

Métodos para Reducción de Ruido (II)

Jorge Torres Gómez¹, Emiliano Alba Blanco²

¹ *Departamento de Telecomunicaciones y Telemática. CUJAE jorge.tg@electrica.cujae.edu.cu*

² *Departamento de Bioingeniería. CEBIO.CUJAE. Grado científico del primer autor emiliano@electrica.cujae.edu.cu*

RESUMEN / ABSTRACT

El presente artículo muestra un análisis de los métodos empleados para la reducción de ruido en los canales de comunicación. Se muestran las ventajas y desventajas de cada variante y sus límites de aplicación.

Palabras claves: Filtro, Filtro Adaptativo, Reducción de Ruido.

Methods for Noise Reduction

The present article shows an analysis for methods used in noise reduction for communications channels. It shows the advantages and disadvantages of each variant and theirs limits of application.

Key words: Adaptive Filter, Filter, Noise Reduction.

Introducción

Comúnmente en las aplicaciones de reconocimiento del habla se necesita de la reducción de ruido para lograr una identificación más eficiente. Los primeros pasos están dados en el filtrado común de la señal contaminada, no obstante se hizo cada vez más necesario la reducción del ruido que aún queda compartiendo el ancho de banda de la señal útil.

Una señal recibida, contaminada con Ruido Aditivo, por ejemplo con Ruido Blanco, es afectada en su banda de frecuencia y fuera de esta por las componentes de ruido. Los filtros de banda analógicos y digitales separan intervalos de frecuencia, así lo hace un filtro pasabanda, permite el paso de una banda de frecuencia y el resto la rechaza. Este tipo de filtro diferencia la señal con respecto al ruido fuera de su banda útil y es capaz de separarla, ver figura 1. Es decir, se basa en el contenido de frecuencia diferente que hay entre el ruido fuera de banda y la señal útil.

No obstante, ¿qué sucede con el ruido en banda que aún sigue contaminando la señal útil? Claramente los filtros por sólo “saber” diferenciar señales por su contenido de frecuencia, no son capaces de seguir separando el ruido en banda de la señal útil. Ambos están ubicados en el mismo intervalo de frecuencia, hasta este punto llega la capacidad de eliminar ruido por los filtros de banda.

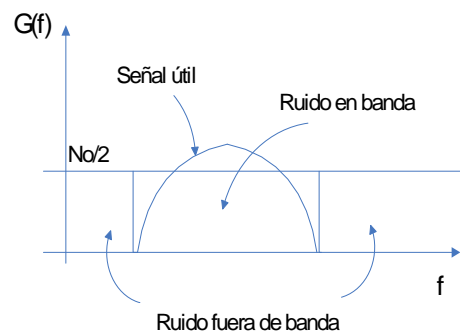


Figura 1. Gráfico de densidad espectral de Ruido Blanco Gaussiano y señal útil

El presente artículo da continuidad a lo explicado en Método de reducción de Ruido (I), ahora con una técnica basada en el empleo de filtros adaptativos.

MÉTODOS PARA DISMINUIR EL RUIDO EN BANDA

La cancelación de ruido también puede ser basada en la descripción estadística del modelo de ruido y de señal útil (1). Una característica general de las señales involucradas, señal útil y ruido, es su estadística de segundo orden. El filtrado adaptativo Wiener se basa en separar ambas señales por su diferencia en cuanto a la correlación cruzada de ambas. Si este cálculo da cero entonces se está en presencia de una condición idónea para emplear el método.

El principio de funcionamiento del filtrado adaptativo Wiener se explica a partir de cinco magnitudes, ver figura 2:

1. $u(0), u(1), u(2), \dots$: entrada del filtro.

2. w_0, w_1, w_2, \dots : coeficientes del filtro.
3. $d(n)$: denominada señal de referencia, es la respuesta deseada del filtro.
4. $y(n)$: respuesta del filtro.
5. $e(n)$: es la diferencia entre la salida del filtro y la señal deseada.

