

# Localización de Fuente Acústica por ERT

G. Miramontes de L., E. García D., I. de la Rosa V., M. Araiza E.,

J. J. Villa H., C. Sifuentes G., S. Mandujano S., y A. Hernández F.

Laboratorio de Procesamiento Digital de Señales,

Universidad Autónoma de Zacatecas, Zacatecas, Zac.-98000.

Tel: +(492)9239407, ext 1518,

correo-e: gmiram@cantera.reduaz.mx, egarcia@cantera.reduaz.mx

**Resumen** – Se describe la realización de un sistema para la localización de una fuente acústica, también conocida como estimación de dirección de arribo. El algoritmo de estimación supone una sola fuente y está basado sólo en la estimación del retardo temporal ERT, el cual es detectado por el arreglo de un par de micrófonos. Se hicieron pruebas de adquisición y estimación del ángulo de arribo, bajo el ambiente MATLAB y se desarrolló una aplicación y una interfase para controlar la orientación de un motor de pasos, bajo el ambiente Windows. Como una aplicación se propone el control de la orientación de una cámara de video en salas para video conferencia.

**Abstract**— This document describes the implementation of an acoustic direction finder system, also known as direction of arrival estimator. Its main application is devised to control the orientation of a video camera for video conference rooms. The algorithm assumes a single source and it is based only on the time delay, which is detected by an array of two microphones. Acquisition tests and angle of arrival estimation were done under MATLAB, then a Windows application was developed together with an interfase to control the camera orientation attached to a stepping motor.

**Palabras Clave** – *Ángulo de arribo, arreglo de micrófonos, conformación de haz, dirección de llegada.*

## I. INTRODUCCIÓN

EL aumento en las prestaciones de los sistemas de comunicación vía internet ofrece la posibilidad de instalar una sala de video conferencia prácticamente en cualquier lugar, es decir, sala audio-visual, cuarto, o salón de clase. Una característica de las cámaras para video conferencia es la capacidad de dirigir su lente hacia la persona que habla. Esto se logra estimando el ángulo de arribo (AdA) de la señal al sistema de captura. Existen en el mercado sistemas para video conferencia que no siempre son accesibles para pequeñas instituciones, y menos para uso personal.

Para la estimación del (AdA), i.e, ubicar la dirección en donde se encuentra una fuente acústica, se usa un conjunto o arreglo de micrófonos separados espacialmente. Este arreglo se usa para dirigir automáticamente una cámara hacia un locutor activo en una sala de video conferencia [1]. La salida del arreglo de micrófonos también se puede procesar de modo que tenga mejor calidad la señal recibida en cierta dirección, suprimiendo las señales provenientes de otras direcciones, a este tipo de procesamiento de la señal se le llama conformación de haz (beam forming). Este último aspecto no se cubre en este trabajo, aunque la conformación de haz y la estimación del AdA se pueden considerar como parte de un mismo problema. Para la conformación de haz, se requiere conocer el AdA, y por otra parte, algunos algoritmos para la estimación del AdA incluyen la conformación de haz.

Las operaciones que se pueden realizar con un arreglo de micrófonos, en general con un arreglo de sensores separados espacialmente, son: realce, localización de fuente, separación de fuentes, entre otras. En el caso de una sola fuente, las operaciones de cambio de fase y suma de las señales de todos los micrófonos dan como resultado una mejora en la relación señal a ruido (SNR) [2]. En los casos de cancelación de fuentes de interferencia, la mejora depende del número de micrófonos en el arreglo [3]. Con esta misma idea se puede usar el arreglo en aplicaciones de localización de fuente, en donde se usa la diferencia de tiempo en la llegada de la señal a cada micrófono, para estimar el ángulo de llegada en el caso de un plano.

