

IMPLEMENTACIÓN DE UN ARREGLO SUPERDIRECTIVO DE MICRÓFONOS CON MÚLTIPLES LÍNEAS DE RETARDO

C. A. Medina S.
cmedina@lenguaje.com

C. V. Rodríguez R.
crodrigu@bellsouth.com.ec

J. A. Apolinário Jr., D.Sc.
apolin@espe.edu.ec

R. D. LeónV., M.Sc
rleon@li.fie-espe.edu.ec

FACULTAD DE INGENIERÍA ELECTRÓNICA
ESCUELA POLITECNICA DEL EJÉRCITO

Resumen

Temas tradicionalmente estudiados en antenas, tales como “beamforming” y estimación de dirección de llegada, han sido blancos de múltiples investigaciones en los sistemas de procesamiento de señales usando arreglos de micrófonos que presentan varias aplicaciones en diferentes campos como teleconferencia, comunicaciones de manos libres y “hearing aids”.

Este trabajo presenta el estudio, desempeño e implementación de un “beamformer” adaptativo basado en un arreglo de micrófonos con múltiples retardos, en donde se busca el número óptimo de retardos por línea. El análisis del desempeño se hace basándose en la distorsión logarítmica causada por el arreglo, respecto a la señal de referencia. En el trabajo se ocupa los algoritmos adaptativos: “Constrained LMS” (Frost) y “Constrained Normalized LMS” (CNLMS) con tamaño de paso variable. El estudio del arreglo y las simulaciones en la computadora se realiza para el caso “far-field”.

Abstract

Topics traditionally studied in antennas, such as “beamforming” and direction of arrival estimation, have been targets of multiple research in signal processing systems using microphone arrays that present several applications in fields like teleconferencing, hands free communications and hearing aids.

This work presents the study, performance and implementation of an adaptive beamformer based on a microphone array with multiple taps, searching for the optimal number of delays per line. Performance analysis is carried out based on the logarithmic distortion produced by the array with respect to the reference signal. On this paper we

work with the following adaptive algorithms: “Constrained LMS” (Frost) and “Constrained Normalized LMS” (CNLMS) with optimized step size. The study of the array and the computer simulations were carried out for the “far field” case.

1. Introducción

Los arreglos de sensores tienen varios campos de aplicación, tales como, cancelación de interferencias espaciales y multipath en comunicaciones inalámbricas, atenuación de jamming, detección de la dirección de arribo de una señal (DOA), teléfono de manos libres, vídeo conferencia y “hearing aids”, entre otras[1].

Varias décadas atrás las técnicas de beamforming eran ocupadas principalmente en aplicaciones de radio frecuencia y sonar [2], en la actualidad su campo de aplicación se ha expandido también a señales de audio ocupando arreglos de micrófonos. Debido al alto costo de micrófonos superdirectivos, estos arreglos se han vuelto populares en reconocimiento de voz y vídeo conferencia, ya que además ofrecen comodidad al locutor.

Los algoritmos adaptativos que ocupamos son el “Constrained LMS” y el “Constrained Normalized LMS” con tamaño de paso variable, estos algoritmos son muy populares debido a su facilidad de implementación y bajo costo computacional.

Debido al pobre desempeño de los filtros espaciales con un solo retardo por sensor, se tiende a aumentar el número de sensores. La estructura ocupada es de un filtro espacial con varios retardos por cada sensor, como se indica en la Figura 1. Esta estructura ofrece mejoras en el desempeño del filtro a costa de un aumento de complejidad computacional.

2. Métodos

Como “beamforming” entendemos al proceso mediante el cual se obtiene la cancelación de interferencias modelando la forma del patrón de directividad.

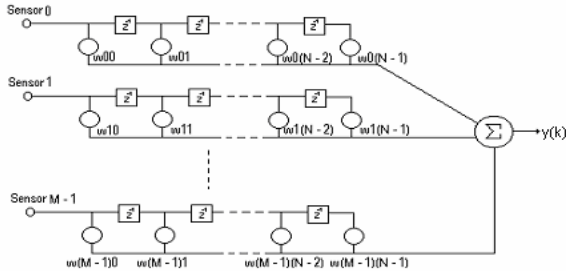


Fig. 1. Arreglo de M micrófonos y N líneas de retardo.