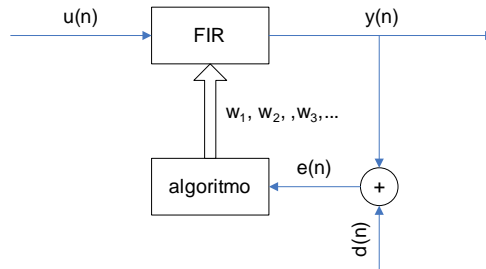


Figura 2. Esquema del filtro adaptativo.

La idea del filtrado es hacer $d(n)$ lo más parecido posible a $y(n)$. En este sentido es de esperar que con un buen diseño, calculando correctamente los coeficientes del filtro por parte del bloque algoritmo en la figura 2, para cada instante, $y(n)$ se aproxime a $d(n)$, de modo que la medida de similitud entre las señales estará dada por comparar dos valores: $y(n)$ y $d(n)$.

Para justificar la configuración mostrada en la figura 2, como método para aproximar $y(n)$ a $d(n)$, se considera lo siguiente: Todo proceso estocástico estacionario puede descomponerse en la suma de dos procesos: un primer proceso descrito por un modelo autoregresivo AR, portando la característica aleatoria, y un segundo que representa un proceso predecible, sus muestras futuras pueden estimarse de sus muestras pasadas (descomposición de Wold) (2), ver ecuación. (1)

$$x(n) = u(n) + s(n) \quad 1)$$

Donde: $x(n)$ denota el proceso estocástico en cuestión.

$u(n)$ denota el proceso descrito por un modelo AR,

$s(n)$ denota un proceso predecible.

El proceso $u(n)$ es descrito por un modelo AR:

$$u(n) = \sum_{k=1}^M w_k u(n-k) + v(n) \quad 2)$$

donde: $v(n)$ denota las muestras de Ruido Blanco Aditivo Gaussiano,

w_k denota los coeficientes del modelo.

Además, los coeficientes w_k se pueden encontrar mediante las ecuaciones de Yule-Walker2:

$$R\vec{w} = \vec{r} \quad 3)$$

Donde: R matriz de autocorrelación del proceso $u(n)$,

vector que representa las columnas de la matriz R.

La configuración que describe el funcionamiento de un filtro adaptativo, ver figura 2, apuesta por igualar dos procesos $y(n)$ y $d(n)$ por variación de los coeficientes del filtro FIR, lo cual está amparado en la ecuación (2), la idea es modificar el conjunto de coeficientes de $x(n)$ para obtener $d(n)$. Por ello la operación del filtro solo es válida cuando cierto grado de igualdad entre los procesos de entrada y señal deseada, lo cual queda identificado en la correlación cruzada de ambos.

Debido a que se trabaja en el conjunto de los números reales, este es un espacio métrico y por tanto se puede definir una distancia, comúnmente se trabaja con la distancia euclidiana:

$$e(n) = |d(n) - y(n)| \quad 4)$$

donde: $e(n)$ se le denomina señal error y da una medida de la operación del filtro,

$d(n)$ denota la señal deseada del filtro adaptativo,

$y(n)$ salida del filtro FIR.

Debido a que $d(n)$ e $y(n)$ son procesos aleatorios sería más apropiado especificar la condición (4) en términos de una característica general de ambos procesos, en vez del valor instantáneo de ambos. Una de estas características pudiera ser el valor esperado del error, pero esto traería como consecuencia que el criterio de comparación consistiría en decir: Cuando ambos procesos tengan igual valor esperado entonces compartirán la misma forma. Esta aseveración es falsa dado que dos procesos con diferente estadística pueden tener igual valor esperado siendo diferentes sus funciones de densidad de probabilidad, y esto dista mucho de encontrar que ambas señales se parezcan en el tiempo.

Así se puede calcular la estadística de segundo orden del error, como otra característica general de este proceso. Es decir, minimizar:

$$E[e(n)]^2 = E[d(n) - y(n)]^2 \quad 5)$$

Lo cual conduce a resultados más apropiados que (4) y que calcular el valor esperado de (4), dado que (5) además del comportamiento en la media, también responde por el comportamiento en la frecuencia 3.

