

Capítulo 5

Pruebas y resultados

En este capítulo se presentan las pruebas realizadas para la generación y evaluación del sistema completo de generación de fuentes de sonidos virtuales en audífonos. Las pruebas se dividen en tres grandes secciones: Las pruebas de audio, las pruebas de video y las pruebas del sistema completo.

5.1 Pruebas de audio realizadas

La primera prueba de audio consiste en la medición de laboratorio de los coeficientes reales de atenuación y de fase de cada canal auditivo usando dos pequeños micrófonos montados en los oídos de un voluntario. El procedimiento y los resultados de este experimento se encuentran en el Apéndice A.

Dentro de las pruebas de audio realizadas en LabVIEW, se crearon tres programas que calculan aproximaciones de ITD, ILD y ATF usando métodos diferentes.

En el primer programa se utilizan tonos senoidales puros que son enviados a los audífonos en formato *stereo* a 44.1KHz y 16 bits. En la Figura 5.1 se muestra la pantalla principal de este programa.

El programa permite ingresar la distancia entre oídos, la temperatura (para determinar la velocidad del sonido), la frecuencia del tono a generar y la ganancia o intensidad del sonido virtual.

Los dos controles principales permiten colocar la fuente virtual de sonido a una distancia d y con un ángulo determinado θ .

El primer programa calcula el ITD y el ILD usando la velocidad del sonido y las distancias de la fuente a cada oído (Distancia R y Distancia L).

Sistema de Audio 1



Figura 5.1.- Panel frontal del programa de audio #1.

Las ecuaciones usadas para determinar la Distancia L y la Distancia R localizadas en el diagrama de la Figura 5.1 son las siguientes:

$$\begin{aligned} \text{Distancia } L &= \sqrt{(\sin \theta \cdot d)^2 + \left(\cos \theta \cdot d + \frac{O}{2}\right)^2} & d = \text{Distancia} & (5.1) \\ & & \text{del centro de la} & \\ & & \text{cabeza a la} & \\ & & \text{fuente de} & \\ \text{Distancia } R &= \sqrt{(\sin \theta \cdot d)^2 + \left(\cos \theta \cdot d - \frac{O}{2}\right)^2} & \theta = \text{Ángulo} & \\ & & O = \text{Distancia} & \\ & & \text{entre oídos} & \end{aligned}$$

Entonces el cálculo de la ITD para el programa #1 se realiza como sigue:

$$ITD = \frac{c}{\text{Distancia } L - \text{Distancia } R}, \quad (5.2)$$

donde c = Velocidad del sonido

Para realizar los cálculos de la ILD se usa la ley inverso cuadrada que se presenta en modos de propagación esféricos sin pérdidas como sigue:

$$ILD = \frac{1}{(\text{Distancia } R)^2} - \frac{1}{(\text{Distancia } L)^2}, \quad (5.3)$$

En el segundo programa, se generan los mismos tonos senoidales puros que en el programa #1 a 44.1KHz y 16 bits. Los datos de entrada permanecen constantes y el control de distancia desaparece debido a que el segundo modelo para el cálculo de ITD asume que la distancia d es mucho mayor al radio de la cabeza r , por lo que se puede descartar.

Para medir la ITD en el programa #2, se utiliza un modelo matemático en donde la cabeza es esférica de radio r y la fuente de sonido esta a una distancia $d \gg r$, por lo tanto se asume que las líneas de audio que inciden en la cabeza son paralelas.

En la Figura 5.2, se muestra el la pantalla principal del programa #2 con su diagrama correspondiente.



Figura 5.2.- Panel frontal del programa de audio #2.

Debido a que la distancia d no será usada, sólo tenemos el control del ángulo θ que en esta ocasión se mide de -90° a 90° .