La dirección de arribo de una señal al beamformer causa que la misma llegue en un instante diferente a los sensores. Este retraso producido por la dirección de arribo de la señal y la distancia entre los micrófonos, es similar a los retardos producidos en los taps de un filtro temporal; esta analogía permite ocupar los mismos algoritmos en los dos casos.

2.1. Técnicas de Comparación de Patrones

La medición del desempeño del “beamformer” se puede hacer de una manera subjetiva, sin embargo, esta no representa una medida de fácil manipulación matemática. Existen métodos matemáticos que pueden dar una mejor idea del desempeño, midiendo la disimilitud entre dos patrones de voz, de acuerdo a un algoritmo que debe ajustarse a parámetros que modifican, o al contrario, no modifican esta señal. Parámetros que no modifican el sonido percibido son [3]:

- Desplazamiento espectral, es decir, $S'(w) = S(w) w^\alpha$, donde α es el desplazamiento espectral.
- Filtrado pasabajos, $S'(w) = S(w) |H(e^{jw})|^2$, donde $H(e^{jw})$ es la función de transferencia de un filtro pasabajos.
- Filtrado pasaaltos, $S'(w) = S(w) |H(e^{jw})|^2$, donde $H(e^{jw})$ es la función de transferencia de un filtro pasaaltos.
- Filtrado de banda, $S'(w) = S(w) |H(e^{jw})|^2$, donde $H(e^{jw})$ es la función de transferencia de un filtro rechaza banda.

Para los casos anteriores la medida de disimilitud deberá ser baja, al contrario, para los siguientes casos la medida de disimilitud deberá ser alta,

debido a que son parámetros que cambian de mayor manera el sonido percibido.

- Diferencias significativas en la posición de las frecuencias formantes, esto es, las frecuencias resonantes de $S(w)$ y $S'(w)$ ocurren a frecuencias distantes.
- Diferencias significativas en el ancho de banda de las frecuencias formantes, esto es, las frecuencias resonantes de $S(w)$ y $S'(w)$ tienen muy distintos anchos de banda.

2.2 Distancia Espectral Logarítmica

Esta forma de medir la distorsión es perceptualmente relevante, por esto ha sido escogida para medir el desempeño del “beamformer”. Considérese $S(w)$ y $S'(w)$ los espectros de las señales, la diferencia entre los dos espectros medidos en magnitud logarítmica versus la frecuencia es:

$$V(w) = \log S(w) - \log S'(w) \quad (1)$$

La medida de distorsión o distancia entre los espectros está dada por la familia de normas definida por:

$$d(S, S') = \int_{-\pi}^{\pi} |V(w)|^p \frac{dw}{2\pi} \quad (2)$$

Para $p = 1$, (2) define el valor medio absoluto de la distorsión espectral, para $p = 2$ la misma ecuación define el valor cuadrático medio (rms), cuando p tiende al infinito (2) se reduce a la medida del pico de la distorsión espectral.

Debido a que la integral es expresada en logaritmos, ésta se mide en dB si se multiplica por el factor 4,34.

3. Algoritmos de Adaptación

3.1. Criterio de Desempeño

Para que se pueda entender mejor el problema, representamos matemáticamente el arreglo de micrófonos por:

$$Y(n) = W^H X(n)$$

las matrices W y X son:

$$\begin{aligned}
W_0^T &= [w_{00} \ w_{01} \ w_{02} \ \dots \ w_{0(N-1)}] \\
W_1^T &= [w_{10} \ w_{11} \ w_{12} \ \dots \ w_{1(N-1)}] \\
&\vdots \\
&\vdots \\
W_{M-1}^T &= [w_{(M-1)0} \ w_{(M-1)1} \ w_{(M-1)2} \ \dots \ w_{(M-1)(N-1)}] \\
W &= \begin{bmatrix} W_0 \\ W_1 \\ W_2 \\ \dots \\ W_{M-1} \end{bmatrix} \quad (3)
\end{aligned}$$

