2012 XVII Symposium of Image, Signal Processing, and Artificial Vision (STSIVA)

COMPENSATION FOR VOICE-PRODUCING DEVICES FOR PEOPLE WITH ORAL COMMUNICATION DIFFICULTIES

Javier Gonzàlez[1], Paula Murgas[2]
1,2 Universidad Santo Tomás
Cra 9 N°51 – 11. Bogotá Colombia
Javier.gonzalez@ieee.org
Paula_murgas@hotmail.com

***Abstract*- This article presents the results obtained using adaptive digital filter FIR to compensate the spectral components of the human voice signal attenuated to be acquired in experiments that simulate amplification devices for people with oral communication difficulties. The signals used are human voice recordings taken at the level of the throat in young subjects as a reference signal and a voice signal acquired from the same subject at the mouth. The results obtained show that the adaptation process provides a set of coefficients that are the basis for designing FIR digital filter with the ability to make the voice signal level of the throat and compensate for attenuated components.**

***Keys*—adaptive filter, voice processing.**

I. INTRODUCCIÓN

La voz es el producto de la interacción de diferentes elementos que componen el sistema respiratorio humano, compuesto principalmente por los pulmones y las vías respiratorias [1,2,3]. El margen habitual para locutores masculinos adultos del valor del pitch es de 50 a 250 Hz, mientras que para el género femenino se tienen valores entre 120 y 500 Hz [4]. Estos parámetros, han sido caracterizados para locutores de procedencia rusa, suiza y norteamericana desde 1965 [5] y estos resultados han sido de referencia para múltiples estudios de características de la voz humana [6,7].

Para el análisis matemático para la extracción y estudio de los parámetros de la voz, se han registrado diversas metodologías [8,9,10,11], las cuales juegan un papel importante en el diseño e implementación de dispositivos de apoyo a personas con dificultades de comunicación oral [12]. Dentro de estos dispositivos se encuentran los micrófonos de garganta, que son diseñados a partir de elementos con componentes resistivas e inductivas, los cuales han sido utilizados para estudiar parámetros de la voz humana [13] y los resultados de estos estudios han sido comparados con los obtenidos por medio de micrófonos convencionales, que adquieren la señal de voz a nivel de la garganta de los sujetos bajo estudio [14,15,16].

El uso de micrófonos de garganta, pueden provocar distorsiones de la señal de voz [17], por lo cual el presente trabajo se enfoca en el diseño de un algoritmo para la compensación de las componentes atenuadas por estos dispositivos.

II. MATERIALES Y MÉTODOS

La recolección de muestras fue tomada de una población de 10 jóvenes, de ambos géneros entre 20 y 28 años. La población contiene igual cantidad de sujetos del género femenino y masculino. Los datos fueron recolectados a través de la tarjeta de sonido de un computador personal con frecuencia de muestreo de 44.100 Hz y con resolución de 16 bits. Las muestras adquiridas fueron tomadas en la ciudad de Bogotá a una temperatura promedio de 15 grados centígrados y en un recinto aislado del ruido. Para la recolección de las muestras de voz se utilizaron dos modos diferentes para cada sujeto: el primero a través de un micrófono que capte la señal de la boca directamente del sujeto (método directo) y el segundo modo se realizó con el mismo micrófono pero la señal fue tomada a la altura de la garganta del sujeto (método indirecto) mediante un fonendoscopio, con la finalidad de simular las condiciones similares a las de un micrófono de garganta.

El segundo modo permitió captar la información en medio similar al utilizado por sistemas de apoyo denominados: micrófono de garganta. Estos dispositivos se encargan de amplificar las vibraciones captadas a la altura de la garganta para mejorar el desempeño de la comunicación oral.

Los espectros calculados fueron promediados con la finalidad de obtener la caracterización para cada vocal en el dominio de la frecuencia. Con estos espectros promediados se obtuvo la herramienta para la realización de comparaciones entre la muestra de género femenino y masculino, con las respectivas diferencias de información para el modo directo e indirecto de adquisición.