El artículo se presenta de la siguiente manera, en la sección II se aborda el marco teórico, en donde se describe el principio básico tanto del funcionamiento del arreglo como de los algoritmos de estimación del AdA. En la Sección III, se presenta la descripción del arreglo propuesto en este trabajo y el desarrollo en el ambiente Matlab. En la sección IV se presenta la descripción de la principal aplicación que se propone, la cual incluye la interfase de control de un motor de pasos para dirigir la cámara hacia el locutor activo. En la sección IV se dan algunos resultados. Finalmente, en la sección V se incluyen algunas conclusiones y se describe el trabajo a futuro.

## II. PRINCIPIO BÁSICO DE FUNCIONAMIENTO

En la Figura 1, se muestra esquemáticamente la configuración espacial del arreglo de micrófonos, en este caso dos, con una sola fuente acústica. El AdA desde la fuente hasta el primer micrófono  $M_0$ , está denotado por  $\theta_0$  a una distancia  $\rho_0$ ,  $d$  es la distancia, en metros, entre micrófonos.

Como la señal de voz tiene una gama de frecuencias que va desde los 100 Hz hasta los 5 kHz, aproximadamente, las frecuencias fundamentales o formantes oscilan en promedio entre los 200 y los 1000 Hz. Estas formantes, que se deben a la resonancia del tracto vocal, contienen la mayor parte de la energía. Las formantes pueden cubrir varias frecuencias, y el algoritmo para la estimación del AdA debe extraer la información del retardo de tiempo en todas estas frecuencias. Así pues, se requiere un algoritmo de banda amplia. Suponiendo una frecuencia de 1000 Hz, tomada aproximadamente como el promedio entre las formantes para sonidos vocalizados, se obtiene una longitud de onda  $\lambda$  de 0.342 m.

La separación entre los micrófonos no debe exceder  $\lambda/2$  [4], así pues  $d_{max} \leq \lambda/2$ . Suponiendo

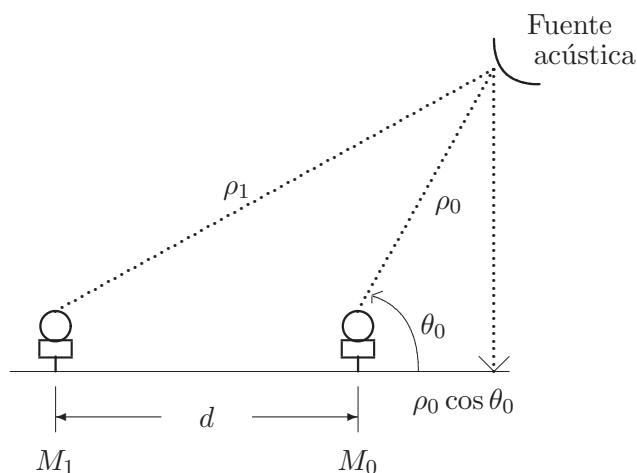


Figura 1. Principio de operación del arreglo para la estimación del AdA.

una distancia  $d = 0.15$  m, y un ángulo  $\theta_0 = 60$  grados, para el arreglo de la Figura 1, se obtiene una  $\rho_0$  de 0.1315 m. Entonces,

$$\begin{aligned}\rho_1 &= \sqrt{(\rho_0 \operatorname{sen}\theta_0)^2 + (d + \rho_0 \cos\theta_0)^2} \\ &= 0.2511 \text{ m.}\end{aligned}\quad (1)$$

El retardo de tiempo de arribo de la señal entre los micrófonos está dado por

$$\tau = \frac{\rho_1 - \rho_0}{C} = 0.325 \text{ ms} \quad (2)$$

en donde se ha empleado el valor de  $C = 324$  m/s.

En la práctica, se requiere medir el retardo  $\tau$  para estimar  $\theta_0$ , es decir el AdA. En la estimación del AdA, las señales recibidas por los micrófonos son versiones retrasadas de la misma señal, y ese retardo está en función del ángulo de incidencia y la separación  $d$ .

