

Una aproximación al filtrado adaptativo para la cancelación de ruidos en señales de voz monofónicas

David Ortiz Puerta*. Olga L. Quintero Montoya**

*Departamento de ciencias Básicas, Universidad EAFIT, Medellín, Antioquia
e-mail: dpuerta1@eafit.edu.co.

** Departamento de ciencias Básicas, Universidad EAFIT, Medellín, Antioquia
e-mail: oquinte1@eafit.edu.co.

Resumen: Generalmente, el uso de filtros adaptativos para la cancelación de ruidos presentes en señales de voz implica el uso de dos canales de audio donde se presente la señal a filtrar y una referencia del ruido. Dentro del trabajo se propuso una metodología de filtrado de señales monofónicas donde a partir de la transformada de Hilbert se extraen los ruidos presentes en los silencios de la prosodia y se utilizan como referencia para la adaptación del filtro utilizando gradiente descendente. Finalmente se realizaron dos pruebas agregando ruidos artificiales adicionales a los ruidos naturales sobre los audios a filtrar.

Palabras Clave: Cancelación de ruidos, Filtros adaptativos, Transformada de Hilbert, señal monofónica.

1. INTRODUCCIÓN

Un ruido se puede describir como una perturbación que interfiere dentro de la señal muestreada alterando sus valores reales. Los ruidos pueden ser producidos por factores que van desde cortos en la electrónica de los micrófonos hasta ruidos producidos por emisores cercanos como la voz de otros hablantes a la hora de la obtención de la señal. Debido a que los ruidos pueden cambiar dependiendo del ambiente donde fueron obtenidos, es difícil diseñar filtros duros que eliminen completamente las frecuencias dentro del espectro de la perturbación, esto se debe a que los filtros pueden dejar intacta las amplitudes de las frecuencias indeseadas, fenómeno conocido como solapamiento espectral [1]. Es necesario diseñar filtros que se adapten únicamente a las características espectrales de las perturbaciones dentro de la señal obtenida de tal forma que la señal original quede intacta. Este proceso se conoce como cancelación de ruido.

Un filtro se conoce como un dispositivo el cual procesa señales de entrada modificando ciertas propiedades dentro de esta. El proceso de adaptación se puede definir como el cambio de las características propias de un mecanismo a partir de un conocimiento adquirido en el ambiente en el cual se desempeña. El mezclar estas dos terminologías apunta a un dispositivo que puede modificar sus características basado en información específica de las entradas, de tal forma que las características de la señal de salida se aproximen a las de una señal de referencia deseada. El campo de los filtros adaptativos hace parte del campo de la teoría de optimización [2].

En un filtro adaptativo hay dos procesos básicos, filtrado de la señal de referencia del ruido y la adaptación del filtro. Generalmente en el proceso de filtrado se utilizan filtros con estructuras *Finite Impulse Response* (FIR) por sus siglas en inglés, que son simples y estables. El proceso de adaptación ajusta los parámetros del filtro FIR de acuerdo a un algoritmo

de adaptación el cual depende del error entre la señal de entrada del filtro y la deseada [2][3]. Existen muchos algoritmos de adaptación para la cancelación de ruidos. El más utilizado por su simpleza y su baja complejidad computacional es mínimos cuadrados (*Least Mean Square*, LMS) por sus siglas en inglés [4]. Otros métodos de adaptación utilizados son los mínimos cuadrados recursivos (*Recursive Least Square*, RLS), Algoritmos de proyecciones afines rápidas (*Fast Affine Projection*, FAP) y algoritmo con dirección euclidiana de búsqueda rápida (*Fast Euclidean Direction Search*, FEDS), que son versiones modificadas del método LMS. Estos métodos dan ventajas y desventajas sobre la velocidad de convergencia y su complejidad computacional [5][6][7].