$$\begin{aligned}
X_0^T &= [x_{00} \ x_{01} \ x_{02} \ \dots \ x_{0(N-1)}] \\
X_1^T &= [x_{10} \ x_{11} \ x_{12} \ \dots \ x_{1(N-1)}] \\
&\vdots \\
&\vdots \\
X_{M-1}^T &= [x_{(M-1)0} \ x_{(M-1)1} \ x_{(M-1)2} \ \dots \ x_{(M-1)(N-1)}] \\
X &= \begin{bmatrix} X_0 \\ X_1 \\ X_2 \\ \dots \\ X_{M-1} \end{bmatrix} \quad (4)
\end{aligned}$$

donde $(\cdot)^T$ representa la transpuesta de la matriz y $(\cdot)^H$ representa el Hermitiano o conjugado transpuesto.

$X(n)$ es la matriz de muestras de la señal recogida por cada sensor, esta matriz debe ser de dimensiones $(MN) \times I$.

La restricción indicada obedece a las ecuaciones:

$$\begin{aligned}
G &= 1 \\
C &= \begin{bmatrix} S_0 \\ S_2 \\ S_1 \\ \dots \\ S_{M-1} \end{bmatrix} \quad (5)
\end{aligned}$$

Para un mejor entendimiento se ha expresado el vector C en función de los vectores $S_0, S_1, S_2, \dots, S_{M-1}$, estos representan los retardos respectivos de la señal de entrada en cada línea de retardos, de la siguiente manera:

$$\begin{aligned}
S_0^T &= [1 \ e^{-j\phi'} \ \dots \ e^{-j(N-1)\phi'}] \\
S_1^T &= [e^{-j\phi} \ e^{-j(\phi+\phi')} \ \dots \ e^{-j[\phi+(N-1)\phi']}] \\
&\vdots \\
&\vdots \\
S_{M-1}^T &= [e^{-j(M-1)\phi} \ e^{-j[(M-1)\phi+\phi']} \ \dots \ e^{-j[(M-1)\phi+(N-1)\phi']}]
\end{aligned} \quad (6)$$

Como se puede observar en las ecuaciones anteriores, ϕ' corresponde al retardo producido entre taps dentro del filtro transversal y cuyo valor ideal es igual a:

$$\phi' = \frac{2\pi v T_s}{\lambda} \quad (7)$$

donde; v es la velocidad del sonido (340 m/s), T_s es el período de muestreo y λ es la longitud de onda de la señal de arribo.

En (6), ϕ corresponde al retardo eléctrico producido en la señal que llega en una dirección deseada [4].

$$\phi = \frac{2\pi d}{\lambda} \sin(\theta) \quad (8)$$

siendo d la distancia entre micrófonos y θ es la dirección de arribo de la señal deseada.

Es fácil demostrar que la distancia máxima entre micrófonos para que no se produzca ambigüedad es de $\lambda/2$ [4].

El criterio de desempeño que se considera para reducir las interferencias direccionales es el “**Linearly Constrained Minimum Variance (LCMV)**”, el fundamento básico de este algoritmo es restringir, en amplitud y fase, la salida del beamformer a señales que llegan de una o más direcciones de interés. Los pesos del filtro son seleccionados de tal forma que se minimice la varianza o la potencia del error. Es decir, el problema puede ser representado como sigue:

$$\min_w E[e(n)^2] \quad \text{sujeto a } C^H W = G \quad (9)$$

donde $e(n)$ es el error de la salida del filtro:

$$e(n) = d(n) - y(n)$$

$$d(n) = 0 \quad (\text{sujeto a restricciones})$$

$$y(n) = W^H X(n)$$

$$\begin{aligned}
E[e(n)^2] &= E[y(n)y^H(n)] = E[W^H X(n)X^H(n)W] \\
&= W^H E[X(n)X^H(n)]W = W^H R W
\end{aligned}$$

por lo tanto, (9) se reduce a:

$$\min_w W^H R W \text{ sujeto a } C^H W = G \quad (10)$$

donde \mathbf{R} , la matriz de autocorrelación de los datos de entrada, tiene dimensiones $(M \times N) \times (M \times N)$, \mathbf{C} es un vector dirección, de dimensiones $(M \times N) \times L$, \mathbf{W} , el vector de coeficientes del filtro, es de $(M \times N) \times L$ y \mathbf{G} es de $L \times 1$. Donde M es el número de sensores, N el número de taps y L es el número de restricciones, cada restricción disminuye un grado de libertad al filtro para atenuar la interferencia, nosotros ocupamos una sola restricción.