La figura 1 ilustra el espectro promediado de la señal de voz obtenida al pronunciar la vocal a por parte de un sujeto femenino para los dos casos de adquisición. El resultado contenido en la Fig. 1, deja evidenciar que al realizar la adquisición por medio del método indirecto, se experimenta un efecto pasa bajos que atenúa las componentes espectrales

978-1-4673-2761-9/12/$31.00 ©2012 IEEE

superiores a 400 Hz. Estas componentes son de gran importancia porque hacen parte de los armónicos relacionados con la frecuencia fundamental de los fonemas pronunciados.

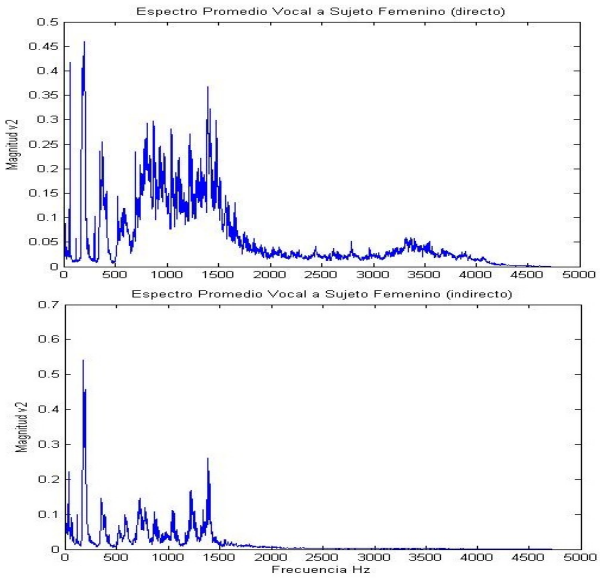

Fig. 1. Espectro en magnitud promedio para la vocal a de sujetos femeninos con los dos métodos de adquisición.

Para el caso de sujetos masculinos, también se evidenció el efecto de atenuación de componentes espectrales por encima de 400 Hz. El resultado obtenido se ilustra en Fig 2.

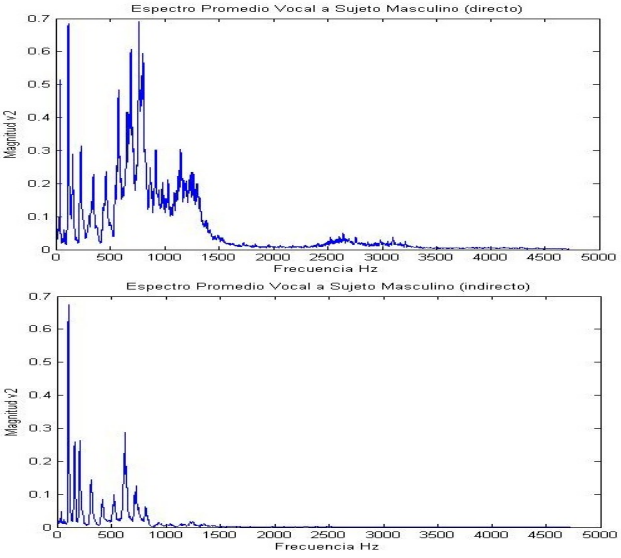

Fig. 2. Espectro en magnitud promedio para la vocal a de sujetos masculinos con los dos métodos de adquisición.

Con la finalidad de poder compensar las atenuaciones provocadas por la adquisición de voz a través del método indirecto, se propone la aplicación basada en la arquitectura de un filtro FIR adaptativo (Fig 3).

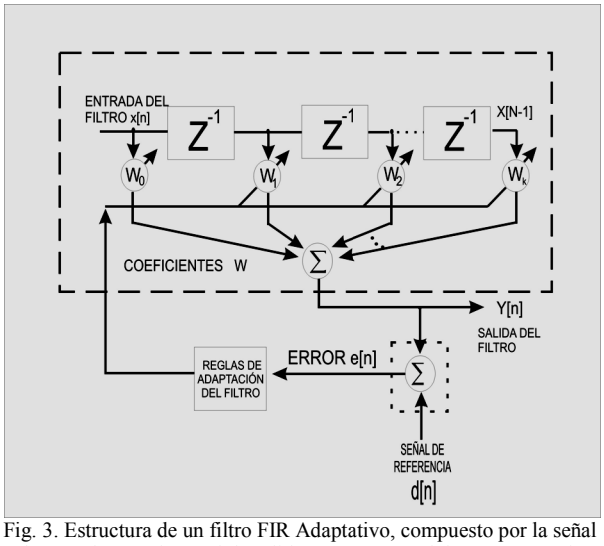

Fig. 3. Estructura de un filtro FIR Adaptativo, compuesto por la señal de entrada x[n], la señal de salida y[n], una señal de error e[n] y la señal deseada d[n].