En general, los filtros adaptativos para cancelación de ruidos utilizan la estructura que se presenta en la Fig. 1 donde al sistema de filtrado entran dos señales, una que será la señal principal a filtrar y la otra será una señal de referencia del ruido presente en la señal principal [8]. Conociendo esta estructura general de filtrado, se plantean las siguientes hipótesis: ¿Es posible adaptar la estructura de un filtro de tal forma que se logren eliminar los ruidos presentes en la señal? Dado el caso de no tener una señal de referencia del ruido ¿Es posible realizar una adaptación exitosa del filtro aun cuando la información espectral del ruido es poca o nula?

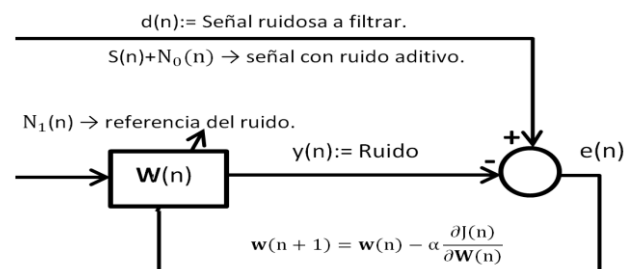


Fig. 1. Estructura general del filtro adaptativo.

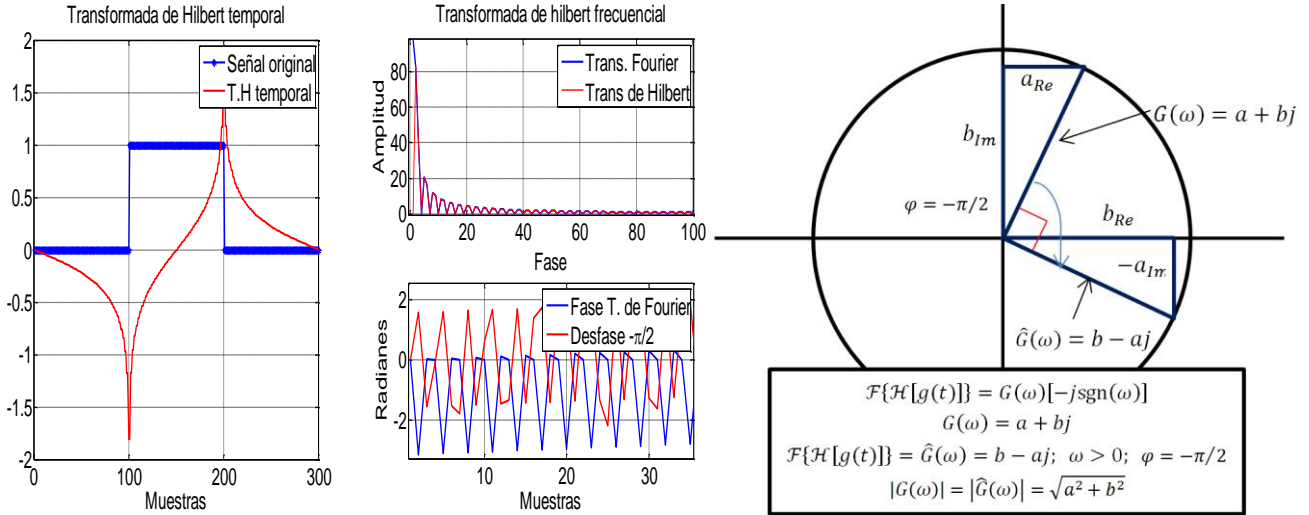


Fig. 2. A la izquierda transformada de Hilbert de una señal cuadrada, a la derecha, fenómeno de desfase producido por la transformada de Hilbert.

Este trabajo está centrado en el filtrado de señales de conversaciones telefónicas las cuales están formadas por un solo canal de información. Teniendo como factor limitante la característica monofónica de estas señales, ¿Es posible diseñar un filtro el cual sea capaz de extraer información del espectro del ruido dentro de la conversación y adaptar su estructura para lograr eliminar el ruido sin modificar la señal real? Dentro del trabajo se propone la extracción de los silencios presentes en las unidades prosódicas de la conversación los cuales contienen información sobre los ruidos adquiridos en la señal. De esta forma adaptar la estructura del filtro utilizando el método de mínimos cuadrados.

