Ingeniería Industrial
ISSN: 1025-9929
fondo_ed@ulima.edu.pe
Universidad de Lima
Perú

Arbildo López, Aurelio; Bigio, José
Codificación de imágenes en sonido como ayuda al invidente
Ingeniería Industrial, núm. 31, enero-diciembre, 2013, pp. 239-265
Universidad de Lima
Lima, Perú

Codificación de imágenes en sonido como ayuda al invidente*

Aurelio Arbildo López, José Bigio
Universidad de Lima. Lima, Perú
Correos electrónicos: aarbildo@ulima.edu.pe, sir.owns.a.lot@gmail.com

Recibido: 6/2/2013 / Aprobado: 20/6/2013

Resumen: Mediante el uso de la tarjeta de sonido y la cámara web del computador personal se ha investigado la codificación de imágenes como patrones de sonido para permitir que las personas invidentes puedan interpretar la información de su entorno y desplazarse con mayor facilidad en esta. En todas las propuestas se ha codificado el tono de gris como volumen y las posiciones de los pixeles como frecuencias. Uno de los desarrollos denominado barrido manual equivale a «ver con el dedo» y permite que el invidente navegue sobre una pantalla sensible al tacto, auscultando la imagen para identificar su estructura.

Palabras clave: Codificación de imágenes / ayudas al invidente / discapacitados

Sound coding of images as an aid to the blind

Abstract: Using the standard computer sound card as well as the web cam, it has been explored the coding of images as sound patterns to allow blind people to interpret information from their surroundings in order to help them to walk around. In all of the cases, the gray level has been coded as a volume level and the (x,y) position of the pixel as frequency values. One of the modes developed, named manual scan, is equivalent to «finger seeing». It gets the image from the user's web cam and permits to finger surf on a touch screen in order to identify the image structure.

Keywords: Image coding / aids for the blind / disabled

* Este trabajo forma parte de la investigación «Codificación de imágenes en sonido como ayuda al invidente», desarrollada con el patrocinio del Instituto de Investigación Científica de la Universidad de Lima. Debemos también reconocer la colaboración del estudiante Álvaro Gloria, quien participó en la etapa inicial del proyecto.

1. INTRODUCCIÓN

El más importante de los sentidos es, sin lugar a dudas, el de la visión, ya que permite tener información completa no solo del entorno en el que nos desenvolvemos sino también información adicional como distancias y velocidades, entre otras.

La pérdida de la visión constituye una desventaja muy grande para quienes la sufren. La mayoría de las ayudas que los ciegos tienen para suplir la carencia de visión son muy rudimentarias, pero terminan siendo muy útiles gracias a que, a través de los otros sentidos, se suple en cierta medida su carencia. Los sentidos que más usan son el tacto y el oído.

En los últimos años han sido muy publicitados los casos denominados «ecolocalización», que captando el rebote de las ondas sonoras, en la mayoría de casos emitida por la persona, puede reconstruir a través de un proceso mental bastante complejo, el espacio físico alrededor suyo. Al existir varios casos de personas capaces de aprender la «ecolocalización», está claro que el oído puede ser utilizado para entrenar el cerebro con el fin de visualizar el entorno. Sobre esta base se sustenta la hipótesis de que es más fácil, para el cerebro humano, interpretar una imagen codificada en sonido si recibe un patrón muy preciso de la imagen que si recibe un patrón muy complejo de reflexión de ondas sonoras.

Como la carencia de visión es una discapacidad que afecta a muchas personas en el mundo, todos los aspectos relacionados con esta han sido estudiados por varios autores. La compilación de Blash, Wiener y Welsh (1997) trata esta discapacidad así como las numerosas ayudas disponibles para los invidentes. Puede verse también información relacionada en otras publicaciones como Wilson et. al. (2009) y en Zeki (2005).

Las investigaciones incluyen técnicas no invasivas e invasivas. Dependiendo de la localización del daño, se está experimentando con implantes de cámaras o electrodos para transmitir imágenes a través del nervio óptico, como en el caso presentado por Caspi et al. (2009), que han explorado la factibilidad de prótesis retinales en humanos, y Szurman et al. (2005), quienes han experimentado y hecho pruebas de larga duración con implantes oculares en conejos.

