuriz@conicet.gov.ar, 0223-4816600, Juan B. Justo 4302 (7600) Mar del Plata.

[3]*Facultad de Ciencias Médicas- Universidad FASTA*
0223-475-7076, Avellaneda 3341(7600) Mar del Plata. Mar del Plata

***Resumen*—Hoy en día muchos especialistas utilizan para el análisis de la voz medidas objetivas de la misma. Estas medidas, que generalmente son calculadas mediante programas de computadora, suelen ser complementarias a las medidas subjetivas de la voz y se utilizan para analizar la evolución del paciente a lo largo del tiempo. La principal limitación que aparece a la hora de calcular dichos parámetros es que las señales de voz son grabadas en el consultorio del especialista, donde rara vez se dispone de una cámara anecoica e insonorizada. Por ello, pueden aparecer señales espúreas que provocan mediciones erróneas. Por lo tanto, se desea desarrollar un sistema capaz de reducir el impacto de dichas señales en el resultado del análisis. Si bien hay diversas formas de eliminar componentes indeseadas de una señal, debe tenerse la precaución de no afectar las señales de interés de forma tal de que del resultado del análisis posterior no se obtengan resultados erróneos.**

Por ello, se propone desarrollar un sistema capaz de adquirir sonidos mediante un arreglo de micrófonos. El equipo funciona en dos configuraciones básicas: promediación y substracción. En el primero de los casos, un arreglo de hasta 4 micrófonos recibe la misma señal y las promedian, y el resultado de dicha señal es utilizado para el análisis. Esta topología permite reducir el nivel de ruido blanco gaussiano en la señal adquirida. En el caso en que el sistema se utilice para substraer componentes indeseados de la señal, se utilizan dos micrófonos, el primero adquiere la señal de voz sumada a los sonidos ambientales, mientras que el segundo registra los sonidos del ambiente. Luego, dichas señales se restan con el fin de reducir los sonidos del entorno en la señal resultante, realzando así la señal de voz. Se realizan simulaciones para analizar el rendimiento de ambos sistemas.

***Palabras clave*— Adquisición de sonidos, reducción de ruido, fonoaudiología.**

I. INTRODUCCIÓN

En las últimas décadas las herramientas objetivas de análisis de la voz han proliferado a la par de las computadoras. Este tipo de técnicas procesan grabaciones de señales de voz, y mediante un proceso de análisis, permiten obtener medidas objetivas de la misma. Algunos ejemplos de estas herramientas son el Praat[1], el *Multi-Dimensional Voice Program* (MDVP)[2] y el Sistema de Análisis de la Voz (SAV) [3] desarrollado por el Laboratorio de Comunicaciones de la Facultad de Ingeniería de la Universidad Nacional de Mar del Plata. Tal como se mencionó previamente, este tipo de programas procesan registros de señales de voz y obtienen una serie de medidas objetivas de acuerdo a los requerimientos del usuario.

Si bien este tipo de aplicaciones permiten obtener muy buenos resultados, estos pueden verse afectados por la calidad de las señales de voz que son ingresadas al sistema. Es decir, si la señal tiene ruido o efectos sonoros indeseados pueden generarse resultados erróneos en el análisis. Existe una serie de métodos desarrollados para procesar la señal de voz de forma tal que se obtengan resultados mas exactos[4]. Este tipo de técnicas debe considerar en primer lugar la naturaleza de la señal, ya que si a la señal de voz se le realiza un procesamiento inadecuado se podrían perder datos necesarios para realizar de forma apropiada el análisis.

En este Trabajo se desarrolla la implementación de un sistema que permite mejorar grabaciones realizadas en el campo. Este sistema, permite a especialistas, como por ejemplo fonoaudiólogos, grabar señales de voz en su consultorio, donde generalmente se carece de una cámara anecoica o las condiciones no son las apropiadas para una correcta grabación. De esta forma, este sistema es de suma utilidad para especialistas que desean realizar grabaciones de sus pacientes para luego realizar un análisis objetivo que permita representar la evolución del paciente a lo largo del tiempo.