Si se suponen condiciones de campo lejano, el frente de onda es plano, y la diferencia de tiempo se puede aproximar por

$$\Delta\rho_{01} = \rho_1 - \rho_0 \approx d \cos\theta, \quad (3)$$

de donde

$$\tau = \frac{\Delta\rho_{01}}{C} \approx \frac{d \cos\theta}{C}. \quad (4)$$

Finalmente, se puede calcular el AdA por

$$\hat{\theta} = \cos^{-1}\left(\frac{\hat{\tau}C}{d}\right), \quad \text{con } -\pi/2 \leq \theta \leq \pi/2. \quad (5)$$

### A. Estrategias de localización de fuente

Las estrategias para localización de fuente se pueden clasificar en tres grandes categorías [5] *i)* por la potencia de la respuesta dirigida, *ii)* por el conformador de haz de alta resolución, y *iii)* las basadas en la estimación del retardo de tiempo (ERT).

En la primera categoría se trata de ponderar y sumar las señales de diferentes micrófonos para obtener la potencia máxima. El estimador óptimo de máxima verosimilitud es un ejemplo de éste. Su desventaja es la complejidad computacional y la estimación del AdA involucra la solución de ecuaciones no lineales. El más simple es el conformador de haz de retraso y suma (DSB de sus siglas en inglés).

La segunda categoría está basada en el análisis espectral de alta resolución que utiliza la matriz de correlación cruzada y la divide en subespacios de señal y ruido usando descomposición en valores singulares para estimar el AdA. Un ejemplo de éste método es el algoritmo MUSIC (Multiple Signal Classification).

El tercer método estima la diferencia de tiempo de arribo entre pares de micrófonos para estimar el AdA. Hay varios métodos que se pueden usar para calcular los retardos de tiempo entre pares de micrófonos, tal como el método de la correlación cruzada generalizada [6] o filtrado de banda estrecha seguido de la estimación de la diferencia de fase de sinusoides, entre otros. La simplicidad de los algoritmos de estimación de retardo de tiempo (ERT) ha hecho que éste sea el método preferido para la estimación de AdA que usa arreglos de micrófonos.

Por razones de espacio, no se puede incluir la discusión de cada caso. Se presenta la estimación que utiliza la correlación generalizada.

#### A.1 Estimación de $\tau$

Se sabe que la correlación cruzada de dos señales  $x_1(t)$  y  $x_2(t)$  está relacionada a la transformada inversa de Fourier de la densidad espectral de potencia cruzada  $G_{x_1,x_2}$ , es decir

$$R_{x_1,x_2} = \int_{-\infty}^{\infty} G_{x_1,x_2}(f) e^{j2\pi f\tau} df \quad (6)$$

en donde

$$R_{x_1,x_2} = \frac{1}{T-\tau} \int_{\tau}^T x_1(t)x_2(t-\tau)dt \quad (7)$$

y

$$G_{x_1,x_2}(f) = X_1(f)X_2^*(f). \quad (8)$$

Pre-filtrando  $x_1(t)$  y  $x_2(t)$  por  $H_1(f)$  y  $H_2(f)$  para tener una salida  $Y_1(f)$  y  $Y_2(f)$ , entonces el espectro cruzado de las salidas filtradas está dado por

$$G_{y_1,y_2}(f) = H_1(f)H_2^*(f)G_{x_1,x_2}(f). \quad (9)$$

Así, la correlación cruzada generalizada está dada por

$$\hat{R}_{y_1,y_2}^g = \int_{-\infty}^{\infty} \psi_g(f)G_{x_1,x_2}(f)e^{j2\pi f\tau} df, \quad (10)$$

en donde  $\psi = H_1(f)H_2^*(f)$ . La estimación de  $\tau$  es aquel valor que maximiza  $R_{x_1,x_2}$ . El papel de la función de ponderación  $\psi$  es acentuar el pico en  $R_{x_1,x_2}$  para mejorar la resolución de retardo de tiempo. Así,

$$\hat{\tau} = \max_{\tau} \int_{-\infty}^{\infty} \frac{G_{x_1,x_2}(f)}{|G_{x_1,x_2}(f)|} e^{j2\pi f\tau} df, \quad (11)$$

en donde  $\hat{\tau}$  es el tiempo de retardo, calculado como el índice del valor máximo de la función de correlación cruzada normalizada.

