

# CANCELACIÓN DE ECO

**José Adrián Fernández**  
**a.fernandez@adinet.com.uy**

## CONTENIDO

RESUMEN .....	2
INTRODUCCIÓN .....	2
SEÑALES .....	4
<i>Señales de voz</i> .....	4
<i>Ruido de fondo</i> .....	5
CANCELADORES DE ECO .....	6
<i>Sistema Mono</i> .....	6
Algoritmos .....	9
NLMS .....	10
Paso de adaptación .....	10
<i>Sistema Estéreo</i> .....	13
Problemas Fundamentales .....	14
Problemas de unicidad de la solución.....	14
Problema de desajuste .....	17
El Efecto Tail.....	19
Solución para reducir la Correlación .....	20
Función de Coherencia .....	20
Método de Transformación no-Lineal .....	20
Filtros Adaptivos de Dos Canales.....	21
RLS Doble Canal.....	21
Algoritmo APA doble- Canal .....	22
<i>Sistema Híbrido Mono/Estéreo</i> .....	24
<i>Detector de Doble Conversación (Doubletalk Detector)</i> .....	25
Detector de Geigel (Geigel DTD).....	26
COMENTARIOS FINALES .....	26
BIBLIOGRAFÍA .....	27

## Resumen

En el siguiente trabajo se hace la presentación del tema Cancelación de Eco (Echo Cancellation), en términos de dar una primera introducción al tema. Se expondrán los conceptos básicos de un problema de la realidad actual en el que existe más de una propuesta para alcanzar la solución. Se evidencia que tras cualquier técnica llevada a cabo existe una basta elaboración desde el punto de vista teórico y práctico. Por lo tanto el enfoque de la atención se hará en la forma en que se aplican las herramientas de tratamiento estadístico de señales y los problemas prácticos principales que aparecen en el momento de la implementación.

## Introducción

A pesar de que la cancelación de *eco* corresponde a uno de los problemas tradicionales del procesamiento de señales a partir de su introducción en 1967, este es aún un campo activo de investigación.

El eco es una versión retrasada y posiblemente distorsionada de sonido que se refleja hacia su emisor.

En el sistema telefónico tradicional la generación de *eco* es una característica inherente al mismo. Por definición este sistema es full-duplex. Para conseguir esto utilizando un único par de cobre llegando al abonado, se utiliza un elemento llamado híbrida que hace el acople de los dos flujos sobre el mismo par de cobre, y en cada terminal hace la separación correspondiente de manera que lo que se habla en uno de los extremos, se escuche en el otro. En la práctica esto nunca funciona en forma exacta, dado que se requiere que haya un balance eléctrico perfecto en la red. El resultado es que parte del flujo que se genera en uno de los extremos es reinyectado en la línea por el extremo opuesto, sumado a lo que este último genera. Es decir, si consideramos una conversación entre A y B, lo hablado por A, B lo reenvía hacia A atenuado y retrasado en el tiempo. Esto es denominado efecto local. Sin embargo, para retardos menores a 25ms esto no significa un problema, mas bien significa una realimentación que favorece a la comunicación. Para retardos mayores a 40ms la distorsión que provoca el *eco* se hace tremendamente molesta y se debe tomar alguna medida al respecto.

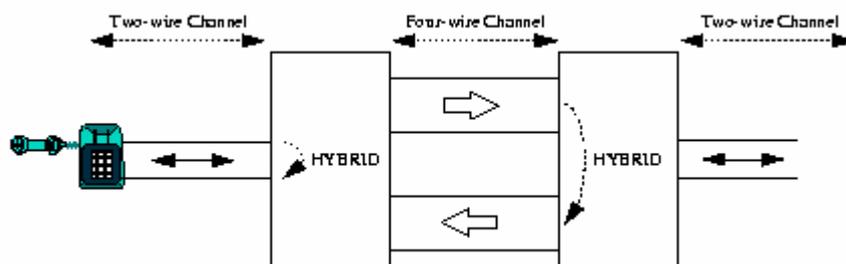


Fig. 1. . Generación de eco en la red telefónica.

En los '90 una nueva atención ha sido dirigida al tema, por dos razones principalmente: Primero, la introducción de sistemas digitales de radiocomunicaciones con sofisticados algoritmos causa largos retardos. En segundo lugar, con el incremento de la utilización de manos libres y la extensión de las teleconferencias el *cancelador de eco* (A.E.C. Acoustic Echo Canceler) pasó a jugar un rol central. En este caso el eco se genera por múltiples reflexiones del audio que proviene de los parlantes hacia el micrófono, así como por acople directo de la señal.

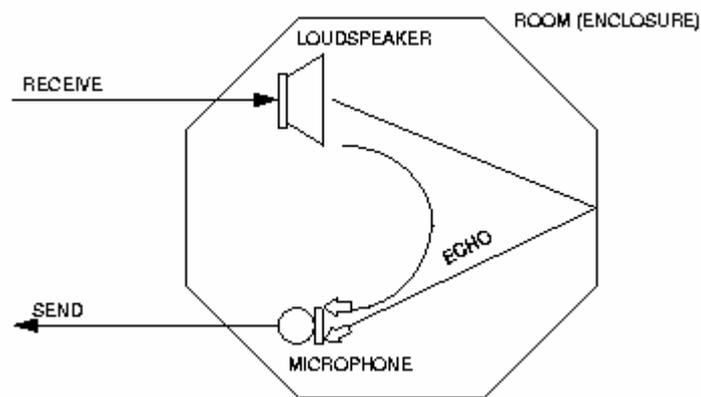


Fig. 2. Generación de eco en espacio cerrado.

En la aplicación a teleconferencias, se busca mejorar la calidad y la localización del sonido. Esto se puede obtener mediante el incremento del ancho de banda y también en la adición de más canales de audio. Esto último implica la necesidad de tener canceladores de eco multi-canal.

En este trabajo se estudia el tema y se exponen los principales aspectos involucrados y diferentes soluciones a la que se ha llegado.

## Señales

### Señales de voz

La performance de algoritmos adaptivos depende en forma crucial de las propiedades de las señales de entrada. En el caso del control de eco acústico, las señales de voz serán la excitación al sistema. La voz humana puede ser caracterizada por tres diferentes excitaciones: segmentos aproximadamente periódicos, segmentos como de murmullo, y pausas.

Podemos asumir estacionalidad solamente en períodos cortos de tiempo, en intervalos de 10 ms a 20 ms. Si se usa una frecuencia de muestreo de 8 KHz. (telefonía), la envolvente del espectro puede tener un rango mayor a 40 dB. Para frecuencias mayores de muestreo, como en el caso de teleconferencias, el rango se verá incrementado.

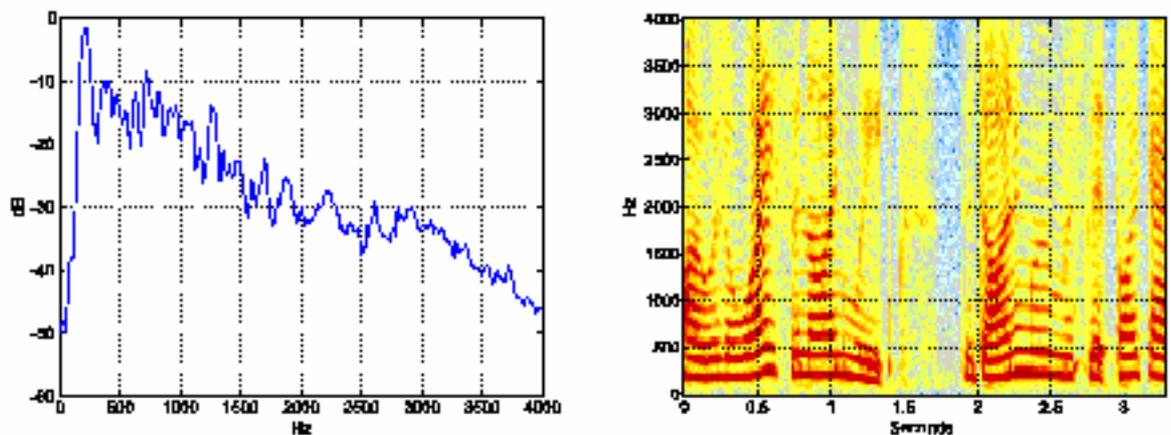


Fig. 3. Propiedades de las señales de voz. A la izquierda, densidad espectral de potencia. A la derecha, análisis tiempo-frecuencia de una secuencia de voz.