La medición del ITD se hace en referencia al oído derecho R . La distancia extra que recorre el sonido para el oído izquierdo L , está dada por $a+b$, como se indica en el diagrama de la Figura 5.2. La distancia a es la parte curva del perímetro y b una distancia recta. Sus cálculos son los siguientes:

$$a = \left(\frac{\theta}{360}\right) \cdot 2\pi r \quad , \quad b = r \cdot \sin \theta \quad , \quad ITD = \frac{c}{a + b} \quad (5.4)$$

$c = \text{Velocidad del sonido}$

Para obtener la ILD en el programa #2, y debido a que no se cuenta con la distancia d , se realiza de la siguiente manera:

$$ILD = \frac{1}{(1)^2} - \frac{1}{(1+(a+b))^2} \quad (5.5)$$

El tercer programa de sonido realizado permite tocar archivos de audio (*wav*) con el fin de reproducir sonidos realistas.

El programa #3 usa el modelo de cálculo de la ITD ocupado en el segundo programa. Esto es debido al buen desempeño demostrado en las comparaciones con los valores medidos en el segundo experimento del Apéndice A.

Por otro lado, para el cálculo de la ILD en el tercer programa, se usa la tabla de valores obtenida experimentalmente en el mismo experimento del Apéndice A. Esta tabla de valores contiene información de los niveles percibidos por cada oído en diferentes ángulos como muestra el diagrama de la Figura 5.3.

Sistema de Audio 3

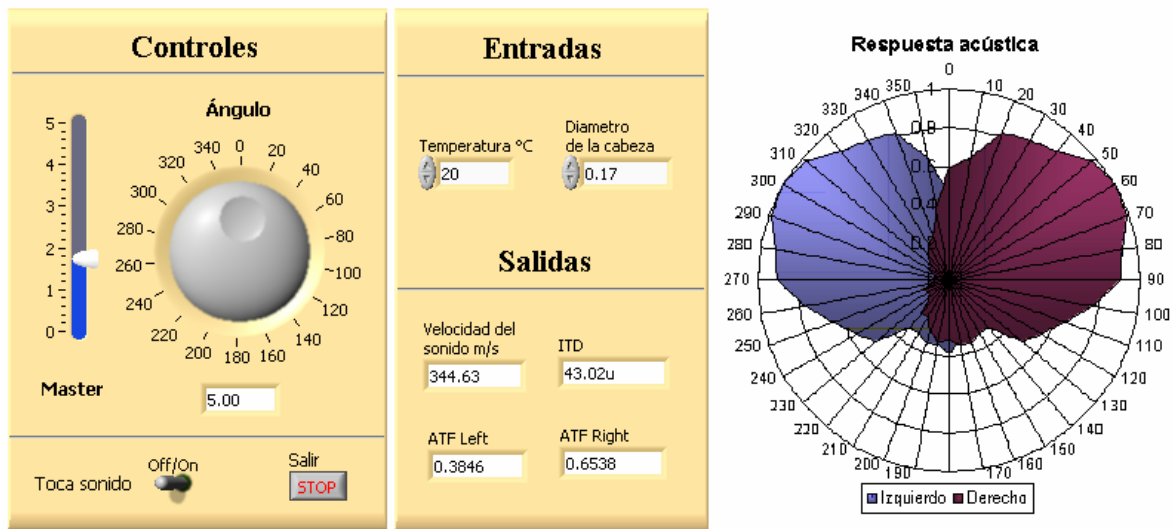


Figura 5.3.- Panel frontal del programa de audio #3.

Es importante aclarar que dentro de los valores de esta tabla están inmersos la forma de la cabeza, las orejas y los rebotes producidos en el cuarto de pruebas. Con lo anterior podemos decir que el programa #3 proporciona al sonido una especie de ATF simplificada.

5.1.1 Evaluación de los programas de sonido

El primer programa obtiene datos aproximados de ITD y de ILD. Esto es debido a que el sistema no considera a la cabeza como un sólido y las líneas de distancia la atraviesan como si ésta no existiera. Del mismo modo, las atenuaciones provocadas por la obstrucción de la cabeza no son consideradas, haciendo que la sensación de realismo sea muy confusa y pobre.

Los sonidos generados en el programa #2, producen sensaciones más acercadas la realidad. En las mediciones prácticas de las ITD medidas en una cabeza de 17cm de diámetro mostradas en la segunda prueba del Apéndice A, se puede observar que la ITD para 90° es de $660\mu\text{s}$; El segundo programa presenta $634\mu\text{s}$, mientras que el programa #1 presenta $435\mu\text{s}$ para este mismo ángulo.