Con el fin de describir el desarrollo, se presentan dos técnicas de reducción de ruido en señales que tienen la propiedad de no afectar la naturaleza de la misma. Por ello, en primer lugar se plantea para el sistema una arquitectura basada en dos micrófonos. Luego, se realiza una combinación lineal de las señales adquiridas por cada micrófono de forma tal de suprimir la las componentes espúreas. En el segundo método bajo estudio[5], el sistema se utiliza para implementar un arreglo de micrófonos, los cuales adquieren la señal y mediante la aplicación de la correlación cruzada y promediación a las señales se puede

reducir el nivel de ruido de la señal de entrada sin generarle distorsiones que afecten el análisis.
El sistema ha sido implementado utilizando CODECs de audio CM108[12], los cuales integran las funcionalidades necesarias para llevar a cabo esta tarea y fundamentalmente tienen una interfaz USB que permite utilizarlos en cualquier computadora sin el agregado de controladores adicionales.
La estructura del Trabajo es la siguiente: en la Sección II se presentan los métodos de reducción de ruido a utilizar en el sistema. La Sección III describe el dispositivo utilizado para la implementación. Por último, en la Sección IV se presentan las conclusiones del trabajo y los lineamientos a seguir en el futuro.

II. TÉCNICAS DE REDUCCIÓN DE RUIDO

Existe una gran variedad de algoritmos que se utilizan para reducir el ruido existente en señales. A la hora de elegir el método de reducción de ruido a utilizar debe tenerse en cuenta la naturaleza de la señal a procesar. Para el caso de señales de voz, dos de las técnicas mas populares son las técnicas basadas en descomposición en valores singulares (SVD) [6,7] y las basadas en la transformada discreta Wavelet (TWD) [8,9]. El primer método, toma la señal de voz a analizar, genera una representación matricial A de la misma. Luego, en base a la descomposición en valores singulares, se modela la señal mediante un producto de tres matrices U, Σ y V, las cuales son utilizadas para obtener A de acuerdo a la Ec. (1).

$$A = U \cdot \Sigma \cdot V^T \qquad (1)$$

Donde U y V son matrices unitarias, y Σ una matriz diagonal, la cual es compuesta por los valores singulares de A. Si se establece que el rango de la matriz *A* es *r*, la Ecuación (1) puede reescribirse como :

$$A = \sum_{k=1}^{r} \sigma_k u_k v_k^T \qquad (2)$$

Donde u_k es la columna *k-ésima* de la matriz U ; v_k, es la columna k-ésima de la matriz V, y σ_k son los primeros r valores singulares ordenados en orden decreciente. Entonces, si el *rango(A)=r*, y $\sigma_{r+1}; \sigma_{r+2}; ...; \sigma_m$ son nulos. Si se trunca la Ecuación (2) asumiendo que los últimos (r-p) valores singulares son despreciables, sale:

$$p < r: \qquad \hat{A} = \sum_{k=1}^{p} \sigma_k u_k v_k^T \qquad (3)$$

Donde la matriz $\hat{A}$ tiene rango p, y una aproximación de la matriz A. Debido a que los autovalores mas pequeños están asociados a componentes de alta frecuencia, y que en señales de voz, en dicha banda de frecuencias predomina el ruido. Se puede decir que $\hat{A}$ es una versión pasobajos de A. Entonces, dependiendo el grado *p* del truncamiento, se puede controlar el nivel de filtrado de la aproximación $\hat{A}$.

La principal desventaja de este método es que en primer lugar el costo computacional de los cálculos matriciales es elevado y por lo tanto, el sistema podría no operar en tiempo real. Además, el nivel de truncamiento de las matrices debería ajustarse de acuerdo al nivel de ruido ambiente. Por otro lado, este truncamiento genera la pérdida de las características de la voz a analizar. Consecuentemente, este método no resulta conveniente para el tipo de análisis al cual apunta este trabajo.