## Ruido de fondo

En algunos sistemas es conveniente considerar ruidos de fondo que distorsionan la señal de voz. Por ejemplo, en el caso del uso de teléfonos manos libres en una oficina en presencia de ruidos generados por turbinas de PCs. Otra fuente de distorsión puede estar generada en el interior de un automóvil en marcha.

En contraste con la señal de voz, muchas veces el ruido muestra un comportamiento aproximadamente estacionario.

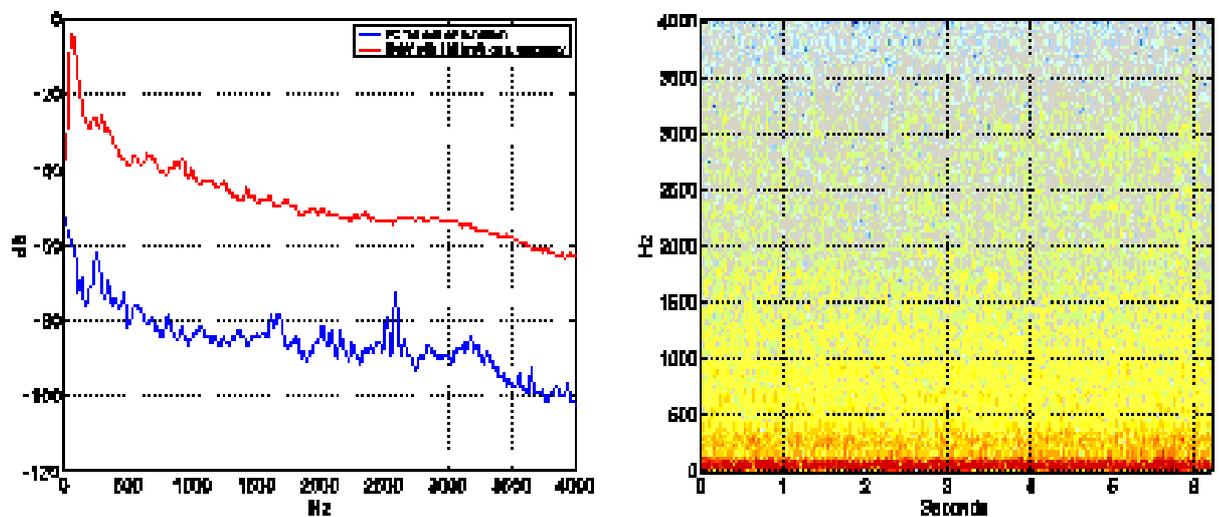


Fig. 4. Propiedades del ruido de fondo. A la izquierda densidad espectral de potencia para dos casos, dentro de un automóvil (rojo) y en una oficina con ruido de PC fan y aire acondicionado (azul). A la derecha, análisis tiempo-frecuencia para el caso del automóvil

## Canceladores de Eco

### Sistema Mono

Los elementos básicos que encontramos en el sistema son: el parlante y el micrófono presentes en un espacio cerrado, como ser una oficina. En el caso de que el sonido esté dentro del rango manejado por el convertor A/D, el sistema puede ser modelado como un sistema lineal tal como se muestra en la figura.

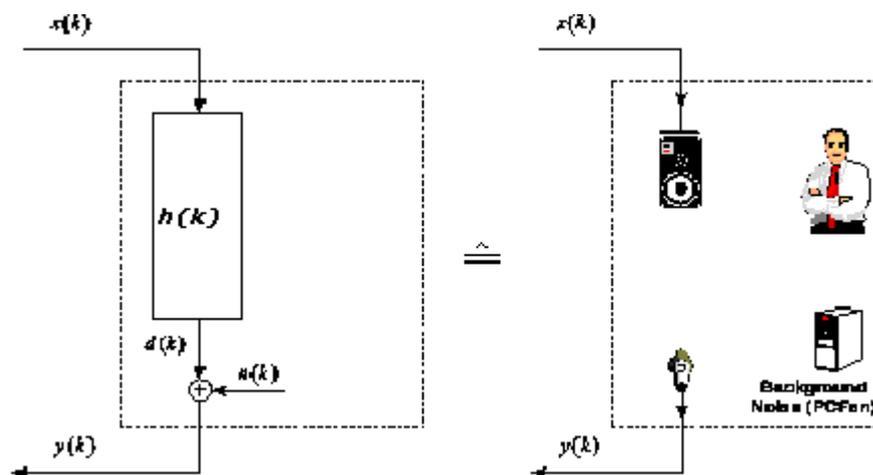


Fig. 5. Modelo del sistema LEM.

Asumiendo que el sistema parlante-micrófono (LEM: Loudspeaker-enclousure-Microphone), tiene respuesta al impulso finita, la señal de eco  $d(k)$  puede ser escrita como:

$$d(k) = \sum_{n=0}^{N-1} h_k(n)x(k-n) = \underline{h}(k)^T(k)\underline{x}(k)$$

(1)

El vector  $\underline{x}(k)$  consiste en las últimas  $N$  muestras de la señal de excitación, mientras que  $\underline{h}(k)$  es el vector de respuesta al impulso del sistema LEM. Los movimientos de personas o cambios de temperatura hacen que la respuesta al impulso sea variante en el tiempo.

La solución al problema se alcanza si tenemos una réplica digital en paralelo del sistema. Si ambos sistemas son excitados por la misma señal  $x(k)$ , él desacople entre el parlante y el micrófono lo podemos lograr simplemente sustrayendo la estimación del eco  $\hat{d}(k)$  de la señal del micrófono  $y(k)$ .

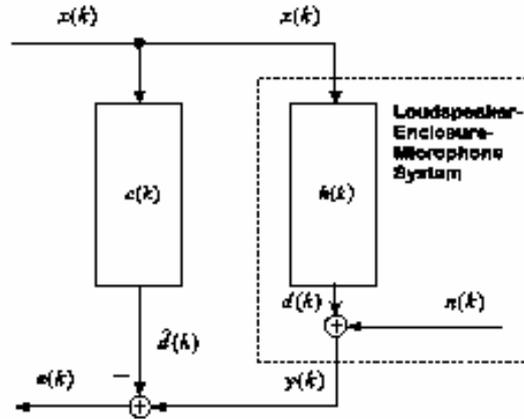


Fig. 6. Estructura del sistema de cancelación de eco.

La estimación de la señal de eco  $\hat{d}(k)$  es calculada usando un filtro adaptativo  $\hat{c}(k)$ . Por razones de estabilidad el filtro será de respuesta finita al impulso, es decir un FIR. El orden  $N - 1$  del filtro debe ser elegido dependiendo del tiempo de reverberación esperado. En oficinas encontramos tiempos de reverberación entre 300 ms a 500 ms, en autos estos son de 50 ms a 100 ms. Si suponemos que los coeficientes del filtro adaptativo se ajustan perfectamente para los primeros  $N$  coeficientes con la respuesta al impulso del LEM, la máxima atenuación del eco se puede calcular en función del orden  $N - 1$  del filtro

$$\frac{E\{e_N^2(k)\}}{E\{d^2(k)\}} = \frac{E\{d(k) - \hat{d}_N(k)\}^2}{E\{d^2(k)\}} = \frac{E\left\{\left(\sum_{i=0}^{\infty} h_i(k)x(k-i) - \sum_{i=0}^{N-1} c_i(k)x(k-i)\right)^2\right\}}{E\left\{\left(\sum_{i=0}^{\infty} h_i(k)x(k-i)\right)^2\right\}} \quad (2)$$

$$= \frac{E\left\{\left(\sum_{i=N}^{\infty} h_i(k)x(k-i)\right)^2\right\}}{E\left\{\left(\sum_{i=0}^{\infty} h_i(k)x(k-i)\right)^2\right\}} = \frac{\sum_{i=N}^{\infty} h_i^2(k)}{\sum_{i=0}^{\infty} h_i^2(k)}$$

en que asumimos ruido blanco de potencia uno como excitación y sin distorsión local. Para lograr una atenuación de 45 dB de acuerdo a la recomendación de ITU, el orden del filtro requiere un orden de alrededor de 1600. La adaptación de filtros de alto orden provoca una gran demanda de potencia de computo en la utilización de DSPs. De esta forma, para implementaciones con procesadores de bajo costo, se requiere de algoritmos adaptivos de baja complejidad o reducción a estructuras menos complejas, como ser procesamiento en sub-bandas.