Pero ¿cómo se pueden configurar la entrada y la respuesta deseada del filtro para obtener una medida satisfactoria?. La respuesta a la pregunta sería determinar donde se conecta la señal más ruido y contra que se va a comparar.

Del gráfico 2 se aprecia que $d(n)$ no es un canal que se modifica, por ello es allí donde debe interconectarse la señal más ruido. De interconectarlo a la entrada del filtro sería modificar al ruido y a

la señal útil, lo cual no tiene ningún interés práctico, sería más complejo recuperar la señal útil. Por otro lado, si se interconecta a la entrada del filtro una señal de ruido distinta r' , dado que es imposible contar con la verdadera señal de ruido que está mezclada, ver figura 3, entonces la ecuación 5) informaría que en el mínimo el filtro modificaría a r' para obtener el siguiente resultado de la función de error(4) :

$$E[e(n)]^2 = E[d(n) - y(n)]^2 = E[x(n) + r(n) - y(n)]^2 \quad 6)$$

$$E[e(n)]^2 = E[x(n)]^2 + 2E[x(n)(r(n) - y(n))] + E[r(n) - y(n)]^2$$

donde: $r(n)$ ruido mezclado con la señal útil,

$x(n)$ señal útil a recuperar.

Por hipótesis la señal de información y ruido son no correlacionadas, por tanto el término que involucre el producto de las mismas es cero, porque el valor esperado de la señal de ruido es cero, la señal de información también está incorrelacionada con la salida del filtro, dado que la correlación cruzada entre la señal útil y el ruido a la entrada del filtro es cero.

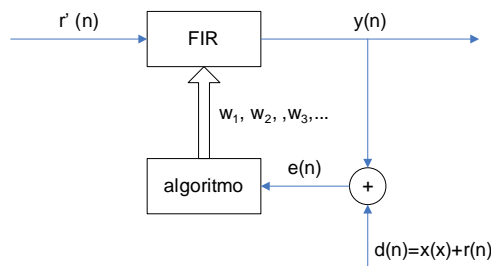


Figura 3. Configuración del filtro para eliminar ruido en banda.

Así la ecuación (6) se convierte en:

$$E[e(n)]^2 = E[x(n)]^2 + E[r(n) - y(n)]^2 \quad 7)$$

Es decir, si la potencia del error tiene un mínimo que estará dado a costa de reducir la diferencia entre el ruido y la salida del filtro, porque la potencia de la señal de información es invariable. Cuando el filtro trabaje en el mínimo de la función error, indicará que está reduciendo el ruido r mezclado con la señal útil, además en el error estará presente la señal útil limpia de ruido.

No obstante cabría preguntarse, ¿qué posibilidades hay de que la función potencia de error tenga un mínimo? Según 2 la 5) tiene un sólo mínimo y su valor se encuentra en la solución del sistema de ecuaciones de Wiener-Hopf. Este sistema determina el juego de valores de los coeficientes del filtro, a partir de los datos de matriz de autocorrelación de los datos de entrada y de la correlación cruzada entre la entrada y la señal deseada.

Cabe aclarar que este mínimo no tiene porqué ser igual al valor de potencia de la señal útil, lo cual significa que operando en esta condición no se igualen las potencias de la salida del filtro y la del ruido

mezclado. Si este mínimo es igual a la potencia de la señal útil, entonces de la ecuación (7) se desprende:

$$0 = E[r(n) - y(n)]^2 = E[r(n)]^2 + E[y(n)]^2 - 2E[r(n)y(n)] \quad 8)$$

Como en esta ecuación están establecidas la potencia del ruido y la de la salida del filtro y son magnitudes positivas, entonces la cantidad mientras más grande, más cancela a la suma de los otros dos términos. De aquí que esta función de correlación cruzada coincida con la función de autocorrelación del ruido y de la salida del filtro. Es decir, con la configuración presentada en el gráfico 3 se tiende a igualar la estadística de segundo orden de ambos procesos, el de r y r' .