La arquitectura del filtro FIR presentado en Fig 3 está compuesta por una señal de entrada x[n] la cual contiene la señal adquirida con componentes de la señal original. La señal d[n], denominada señal de referencia contiene la señal ideal que se desea obtener con el proceso de adaptación y la señal y[n] es la señal de salida del filtro. La regla de adaptación está conformada por la señal de error e[n] que equivale a la diferencia entre la señales y[n] y d[n]. Con base a la señal de error e[n], se realiza la actualización de los coeficientes W_k según la ecuación 1.

$$W(k) = W(k) + \mu.e(n).x(n-k) \qquad (1)$$

Siendo μ el coeficiente de adaptación ($0<\mu<1$), el cual se puede calcular a partir de las muestras de la señal de entrada x(n). Ver ecuación 2.

$$\mu = \frac{1}{2*\sum_{n=0}^{N-1}|x(n)|^2} \qquad (2)$$

El filtro FIR adaptativo utilizado fue diseñado con una cantidad de 20 coeficientes, que es inicializado con el valor de cero para cada uno y el valor de uno para el primer coeficiente. En este caso la señal de entrada x[n] corresponde la señal de voz tomada por el método indirecto y la señal deseada d[n] es la señal tomada por el método directo. La salida del sistema es un conjunto de coeficientes que permiten diseñar un filtro digital FIR que compense las atenuaciones de la voz.

III. RESULTADOS

Tomando las señales de voz, se ha procedido a realizar el proceso de adaptación del filtro. La figura 4 ilustra las señales de entrada y salida utilizadas para el caso de la señal de voz femenina.

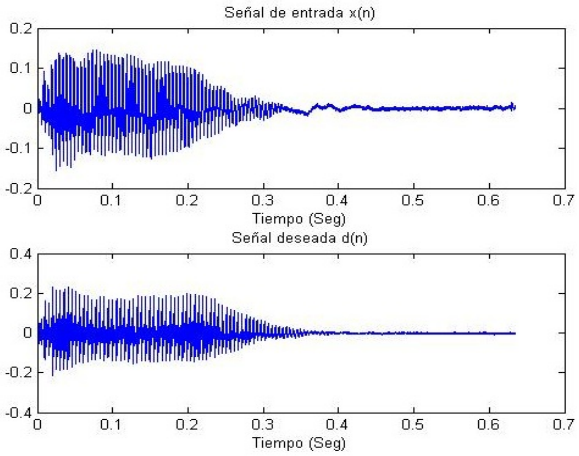

Fig. 4. Ilustración en el dominio del tiempo de la señal de entrada x[n] y señal deseada d[n] utilizadas para el proceso de adaptación del filtro.

El proceso de adaptación ofrece como resultado los coeficientes que permiten la implementación de un filtro digital tipo FIR. La figura 5 contiene el análisis en el dominio de la frecuencia del filtro digital obtenido. El comportamiento en el dominio de la frecuencia del filtro demuestra que el sistema obtenido es similar a un filtro pasa altas con frecuencia de corte igual a 400 Hz.

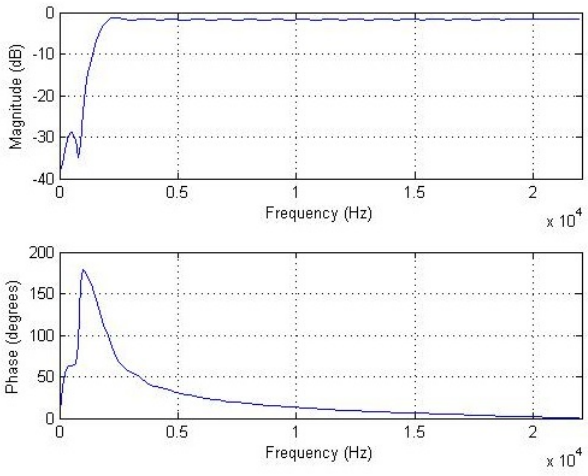

Fig. 5. Respuesta en frecuencia del filtro digital FIR con los coeficientes que se han obtenido en el proceso de adaptación.