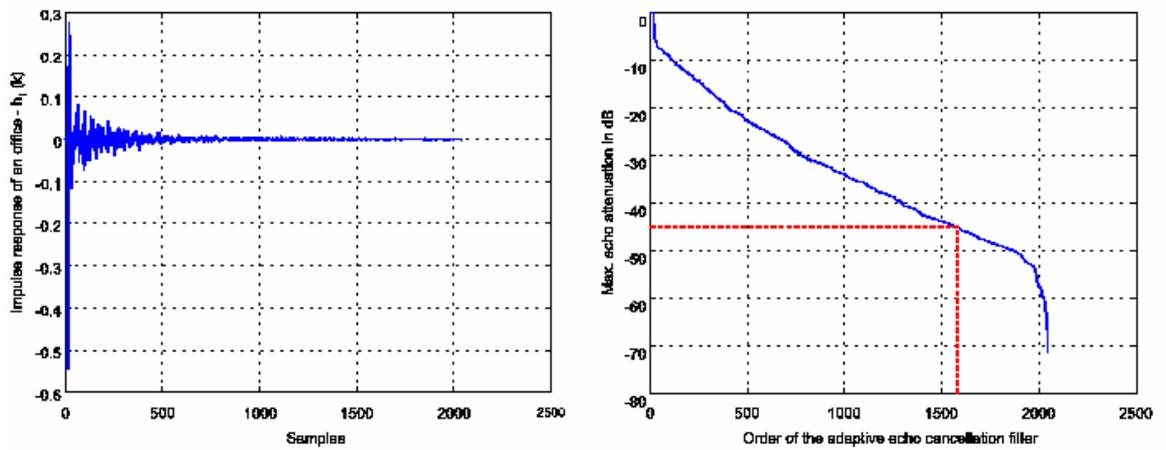


Fig. 7. Respuesta al impulso en una oficina con tiempo de reverberación de 300ms (izq.) y atenuación máxima del eco en función del largo del filtro (der.).

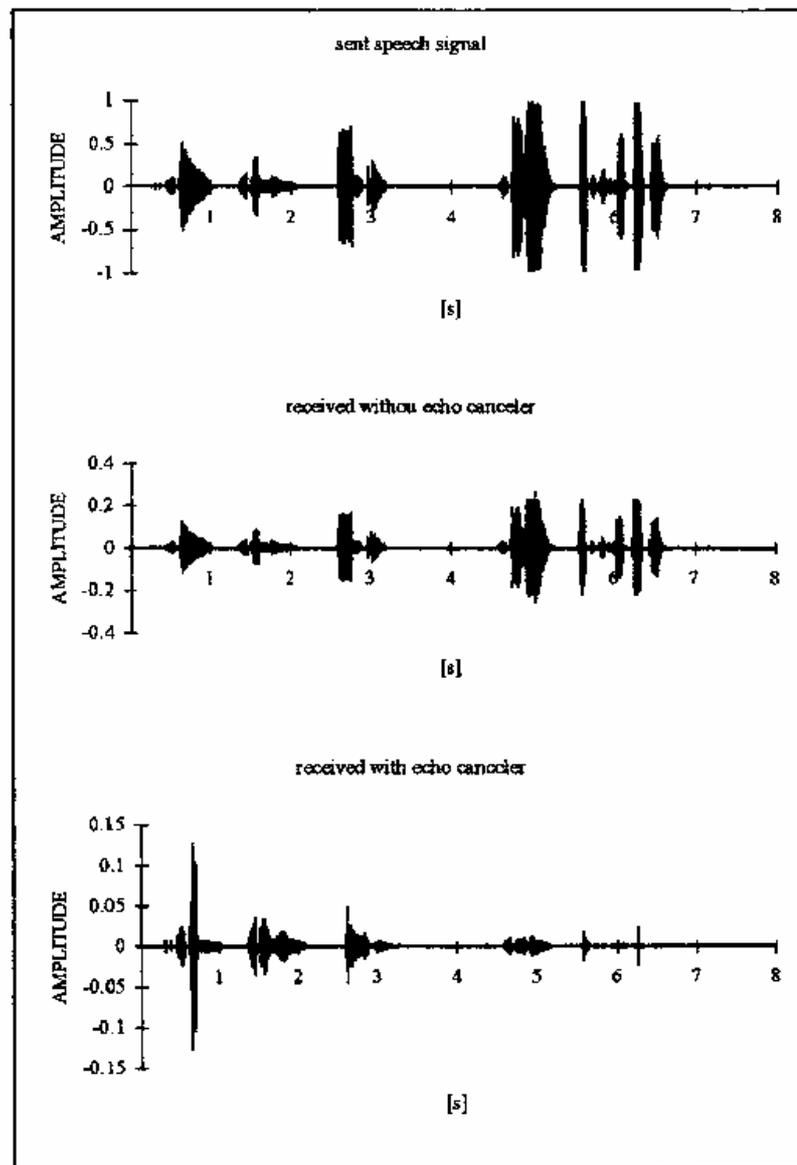


Fig. 8.

## Algoritmos

En el diseño del filtro  $\underline{c}(k)$ , tres de los algoritmos adaptivos más populares utilizados son:

- NLMS (Normalized Least Mean Square),
- APA (Affine Projection Algorithm), y
- RLS (Recursive Least Square).

Para el caso monofónico, el algoritmo NLMS presenta ventajas en cuanto a su baja complejidad, que lo harán preferible en muchas de las implementaciones y es usualmente elegido en casos de tiempo real. Este requiere solo una multiplicación y una adición por coeficiente, y es robusto respecto a efectos de precisión numérica y largo de palabra. A pesar de estas ventajas, la convergencia es más lenta a la de los otros algoritmos. Cuando se vea el caso estéreo se verá algún detalle más de los algoritmos RLS y APA. Como comentario decimos que el algoritmo RLS presenta una considerable mayor velocidad de convergencia, aunque requiere alto procesamiento y presenta problemas de estabilidad que lo hacen prohibitivo para implementaciones de bajo costo. Por su lado el algoritmo APA tiene una velocidad de convergencia cercana al RLS si se usa señal de voz como entrada y su complejidad es intermedia.

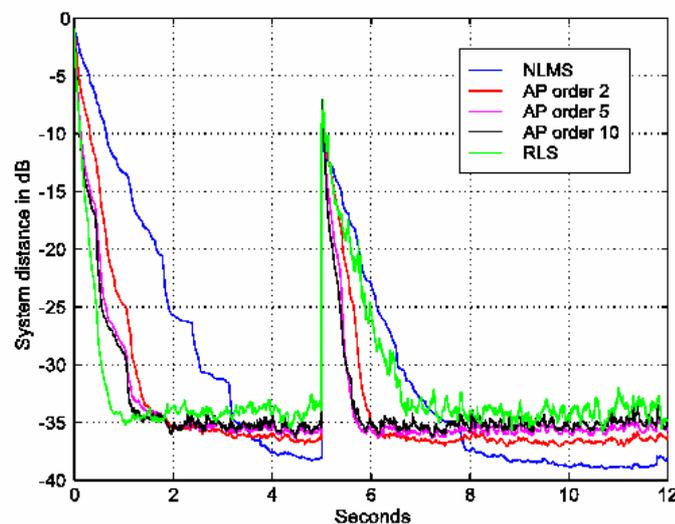


Fig. 9. Convergencia de los algoritmos NLMS, AP, y RLS.

NLMSPaso de adaptación

Recordando la expresión del algoritmo NLMS, la ecuación de actualización es:

$$\underline{c}(k+1) = \underline{c}(k) + \alpha(k) \frac{\underline{x}(k)e(k)}{\underline{x}^T(k)\underline{x}(k)}, \quad (3)$$

siendo  $\alpha(k)$  el paso de adaptación.

Buscaremos una forma de obtener el valor óptimo del paso de adaptación.