3.2. Matriz de pesos óptimos.

Una vez construida la ecuación de restricciones, procedemos a derivar la matriz de pesos óptimos, asumiendo que los pesos son fijos, constantes y complejos. Ocupando el método de los multiplicadores de Lagrange, la función objetivo a minimizar resulta ser:

$$J(w) = \frac{1}{2} W^H R W + \gamma_1^T [\text{Re}(C^H W - G)] + \gamma_2^T [\text{Im}(C^H W - G)] \quad (11)$$

donde; γ_1 y γ_2 son vectores columna (reales) de los multiplicadores de Lagrange.

Obteniendo el gradiente de $J(W)$ e igualando a cero obtenemos W óptimo para el caso complejo.

$$W_{opt} = R^{-1} C (C^H R^{-1} C)^{-1} G \quad (12)$$

donde la inversa existe debido a que R es positiva definida.

3.3. Constrained LMS

La ecuación (12) es la matriz de pesos óptimos, sin embargo esta matriz es útil solo cuando la señal de entrada es estacionaria. En casos reales es más común encontrar señales que conservan sus estadísticas por sectores, por eso es necesario desarrollar algoritmos que se adapten a estos cambios, uno de ellos es el algoritmo de Frost o Constrained Least Mean Square (CLMS).

El CLMS es una derivación del LMS y se basa en la aproximación de los pesos óptimos en el caso de haber restricciones, esta aproximación es realizada siguiendo el método de Steepest Descent.

Para el caso complejo, el proceso de actualización de W , en este caso, es:

$$W(n+1) = W(n) - \frac{\mu}{2} \nabla_w [J(W)] \quad (13)$$

Reemplazando adecuadamente se puede demostrar que la fórmula de actualización de la matriz de pesos es:

$$\begin{aligned} W(n+1) &= P[W(n) - \mu R W(n)] + F \\ P &= I - C(C^H C)^{-1} C^H \\ F &= C(C^H C)^{-1} G \end{aligned} \quad (14)$$

Aproximando $R = X(n)X(n)^H$ y usando $y = W^H X$ en (14) resulta:

$$W(n+1) = P[W(n) - \mu X(n)y^*] + F \quad (15)$$

El término μ debe ser escogido de tal forma que permita la convergencia de (15), para esto μ debe cumplir [4]:

$$0 < \mu < \frac{2}{\text{tr}(R)} \quad (16)$$

donde $\text{tr}(R)$ es la traza de la matriz de autocorrelación de la señal de entrada.

3.4. Constrained Normalized LMS

Este algoritmo es ampliamente ocupado en filtrado espacial debido a su simplicidad y rápida convergencia, sin embargo su desempeño depende del esparcimiento de los valores propios de la matriz de autocorrelación de la señal de entrada [5].

Los siguientes pasos constituyen el CNLMS:

$$\begin{aligned} e(n) &= d(n) - x^H(n)w(n) \\ w(n+1) &= P \left[w(n) + \mu \frac{e(n)x(n)}{x^H(n)Px(n)} \right] + F \end{aligned} \quad (17)$$

Este algoritmo presenta una mejor convergencia y un mejor desempeño si el tamaño de paso se adapta de acuerdo a las siguientes fórmulas [6]:

$$\mu(n) = \mu(n+1) \left(\frac{1 - \frac{\mu(n-1)}{M * N}}{1 - \frac{\mu^2(n-1)}{M * N}} \right) \quad (18)$$

donde $(M*N)$ es el orden del filtro.

4. Simulación

Para obtener una medida de desempeño del beamformer se ocupó dos señales, una proveniente de 0 rad y otra proveniente de los $\pi/3$ rad, a cada señal se le adicionó ruido blanco. La respuesta del arreglo se fijó en 0 rad.