2. METODOLOGÍA

Como se mencionó anteriormente, la estructura general de los filtros adaptativos está conformada por dos señales de entrada, una de ellas contiene información relacionada con las perturbaciones presentes en la señal original. Teniendo en cuenta que el enfoque del trabajo son señales monofónicas donde no se tiene información directa del ruido presente en la señal original, se propuso una estrategia de filtrado en dos pasos, primero se identifica los silencios sobre la señal de audio y luego a partir de esta información realizar la adaptación del filtro.

2.1 Obtención de silencios:

Para la obtención de los silencios, se utilizó la transformada de Hilbert como herramienta para definir el cubrimiento de una señal. La transformada de Hilbert de una función $g(t)$ se define como

$$\mathcal{H}[g(t)] = g(t) * \frac{1}{\pi t} = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{g(t - \tau)}{\tau} d\tau \quad (1)$$

Utilizando la transformada de Fourier de (1), la transformada de Hilbert en frecuencia se define como

$$\mathcal{F}\{\mathcal{H}[g(t)]\} = \mathcal{F}\left\{g(t) * \frac{1}{\pi t}\right\} = \mathcal{F}\{g(t)\}\mathcal{F}\left\{\frac{1}{\pi t}\right\} \quad (2)$$

De (2) se puede ver que

$$\mathcal{F}\left\{\frac{1}{\pi t}\right\} = -j\text{sgn}(\omega) = \begin{cases} -j & \omega > 0 \\ 0 & \omega = 0 \\ j & \omega < 0 \end{cases} \quad (3)$$

Reemplazando (3) en (2)

$$\mathcal{F}\{\mathcal{H}[g(t)]\} = G(\omega)[-j\text{sgn}(\omega)] \quad (4)$$

La ecuación (4) genera un desfase de $-\pi/2$ en el dominio de la frecuencia de la función $g(t)$ sin modificar la amplitud como se puede ver en la Fig 2. La transformada de Fourier permite disminuir costo computacional de la convolución en la transformada de Hilbert, simplificándola en una operación de producto. Una vez realizado el cálculo se aplica transformada inversa de Fourier para obtener la señal en el dominio temporal nuevamente.

$$\hat{g}(t) = \mathcal{F}^{-1}\{G(\omega)[-j\text{sgn}(\omega)]\} \quad (5)$$

Una vez definida la transformada de Hilbert temporal, se construye la señal analítica definida como

$$\psi(t) = g(t) + j\hat{g}(t) \quad (6)$$

Donde $g(t)$ es la señal original y $\hat{g}(t)$ es la transformada de Hilbert de la señal. Esta señal analítica contiene la información de la señal original y el desfase producido por la transformada de Hilbert.

El cubrimiento de la señal se puede obtener como el módulo de la señal analítica (6)

$$|\psi(t)| = \sqrt{g(t)^2 + \hat{g}(t)^2} \quad (7)$$

Finalmente realizando una decimación sobre (7) disminuye la resolución sobre el cubrimiento de la señal analítica