### III. DESCRIPCIÓN DEL ARREGLO

El arreglo de micrófonos empleado en el sistema consta de dos micrófonos del tipo electret con una etapa de amplificación a base de un amplificador operacional doble LM358 alimentado por una batería de 9 Volts. La distancia entre los micrófonos es de 13.5 cm, y están colocados sobre una barra de modo que presenten una buena alineación. En la Figura 2, se muestra el montaje del sistema con el arreglo de dos micrófonos al frente, y el amplificador en la parte posterior. El sistema está conectado a la entrada de línea de la tarjeta de sonidos.

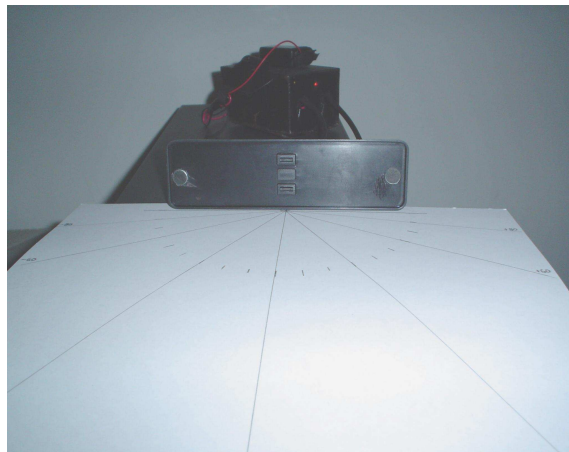


Figura 2. Arreglo experimental de dos micrófonos.

En el cálculo del AdA se empleó el ambiente de trabajo MATLAB, para realizar la correlación entre las señales del canal izquierdo y derecho, correspondientes a los micrófonos M0 y M1, las cuales son entregadas por las funciones de captura de audio `analoginput`, vía la tarjeta de sonido en modo estéreo. La frecuencia de muestreo, por omisión, es de 8 kHz y se toman 8,000 muestras en cada captura. En la Figura 3 se muestra el diagrama de flujo del programa de captura y cálculo del AdA. El índice corresponde al número de muestras en donde se encuentra el valor máximo de la función de correlación.

En la Figura 4, la cual se obtuvo con una pequeña variación a la lógica del programa descrita anteriormente, se muestra el resultado del AdA para diferentes posiciones de la fuente acústica. Además M0 y M1 se denotan como M1 y M2. Puede observarse que el AdA varía entre  $-\pi/2 \leq \theta \leq \pi/2$ , es decir abarca sólo la parte frontal del arreglo. No se hicieron provisiones para calcular ángulos mayores, ya que se supone que la cámara se encontrará sólo al frente del salón o de la mesa de trabajo.

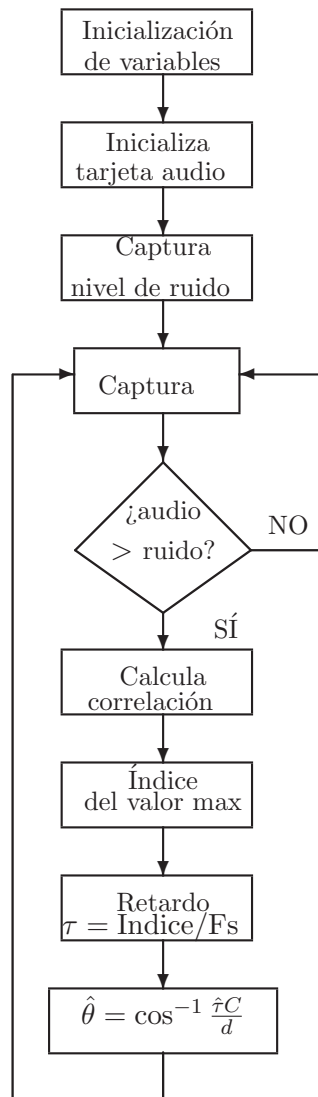


Figura 3. Diagrama de flujo del cálculo del AdA.