Definiendo el vector diferencia del sistema como

$$\underline{m}(k) = \underline{h}(k) - \underline{c}(k),$$

suponiendo que el sistema no cambia ( $\underline{h}(k) = \underline{h}(k-1)$ ), podemos usar lo anterior para obtener:

$$\underline{m}(k+1) = \underline{m}(k) - \alpha(k)e(k) \frac{\underline{x}(k)}{\|\underline{x}(k)\|^2}. \quad (4)$$

Para determinar el paso óptimo, el criterio tomado es que el costo de la función debe reducirse en media por cada iteración. Esto es:

$$E\left\{\|\underline{m}(k+1)\|^2\right\} - E\left\{\|\underline{m}(k)\|^2\right\} < 0. \quad (5)$$

Usando la definición del vector diferencia del sistema y la relación  $\underline{x}^T(k)\underline{m}(k) = \varepsilon(k)$ , la diferencia anterior la podemos describir como:

$$E\left\{\|\underline{m}(k+1)\|^2\right\} - E\left\{\|\underline{m}(k)\|^2\right\} = \alpha^2(k)E\left\{\frac{e^2(k)}{\|\underline{x}(k)\|^2}\right\} - 2\alpha(k)E\left\{\frac{e(k)\varepsilon(k)}{\|\underline{x}(k)\|^2}\right\} < 0 \quad (6)$$

La señal  $\varepsilon(k)$  es la señal de error sin distorsión, ya que esta puede ser calculada como  $\varepsilon(k) = e(k) - \eta(k)$ . Así, el paso de adaptación  $\alpha(k)$  cumplirá la condición

$$0 < \alpha(k) < 2 \frac{E \left\{ \frac{e(k)\varepsilon(k)}{\|\underline{x}(k)\|^2} \right\}}{E \left\{ \frac{e^2(k)}{\|\underline{x}(k)\|^2} \right\}}.$$

(7)

El mayor decremento de la distancia del sistema se consigue en la mitad del intervalo definido. De esta forma el paso óptimo  $\alpha(k)$  se puede dar como

$$\alpha_{opt}(k) = \frac{E \left\{ \frac{e(k)\varepsilon(k)}{\|\underline{x}(k)\|^2} \right\}}{E \left\{ \frac{e^2(k)}{\|\underline{x}(k)\|^2} \right\}}.$$

(8)

Suponiendo que la excitación  $\|\underline{x}(k)\|^2$  puede ser aproximada por una constante, y suponiendo que las señales  $\eta(k)$  y  $\varepsilon(k)$  no son correlacionadas, el paso óptimo puede ser simplificado de la siguiente manera:

$$\alpha_{opt}(k) \approx \frac{E \{ \varepsilon^2(k) \}}{E \{ e^2(k) \}}.$$

(9)

En ausencia de distorsión local, osea  $\eta(k) = 0$ , el error distorsionado  $e(k)$  es igual al error no distorsionado  $\varepsilon(k)$  y el paso óptimo será uno. En este caso, si el filtro cancelador de eco está bien ajustado a la respuesta al impulso del sistema, la potencia del error residual  $\varepsilon(k)$  será muy pequeña. En presencia de distorsión local, la potencia de la señal  $\eta(k)$  y por lo tanto también la potencia de  $e(k)$  se incrementará. En este caso el numerador de la aproximación de arriba es notoriamente menor que el denominador, dando como resultado un paso de adaptación nulo.

Para la estimación  $\alpha_{opt}(k)$  es necesario estimar la potencia de la señal de error no distorsionada, la cuál no es accesible. A partir de la estructura en paralelo que presenta el sistema y el filtro adaptivo FIR, la señal  $\varepsilon(k)$  puede ser escrita como:  $\varepsilon(k) = \underline{x}^T(k) \cdot [\underline{h}(k) - \underline{c}(k)]$ . A partir de esto podemos escribir:

$$E \{ \varepsilon^2(k) \} = E \{ x^2(k) \} \cdot E \{ \|\underline{h}(k) - \underline{c}(k)\|^2 \},$$

(10)

en dónde el segundo factor indica el acoplamiento del eco  $\tilde{\beta}(k) = E\{\|\underline{h}(k) - \underline{c}(k)\|^2\}$  luego del proceso de cancelación. Por lo tanto obtenemos:

$$\alpha(k) \approx \frac{E\{\varepsilon^2(k)\}}{E\{e^2(k)\}} = \frac{E\{x^2(k)\} \cdot \tilde{\beta}(k)}{E\{e^2(k)\}}. \quad (11)$$

## Sistema Estéreo

La *cancelación de eco estereofónica* (SAEC: Stereophonic Acoustic Echo Canceller) es fundamentalmente diferente al sistema mono clásico. Una buena implementación de un SAEC debe tener no solamente seguimiento del cambio de camino del eco en el recinto de recepción, sino también en el de transmisión. Por ejemplo, el cancelador debe reconverger si un locutor para de hablar y comienza otro que está en un lugar diferente en el lugar de transmisión. Esto implica nuevos requerimientos de rapidez, y por lo tanto en este esquema, la generalización del sistema mono resulta en una pobre supresión del eco.

El problema fundamental es que el canal doble puede acarrear relaciones lineales tales que pueden hacer que la ecuación normal que resuelve el problema sea singular. Esto implica que no haya una solución única a la ecuación, y se puede ver que las infinitas soluciones dependen de la sala de transmisión por lo cuál a la que se llegue difícilmente sea la solución física verdadera. Se verá que esta situación se resuelve reduciendo la correlación entre las señales estéreo.

La performance del SAEC es afectada en forma mas severa por la elección del algoritmo que en la contraparte mono. Se puede reconocer que la performance en la mayoría de los algoritmos depende del número de condición de la matriz de correlación de entrada. En el caso estéreo, el número de condición es muy alto, por lo que algoritmos como el LMS, que no consideran la correlación cruzada entre las señales de entrada, convergen muy lentamente a la solución verdadera.

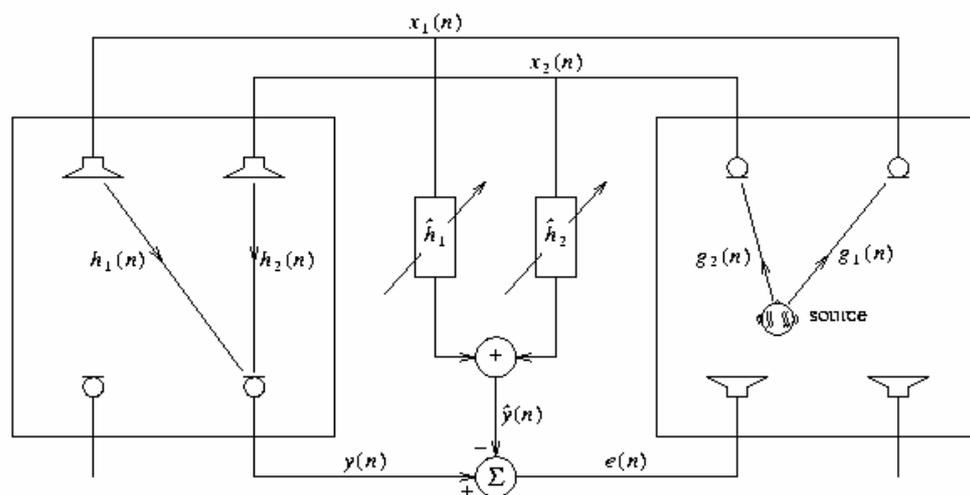


Fig. 10. Diagrama esquemático de cancelación de eco estereofónico

## Problemas Fundamentales

### Problemas de unicidad de la solución

Veremos que la solución a la ecuación normal en el caso estéreo no es tan obvia con respecto al caso del canal simple. En efecto, a partir de que las dos entradas son obtenidas filtrando la señal proveniente de una fuente común, aparecerán problemas nuevos. En la siguiente discusión distinguiremos entre el largo ( $M$ ) de la respuesta al impulso en la sala de transmisión, el largo ( $L$ ) del filtro, y el largo ( $N$ ) de la respuesta al impulso en la sala de recepción.

Se asume que el sistema (sala de transmisión) es lineal e invariante en el tiempo; por lo que se obtiene la siguiente relación:

$$\underline{x}_{1,M}^T(n) \underline{g}_{2,M} = \underline{x}_{2,M}^T(n) \underline{g}_{1,M} \quad (12)$$

dónde

$$\underline{x}_{i,M}^T(n) = [x_i(n) \ x_i(n-1) \ \dots \ x_i(n-M+1)]^T, \quad i = 1,2$$

son los vectores de muestras a la salida del micrófono en la sala de transmisión, y el vector de respuesta al impulso esta definido como:

$$\underline{g}_{i,M} = [g_{i,0} \ g_{i,1} \ \dots \ g_{i,M-1}]^T, \quad i = 1,2.$$

Lo anterior se puede ver en el dominio de la frecuencia, a partir del espectro de la fuente  $S(f)$  tenemos

$$X_1(f) = G_1(f)S(f) \quad (13)$$

$$X_2(f) = G_2(f)S(f); \quad (14)$$

por lo tanto

$$X_1(f)G_2(f) = X_2(f)G_1(f) \quad (15)$$

la cuál es la transformación de Fourier de (12).