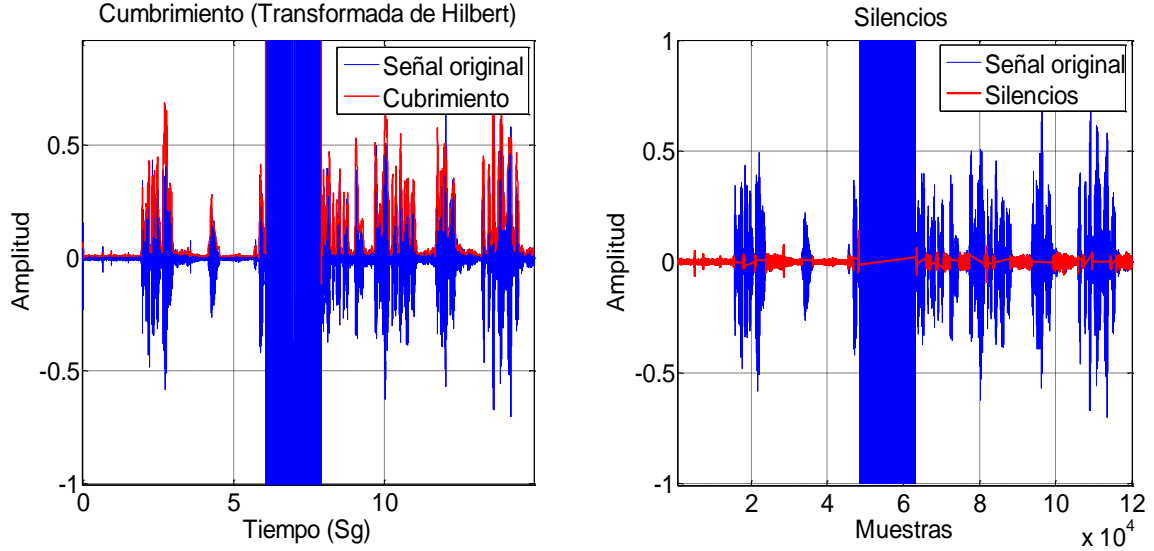


Fig. 3. A la derecha, cubrimiento de la señal de audio, a la izquierda, silencios detectados en la señal de audio utilizando transformada de Hilbert.

permitiendo ver de forma más global el cubrimiento de la señal como se observa en la Fig 3. Una vez obtenido el cubrimiento, se determina un umbral el cual discrimina las unidades prosódicas del hablante con los silencios en la conversación.

2.2 Filtrado Adaptativo

Se define $\mathbf{W}(n)$ como el conjunto de parámetros del filtro

$$\mathbf{W}(n) = [w_0(n) \dots w_{L-1}(n)]$$

Donde $\{w_i(n)\}$ para $0 \leq i \leq L-1$, son los L parámetros a adaptar. Para el filtrado adaptativo se propuso una estructura FIR definida como

$$y(n) = \sum_{i=0}^{L-1} w_i(n)x(n-i) \quad (8)$$

Donde $x(n)$ será la señal de referencia del ruido. La idea del algoritmo es estimar los parámetros w_i de tal forma que la diferencia de la señal de salida del filtro con la señal deseada a filtrar sea mínima. La salida estimada por el filtro se obtiene como

$$\hat{d}(n) = \sum_{i=0}^{L-1} \hat{w}_i(n)x(n-i) \quad (9)$$

el error está determinado por

$$e(n) = d(n) - \hat{d}(n) \quad (10)$$

El objetivo es encontrar los $\hat{w}_i(n)$ que minimicen (10). El objetivo del algoritmo del gradiente descendente es variar los coeficientes del filtro en función del error de tal forma que la función de coste sea mínima, para esto se tiene

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \mu \frac{\partial J(n)}{\partial \mathbf{W}(n)} \quad (11)$$

Donde $\mathbf{w}(n)$ es el vector de parámetros del filtro y $J(n)$ la función del error a minimizar. Esta expresión es general para cualquier función de coste. Para el trabajo se definió como función de coste (10). El vector $\partial J(n)/\partial \mathbf{W}(n)$ es conocido como el gradiente de la superficie del error y la constante μ es la constante de proporcionalidad [9].

Para poder implementar el algoritmo del gradiente descendente es necesario evaluar la ecuación diferencial parcial en (11) con respecto a la función de coste deseada. Para el caso de la función de coste (10), se tiene

$$\begin{aligned} \frac{\partial e^2(n)}{\partial \mathbf{W}(n)} &= \frac{\partial e^2(n)}{\partial \mathbf{W}(n)} \\ &= 2e(n) \frac{\partial e(n)}{\partial \mathbf{W}(n)} \\ &= 2e(n) \frac{\partial (d(n) - \mathbf{W}^T(n)\mathbf{X}(n))}{\partial \mathbf{W}(n)} \\ &= -2e(n)\mathbf{X}(n) \end{aligned} \quad (12)$$