En la Figura 5.4, se presenta una gráfica comparativa de las ITD's medidas en el Apéndice A con respecto a las calculadas usando el modelo del programa #2. Se puede identificar que es una muy buena aproximación a las mediciones reales. Las variaciones máximas presentadas entre las dos curvas son de aproximadamente $60\mu\text{s}$, lo que hace que las variaciones sean despreciables.

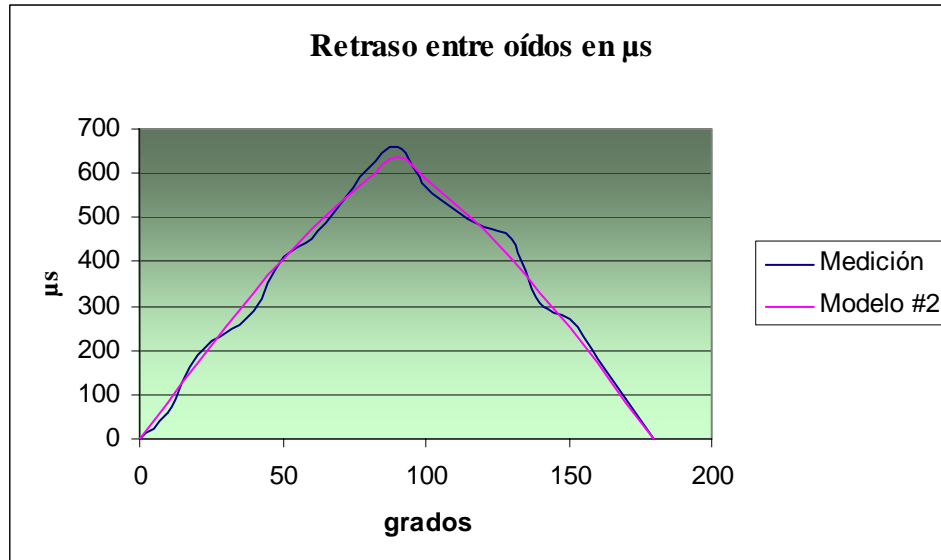


Figura 5.4.- Gráfica comparativa de ITD's.

El desempeño del tercer programa es considerablemente mejor que los dos anteriores. Esto es debido a que al aplicar la tabla de ATF medida en el laboratorio, se puede distinguir con facilidad si el sonido se encuentra enfrente o detrás del espectador, creando una sensación de realismo sorprendente.

Se dice que la ATF usada en el tercer programa es una ATF simplificada por que las funciones de transferencia son dependientes de la frecuencia, y en este caso fueron medidas utilizando señales pulsadas. Esto hace que no se tenga la información de la ATF real para cada frecuencia, si no de una ATF sintetizada o simplificada.

Debido a su notable desempeño, el programa #3 de audio será utilizado para la elaboración del sistema completo de generación de fuentes virtuales de sonidos en audífonos.

5.2 Pruebas de imagen realizadas

La primera prueba realizada en el área de imagen y video consiste en determinar las características de la *webcam* a usar. Esta cámara de marca *CobraCam*, tiene una apertura medida en el eje *z* de 33° y en el eje *y* de 28° como indica la Figura 5.5.

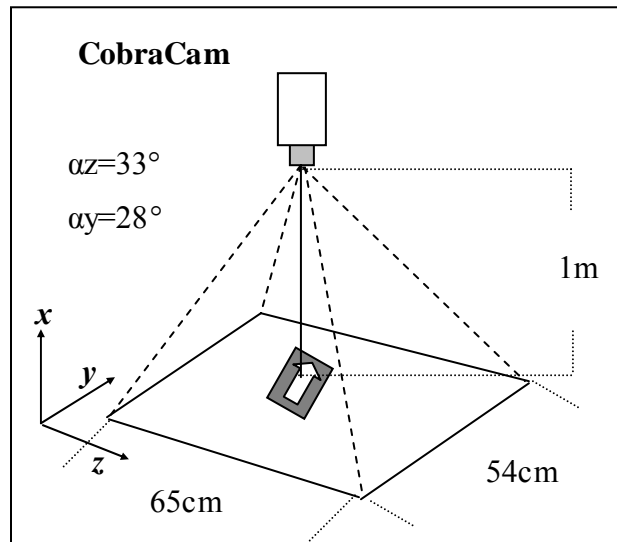


Figura 5.5.- Apertura del lente de la *CobraCam*.