Con la configuración anterior se observó que el AdA presentaba cambios sólo cuando la fuente era desplazada un ángulo suficientemente grande. La razón obvia es que la resolución, en el cálculo del retardo  $\tau$ , depende en gran medida de la frecuencia de muestreo  $F_s$ . Así, si el número de muestras de la posición del pico es igual a uno, es decir tenemos el mínimo retardo, entonces  $\tau = 0.125$  ms, y el mínimo AdA que se puede detectar es  $\theta = 17.45$  grados. Esto se muestra en la Figura 5.

Empleando la función de interpolación (`interp`) de MATLAB, se logró mejorar la resolución al obtener un retardo mínimo de 0.0312 ms, y siendo el AdA de  $\theta = 4.30$  grados.

Finalmente se empleó una frecuencia de muestreo de 44.1 kHz, y se obtuvo un retardo de  $2.267 \times 10^{-5}$  s, con una resolución  $\theta = 3.1197$

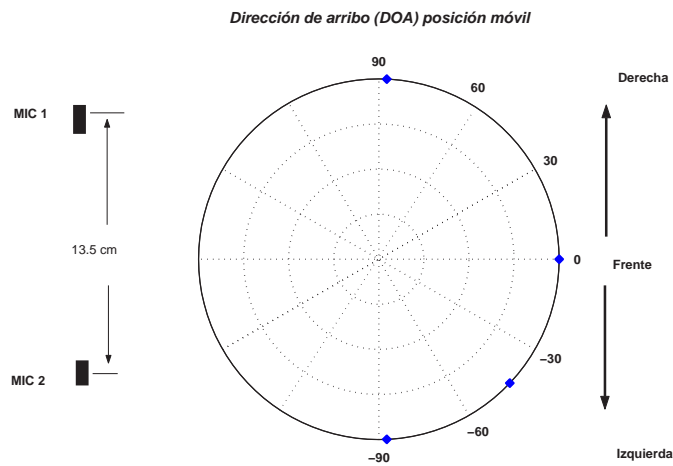


Figura 4. AdA de la fuente acústica en varias posiciones.

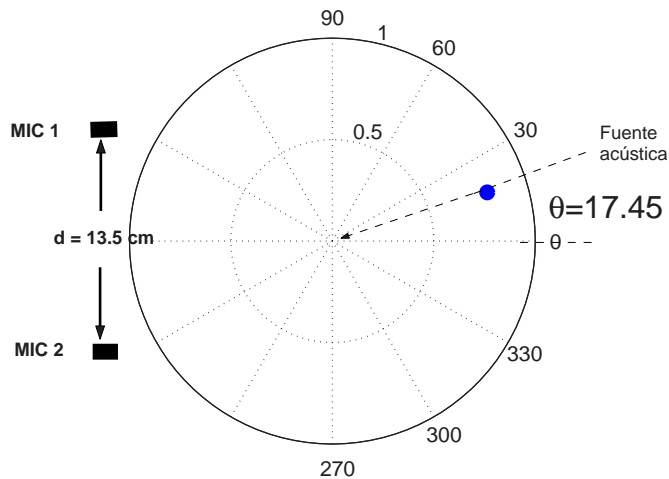


Figura 5. Resolución en el AdA para  $F_s$  de 8000 Hz.