El desarrollo de la estimación del error se hará en términos de mínimos cuadrados, por lo que se define:

$$J(n) = \sum_{l=1}^n \lambda^{n-l} e^2(l) \quad (16)$$

dónde  $\lambda$  ( $0 < \lambda \leq 1$ ) es el factor de olvido.

El error entre la salida del micrófono en la sala de recepción,  $y(n)$  y su estimación, es

$$e(n) = y(n) - \hat{\underline{h}}_{1,L}^T(n) \underline{x}_{1,L}(n) - \hat{\underline{h}}_{2,L}^T(n) \underline{x}_{2,L}(n) \quad (17)$$

dónde

$$\hat{\underline{h}}_{i,L}(n) = \left[ \hat{h}_{i,0}(n) \quad \hat{h}_{i,1}(n) \quad \Lambda \quad \hat{h}_{i,L-1}(n) \right]^T$$

son los filtros modelados,

$$\underline{x}_{i,L}(n) = \left[ x_i(n) \quad x_i(n-1) \quad \text{K} \quad x_i(n-L+1) \right]^T, \quad i = 1,2$$

La salida del micrófono viene dada por

$$y(n) = \underline{h}_{1,N}^T(n) \underline{x}_{1,N}(n) - \underline{h}_{2,N}^T(n) \underline{x}_{2,N}(n) \quad (18)$$

dónde

$$\underline{h}_{i,N}(n) = \left[ h_{i,0}(n) \quad h_{i,1}(n) \quad \Lambda \quad h_{i,N-1}(n) \right]^T$$

es la respuesta al impulso entre los dos parlantes y el micrófono ubicado en el lugar de recepción, y

$$\underline{x}_{i,N}(n) = \left[ x_i(n) \quad x_i(n-1) \quad \text{K} \quad x_i(n-N+1) \right]^T, \quad i = 1,2$$

La minimización de (16) nos conduce a la ecuación normal:

$$\underline{\underline{R}}(n) \begin{bmatrix} \hat{\underline{h}}_{1,L}(n) \\ \hat{\underline{h}}_{2,L}(n) \end{bmatrix} = \underline{r}(n) \quad (19)$$

dónde

$$\underline{\underline{R}}(n) = \sum_{l=1}^n \lambda^{n-l} \begin{bmatrix} \underline{x}_{1,L}(l) \\ \underline{x}_{2,L}(l) \end{bmatrix} \begin{bmatrix} \underline{x}_{1,L}^T(l) & \underline{x}_{2,L}^T(l) \end{bmatrix} = \begin{bmatrix} \underline{\underline{R}}_{11}(n) & \underline{\underline{R}}_{12}(n) \\ \underline{\underline{R}}_{21}(n) & \underline{\underline{R}}_{22}(n) \end{bmatrix} \quad (20)$$

es la estimación de la matriz de correlación de la señal de entrada y la estimación de la correlación cruzada entre el vector de entrada y la señal de salida es

$$\underline{r}(n) = \sum_{l=1}^n \lambda^{n-l} y(l) \begin{bmatrix} x_{1,L}(l) \\ x_{2,L}(l) \end{bmatrix} \quad (21)$$

Ahora bien, la cuestión importante es si  $\underline{R}(n)$  es de rango completo. Si no lo es, entonces la solución al problema no es única y un algoritmo adaptivo derivará en una de las muchas posibles soluciones, la cuál puede ser muy diferente a la solución deseada  $\hat{h}_{1,L} = \underline{h}_{1,L}$  y  $\hat{h}_{2,L} = \underline{h}_{2,L}$ , dónde

$$\underline{h}_{i,L}(n) = [h_{i,0}(n) \quad h_{i,1}(n) \quad \Lambda \quad h_{i,L-1}(n)]^T, \quad i = 1,2$$

Examinando dos posibles casos según el largo del filtro encontramos:

(i)  $L \geq M$

Podemos identificar un vector propio de la matriz de correlación  $\underline{R}(n)$  dado por

$$\underline{u} = \begin{bmatrix} \underline{g}_{2,M}^T & 0 & \Lambda & 0 & -\underline{g}_{1,M}^T & 0 & \Lambda & 0 \end{bmatrix}^T,$$

el cuál contiene  $2 \times (L-M)$  coeficientes nulos. Usando (12) podemos verificar que  $\underline{R}(n)\underline{u} = 0_{2L \times 1}$ , por lo que  $\underline{R}(n)$  no es invertible. A partir de este resultado, si sólo tenemos un vector propio de  $\underline{R}(n)$  con valor propio nulo deducimos que para  $L \geq N$  tenemos

$$\hat{h}_{1,L}(n) = \underline{h}_{1,L} + \beta \begin{bmatrix} \underline{g}_{2,M}^T & 0 & \Lambda & 0 \end{bmatrix}^T \quad (22)$$

y

$$\hat{h}_{2,L}(n) = \underline{h}_{2,L} + \beta \begin{bmatrix} -\underline{g}_{1,M}^T & 0 & \Lambda & 0 \end{bmatrix}^T \quad (23)$$

siendo  $\beta$  una constante arbitraria. La solución es dependiente de la respuesta al impulso de la sala de transmisión. Esto se considera intolerable ya que  $\underline{g}_{i,M}$  pueden cambiar en forma instantánea, por ejemplo cuando una persona deja de hablar y comienza otra.

(ii)  $L < M$

Este caso es el más realista dado que  $\underline{g}_{i,M}$  son de largo infinito. Ahora, (12) puede ser expresada como

$$\underline{x}_{1,L}^T(n) \underline{g}_{2,L} + q_1(n-L) = \underline{x}_{2,L}^T(n) \underline{g}_{1,L} + q_2(n-L) \quad (24)$$

con

$$q_1(n-L) = \sum_{i=L}^{M-1} x_1(n-i)g_{2,i}$$

$$q_2(n-L) = \sum_{i=L}^{M-1} x_2(n-i)g_{1,i}$$

A partir de (12) vemos que  $\underline{x}_{1,M}(n)$  y  $\underline{x}_{2,M}(n)$  están linealmente relacionadas, pero de (24) se ve que lo mismo no ocurre (salvo en casos especiales) para  $\underline{x}_{1,L}(n)$  y  $\underline{x}_{2,L}(n)$ . De aquí que, en principio la matriz  $\underline{R}(n)$  es de rango completo, pero esta muy mal condicionada porque en general  $q_1(n-L)$  y  $q_2(n-L)$  son pequeños comparados con  $\underline{x}_{1,L}^T(n)\underline{g}_{2,L}$  y  $\underline{x}_{2,L}^T(n)\underline{g}_{1,L}$ .

De esta manera, para el caso práctico cuando  $L < M$ , tenemos solución única de la ecuación normal, aunque la matriz de correlación esté muy mal condicionada.

### Problema de desajuste

La diferencia entre la respuesta al impulso truncada en la sala de recepción

$$\underline{h} = \begin{bmatrix} \underline{h}_{1,L}^T & \underline{h}_{2,L}^T \end{bmatrix}^T \quad (25)$$

con el filtro  $\hat{\underline{h}} = \begin{bmatrix} \hat{\underline{h}}_{1,L}^T & \hat{\underline{h}}_{2,L}^T \end{bmatrix}^T$ , es cuantificada por

$$\varepsilon = \frac{\|\underline{h} - \hat{\underline{h}}\|}{\|\underline{h}\|}, \quad (26)$$

que llamaremos desajuste. Es posible tener buena cancelación de eco aún cuando el desajuste sea grande. Sin embargo, en tal caso, la cancelación de será degradada si cambian las  $\underline{g}_{i,M}$ . Un objetivo principal es investigar como podemos eludir este problema.