Reemplazando (12) en (11)

$$\mathbf{w}(n+1) = \mathbf{w}(n) + 2\mu e(n)\mathbf{X}(n) \quad (13)$$

Otro aspecto importante a tener en cuenta en el método del gradiente descendente es su convergencia, esta está dada por el parámetro μ en la ecuación (13). Para un filtro de orden L el algoritmo será estable si se cumple la siguiente condición [3]

$$0 < \mu < 1/(10LP_{xx})$$

Donde

$$P_{xx} \approx \frac{1}{M+1} \sum_{n=0}^{M-1} x^2(n)$$

Es la potencia total de la señal de entrada.

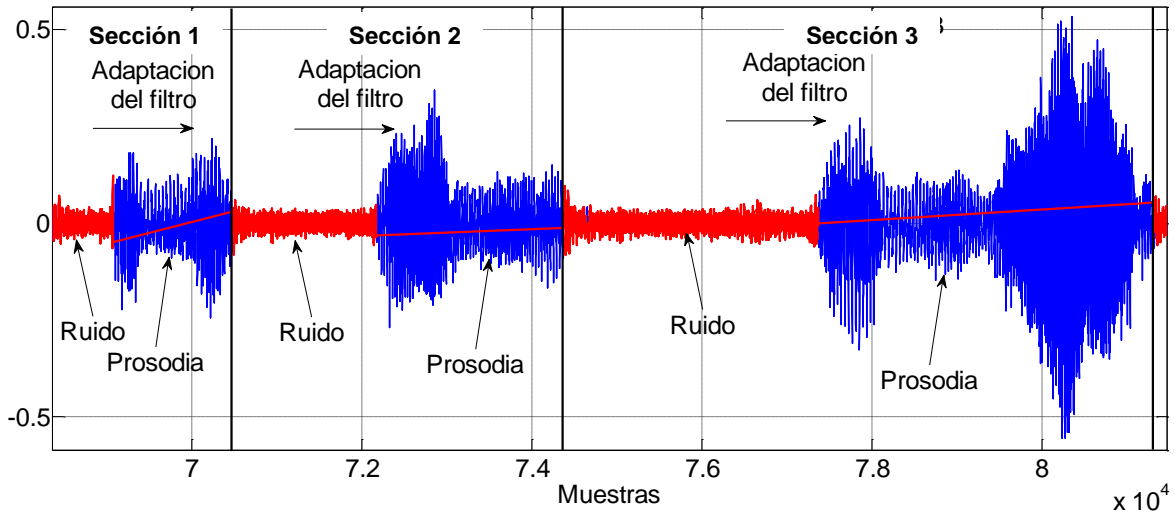


Fig. 4. Metodología de adaptación del filtro para señales monofónicas.

2.3 Procedimiento de filtrado

Una vez definido el método de obtención de silencios y el algoritmo para la adaptación del filtro, se estableció la implementación de los dos.

Dado un análisis global de la señal, se propone hacer un filtrado por secciones donde cada sección está compuesta por un silencio prosódico con la información del ruido y una unidad prosódica con la información de voz (Fig. 4.). Para la formación de estos segmentos se realiza el siguiente procedimiento.

1. Primero se identifican los silencios dentro de la señal de audio y las unidades de la prosodia que contienen la información de voz de los emisores.
2. Cuando se identifica un intervalo de silencio, se busca la unidad prosódica que le sigue y se adapta el filtro a partir de la información del ruido presente en el silencio. En este caso, la señal deseada $d(n)$ será la unidad prosódica y la señal de silencio será la entrada del filtro. En la Fig. 4 se puede observar la metodología de filtrado, donde los ruidos se identifican de color rojo y las unidades prosódicas a filtrar tienen color azul.
3. En ocasiones la señal de referencia del ruido para adaptar el filtro es menor que la señal de la prosodia, en este caso se repite varias veces la muestra hasta terminar de filtrar la sección.