#### IV. ADA Y CONTROL DEL MOTOR DE PASOS

Para aplicar el cálculo del AdA de una fuente acústica en una sala de video conferencia, se desarrolló una aplicación bajo la plataforma Windows empleando lenguaje C. Se construyó la interfase de control para el motor de pasos lográndose ubicar la posición de la cámara en el ángulo deseado. La aplicación recibe como parámetro el ángulo calculado por el algoritmo de estimación del AdA y envía por el puerto paralelo el número de pasos correspondientes al ángulo deseado. La conexión entre el programa en MATLAB y la aplicación de control de la posición de la cámara está en desarrollo. Además se desarrollaron bibliotecas de enlace dinámico (dynamic link library, dll) para acelerar el cálculo de la correlación y poder aplicar el algoritmo en tiempo real.

## V. RESULTADOS

Bajo el ambiente de MATLAB la mejor resolución resultó ser de aproximadamente 3 grados. Esto se debe principalmente a la frecuencia de muestreo. En nuestro caso se emplearon dos frecuencias de muestreo, 8 kHz y 44.1 kHz.

Una estimación del error en el cálculo del AdA se da en la Tabla I. Cabe señalar que la medición del ángulo de la posición de la fuente no es muy precisa, ya que fácilmente hay variaciones en la posición real de la fuente acústica respecto de la línea que describe el ángulo medido. Así pues, si el ángulo medido era de 30 grados, al colocarse frente a los micrófonos, ese ángulo fue en realidad aproximado. Por tal motivo, se puede considerar una incertidumbre de aproximadamente  $\pm 5$  grados en cada medición, ya que fácilmente puede haber una variación de 5 grados a cada lado del punto estimado. Como puede observarse, con ciertas limitaciones, el AdA estimado puede ser aceptable para la aplicación propuesta. Además, en el caso del sistema operando bajo el ambiente MATLAB, no se tiene el uso del nivel de la señal de entrada como disparo para iniciar la adquisición, así que cuando no se logra hacer la captura completa, el AdA encontrado está fuera de toda proporción y entrega un número complejo (ver la tercera columna de la Tabla I).

TABLA I  
RESULTADOS DE LA ESTIMACIÓN DEL ADA.

Iter. No.	AdA estimado 1	AdA estimado 2	AdA estimado 3	AdA estimado 4
1	-27.859	-44.504	4.466	44.504
2	-51.155	-33.037	0	51.155
3	-44.504	-44.504	4.466	4.5+1.97i
4	-38.541	-38.541	4.466	51.155
5	-44.504	-38.541	4.466	44.504
6	-44.504	-27.859	8.961	58.951
7	-44.504	-38.541	4.466	44.504
8	-51.155	-33.037	8.961	44.504
9	-51.155	-18.151	4.466	44.504
Promedio	-44.209	-35.19	4.96	47.97
AdA real	-50 $\pm$ 5	-40 $\pm$ 5	0 $\pm$ 5	50 $\pm$ 5

En la Figura 6 se muestra el AdA estimado para una sola posición de la fuente acústica, manteniéndola al frente (0 grados). Las variaciones se deben principalmente al error en el disparo de captura de audio, ya que se da inicio a la captura independientemente si hay o no señal acústica, si la captura alcanza sólo parte de esta señal, entonces la energía es mayor al nivel de ruido detectado inicialmente y con esa señal se calcula el AdA. Para este caso el valor promedio es cero con una varianza igual a 5.73 grados.

Finalmente, en la Figura 7, se comparan los valores promedios del AdA para diferentes ángulos de la fuente acústica empleando en cada caso 50 iteraciones y el AdA real usado en cada prueba.

Cabe señalar que los errores en la estimación del AdA, se supone, se deben principalmente