Por otro lado, los métodos basados en la Transformada Wavelet Discreta (TWD), se basan en una transformación tiempo-frecuencia, que permite llevar a cabo un análisis multiresolución, que es conveniente para señales no estacionarias como la voz. La principal limitación a la hora de implementar este algoritmo, es que de acuerdo al nivel de ruido a filtrar debe determinarse el nivel del análisis a realizarle a la señal.

Debido a los motivos previamente descriptos, a la hora de implementar un sistema para adquisición de señales de voz para fonoaudiología, se utilizarán dos algoritmos que se caracterizan por conservar las características de la señal de la voz necesarias para obtener medidas objetivas de la misma. Estos algoritmos son descriptos en detalle en las siguientes subsecciones.

A. Reducción de ruido mediante sustracción de señales

Este método de reducción se ruido se basa en un arreglo compuesto por dos micrófonos. El primero de estos micrófonos está ubicado frente al orador, mientras que el segundo está ubicado detrás del mismo. De esta forma, el primero de los micrófonos capta la señal de voz a procesar sumada al ruido existente en el ambiente $(x+n)$, mientras que el segundo capta solo el ruido ambiente n'. Una representación de esta topología puede apreciarse en la Figura 1.

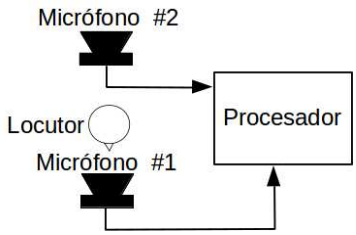

Fig. 1: Esquema del reductor de ruido mediante substracción.

La señal adquirida por ambos transductores es luego procesada por la computadora, la cual realiza la diferencia entre ambas señales, de forma tal de obtener una nueva, compuesta solo por la señal de voz a analizar.

La principal desventaja de este método es que la amplitud de la señal n' del micrófono 2, que captura el ruido ambiente debe ser afectada por un factor α de forma tal que al calcular y, la diferencia entre ambas señales, se reduzca lo suficiente las señales indeseadas.

$$y = (x+n) - \alpha.n' \qquad (4)$$

En la Subsección siguiente se desarrolla un método en el cual se utiliza un arreglo compuesto por dos micrófonos para reducir el nivel de ruido en la señal adquirida, pero utilizando el algoritmo de la correlación cruzada.

B. Reducción de ruido mediante correlación cruzada

De la misma forma que en el caso anterior, este método[10,11] está basado en un arreglo de micrófonos.

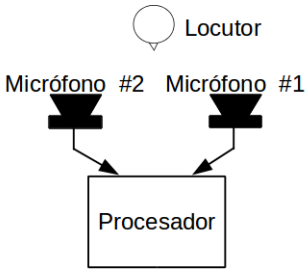

Fig. 2: Esquema del reductor de ruido basado en correlación cruzada.

Pero, tal como se aprecia en la Figura 2, en este caso los micrófonos se encuentran ubicados uno al lado de otro. La Figura 3 presenta la geometría de este arreglo en el momento que una onda sonora incide a dicho arreglo:

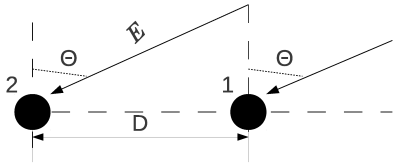

Fig. 3: Esquema del arreglo de micrófonos propuesto.

Este método se basa en el hecho que la señal proveniente de la dirección θ_s arriba primero al micrófono 1, luego recorre una distancia ξ y, es recibida por el micrófono 2. Entonces, si: $\xi = D.sin\theta_s$, la dirección de arribo de la señal puede ser determinada usando la Ecuación 5.

$$\theta_s = sin^{-1}(\frac{\upsilon.\tau}{D}) \qquad (5)$$

Donde υ es la velocidad del sonido, y τ que le lleva a la señal atravesar la distancia υ. Entonces, una vez que un valor de τ del rango $-T < \tau < T$ es obtenido, la dirección de arribo puede ser estimada. Pero, en este trabajo, τ es usado para determinar el desplazamiento temporal entre las señales que son adquiridas por cada micrófono. Entonces, con el fin de ajustar temporalmente ambas señales adquiridas, una de las mismas es desplazada d muestras, donde $d = \frac{\tau}{T_s}$, y T_s el período de muestreo. En este trabajo, el retardo τ es obtenido calculando el valor máximo de la correlación cruzada $\Phi(\tau)$ (ver Ec. 6) entre los vectores de entrada $X_1(t)$ y $X_2(t)$.