3. RESULTADOS Y ANALISIS

3.3 Pruebas

Para las pruebas del filtro se utilizaron audios con calidad telefónica muestreados a 8000 Hz. A pesar que estos audios ya incluyen sus propios ruidos, se le adicionaron ruidos artificiales dentro de la banda de frecuencias principales de la voz (100 Hz a 3200 Hz) para generar de esta forma un

solapamiento espectral, principal problema al eliminar ruidos con filtros duros. Se utilizaron 3 audios de prueba a los cuales se les introdujo una señal senoidal a 300 Hz con una amplitud de -16.98 dB como se puede ver en la Fig. 5.

Se utilizó un filtro adaptativo de orden 100, se estableció como constante de proporcionalidad μ igual a 0.2 y un reinicio en cero de los coeficientes del filtro cada 4 secciones de filtrado. Además de estas condiciones se realizó un pre-filtrado de las señales utilizando un filtro 'butterworth' de orden 6 pasa-banda con frecuencia de corte entre 100 Hz y 3200 Hz con la finalidad de eliminar los posibles ruidos que no se superponen a las frecuencias principales de la voz minimizando el esfuerzo de adaptación del filtro.

Se realizó una segunda prueba donde a las mismas señales filtradas se les adiciono ruido coloreado presentado en la Fig. 6. Se puede observar que las frecuencias del ruido están entre 1200 Hz y 2000 Hz y tiene una amplitud de -10 dB aproximadamente.

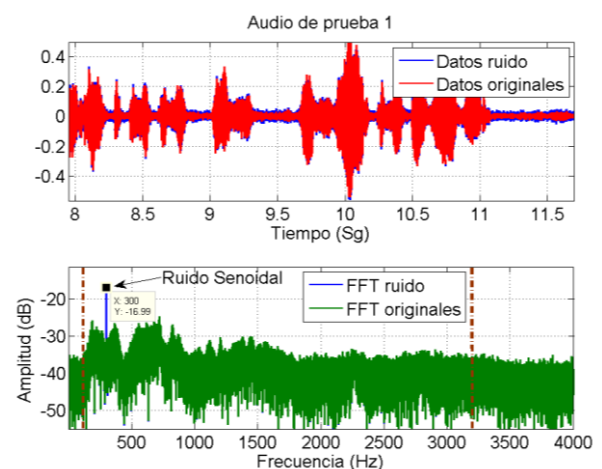


Fig. 5. Audios de prueba, se compara los datos originales con los ruidosos en el dominio temporal y frecuencial.

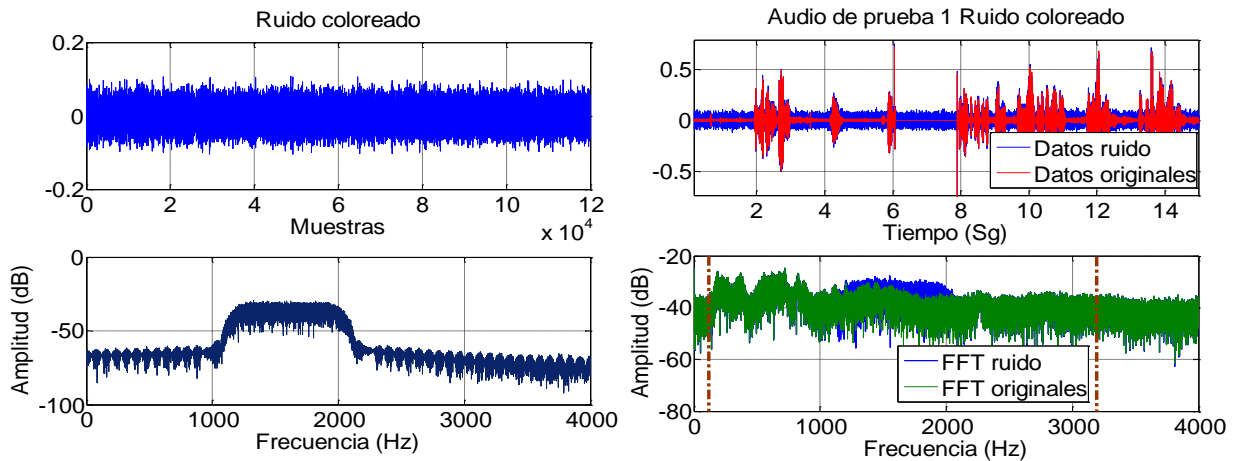


Fig. 6. Ruido coloreado y audio para la segunda prueba.