Se puede ver que pasa con los coeficientes del filtro en la práctica y examinar el vínculo entre la correlación cruzada de las señales de entrada y el desajuste. Primero separemos las dos respuestas al impulso en dos partes cada una:

$$\underline{h}_{i,N} = \begin{bmatrix} \underline{h}_{i,L} \\ \underline{h}_{i,t} \end{bmatrix}, \quad i = 1,2$$

dónde  $\underline{h}_{i,L}$  son vectores de largo  $L$ , y  $\underline{h}_{i,t}$  son la parte que resulta de truncar la respuesta al impulso y que no es modelada por el filtro. En el caso ideal de ausencia de ruido, la señal de salida del micrófono se expresa como

$$\begin{aligned} y(n) &= \sum_{i=1}^2 \underline{h}_{i,N}^T(n) \underline{x}_{i,N}(n) \\ &= \sum_{i=1}^2 \underline{h}_{i,L}^T(n) \underline{x}_{i,L}(n) + \sum_{i=1}^2 \underline{h}_{i,t}^T(n) \underline{x}_{i,t}(n-L) \end{aligned} \quad (27)$$

con

$$\underline{x}_{i,t}(n-L) = [x_i(n-L) \ x_i(n-L-1) \ \dots \ x_i(n-N+1)]^T$$

Ahora, usando (21) y (27) podemos reescribir la ecuación normal (19) como

$$\begin{aligned} \underline{\underline{R}}(n) \begin{bmatrix} \hat{\underline{h}}_{1,L}(n) \\ \hat{\underline{h}}_{2,L}(n) \end{bmatrix} &= \sum_{l=1}^n \lambda^{n-l} \begin{bmatrix} x_{1,L}(l) \\ x_{2,L}(l) \end{bmatrix} \begin{bmatrix} x_{1,L}^T(l) & x_{2,L}^T(l) \end{bmatrix} \begin{bmatrix} \underline{h}_{1,L} \\ \underline{h}_{2,L} \end{bmatrix} \\ &\quad + \sum_{l=1}^n \lambda^{n-l} \begin{bmatrix} x_{1,L}(l) \\ x_{2,L}(l) \end{bmatrix} \begin{bmatrix} x_{1,L}^T(l-L) & x_{2,L}^T(l-L) \end{bmatrix} \begin{bmatrix} \underline{h}_{1,t} \\ \underline{h}_{2,t} \end{bmatrix} \end{aligned} \quad (28)$$

Asumiendo que  $\underline{\underline{R}}(n)$  es invertible, podemos resolver esta ecuación para  $\begin{bmatrix} \hat{\underline{h}}_{1,L}^T \\ \hat{\underline{h}}_{2,L}^T \end{bmatrix}$ ,

$$\begin{bmatrix} \hat{\underline{h}}_{1,L}(n) \\ \hat{\underline{h}}_{2,L}(n) \end{bmatrix} = \begin{bmatrix} \underline{h}_{1,L} \\ \underline{h}_{2,L} \end{bmatrix} + \underline{\underline{R}}^{-1}(n) \underline{\underline{R}}_t(n) \begin{bmatrix} \underline{h}_{1,t} \\ \underline{h}_{2,t} \end{bmatrix} \quad (29)$$

dónde  $\underline{\underline{R}}_t(n)$  puede ser escrito como

$$\underline{\underline{R}}_t(n) = \begin{bmatrix} \underline{\underline{R}}_{11,t}(n) & \underline{\underline{R}}_{12,t}(n) \\ \underline{\underline{R}}_{21,t}(n) & \underline{\underline{R}}_{22,t}(n) \end{bmatrix}. \quad (30)$$

El desajuste que obtenemos al minimizar  $J(n)$  es:

$$\begin{aligned}\varepsilon_{\min}^2(n) &= \sum_{i=1}^2 \left[ \underline{h}_{i,L} - \hat{\underline{h}}_{i,L}(n) \right]^T \left[ \underline{h}_{i,L} - \hat{\underline{h}}_{i,L}(n) \right] / \underline{h}^T \underline{h} \\ &= \underline{h}_t^T \underline{Q}(n) \underline{h}_t / \underline{h}^T \underline{h}\end{aligned}\quad (31)$$

dónde

$$\begin{aligned}\underline{h} &= \begin{bmatrix} \underline{h}_{1,L}^T & \underline{h}_{2,L}^T \end{bmatrix} \\ \underline{h}_t &= \begin{bmatrix} \underline{h}_{1,t}^T & \underline{h}_{2,t}^T \end{bmatrix}\end{aligned}$$

y

$$\underline{Q}(n) = \underline{R}_t^T(n) \underline{R}^{-2}(n) \underline{R}_t(n).$$

En esta formulación, está implícita la asunción realista que  $L < M$  y  $\underline{R}^{-1}$  existe.

Si  $L \geq N$  entonces el mínimo valor para el desajuste es cero. Si  $L < N$ , y para  $L$  suficientemente largo, podemos hacer la siguiente aproximación:  $\underline{R}_{11,t}(n) \approx 0_{L \times (N-L)}$  y  $\underline{R}_{22,t}(n) \approx 0_{L \times (N-L)}$ . Sin embargo, no podemos decir nada sobre los términos  $\underline{R}_{21,t}(n)$  y  $\underline{R}_{12,t}(n)$ , ya que estos valores dependen de la correlación cruzada y el defasaje entre las señales  $x_1$  y  $x_2$ . Además, a partir de que la matriz de correlación no esta bien condicionada,  $\underline{Q}(n)$  no es despreciable. De esta manera, en el caso estéreo, el desajuste puede ser muy alto para  $L < N$ .

Para  $L < N$ , se introduce un error en los coeficientes del filtro en ambos casos, aplicaciones mono y estéreo. En el caso estéreo, el problema se ve amplificado a causa de la fuerte correlación entre las dos señales de entrada. Por lo tanto en la práctica, se puede tener gran desajuste aunque tengamos solución única a la ecuación normal.

### El Efecto Tail

Podemos ver que el efecto que produce la cola de la respuesta al impulso en las salas de transmisión y recepción juega un rol clave. Gracias al residuo en la respuesta al impulso no modelada, en la sala de transmisión, podemos tener solución única. Sin embargo, en la sala de recepción esto hace que tengamos potencialmente un gran desajuste.

Hay dos formas para reducir el desajuste. La primera forma es aumentando el largo del filtro adaptivo. Aunque posible, esto vuelve al algoritmo muy lento en términos de convergencia y es caro de implementar en términos de memoria, complejidad, etc.

La correlación entre los canales estéreos puede ser vinculada con la mala condición de la matriz de correlación por medio de la medida de la coherencia. El problema del mal condicionamiento puede ser monitoreado por la función de coherencia, la cual sirve como medida de la correlación. A continuación se presentará una forma posible para reducir la correlación cruzada.

## Solución para reducir la Correlación

### Función de Coherencia

En general, la coherencia entre dos señales aleatorias  $x_1$  y  $x_2$  está definido en el espacio de frecuencias como

$$\gamma(f) = \frac{S_{x_1 x_2}(f)}{\sqrt{S_{x_1 x_1}(f) S_{x_2 x_2}(f)}} \quad (32)$$

dónde

$$\begin{aligned} S_{x_k x_l}(f) &= \sum_{\tau=-\infty}^{+\infty} E\{x_k(n)x_l(n-\tau)\}e^{-j2\pi f\tau} \\ &= \sum_{\tau=-\infty}^{+\infty} R_{x_k x_l}(\tau)e^{-j2\pi f\tau} \end{aligned} \quad ,k,l=1,2. \quad (33)$$

Se puede demostrar que la función de coherencia está relacionada con la matriz de correlación, de forma que el mínimo valor propio está ligado a  $[1 - |\gamma(f)|^2]$ , y si  $|\gamma(f)| = 1$ , la matriz de correlación es singular.

Podemos ver, a partir de (32), que la magnitud de la coherencia entre dos procesos vale uno si y sólo si están relacionados en forma lineal. De acuerdo a esto, una propuesta es aplicar alguna transformación no lineal, teniendo especial cuidado que esto no sea audible y que no tenga efecto en la percepción estereofónica.

### Método de Transformación no-Lineal

Un método simple de transformación no lineal que tiene buena performance usa un rectificador de media onda,

$$x'_i(n) = x_i(n) + \alpha \frac{x_i(n) + |x_i(n)|}{2}, \quad i = 1, 2 \quad (34)$$

Para este método aún puede haber relación lineal entre los canales si por ejemplo  $x_1$  y  $x_2$  son ambas positivas. En la práctica este caso nunca ocurre ya que las señales son siempre de media nula.

Una versión mejorada de esta técnica es usando un rectificador positivo y otro negativo para cada señal respectivamente,

$$x'_1(n) = x_1(n) + \alpha \frac{x_1(n) + |x_1(n)|}{2} \quad (35)$$

$$x'_2(n) = x_2(n) + \alpha \frac{x_2(n) - |x_2(n)|}{2} \quad (36)$$

Los experimentos muestran que la percepción estéreo no es afectada por este método con valores de  $\alpha$  menores a 0.5. La distorsión en la voz es apenas perceptible dada la naturaleza de la percepción humana. Esta distorsión se reconoce aceptable para señales de música pero se hacen objeciones en el caso de tonos puros.