En cada simulación se varió λ , es decir la matriz de restricción y la distancia entre micrófonos, y μ de forma tal que se obtenga la mejor respuesta del arreglo en cada caso.

Para la medida de distorsión se ocupó como señal patrón, la señal proveniente de la dirección deseada, 0 rad, y como señal sujeta a medida, la señal de salida del beamformer.

5. Conclusiones

Como podemos observar en las figuras 2 y 3, a partir de 5 taps el desempeño no mejora notablemente, pero sí aumenta considerablemente el costo computacional de cada algoritmo. Nosotros sugerimos aprovechar este costo computacional aumentando el número de micrófonos y ocupando unos 4 taps por cada uno.

Es de notar también que al ocupar un número mayor de taps, el ajuste de los parámetros es más fino y es más difícil lograr la convergencia del beamformer.

Como se esperaba el CNLMS presenta una más rápida convergencia que el CLMS (figura 4) pero a costa de aumentar la complejidad computacional y el error, por esto el desempeño disminuye al ocupar el CNLMS frente al CLMS.

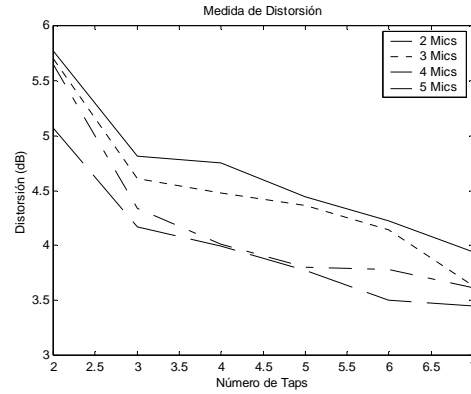


Fig. 2. Variación de la distorsión para $p = 1$ (ecuación 2) ocupando el CLMS.

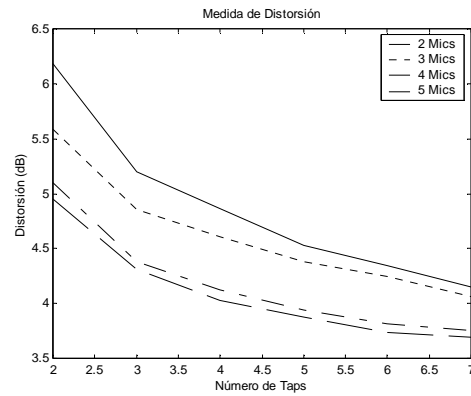


Fig. 3. Variación de la distorsión para $p = 1$ (ecuación 2) ocupando el CNLMS.

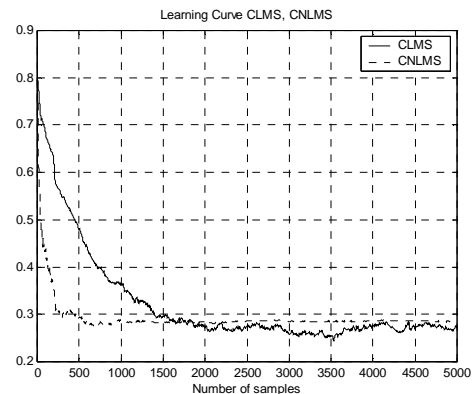
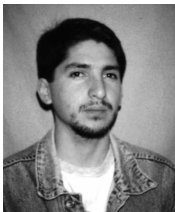


Fig. 4. Curva de Aprendizaje de los dos algoritmos.

Referencias

- [1] B. D. Van Veen and K. M. Buckley, "Beamforming: A Versatile Approach to Spatial Filtering", IEEE ASSP Magazine, pp 4-24, April 1988.
- [2] P. L. Chu, "Superdirective Microphone Array for a Set-Top Video Conferencing System", Proc. ICASSP, pp. 235-238, Munich, Germany, April 1997.
- [3] L. Rabiner and B.-H. Juang, "Fundamentals of Speech Recognition", Prentice - Hall, Signal Processing Series, 1993.
- [4] S. Haykin, "Adaptative Filter Theory", Prentice - Hall, third edition, 1996.
- [5] J. A. Apolinário Jr., S. Werner, P. S. R. Diniz, and T. Laakso, "Constrained Normalized Adaptative Filters for CDMA Mobile Communications", Proc. EUSIPCO, pp.2053-2056, Island of Rhodes, Greece, September 1998.
- [5] J. A. Apolinário Jr., P. S. R. Diniz, T. Laakso, and M. L. R. de Campos, "Step Size Optimization of the BNDR - LMS Algorithm", Proc. EUSIPCO, pp.1429-1432, Island of Rhodes, Greece, September 1998.