3.4 Resultados y análisis

Uno de los criterios que se debe tener en cuenta para la evaluación de la efectividad del filtro adaptativo es la capacidad de dejar intactas las frecuencias que si hacen parte de la señal original. En la Fig. 7 se pueden observar los resultados para la primera prueba realizados sobre el primer audio.

Al analizar el cuadrante inferior izquierdo donde se compara el espectro de la señal ruidosa con la filtrada, es posible ver que este se superpone casi en su totalidad, lo que indica que la variación de las frecuencias importantes durante el filtrado de la señal no es mucha. Se resalta la atenuación en los extremos del espectro producidos por el pre-filtrado. Al

realizar un acercamiento al espectro de las señal ruidosa y filtrada sobre los 300 Hz del ruido senoidal se puede observar que la amplitud se redujo de -16.99 dB a -26.91 dB lo cual indica una atenuación total del ruido de -9.92 dB sobre toda la señal.

En la segunda imagen de la Fig. 7 se puede observar el resultado de la segunda prueba realizada sobre el primer audio. En este caso se observa la misma atenuación en los extremos realizada por el pre-filtrado. Además se observa que en las frecuencias en las que se localiza el ruido coloreado hay una disminución de la amplitud lo que indica que el filtro se logró adaptar a las características espectrales del ruido atenuándolo.