## Filtros Adaptivos de Dos Canales

La optimización del criterio (2) en cancelación de eco estereofónico es hecha por medio de filtros adaptivos de dos canales. Sin embargo, la performance depende en gran medida en la elección del algoritmo. En lo siguiente se presentarán algunos algoritmos clásicos. La generalización a casos multi-canal se pueden derivar de los mismos.

### RLS Doble Canal

La ecuación de actualización clásica para el algoritmo RLS doble-canal viene dado por:

$$e(n) = y(n) - \hat{\underline{h}}^T(n-1)\underline{x}(n), \quad (37)$$

$$\hat{\underline{h}}(n) = \hat{\underline{h}}(n-1) + \underline{\underline{R}}^{-1}(n)\underline{x}(n)e(n) \quad (38)$$

dónde  $\hat{\underline{h}}(n) = \begin{bmatrix} \hat{\underline{h}}_{1,L}^T & \hat{\underline{h}}_{2,L}^T \end{bmatrix}^T$  y  $\underline{x} = \begin{bmatrix} \underline{x}_{1,L}^T & \underline{x}_{2,L}^T \end{bmatrix}^T$ . El índice L lo excluimos en adelante por simplicidad. Se puede obtener ventajas al factorizar la matriz de correlación, en términos de simplificación de cálculo. De acuerdo a esto,

$$\underline{\underline{R}}^{-1}(n) = \begin{bmatrix} \underline{\underline{R}}_1^{-1}(n) & \underline{\underline{0}}_{L \times L} \\ \underline{\underline{0}}_{L \times L} & \underline{\underline{R}}_2^{-1}(n) \end{bmatrix} \begin{bmatrix} \underline{\underline{I}}_{L \times L} & -\underline{\underline{R}}_{12}(n)\underline{\underline{R}}_{22}^{-1}(n) \\ -\underline{\underline{R}}_{21}(n)\underline{\underline{R}}_{11}^{-1}(n) & \underline{\underline{I}}_{L \times L} \end{bmatrix}, \quad (39)$$

dónde

$$\underline{\underline{R}}_1(n) = \underline{\underline{R}}_{11} - \underline{\underline{R}}_{12}(n)\underline{\underline{R}}_{22}^{-1}(n)\underline{\underline{R}}_{21}(n) \quad (40)$$

$$\underline{\underline{R}}_2(n) = \underline{\underline{R}}_{22} - \underline{\underline{R}}_{21}(n)\underline{\underline{R}}_{11}^{-1}(n)\underline{\underline{R}}_{12}(n) \quad (41)$$

Podemos escribir (38) como

$$\hat{\underline{h}}_1(n) = \hat{\underline{h}}_1(n-1) + \underline{\underline{R}}_1^{-1}(n)\underline{z}_1(n)e(n) \quad (42)$$

$$\hat{\underline{h}}_2(n) = \hat{\underline{h}}_2(n-1) + \underline{\underline{R}}_2^{-1}(n)\underline{z}_2(n)e(n) \quad (43)$$

dónde

$$\underline{z}_1(n) = \underline{x}_1 - \underline{\underline{R}}_{12}(n)\underline{\underline{R}}_{22}^{-1}(n)\underline{x}_2(n) \quad (44)$$

$$\underline{z}_2(n) = \underline{x}_2 - \underline{\underline{R}}_{21}(n)\underline{\underline{R}}_{11}^{-1}(n)\underline{x}_1(n) \quad (45)$$

Si  $\underline{\underline{R}}(n) = E\{\underline{x}(n)\underline{x}^T(n)\}$ , tenemos la siguiente propiedad ortogonal,

$$E\{\underline{x}_i^T(n)\underline{z}_j(n)\} = 0 \quad (46)$$

$$E\{\underline{z}_i(n)\underline{x}_j^T(n)\} = \underline{\underline{0}}_{L \times L} \quad i, j = 1, 2, i \neq j \quad (47)$$

Podemos ver que esta implementación del RLS remueve la correlación intercanal antes de actualizar los dos filtros RLS de canal simple.

### Algoritmo APA doble- Canal

El algoritmo APA (Affine Projection Algorithm), se ha popularizado debido a que presenta menor complejidad con respecto al RLS mientras en muchas aplicaciones tiene razones de convergencia cercanas.

Definamos los vectores, de error a priori  $\underline{e}(n)$ , y a posteriori  $\underline{e}_a(n)$  como:

$$\underline{e}(n) = \underline{y}(n) - \underline{\underline{X}}^T(n)\hat{\underline{h}}(n-1) \quad (48)$$

$$\underline{e}_a(n) = \underline{y}(n) - \underline{X}^T(n)\hat{\underline{h}}(n), \quad (49)$$

dónde

$$\begin{aligned} \underline{X}(n) &= \begin{bmatrix} \underline{X}_1(n) \\ \underline{X}_2(n) \end{bmatrix} \\ \underline{X}_i(n) &= [x_i(n) \quad x_i(n-1) \quad \Lambda \quad x_i(n-p+1)], \quad i = 1,2 \end{aligned} \quad (50)$$

$\underline{y}(n)$  y  $\underline{e}(n)$  son vectores que contienen las  $p$  últimas muestras de  $y(n)$  y  $e(n)$  respectivamente. El algoritmo APA es derivado a partir de la restricción

$$\underline{e}_a(n) = 0_{px1} \quad (51)$$

Usando las ecuaciones (48) y (49),

$$\underline{X}^T(n) [\hat{\underline{h}}(n) - \hat{\underline{h}}(n-1)] = \underline{e}(n) \quad (52)$$

Entonces, la solución de norma mínima del sistema indeterminado (52), corresponde al algoritmo estándar APA.

A partir de los presentados, derivan varios algoritmos mejorados (por ej. ELMS, FLRS, PNLMS etc), que se busca, en algunos casos, desarrollar de manera de explotar la correlación intercanal.

### Sistema Híbrido Mono/Estéreo

Cómo se había mencionado, el procesamiento en sub-bandas es una técnica a tener en cuenta en cancelación de eco. La aplicación de la misma puede reducir el peso de cómputo, por un lado haciendo posible la reducción del largo de los filtros al submuestrear, y también por permitir aplicar distintos tipos de filtro según la banda de frecuencia a tratar. A continuación se presenta un caso particular de procesamiento en sub-bandas por el cuál se puede disminuir la complejidad del sistema. Por supuesto que este tema es mucho más amplio e implica aspectos no abarcados en este ejemplo.

Esta implementación de cancelador de eco se basa fundamentalmente en la observación experimental de que el efecto estereofónico es principalmente importante en frecuencias por debajo del kHz. El principio de funcionamiento del AEC mono/estéreo es el uso de sonido estéreo con un AEC estéreo para frecuencias bajas (por ejemplo, menores a 1kHz.), y sonido monofónico con un AEC mono convencional a frecuencias altas. Esta solución es un buen compromiso entre complejidad y realismo espacial.

En la figura podemos ver el funcionamiento de este sistema. Primero las dos señales  $x_1$  y  $x_2$  (canales derecho e izquierdo) pasan por un filtro pasabajos de frecuencia de corte  $f_c$ , en el orden de 1kHz, correspondiendo a la señal que efectúa el efecto estéreo. La suma de los dos canales es pasada por un filtro pasaltos obteniendo la señal mono. El factor de ganancia  $A$  es incluido para ajustar el balance. Finalmente cada parlante recibe la suma de las frecuencias bajas de la señal de su correspondiente canal, más la suma de las frecuencias altas de ambos canales.