Si estos procesos son estacionarios, entonces es aplicable el teorema de Wiener-Kichinne 3. Esto conduce a que ambos procesos tienen la misma distribución de densidad espectral de energía, se diferenciarán a lo sumo en fase, por ser esta cantidad una magnitud real siempre positiva y carente de parte imaginaria. No obstante en el estudio del ruido AWGN, se especifica que la función de densidad de probabilidad del ruido es gaussiana, y esta función tiene la propiedad de que su matriz de covarianza determina toda la estadística del proceso, es decir, más allá de esta matriz no hay más información(5).

No obstante, al igual que en el método de sustracción espectral, el filtrado adaptativo produce ruido musical (6). En parte debido a la imposibilidad de igualar totalmente en cuanto al contenido de frecuencia, por otra parte debido al rápido cambio de los coeficientes del filtro.

Para contrarrestar el rápido cambio de los coeficientes en (6) se propone promediar de la función transferencial del filtro mediante la siguiente relación:

$$H(k, p) = \frac{1}{f(\sigma, M)} \sum_{j=-M}^M \sigma^{|j|} H_0(k, p + j) \quad 9)$$

donde: los índices k y p denotan muestra y trama respectivamente (los coeficientes y por ello la función transferencial, pueden cambiar en cada muestra y por tramas).

H_0 : denota el estimado inicial en cada trama,

H : denota la promediación de los estimados,

f : denota una constante de normalización.

La constante f se determina por:

$$f(\sigma, M) = \sum_{j=-M}^M \sigma^{|j|} \quad 10)$$

También se define un valor, que determina una cota inferior para los coeficientes del filtro. En (6) se plantea variar σ en función de la presencia de ruido o de voz más ruido. Cuando se está en presencia de ruido solamente se plantea tomar un valor de $\sigma = 1$ para dar más peso a esa trama en la promediación y lograr una copia más fiel del ruido. Cuando se está en presencia de voz más ruido se plantea reducir la

constante a 0.8. Por otro lado el uso de la constante μ permite disminuir el nivel de ruido o mantener un ruido de fondo para confort del sonido.

VARIANTES DE FILTRADO ADAPTATIVO

La búsqueda de los coeficientes del filtro está fijada por varios algoritmos, en (2) se presenta el LMS (Least Mean Square), el cual forma parte de la familia algoritmos de gradientes estocásticos, dicho así para diferenciarlo del método de paso descendente en el cual se emplea el cálculo del gradiente de forma determinística. Una de las ventajas del algoritmo LMS es su simplicidad, no requiere del cálculo de la función de correlación.

El operador E en comparación con el método de paso descendente se elimina, por ello el cálculo del vector $w(n)$ “sufrir de ruido de gradiente”. Debido a este ruido el vector $w(n)$ según LMS ejecuta un movimiento aleatorio alrededor del punto mínimo del error, no así por el método de paso descendente que sigue una trayectoria determinística hacia la solución del filtro Wiener.

La ecuación que describe el algoritmo es (2):

$$\hat{w}(n + 1) = \hat{w}(n) + \mu \hat{u}(n)[d^*(n) - \hat{u}^H(n) \hat{w}(n)] \tag{11}$$

Donde: $\hat{w}(n)$ denota los valores estimados para el filtrado en el instante $n+1$,

$\hat{u}(n)$ vector calculado para el instante actual,

$u(n)$ señal de entrada,

$d(n)$ respuesta deseada.

En la expresión están definidos los datos de entrada y la respuesta deseada, excepto por la constante μ . Según 2, el valor de μ para garantizar convergencia debe ser tal que:

$$0 < \mu < \frac{2}{\lambda_{max}} \tag{12}$$

Donde: λ_{max} denota el mayor valor propio de la matriz de autocorrelación de los datos de entrada.

Se demuestra que la rapidez de convergencia está dada por:

$$\frac{-1}{\ln(1 - \mu \lambda_{max})} \leq \tau \leq \frac{-1}{\ln(1 - \mu \lambda_{min})} \tag{13}$$

Donde: τ es la constante de tiempo para la convergencia del filtrado.