César A. Medina S.- Nació en Cuenca - Ecuador el 17 de diciembre de 1976, bachiller en Humanidades Modernas especialidad Físico Matemáticas en el Instituto Técnico Superior Daniel Álvarez Burneo (Loja - 1994), Segundo Premio Nacional de Física 1994, está actualmente culminando los estudios de Ingeniería Electrónica especialidad en Telecomunicaciones en la Escuela Politécnica del Ejército (Quito Ecuador), trabaja en SIGNUM Cia. Ltda. en el Departamento de Investigación y Desarrollo en el Area de Lingüística Computacional. Sus áreas de interés son las redes neuronales, procesamiento digital de señales, desarrollo de software.



Carlos V. Rodríguez R.- Nació en Quito el 23 de Julio de 1976, bachiller en humanidades modernas con la especialidad de Físico Matemático en el Colegio Municipal Experimental "Sebastián de Benalcázar", actualmente culminando los estudios superiores en la Facultad de Ingeniería Electrónica de la Escuela Politécnica del Ejército (ESPE) en la especialidad

de Telecomunicaciones. Presta sus servicios en Otecel S.A. Ecuador como Ingeniero de Calidad de Red y CDPD en el área de Radio Frecuencia desde Julio de 1999. Las áreas de interés son el procesamiento digital de señales y las comunicaciones móviles.



Ing. José Antonio Apolinário Jr.- Nació en Taubaté, Brasil, en Enero de 1960. Egresó de la escuela de formación de oficiales del Ejército de Brasil, la Academia Militar de Agulhas Negras (AMAN), Resende, Brasil, en 1981 y se encuentra ahora con el rango de teniente coronel. Obtuvo sus títulos de ingeniero electrónico en el Instituto Militar de Engenharia (IME), Rio de Janeiro, Brasil, 1988, M.Sc. en la Universidade de Brasília (UnB), Brasília, Brasil, 1993 y D.Sc. en la Universidade Federal do Rio de Janeiro (COPPE/UFRJ), Rio de Janeiro, Brasil, 1998. Tiene especialización en Guerra Electrónica en Alemania (Rohde & Schwarz, 1990) y pos-grado "lato sensu" en análisis de sistemas (Rio de Janeiro, 1987). Ha trabajado en el Centro de Instrução de Guerra Eletrônica (CIGE) del Ejército de Brasil de 1989 hasta 1994. Estuvo como investigador visitante en la "Helsinki University of Technology" (HUT), Finlandia, en 1997 donde trabajó bajo un convenio entre la COPPE/UFRJ y la HUT. En 1998 fue nombrado profesor del IME y desde noviembre de 1998 está como asesor y profesor visitante en la Escuela Politécnica del Ejército (ESPE), Ecuador. El tiene decenas de papers publicados en revistas y conferencias brasileñas e internacionales. Doctor Apolinário es miembro del IEEE.

Rubén D. León V.- Lugar y Fecha de Nacimiento: Ambato, 30 de abril de 1962. Master en Ciencias: Instituto Tecnológico de Aeronáutica, BRASIL, 1992, Ingeniero Electrónico: ESPE, 1985, Anillo de Oro, Primera Antigüedad. Funciones Actuales: Subdecano de la Facultad de Ingeniería Electrónica de la ESPE, Asesor en Telecomunicaciones del Jefe del Comando Conjunto de las Fuerzas Armadas. Areas de Interés: Procesamiento Digital Adaptativo de Señales de Radar, Gestión y Administración del espectro electromagnético, Procesamiento de señales aplicado en telecomunicaciones, Modelamiento estocástico de señales.