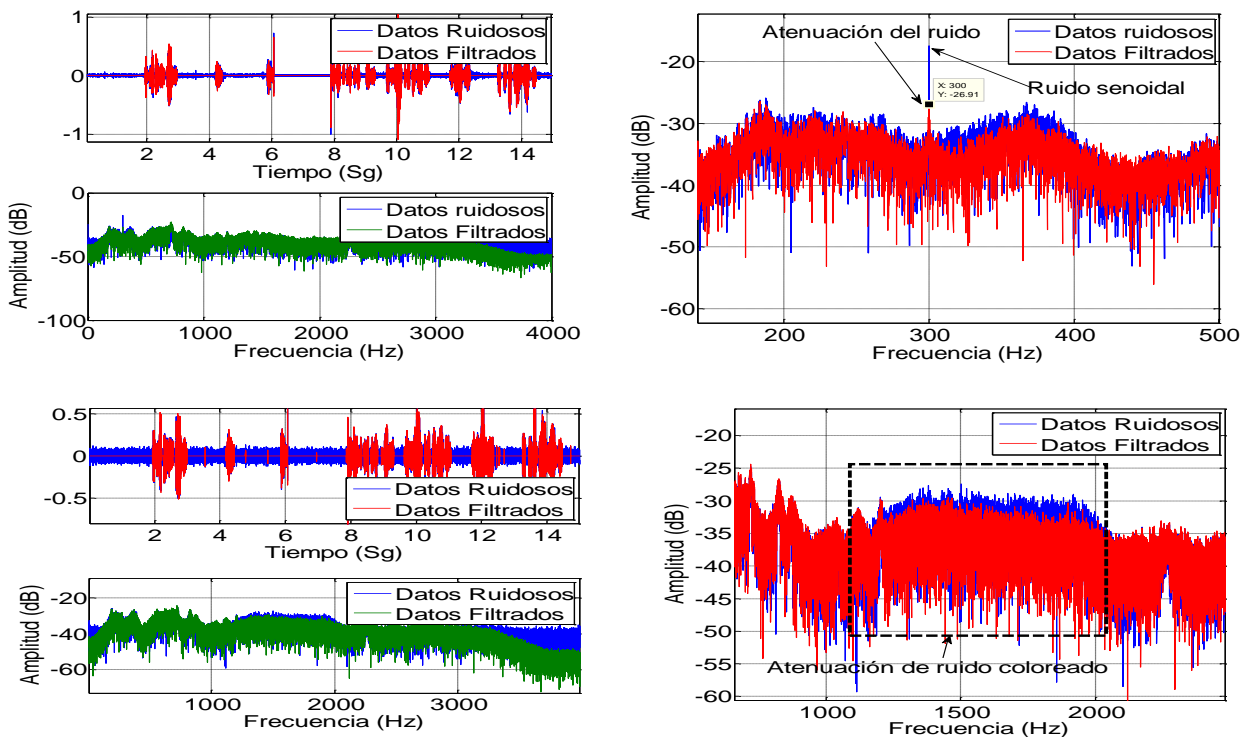


Fig. 7. Parte superior, resultados de primer audio para primera prueba, parte inferior, resultados para segunda prueba

Es posible que las otras secciones del espectro de la señal filtrada donde se observa disminución de la amplitud sean por los ruidos inherentes a la toma de los datos, esta atenuación indica que el filtro no solo se adaptó al espectro del ruido introducido artificialmente, sino que también se adaptó al ruido natural que se presenta.

A pesar de que el método del gradiente se presenta como un buen método de estimación para los coeficientes del filtro, se ha encontrado que no es el más rápido para converger [4]. Debido es posible que el filtro no se logre adaptar totalmente a las condiciones del ruido y es por tal motivo que la atenuación de los ruidos no es total.

4. CONCLUSIONES

Dados los resultados obtenidos, es posible realizar un filtrado exitoso de señales de audio monofónicas realizando la extracción de ruidos directamente dentro de la señal, de esta forma se puede responder a las hipótesis planteadas anteriormente.

A pesar que el método de mínimos cuadrados es una buena herramienta para la adaptación de los coeficientes del filtro, su velocidad de convergencia entra a jugar un papel importante dentro de la estructura de las señales a filtrar, es decir, señales con pocos silencios y unidades prosódicas extensas pueden generar que el filtro no se adapte correctamente a las características espectrales del ruido en la señal.

5. REFERENCIAS

- [1] M. A. Garcés Correa and E. Laciari Leber, "Noise removal from EEG signals in polysomnographic records applying adaptive filters in cascade," 2011.
- [2] V. K. Madiseti, *Adaptive Filtering*, Second Edi. Boca Raton: CRC Press, 2010, p. 906.
- [3] a G. Correa, E. Laciari, H. D. Patiño, and M. E. Valentinuzzi, "Artifact removal from EEG signals using adaptive filters in cascade," *J. Phys. Conf. Ser.*, vol. 90, p. 012081, Nov. 2007.
- [4] S. A. Hadei and M. Lotfizad, "A Family of Adaptive Filter Algorithms in Noise Cancellation for Speech Enhancement," *Int. J. Comput. Electr. Eng.*, vol. 2, no. 2, 2010.
- [5] J. Zeng and R. A. de Callafon, "RECURSIVE LEAST SQUARES GENERALIZED FIR CANCELLATION," *IFAC*, pp. 327–332, 2004.
- [6] S. L. Gay and S. Tavathia, "The fast affine projection algorithm," Murray Hill.
- [7] S. A. Hosseini, S. A. Hadei, M. B. Menhaj, and P. Azmi, "Fast Euclidean Direction Search Algorithm in Adaptive Noise Cancellation And System Identification.," *Int. J. Innov. Comput. Inf. Control*, vol. 9, 2013.
- [8] T. Drumright, "Adaptive Filtering," 1998. [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.192.1477&rep=rep1&type=pdf>.
- [9] V. J. Mathews and S. C. Douglas, "Adaptive Filters.," University of Utah, Salt Lake City.