En la ecuación 13) se muestra que de obtener un valor pequeño de μ , la rapidez de convergencia empeoraría, mientras que si este valor multiplicado por μ se aproxima a 1, menor será τ , siempre y cuando λ_{min} también se aproxime a λ_{max} . Es decir, mientras menos se diferencien estos valores y multiplicados por μ más se aproximen a la unidad, más favorable será la convergencia.

Por esta vía, en (7) se presenta un método para incrementar la rapidez de convergencia del algoritmo LMS. Se busca encontrar una matriz de autocorrelación que se parezca lo más posible a la matriz idéntica. Con ello se pasan las muestras temporales al plano de frecuencia, utilizando la transformada discreta de Fourier (DFT) o por la transformada discreta del coseno a través de las relaciones siguientes:

$$\begin{aligned}
 u_k(i) &= \sum_{l=0}^{N-1} T_N(i, l) x_{k-l} & 14) \\
 v_k(i) &= \frac{u_k(i)}{\sigma_k(i) + \gamma} \\
 T_N(i, l) &= \begin{cases} F_N(i, l) = \sqrt{\frac{1}{N}} e^{j \frac{2\pi i l}{N}} & DFT \\ T_N(i, l) = \sqrt{\frac{2}{N}} k_i \cos\left(\frac{i(l + \frac{1}{2})\pi}{N}\right) & DCT \end{cases}
 \end{aligned}$$

donde: x_{k-l} : es la señal de entrada al filtro, constituida por un vector de dimensión N , i, l

$u_k(i)$: denota la transformación de la señal de entrada según la matriz que se haya escogido, DFT o DCT,

$V_k(i)$: representa una normalización de potencia de la transformada de la señal de entrada.

La señal $u_k(i)$ es un vector para cada valor de k de dimensión N , y estará conformado en cada coordenada por funciones bases ortogonales. Así al calcular la matriz de correlación de este vector debe tender a una matriz idéntica por una constante, que se cancela al normalizar en el nuevo vector. El método constituye un decorrelador de la señal de entrada que hace tender la matriz de autocorrelación a la matriz idéntica.

De este modo los valores propios de la matriz de autocorrelación tienden a 1 y esto favorece la rapidez de convergencia según la ecuación 13). No obstante, la gran desventaja del método radica en realizar por la parte de DFT cálculos para la parte real y para la parte imaginaria, por la parte del DCT se necesita calcular el valor del coseno, que se debe hallar en una solución con FPGA mediante su serie de potencias, lo cual introduce más operaciones.

Con vistas a disminuir los cálculos presentados en la ecuación (11) se presenta en (8) un método que actualiza los coeficientes no cada vez que se toma una muestra, sino después de cierta cantidad de pasos de tiempo s . El método denominado PBS_LMS, aunque presenta una gran ventaja al disminuir el número de veces a calcular los coeficientes del filtro, introduce indudablemente un retardo en la convergencia de la solución en función del salto s . Por ello en esta vía se llega a una solución de compromiso, entre disminuir el cálculo computacional o disminuir la velocidad de convergencia.

No obstante el método sería muy útil para casos donde se cuente con un hardware "lento" para realizar los cálculos, pero con un conversor A/D más rápido. Se puede muestrear la señal más rápido de lo habitual y realizar los cálculos a un paso más lento. Esto conllevaría a compensar el tiempo de convergencia con respecto al método LMS si este actualiza su entrada a la velocidad que actualiza los cálculos el método PBS_LMS.

En (9) se especifica otro método de búsqueda de los coeficientes del filtro que minimicen la función error. Mediante un algoritmo que emplea una población de partículas PSO (Particle Swarm Optimization) se definen reglas a seguir por las partículas en su “movimiento”, hasta alcanzar el valor óptimo.