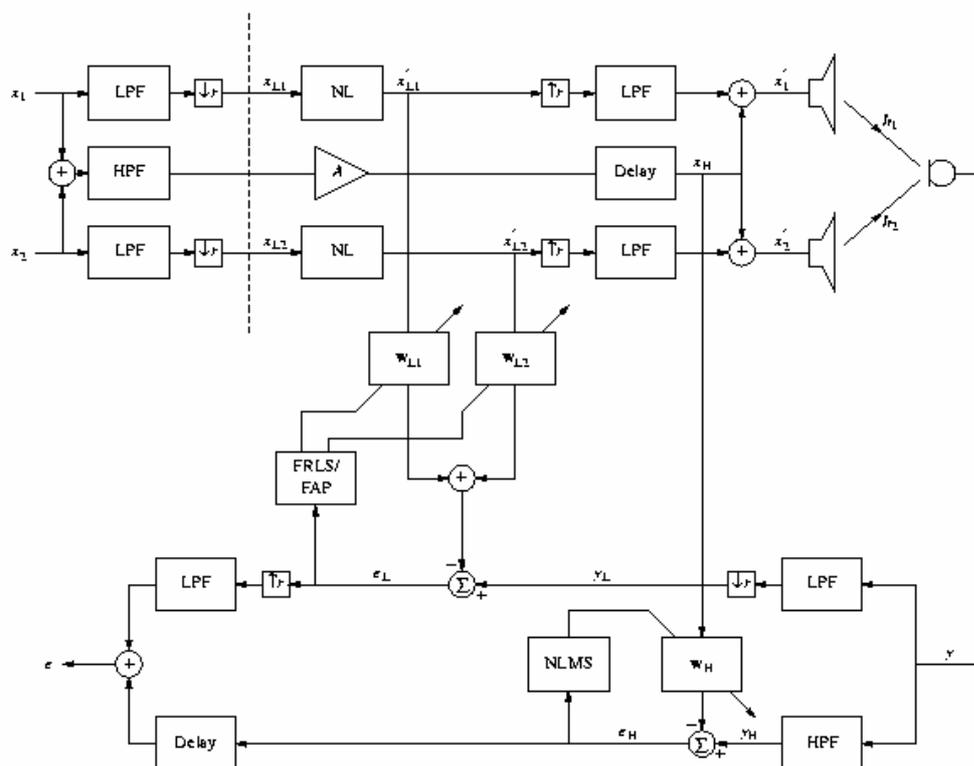


Fig. 11. Cancelador de eco híbrido mono/estéreo.

La estructura anterior es mucho más eficiente que el sistema de banda completa, a pesar del hecho de tener dos AECs diferentes. Por el lado del AEC estéreo, a partir de que la máxima frecuencia es  $f_c$ , podemos submuestrear la señal en un factor  $r = f_s / (2f_c)$ , donde  $f_s$  es la frecuencia de muestreo del sistema. De esta forma la complejidad aritmética se divide por  $r^2$  en comparación con la implementación de banda completa. En este caso podemos usar algoritmos de rápida convergencia como el FRLS de dos canales. Por otro lado, el algoritmo simple NLMS puede ser usado en el filtro de frecuencias altas. La convergencia puede ser lenta, pero esto es de poco interés ya que la mayor energía está en bajas frecuencias como podemos ver en la [Fig. 3](#).

### Detector de Doble Conversación (Doubletalk Detector)

El detector de doble conversación (DTD) es un elemento importante, que es comúnmente empleado en canceladores de eco, el cuál es esencial en el resguardo del filtro adaptivo de divergencias provocadas por la presencia de señal de voz en el lugar de recepción, y otros ruidos distorsionantes. La alta convergencia de los algoritmos adaptivos es generalmente acompañada por alta divergencia en presencia de doble conversación. Cuando es detectada la doble conversación, el paso de adaptación del filtro adaptivo es seteado a cero inhibiendo la adaptación. Es decir, los coeficientes del filtro son congelados hasta que la adaptación pueda seguir.

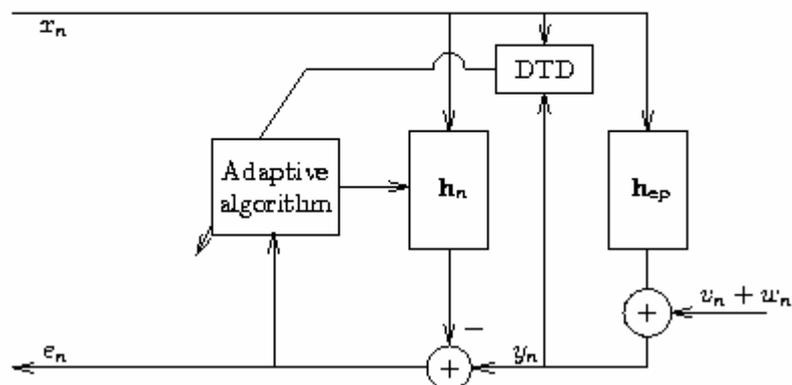


Fig. 12. Diagrama de bloques de cancelador de eco y detector de doble conversación

## Detector de Geigel (Geigel DTD)

Una manera simple y eficiente de detección de doble conversación (DT) es comparando la magnitud de las señales de transmisión y recepción, y declarar la DT si el valor de la señal generada en el lugar de recepción se hace mayor a cierto valor determinado por la señal generada en el lugar de transmisión. Un algoritmo probado, de uso comercial durante muchos años, es el detector de Geigel. En este algoritmo la doble conversación es declarada si

$$|y_n| \geq \mathcal{G} \max\{|x_n|, |x_{n-1}|, \Lambda, |x_{n-L+1}|\}$$

El umbral de detección  $\mathcal{G}$ , es seteado en 0.5 si se supone atenuación de 6 dB, y 0.7 si se supone atenuación de 3 dB. El llamado tiempo de *hangover*,  $T_{hold}$ , es definido como el mínimo tiempo por el cuál la adaptación es inhibida luego de la detección de doble conversación. Luego de detectada la finalización de la doble conversación, la adaptación sigue inhibida durante el tiempo  $T_{hold}$ .

## Comentarios Finales

En *cancelación de eco* podemos encontrar más de una propuesta de uso comercial que resuelve el problema. De todos modos existen nuevos desarrollos tendientes a optimizar los diseños y mejorar la calidad. Aparecen nuevas propuestas, como ser algoritmos aplicados en el espacio de frecuencias, como modo de explotar eficientemente la coherencia entre canales estéreo, al hacer uso de herramientas de estimación espectral. El compromiso entre complejidad (por lo tanto costo) y eficiencia siempre está presente, así como problemas de estabilidad que hay que tener en cuenta.

## Bibliografía

1. T. Gaensler, S. L. Gay, M. M. Sondhi, and J. Benesty, "Double-talk robust fast converging algorithms for network echo cancellation," *IEEE Trans. Speech Audio Processing*, to appear.
2. S. Haykin, *Adaptive Filter Theory*, Prentice Hall, 1996.
3. M. H. Hayes, *Statistical Digital Signal Processing and Modeling*, John Wiley & Sons, 1996
4. T. Gaensler and J. Benesty, "Stereophonic acoustic echo cancellation and two-channel adaptive filtering: an overview," *International Journal of Adaptive Control and Signal Processing*, Feb. 2000.
5. J. H. Cho, D. R. Morgan, and J. Benesty, "An objective technique for evaluating doubletalk detectors in acoustic echo cancelers," *IEEE Trans. Speech Audio Processing*, vol. 7, pp. 718-724, Nov. 1999.
6. P. Eneroth, S. L. Gay, T. Gaensler, and J. Benesty, "An implementation of a stereophonic acoustic echo canceler on a general purpose DSP," in *Proc. ICSPAT*, 1999.
7. T. Gaensler, S. L. Gay, M. M. Sondhi, and J. Benesty, "Double-talk robust fast converging algorithms for network echo cancellation," in *Proc. IEEE ASSP Workshop Appls. Signal Processing Audio Acoustics*, 1999.
8. J. Benesty, A. Gilloire, and Y. Grenier, "A frequency domain stereophonic acoustic echo canceler exploiting the coherence between the channels," published 21 July 1999. Also in *J. Acoust. Soc. Am.*, vol. 106, pp. L30-L35, Sep. 1999.
9. J. Benesty, D. R. Morgan, and M. M. Sondhi, "A better understanding and an improved solution to the specific problems of stereophonic acoustic echo cancellation," *IEEE Trans. Speech Audio Processing*, vol. 6, pp. 156-165, Mar. 1998
10. J. Benesty, D. R. Morgan, and M. M. Sondhi, "A hybrid mono/stereo acoustic echo canceler," in *Proc. IEEE ASSP Workshop Appls. Signal Processing Audio Acoustics*, 1997.
11. J. Benesty, D. R. Morgan, and M. M. Sondhi, "A better understanding and an improved solution to the problems of stereophonic acoustic echo cancellation," in *Proc. IEEE ICASSP*, 1997, pp. 303-306.
12. J. Benesty, P. Duhamel, and Y. Grenier, "Multichannel adaptive filtering applied to multichannel acoustic echo cancellation," in *Proc. EUSIPCO*, 1996.
13. DSP World Spring Design Conference 2000, San José, California, USA, April 2000, Conference Proceedings, pp. 1 – 28.