La cámara *CobraCam* cuenta con una interfase USB y tiene una resolución máxima de 640x480 píxeles.

La primera prueba realizada en el área de imagen fue un programa en LabVIEW que capturara imágenes estáticas y de video de la cámara *web CobreCam*.

Para lograr esta tarea, se utilizó una librería externa de LabVIEW creada por Peter Parente obtenida de su página de Internet: <http://www.cs.unc.edu/~parente/labview/webcam1.4.zip>

La resolución usada en este caso fue de 352x288 píxeles, debido a que ésta era la resolución predeterminada de la *CobraCam*.

El diagrama del programa de captura de video se indica en la Figura 5.6.

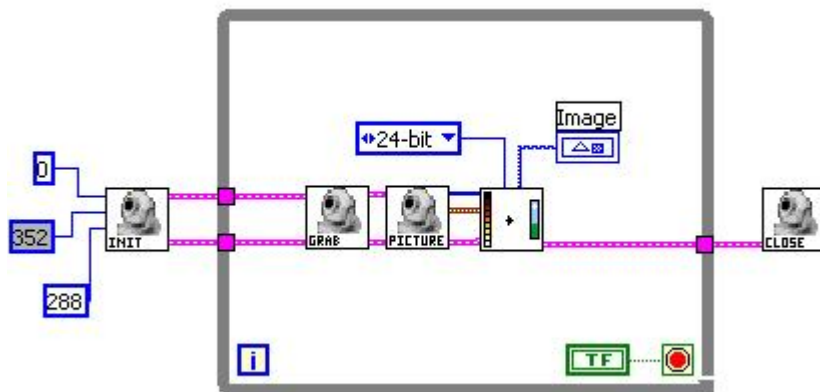


Figura 5.6.- Programa de captura de video continuo.

Todos los *SubVI's* que integran el diagrama de la Figura 5.6 pertenecen a la librería de Peter Parente. En el primero de izquierda a derecha se inicializa el dispositivo USB especificando el puerto de control y la resolución de adquisición.

El segundo bloque inicia una sesión de captura en el dispositivo, mientras que el siguiente captura una imagen única. El cuarto elemento se encarga de convertir el formato de imagen plana RGB a formato de imagen gráfica de 24bits. El último bloque cierra la sesión de captura y libera el puerto de control.

La velocidad del programa de captura de video es un poco lenta. Aproximadamente se cuenta con 8 cuadros por segundo, provocando que la imagen se vea cortada en lugar de fluida. Esto definitivamente afectará a la velocidad de retroalimentación del sonido enviado a los audífonos degradando la calidad del sistema completo.

Se realizó un segundo programa en LabVIEW con el fin de reconocer patrones en imágenes en una ubicación fija en el disco duro de la PC de control.

Los formatos de imágenes que se utilizan en este programa son del tipo JPEG.

El programa presenta como salida el ángulo y la posición (x, y) en píxeles en base al sistema de coordenadas visto en la Figura 4.3 (a). El panel frontal del programa se muestra en la Figura 5.7.

El programa también cuenta con una entrada de calibración, que se utiliza para ingresar un escalamiento lineal con el fin de convertir las coordenadas en píxeles en coordenadas reales en centímetros.



Figura 5.7.- Programa de reconocimiento de imagen.

En la parte inferior del panel de la izquierda, se ingresa la ruta del archivo de la imagen a analizar. El objeto a reconocer, y las salidas de la posición relativa se muestran en el panel de la derecha.

Las herramientas usadas para el análisis de la imagen, son las que presenta la librería de visión IMAQ *Vision* de LabVIEW 7.1.

El diagrama del programa de reconocimiento de imágenes se muestra en la Figura 5.8.