$$\Phi(\tau) = \frac{1}{N}\sum_{t=0}^{N-1} X_1(t)X_2(t+\tau) \qquad (6)$$

Donde $\Phi(\tau)$ es un vector de $2.N-1$ elementos.

Se asume que la señal que llega a ambos micrófonos es exactamente la misma pero desplazada temporalmente τ. Si el elemento de valor máximo del vector $\Phi(\tau)$ coincide con el centro del vector calculado, se concluye que no hay defasaje entre ambas señales, y por lo tanto $\tau = 0$. Por otro lado, si el valor máximo no coincide con el centro del vector obtenido mediante correlación cruzada, se puede concluir que ambas señales han arribado a los micrófonos con un retardo τ. Cabe destacar que el mínimo valor de retardo está asociado al período de muestreo del conversor analógico-digital que adquiere la señal de ambos micrófonos, ya que este intervalo de tiempo está asociado al valor obtenido de τ cuando $d = 1$.

Una vez que se obtiene el retardo entre ambas señales $d = \frac{\tau}{T_s}$, el algoritmo propuesto desplaza una de las señales para compensar dicho retardo y luego promedia ambos vectores. Como consecuencia de esta operación, el ruido blanco contenido en la señal promediada se reduce respecto a cada una de las señales originales tantas veces como micrófonos tenga el arreglo. La Figura 4, presenta el rendimiento del sistema para el caso en que el arreglo se compone de dos micrófonos. La curva de color azul representa la relación señal a ruido de la señal adquirida por un micrófono para niveles de ruido variable. Por otro lado, la curva de color rojo representa la relación señal a ruido de la señal resultante de aplicar el método de reducción señal a ruido basado en la correlación cruzada, para un arreglo compuesto por dos micrófonos. Se puede apreciar una mejora de 3dB en dicha relación señal a ruido, debida a la promediación de ambas señales.

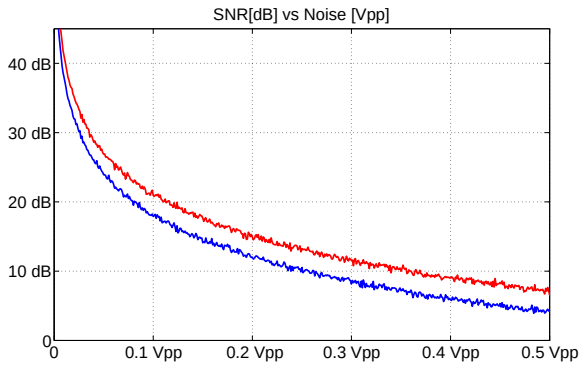

Fig. 4: Rendimiento del algoritmo reductor de ruido basado en la correlación cruzada, para diversos valores de ruido en la entrada.

III. Implementación del Sistema

Con el fin de implementar un sistema que aplique los métodos previamente descriptos, se ha construido un sistema de adquisición de señales capaz de registrar la señal adquirida por hasta cuatro micrófonos. Un esquema del sistema desarrollado se presenta en la Figura 5.

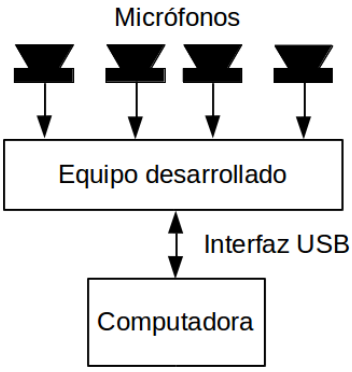

Fig. 5: Esquema del sistema propuesto.