Las partículas son vectores que representan los coeficientes, que inicialmente se distribuyen aleatoriamente sobre un dominio a especificar. Cada partícula evalúa la función de error en el punto que se encuentra y en puntos vecinos al actual. Memoriza los mejores resultados obtenidos, en cuanto a la mejor posición por la que haya pasado y el mejor resultado de sus vecinos. La regla de movimiento se muestra en 15):

$$\Delta \vec{w}_i(k+1) = \Delta \vec{w}_i(k) + c_{1i} \text{rand}_{1i}(k) (\vec{P}_{best} - \vec{w}_i(k)) + c_{2i} \text{rand}_{2i}(k) (\vec{N}_{best} - \vec{w}_i(k)) \quad 15)$$

donde: denota la predicción de los coeficientes del filtro,

valor actual de los coeficientes del filtro,

y constantes de ajuste del método,

posición con el mejor resultado,

posición con el mejor resultado de sus vecinos.

Esta regla de movimiento puede favorecer una de dos direcciones, la del vector que archiva la posición de mejor resultado, o la del vector que registra el mejor resultado de sus vecinos. Así esta regla no permitiría alejarse mucho de la mejor posición con aportes en la dirección de su mejor vecino, de no ser por el valor rand, que modifica aleatoriamente ambos términos, simulando cierto movimiento de exploración alejándose de su mejor posición, en búsqueda de mejores posiciones.

La desventaja del método radica en el cálculo de muchas funciones de error, una en cada iteración por cada partícula que se considere, el orden de los cálculos se multiplica por el total de vectores. En 9 se muestra una tabla de precisión en cuanto al cálculo del mínimo de la función error y se aprecia que los mejores órdenes se obtienen con un total de 200 a 500 iteraciones, con un total de 50 a 100 partículas, es decir luego de 200 muestras es que viene a converger el resultado. Para una señal muestreada a 8 kHz, se deben esperar buenos resultados entre los 25 ms y los 62 ms, multiplicando de 50 a 100 el número de cálculos.

En 10 se presenta un método sin precedentes para la igualación de dos señales. Las condiciones de igualdad no contemplan la naturaleza de las mismas, es decir, no cuenta con la necesidad de que la señal deseada y la de entrada al filtro compartan una misma estadística. No se hace necesario que tengan o no igual contenido de frecuencia, que tengan o no igual media o igual función de autocorrelación. Simplemente entrega un valor de error cero independientemente de las señales que esté considerando, en una configuración análoga a la mostrada en la figura 2.

Se basa en la teoría de la estabilidad de Lyapunov, esta teoría asegura convergencia en la solución y convergencia hacia el valor deseado. Es decir el valor del error puede garantizarse convergente y hacia el valor cero. El principio de funcionamiento de esta teoría se explica por la actualización que realiza a los coeficientes del filtro, figura 2:

$$\begin{aligned}
 y(k) &= H^T(k)U(k) & 16) \\
 H(k) &= [h_k(0), h_k(1), h_k(1), \dots, h_k(n-1)]^T \\
 H(k) &= [h_k(0), h_k(1), h_k(1), \dots, h_k(n-1)]^T \\
 H(k) &= H(k-1) + g(k)\alpha(k) \\
 g(k) &= \frac{U(k)}{\lambda_1 + \|U(k)\|^2} \left(1 - \chi \frac{|e(k-1)|}{\lambda_2 + |\alpha(k)|} \right) \\
 \alpha(k) &= d(k) - H^T(k-1)U(k)
 \end{aligned}$$

donde: $y(k)$ denota la salida del filtro FIR,

$U(k)$ denota la entrada al filtro,

$H_i(k)$ denota los coeficientes del filtro,

$d(k)$ denota la respuesta deseada,

χ , λ_1 y λ_2 denotan coeficientes de diseño.

Según (10) el error converge hacia una bola con centro en cero con un radio que depende de χ , λ_1 y λ_2 , Estos son parámetros que se fijan en el diseño, ecuaciones 16). Donde χ determina la rapidez de convergencia, y λ_1 , λ_2 se emplean para evitar singularidades para la función $g(k)$.

En el gráfico 3 se grafican diferentes curvas para mostrar la bola a la que converge el error en función de los parámetros χ , λ_1 y λ_2 . El gráfico se calculó en Scilab 5.1 y muestra por el eje de las abcisas el valor de χ , y por las imágenes aparecen varias curvas para iguales valores de λ_1 y λ_2 entre cero y uno.