Con los coeficientes se ha implementado el filtro digital y se ha puesto a prueba con la señal de entrada. La figura 6 ilustra la señal filtrada, en el dominio del tiempo.

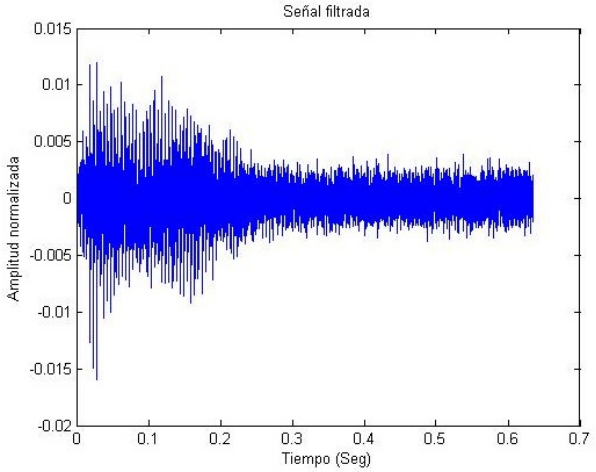

Fig. 6. Señal filtrada y[n], en el dominio del tiempo.

La figura 7 ilustra una comparación del espectro en magnitud de la señal de entrada x[n], la señal deseada d[n] y la señal filtrada y[n]. En esta comparación se puede ver que la señal de entrada x[n] posee atenuaciones en las componentes superiores a 400 Hz, lo cual es consecuencia del método indirecto que fue utilizado para su adquisición.

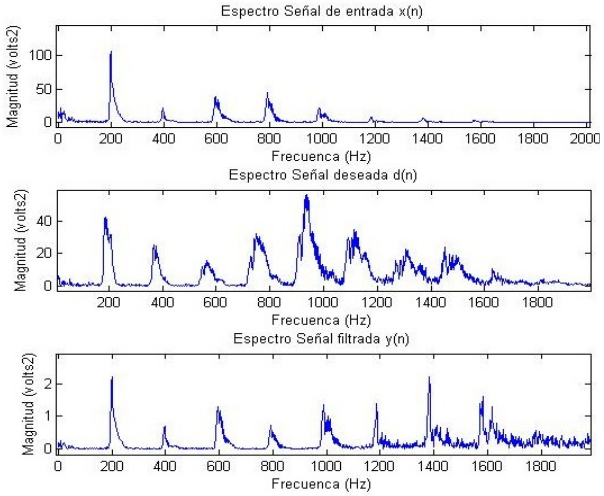

Fig. 7. Espectros en magnitud de las señales de entrada x[n], deseada d[n] y filtrada y[n].

Como se puede ver en la figura 7, la señal filtrada y[n] posee un proceso de compensación en las componentes superiores a 400 Hz. Esto evidencia el funcionamiento del filtro digital FIR como sistema digital de compensación de las componentes espectrales atenuadas.

IV. CONCLUSIONES

La importancia de la realización de este trabajo radica en estudiar el comportamiento de las componentes espectrales de la voz humana dependiendo del método de adquisición. Se ha evidenciado que los dispositivos para amplificar la voz (micrófonos de garganta) alteran las componentes espectrales actuando como un filtro pasa bajos. Para poder conservar las componentes espectrales de la voz, es necesario implementar una estrategia que permita compensar estas atenuaciones.

El resultado de este trabajo presenta la implementación de un algoritmo para la compensación de estas atenuaciones, diseñado a partir de la arquitectura de un filtro adaptativo tipo FIR. Los resultados evidenciaron la efectividad del método propuesto al lograr la compensación de las componentes atenuadas y la recuperación del espectro original de la seña de voz. El método presentado en este trabajo, es de gran utilidad por ser un filtro digital FIR y de fácil implementación en un sistema digital programable.