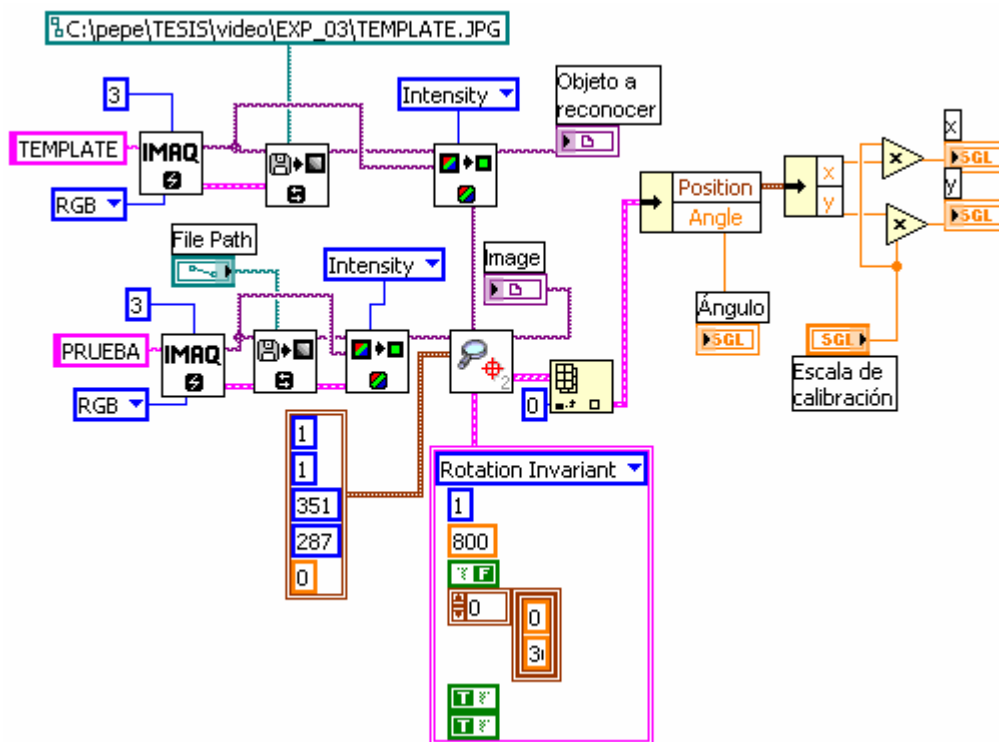


Figura 5.8.- Diagrama del programa de reconocimiento de imagen.

El desempeño del programa es sorprendente, las imágenes son reconocidas en cualquier posición y ángulo dentro de la imagen. El único problema es que en este caso el programa puede presentar un error de 180° en la medición del ángulo debido a la simetría en un eje que presenta el objeto a reconocer. Este problema se puede solucionar definiendo otra marca a reconocer que sea totalmente asimétrica.

5.3 Evaluación del sistema completo

Una vez que se tienen todos los bloques necesarios para el funcionamiento del sistema de generación de sonidos virtuales en audífonos, sólo resta ensamblarlos en conjunto y analizar el desempeño del sistema completo.

Los bloques que conforman el sistema completo son; El programa #3 de sonido, el programa de captura de video de la *webcam* y el programa de reconocimiento de imágenes como se muestra en la Figura 5.9.

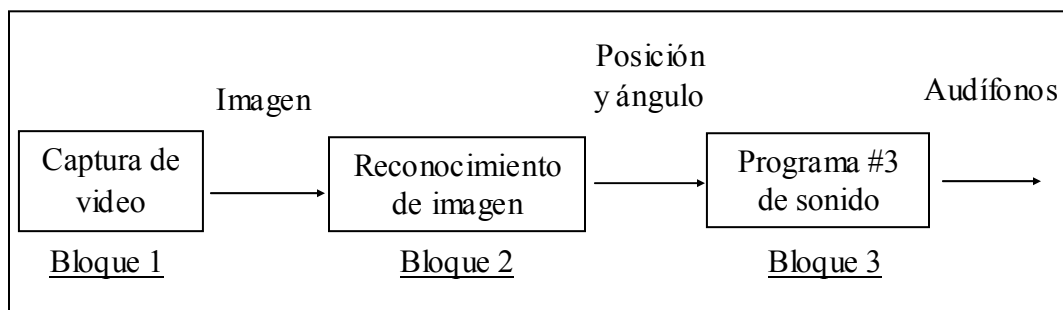


Figura 5.9.- Diagrama a bloques del sistema completo.