Se puede apreciar que efectivamente se puede obtener un error de cero y para valores de χ pequeños, lo cual garantiza una velocidad de convergencia tan rápida como se desee. Pero el método carece de utilidad práctica para separar dos señales, si siempre puede convertir una señal en otra, entonces ¿cómo convertir r' en r si no se cuenta con r ? Este método tendría un excelente desempeño en la ecualización de canal con la descripción de sistemas no lineales, característica que aventaja mucho al filtro Wiener.

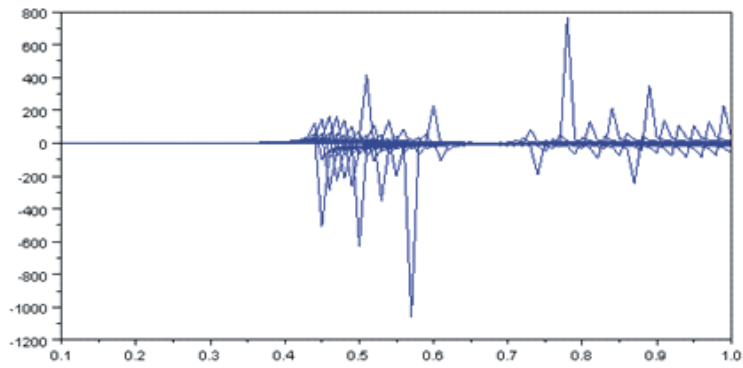


Figura 3. Valores de la bola de error del método de Lyapunov.

CONCLUSIONES

A partir de lo presentado en este artículo, y dando continuidad a Métodos de Reducción de Ruido (I), se puede añadir a la clasificación de los métodos un tercer grupo. En este se incluyen las técnicas que trabajan con criterios estadísticos.

Las técnicas del primer grupo no cancelan de forma efectiva el ruido de la señal en banda, mientras que lo presentado en el segundo grupo la factorización de matrices no negativas, necesita de un procesamiento difícil de hacer en tiempo real, cuando se cuenta con un solo canal no es posible obtener un modelo de señal y otro de ruido por sí solos. Por esto el empleo de Filtros Adaptativos para la compensación de Ruido representa una solución a la vez más efectiva y más práctica.

REFERENCIAS

1. Takuya Yoshioka, M.M.: "Adaptive Supression of Non-Stationary Noise by using de Variational Bayesian Method.": ICASSP 2008, 2008.
2. Haykin, S.: Adaptative Filter Theory, 3ra edición ed: Prentice Hall. 997.
3. Peebles, P.Z. : Probability, Random Variables, and Random Signal Principles, Second ed, ed. McGraw-Hill. 1987.
4. Paris Mastorocostas, D.V., Constantino Mastorocostas, Constatinos Hilas: "A Recurrent Fuzzy Filter for Adaptative Noise Cancellation". IEEE, 2005.
5. Dimitris G. Manolakis, V.K.I., Stephen M. Kogon: "Statistical and Adaptive Signal Proccesing". 2005, Boston: Artech House. 787.
6. Phil S. Whitehead, D.V.A., Mark A. Clements, Adaptive, Acoustic Noise Suppression for Speech Enhancement. ICME 2003, 2003.
7. Zhao Shengkui, M.Z., Khoo Suiyang,: "On Performance of Transform Domain Adaptive Filters with Markov-2 Inputs." IEEE Trans. circuits Syst. I, 2008.
8. REZA Seifi Majdar, M.E.: "Implementation of PBS_LMS algorithm for Adaptive filters on FPGA". 2004.
9. Xiuhong Wang, Q.G., Qiqiang Li, Jinsong Zhang: "High-order Cumulant-based Adaptive Filter using Particle Swarm Optimization ". CCDC 2008 Volume.
10. Kah Phooi Seng, Z.M., Hong Ren Wu, "Lyapunov-Theory-Based Radial Basis Function Networks for Adaptive Filtering" IEEE Trans. circuits Syst. I: Fundamental Theory and Applications, 2002. 49(8).