Existen casos de personas que han perdido la visión y que han utilizado la ecolocalización, como en el caso de Ben Underwood, un joven

que perdió la visión a muy temprana edad; podía movilizarse con soltura y sobre quien existe amplia información en Internet (www.reportajes.org), páginas con videos como http://www.zappinternet.com, así como una página administrada por la fundación que lleva su nombre. Hay otros casos, como el de Tom De Witte, en Bélgica, y Daniel Kish, quien no solo aplica la ecolocalización sino que la enseña (http://www.worldaccessfortheblind.org).

La idea de codificar imágenes en sonido no es nueva. Hace varios años que se vienen ensayando diversas metodologías. Muchas de ellas han estado limitadas en el pasado por la complejidad, capacidad de almacenamiento, velocidad y portabilidad. Cronly-Dillon, Persaud y Blore (2000) han utilizado la codificación musical de escenas, y Arno y colaboradores (2001) han estudiado el reconocimiento de patrones usando el oído. Existen varias patentes registradas y otras solicitadas sobre metodologías y equipos que codifican imágenes como sonido, como las de Nakanishi (2008), Okada y Taketa (2008) y Tsuji, Suzuki y Toyama (2008).

Entre los diversos grupos que están trabajando actualmente en el tema, quien aparentemente ha tenido un éxito sostenido es el grupo liderado por Peter Bartus Leonard Meijer (1992), que ha desarrollado un sistema computarizado para codificar las imágenes y ha logrado resultados de aprendizaje con algunos pacientes desde la publicación de su artículo más representativo.

Una gran parte de la información disponible no detalla la forma precisa como se codifica la información ni los detalles constructivos de los dispositivos que están utilizando; por ello se amerita el presente trabajo, toda vez que se exploran metodologías alternativas a las existentes mediante el uso de computadores personales para migrar al *hardware* comercial de última generación que promete desempeños acordes con las necesidades de la visión por codificación de imágenes en sonido.

2. *LUCES Y SONIDO*

El sentido de la vista está constituido por un conjunto de órganos capaces de captar información de la reflexión de la luz sobre los objetos para interpretar el conjunto completo de estos en el entorno, así como las relaciones e interacciones entre ellos y el observador. El ojo, en su parte frontal, tiene un sistema de lentes que focaliza la imagen en la retina,

está compuesta de conos y bastones para captar la luz de cada punto de la escena usando más de 5 millones de células en tres rangos de frecuencia o colores y unos 100 millones de células sensibles a la intensidad. La información de tales células son acondicionadas y transmitidas en paralelo a través del nervio óptico hacia las regiones de la corteza cerebral, donde acoplan la información de la señal a las redes neuronales que analizan las imágenes y las asocian a conceptos complejos de color, forma y movimiento. Cada uno de los sentidos capta información complementaria del entorno y nos permiten actuar dentro de este. El sentido del oído está orientado a identificar sonidos del entorno como complemento de la vista; en el caso de los humanos ha evolucionado como soporte al lenguaje. La pérdida de la audición afecta sobre todo la relación con los congéneres, a diferencia de la visión, donde la mayor pérdida es la relación con el entorno. En el caso del sentido del oído, la interpretación especializada del cerebro está orientada a diferenciar las diversas frecuencias del sonido y su estructura temporal en lo que podríamos llamar un analizador simultáneo de frecuencia/tiempo.

3. LA INVIDENCIA

La pérdida de la visión, ya sea antes o después del nacimiento, origina la mayor de las pérdidas de relación con el entorno y constituye una de las discapacidades más limitantes en relación con la pérdida de alguno de los sentidos.

3.1 Ayudas para invidentes

La mayoría de ayudas para invidentes pueden ser interpretadas como extensiones de los otros sentidos para captar información del entorno que normalmente sería captado con la vista. Las ayudas más útiles son aquellas que contribuyen a que el invidente pueda desplazarse; entre las cuales destaca el bastón, que a través del tacto permite captar información del entorno cercano (dentro del alcance físico del bastón). El bastón puede considerarse un instrumento de medición que hace posible identificar obstáculos y características morfológicas de los objetos que ausculta; a través de las vibraciones que transmite, envía información adicional del entorno sobre el que se desplaza, como podría ser la textura y, eventualmente, información que proviene de una distancia mayor.