Los tres bloques fueron ensamblados en un programa principal en LabVIEW. El sistema de sonido fue encapsulado en un *SubVI* y se le asignó un icono que presenta la figura de una pantalla como se observa en el extremo derecho del diagrama del sistema completo de la Figura 5.10.

Cada uno de los tres bloques es diferenciado por su etiqueta correspondiente. El de captura de video se encuentra en la parte izquierda del diagrama, mientras que el de identificación se encuentra en el centro y el de sonido a la derecha.

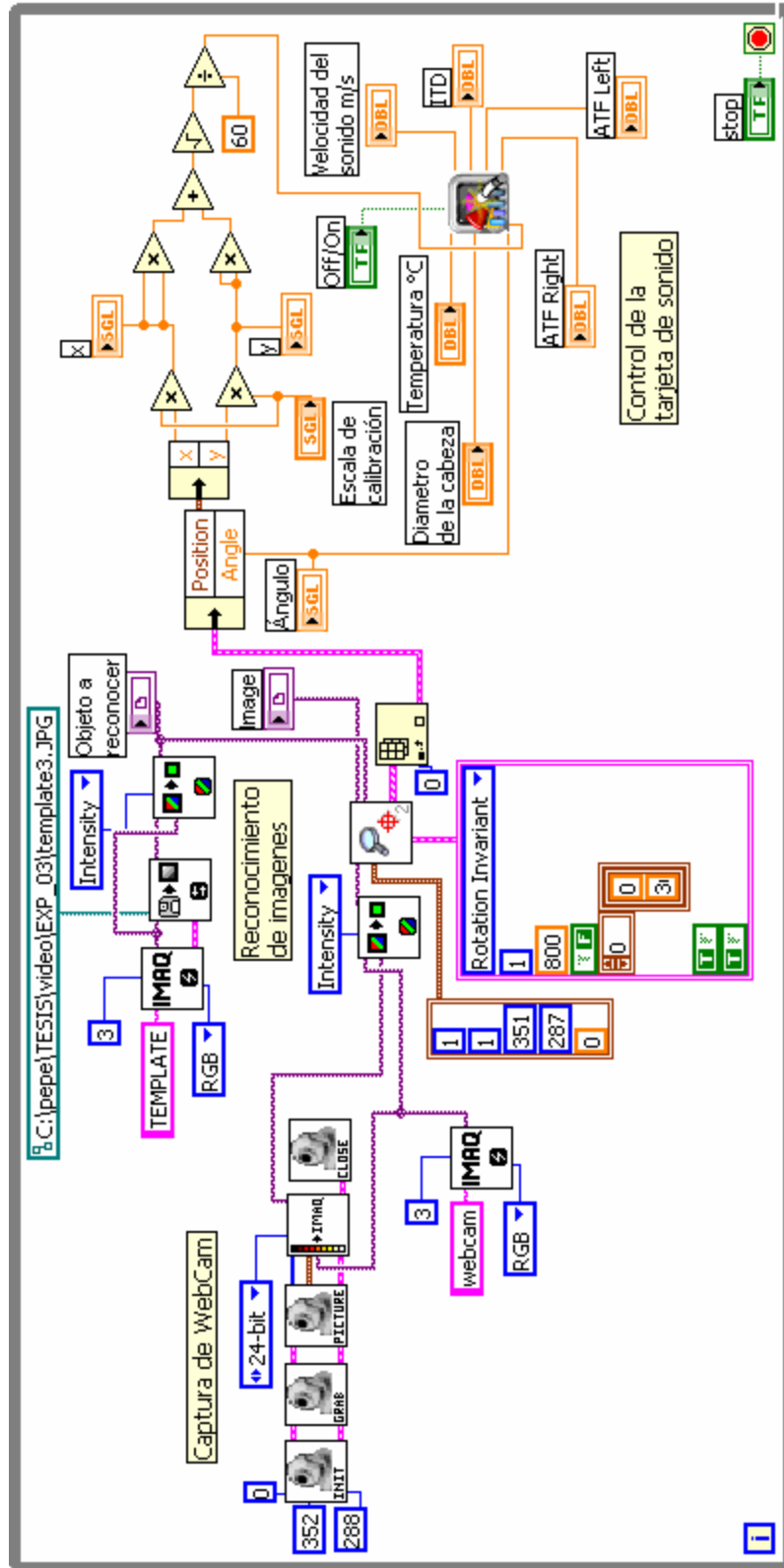


Figura 5.10.- Diagrama del sistema completo.