El dispositivo desarrollado adquiere las señales provenientes de los micrófonos y las transmite a través de una interfaz USB a una computadora. Cabe destacar, que al operar en esta configuración, la computadora reconoce al dispositivo desarrollado como cuatro placas de sonido independientes. Por este motivo, la señal entregada por el dispositivo puede ser adquirida por cualquier programa para edición de sonido.

A. Implementación del dispositivo

Debido a que la finalidad del dispositivo exige que la relación señal a ruido de las señales sea lo mas elevada posible, se optó por realizar la adquisición de la señal sonora utilizando CODECs de audio integrados. En particular se optó por el circuito integrado CM108[12]. Este circuito presenta las siguientes funcionalidades:

- Encapsulado LQFP de 48 pines.
- Compatible con Windows, Linux y Mac OS9 / OS X sin controladores adicionales.
- Reguladores de tensión integrados, que permiten al sistema operar con solo 5V.
- Conversor analógico a digital de 16 bits embebido.
- Sistema de reducción de ruido integrado.
- Conversor digital a analógico de 16 bits integrado.

Además, estos CODECs requieren muy pocos componentes externos para su funcionamiento, lo que los hace ideal para implementaciones que requieran un alto nivel de integración.

Otra ventaja es que disponen de interfaz USB, y por lo tanto pueden ser conectados directamente a una computadora. Además, no requieren controladores adicionales en la mayoría de los sistemas operativos, por lo que el sistema puede ser utilizado bajo casi cualquier plataforma.

Debido a que el sistema a implementar tiene cuatro canales de entrada, se utilizan cuatro circuitos integrados CM108 (uno por cada canal). Con el fin de aumentar el nivel de integración del dispositivo, se agregó un concentrador (USB) HUB, de forma tal que al mismo se conecten los cuatro CODECs y sólo deba conectarse un cable USB a la computadora. Un diagrama de bloques del dispositivo implementado se presenta en la Figura 6.

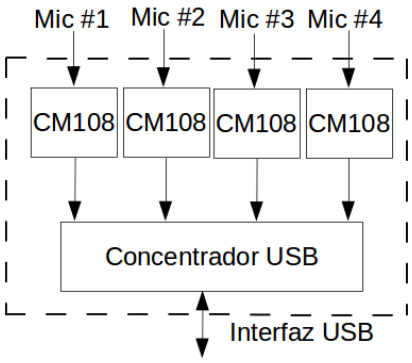

Fig. 6: Diagrama de bloques del dispositivo implementado.

Cabe destacar que para lograr el máximo rendimiento del sistema, fue necesario realizar las grabaciones utilizando una computadora portátil alimentada a batería (sin conexión a la línea eléctrica). De esta forma, se evitó que interferencias debidas a la frecuencia de la línea eléctrica deterioraran la calidad de la señal adquirida.

B. Aplicación desarrollada

Las señales de voz adquiridas por el equipo deben ser grabadas en la computadora utilizando un programa que permita almacenar en forma sincronizada la señal recibida por cada una de las fuentes. Luego, con el fin de procesar la información adquirida, se desarrolló una aplicación en MATLAB que permite aplicar las operaciones descriptas en la sección previa.

C. Simulaciones realizadas

En primer lugar, se simuló el rendimiento del algoritmo descripto en la Subsección II-A, y se verificó el rendimiento del mismo. La principal desventaja de este método es que el factor α debe ajustarse de acuerdo a las condiciones de la sala de grabación.

Por otro lado, con el fin de validar el funcionamiento del algoritmo presentado, se generaron señales de voz a las cuales se contaminó con ruido blanco gaussiano (AWGN) con una potencia conocida. Para dichas señales se simuló el caso implementado en el sistema. En la Figura 6 se presentan los resultados obtenidos en el experimento. Puede verse que las señales procesadas con el algoritmo seleccionado (trazo color rojo) obtienen una mejora de 6dB respecto a las señales sin ningún procesamiento (trazo color azul).