RECONOCIMIENTOS

Este trabajo ha sido el logro de estudiantes miembros del Semillero del Grupo de Investigación en Procesamiento Digital de Señales de la Facultad de Ingeniería Electrónica de la Universidad Santo Tomás.

REFERENCIAS

[1] Koeppen B, Stanton B. Fisiología. Elsevier Mosby Ed. España 2009. 833, pp417429.

[2] Guyton A, Hall J. Physiology. Elsevier Mosby Ed. China 2006. 1052: pp471-501.

[3] Purves D. Neurosciencie. Sinauer Associetes Ed. USA 2004. 710, p637-p641.

[4] Faúndez M. Tratamiento Digital de Voz e Imagen. Alga Omega Ed. México 2001. 271, pp31-32.

[5] Ohman, S.E.Coarticulation in VCV utterances: spectrographic measurements.(1966) Journal of the Acoustical Society of America, 39 (1), pp. 151-168.

[6] Megha Sundara, Linda Polka, Discrimination of coronal stops by bilingual adults: The timing and nature of language interaction, Cognition, Volume 106, Issue 1, January 2008, Pages 234-258.

[7] GibsonT.,Ohde R. F2 Locus Equations: Phonetic Descriptors of Coarticulation in 17- to 22-Month-Old Children. Journal of Speech, Language, and Hearing Research. February 2007. Vol. 50 • 97–108.

[8] Umesh, S.; Sinha, R.; , "A Study of Filter Bank Smoothing in MFCC Features for Recognition of Children's Speech," Audio, Speech, and Language Processing, IEEE Transactions on , vol.15, no.8, pp.2418-2430, Nov. 2007.

[9] de la Torre, A.; Segura, J.C.; Benitez, C.; Peinado, A.M.; Rubio, A.J.; , "Non-linear transformations of the feature space for robust speech recognition," Acoustics, Speech, and Signal Processing, 2002. Proceedings. (ICASSP '02). IEEE International Conference on , vol.1, no., pp. I-401- I-404 vol.1, 2002.

[10] Xiaodong Cui; Yifan Gong; , "Variable parameter Gaussian mixture hidden Markov modeling for speech recognition," Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03). 2003 IEEE International Conference on , vol.1, no., pp. I-12- I-15 vol.1, 6-10 April 2003.

[11] Padmanabhan, M.; Saon, G.; Zweig, G.; Huang, J.; Kingsbury, B.; Mangu, L.; , "Evolution of the performance of automatic speech recognition algorithms in transcribing conversational telephone speech," Instrumentation and Measurement Technology Conference, 2001. IMTC 2001. Proceedings of the 18th IEEE , vol.3, no., pp.1926-1931 vol.3, 2001.

[12] Ben Mosbah, B.; "Speech Recognition for Disabilities People," Information and Communication Technologies, 2006. ICTTA '06. 2nd , vol.1, pp.864-869.

[13] Mubeen, Nafeesa; Shahina, A.; Khan, A. Nayeemulla; Vinoth, G.; , "Combining spectral features of standard and Throat Microphones for speaker identification," Recent Trends In Information Technology (ICRTIT), 2012 International Conference on , pp.119-122, 19-21 April 2012.

[14] Graciarena, M.; Franco, H.; Sonmez, K.; Bratt, H.; , "Combining standard and throat microphones for robust speech recognition," Signal Processing Letters, IEEE , vol.10, no.3, pp.72-74, March 2003.

[15] Akargun, U.C.; Erzin, E.; , "Estimation of Acoustic Microphone Vocal Tract Parameters from Throat Microphone Recordings," Signal Processing and Communications Applications, 2007. SIU 2007. IEEE 15th , pp.1-4, 11-13 June 2007.

[16] Erzin, E.; , "Improving Throat Microphone Speech Recognition by Joint Analysis of Throat and Acoustic Microphone Recordings," Audio, Speech, and Language Processing, IEEE Transactions on , vol.17, no.7, pp.1316-1324, Sept. 2009.

[17] Shahina, A.; Yegnanarayana, B.; , "Language identification in noisy environments using throat microphone signals," Intelligent Sensing and Information Processing, 2005. Proceedings of 2005 International Conference, pp. 400- 403, 4-7 Jan. 2005.