El panel frontal del programa completo se ilustra en la Figura 5.11. En este panel se observa la imagen capturada de la *webcam* y sus propiedades en la sección de la izquierda. En la sección del centro se muestra el patrón u objeto a reconocer, las salidas de posición, ángulo y la escala de calibración. En la sección de la derecha se encuentra el área de sonido. En esta área se puede modificar la temperatura y el diámetro de la cabeza. También en esta sección se despliega la velocidad del sonido calculada, la diferencia de fase o ITD, y los coeficientes de ATF para cada oído obtenidos en la interpolación de las mediciones de ATF realizadas en el Apéndice A.

El botón de stop detiene la captura de imágenes mientras que el *switch* ON/OFF enciende o apaga el sonido.



Figura 5.11.- Panel frontal del programa principal.

5.3.1 Desempeño y evaluación de percepción de localización

El desempeño de cada bloque independiente ha sido evaluado en su sección correspondiente.

Para la evaluación del desempeño del sistema completo, es necesario visualizarlo como un programa secuencial. Cada bloque se ejecuta a la vez, por lo que la velocidad total se reduce considerablemente. El tiempo que se demora el sistema completo en realizar una iteración, es la suma del tiempo de demora de cada bloque más el tiempo de transición entre bloques.

El procesado digital de imágenes usado en el reconocimiento de patrones complejos, requiere de procesadores de alta velocidad y de gran memoria para su buen desempeño. Por otro lado el procesado digital de audio requiere mucho menos velocidad en comparación con el caso anterior. Aún así, el procesar video y audio en tiempo real, es una tarea que demanda un gran nivel de procesamiento.

Todas las pruebas realizadas se hicieron en una *LapTop* COMPAQ Presario 700, con un procesador AMD Duron a 489MHz y 240MB de RAM.

Con esta velocidad de procesamiento no se alcanzó el desempeño deseado. El sistema tarda aproximadamente dos segundos en realizar una iteración. Entiéndase una iteración como la siguiente secuencia; Capturar una imagen de la cámara *web*, procesarla, identificar el patrón complejo, calcular posición y ángulo, calcular velocidad del sonido, calcular ITD, interpolar el ATF, leer el archivo de audio, modificar el archivo de audio y mandarlo a la tarjeta de sonido.

Es evidente que con una velocidad de medio cuadro por segundo el sistema no es capaz de brindar fluidez o continuidad en la percepción de realidad.

La velocidad deseada para que el sistema funcionara de manera correcta es de unas tres o cuatro iteraciones por segundo.

Un factor importante en donde se perdió gran parte de la velocidad del sistema completo, es el tipo de cámara y el controlador usados. LabVIEW cuenta con una cámara especial con interfase FIRE-WIRE que permite de manera más eficiente realizar aplicaciones de video y sistemas de visión inteligente en tiempo real.

En esta tesis se decidió no usar esta cámara especial debido a su alto costo. Se escogió una cámara *web* con el fin de demostrar la factibilidad de construir un sistema de audio virtual y se cumplió con el objetivo a un menor costo. Es importante aclarar que para que tener la velocidad deseada en el sistema completo, sólo basta con tener un procesador más grande y una cámara de captura de video digital apropiada como la de *Nacional Instruments*.

La evaluación de percepción de localización es un tema un tanto subjetivo, esto es debido a que intervienen varios factores como el nivel de audición del espectador y la memoria auditiva entre otros.

Aún así, la sensación provocada por el programa #3 de audio es sorprendente. La imagen tridimensional que se crea supera por mucho a lo alcanzado con una señal *stereo* convencional. Los sonidos presentan una pista extra que permite engañar al cerebro y crear ambientes virtuales.