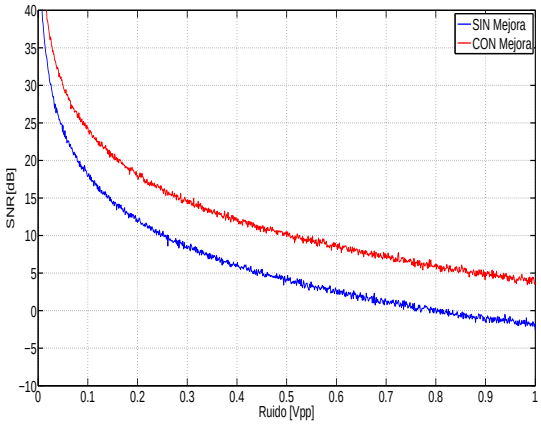

Fig. 7: Rendimiento del algoritmo de correlación cruzada para el caso implementado.

Por último, se analizó el rendimiento del dispositivo implementado. Se adquirieron señales de voz grabadas en condiciones de campo. Se grabaron las señales en formato .WAV, con una longitud de pablabra de datos de 16-bits y con una frecuencia de muestreo de 44100Hz. A las señales registradas se les aplicaron las técnicas presentadas y los resultados obtenidos se ajustaron a lo esperado en las simulaciones.

IV. CONCLUSIONES

En este trabajo se presentó un sistema de adquisición de sonidos para señales de voz destinado a obtener medidas objetivas de la misma.

Con el fin de mejorar la relación señal a ruido de la voz, se estudiaron algoritmos para reducción de ruido en señales de voz, dos de los cuales fueron exitosamente implementados en una aplicación en entorno MATLAB.

En el futuro se desea profundizar el estudio de los algoritmos de reducción de ruido para este propósito. Además se desean estudiar las condiciones físicas necesarias para acondicionar el consultorio del especialista para las tareas a realizar.

REFERENCIAS

[1] http://www.fon.hum.uva.nl/praat/

[2] http://www.kayelemetrics.com/

[3] P.D. Agüero, J.C. Tulli, G. Moscardi, E. González y A.J. Uriz, "Estimating RASATI scores using acoustical parameters". *Journal of Physics: Conference Series (JPCS)*. Diciembre 2011.

[4] M.G. Kuzman, P.D. Agüero, J.C. Tulli, E.L. González, A.J. Uriz y M.P. Cervellini. "Development of a voice database to aid children with hearing impairments". *Journal of Physics: Conference Series (JPCS)*. Diciembre 2011.

[5] A.J. Uriz, P.D. Agüero, J.C Tulli, J. Castiñeira Moreira y E.L. González, "Implementation of a noise reduction algorithm in a hearing aid device based on a dsPIC" . In Proceedings of IEEE ARGENCON 2012. 2012.

[6] H. G. Gauch, "Noise Reduction By Eigenvector Ordinations", *Ecological Society of America*, Vol. 63, No. 6, pp. 1643-1649, 1982.

[7] E. V. de Payer, "Preprocesado de la señal de voz: el método de la descomposición de subespacios", *Revista de la Sociedad Argentina de Bioingeniería*, Vol.16, No.1. June 2010.

[8] V. Balakrishnan, N. Borgesa and L. Parchment, "Wavelet Denoising and Speech Enhancement", 2006.

[9] T. Young y W. Qiang, "The realization of Wavelet Threshold noise filtering Algorithm", *In Proc. of 2010 Conference on Measuring Technology and Mechatronics Automation.* pp 953-956. 2010.

[10] Ch. Dolabdjian, J. Fadili y E. Huertas Leyva, "Classical low-pass filter and real-time wavelet-based denoising technique implemented on a DSP: a comparison study", *The European Physical Journal Applied Physics*. Vol.20, pp 135-140. 2002.

[11] A. K. Tellakula, "Acoustic Source Localization Using Time Delay Estimation", Degree Thesis. Bangalore, India: Supercomputer Education and Research Centre Indian Institute of Science, 2007.

[12] C-MEDIA Electronics Inc. "CM108 High Integrated USB Audio I/O Controller DataSheet" *http://www.cmedia.com.tw/*