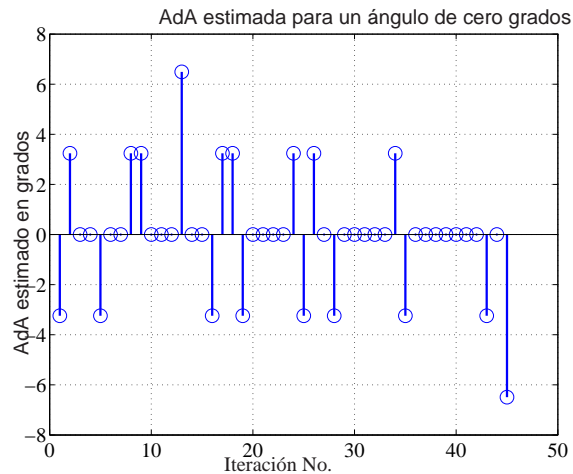
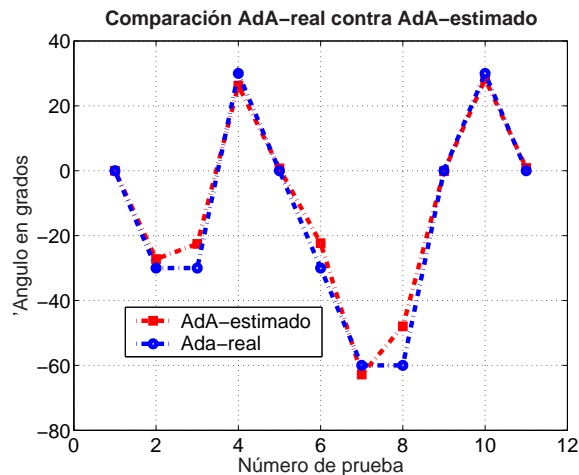
Figura 6. AdA para cero grados en  $n$  iteraciones.

Figura 7. Error de estimación del AdA

a la falta de sincronía en la captura de la señal acústica, y no al algoritmo de estimación, aunque se deberán hacer más pruebas para comprobarlo.

## VI. CONCLUSIONES Y TRABAJO A FUTURO

Se obtuvo el desarrollo de un sistema simple de estimación del AdA de una señal acústica a un arreglo de micrófonos. Se espera que el sistema sea útil para la orientación automática de una video cámara, para aplicaciones de video conferencia. El algoritmo de estimación del AdA está basado sólo en la estimación del tiempo de retardo. Aunque el algoritmo resulta ser de los más simples, los resultados obtenidos son aceptables para esta aplicación.

Algunos aspectos interesantes por investigar son: el cálculo de la posible velocidad de desplazamiento de un locutor activo en una sala de video-conferencia y su relación con la velocidad de procesamiento requerida para ofrecer, sin pérdida apreciable de información, un sistema en tiempo real, determinar la máxima resolución requerida para este tipo de aplicaciones.

Como trabajo por desarrollar se están haciendo pruebas preliminares en el desarrollo de una aplicación que incluye la captura, la estimación del AdA, y el control de la posición de la video cámara. Además se está trabajando en la implementación del algoritmo de estimación en un procesador de señales de Texas Instruments, el TMS320C6701.

### RECONOCIMIENTOS

Deseamos agradecer a la Compañía Texas Instruments, especialmente al Ing. Alex Porter y a la Ing. Verónica Vázquez por su constante apoyo al Cuerpo Académico de Procesamiento Digital de Señales.

### REFERENCIAS

- [1] Y. Huang, J. Benesty, and G. W. Elko, "Microphone Arrays for Video Camera Steering," *Acoust. Signal Processing for Telecommunications*, ed. S. L. Gay and J. Benesty, Kluwer Academic Publishing, 2000
- [2] S. U. Pillai, *Array Signal Processing*, Springer-Verlag, 1989
- [3] B. Widrow and S. Stearns, *Adaptive Signal Processing*, Prentice Hall Inc. 1985
- [4] S. Haykin, *Adaptive Filter Theory*, Third Edition, Prentice Hall, Inc., 1996.
- [5] J. H. DiBiase, H.F. Silverman, and M. Brandstein, "Robust Localization in Reverberant Rooms," *Microphone Arrays*, Springer Verlag, 2001
- [6] C. H. Knapp and G. C. Carter, "The Generalized Correlation Method for Estimation of Time delay," *IEEE Trans. Acoustics, Speech and Signal Proc.*, Vol. ASSP-24, No. 4, AUG, 1976.