El modelo del bastón nos permite generalizar la naturaleza de la mayoría de ayudas para invidentes, que son instrumentos diseñados para transmitir, a través de cualquiera de los sentidos, la información del entorno. Cualquier instrumento que entregue información del entorno a través de cualquiera de los sentidos a un invidente, será una ayuda para este. Los otros sentidos son entrenados para incorporar información adicional que en los no invidentes carece de importancia, pues la vista los proporciona con exactitud y precisión; así, por ejemplo, la dirección de la brisa sobre el rostro brinda información al invidente sobre los espacios libres durante una caminata a campo abierto.

Existe una gran cantidad de implementos de ayuda diseñados para invidentes, incluyendo muchas que utilizan dispositivos electrónicos de menor o mayor sofisticación.

El avance en los sintetizadores de voz desarrollados para computadores y para dispositivos portátiles de uso masivo ha hecho que aparezcan ayudas de lectura que prescindan del código braille al pasar de un texto codificado en braille a un texto sintetizado como voz. En este grupo se encuentran muchas aplicaciones que permiten a los invidentes navegar en internet con bastante solvencia.

4. *PLASTICIDAD*

El cerebro puede ser analizado desde dos aspectos muy diferenciados que podemos asociar con el *hardware* y el *software*. El *hardware* está constituido por el arreglo de neuronas que a su vez podemos separar en un conjunto dedicado a la captación de información, un conjunto de neuronas dedicado a la generación de señales para acciones externas (locomoción, generación de voz, etcétera) y un conjunto dedicado a la generación de pensamiento y relaciones entre las dos primeras. Las regiones del cerebro dedicadas a los sentidos son más o menos extensas en función de la complejidad de información que captan y analizan. Así, el sentido de la vista es el que monopoliza la mayor parte de la capacidad del cerebro. Contiguo a las zonas del cerebro que reciben información de los sentidos se encuentran las zonas dedicadas al procesamiento primario de dicha información, así como las regiones que realizan un procesamiento de información más complejo.

En el caso de la invidencia implica que una parte importante del cerebro prevista para realizar las funciones asociadas a la visión quedan

disponibles para procesar información asociada a otros sentidos o susceptibles de conectarse con otras subredes que procesan información proveniente de otros sentidos. La capacidad del cerebro de adaptar sus funciones para suplir la deficiencia de un sentido a través de otro sentido está dentro de lo que se conoce como plasticidad.

El tema de plasticidad del cerebro ha sido estudiado extensivamente por varios autores, como Neville et al. (2008), Shouval (2011) y Maurice et al. (2005). Sustentado en la plasticidad se han desarrollado varias aplicaciones para captar información de imágenes a través de los otros sentidos.

4.1 *The vOICe*

La investigación de más trascendencia es probablemente la iniciada por Peter Meijer (1992), quien por muchos años se dedicó a la codificación de imágenes como patrones de sonido, utilizando en sus primeros diseños circuitos electrónicos que han evolucionado hasta una aplicación de computador que está disponible en internet con el nombre de The vOICe (OIC en inglés suena como Oh I See). Las metodologías de barrido que describiremos más adelante son muy similares a las que sustentan la aplicación denominada The vOICe, con la diferencia de las herramientas utilizadas y alternativas propuestas como la del barrido aleatorio y la del patrón complejo multiparlante.

5. *METODOLOGÍAS ENSAYADAS*

Se han ensayado varias alternativas (véanse los incisos 5.4, 5.5 y 5.6) de codificación de imágenes basadas en criterios de plasticidad cerebral, tomando en cuenta las asociaciones que por estar basadas en conceptos simples deberían ser asimiladas con mayor facilidad por usuarios invidentes. Los conceptos implícitos en todas las codificaciones están asociados a los conceptos arriba-abajo y derecha-izquierda. Toda vez que para invidentes de nacimiento el concepto de claroscuro o el concepto mismo de luz que va asociado con las imágenes son ajenos a ellos, estos tendrán que ser asimilados durante el entrenamiento.

Tomando en cuenta que el sentido de la vista procesa mucho más información que los otros sentidos, y que por otro lado el oído tendrá que compartir información tanto de voz como de imagen, se ha tratado

de simplificar la complejidad de las imágenes y la codificación considerando los siguientes aspectos:

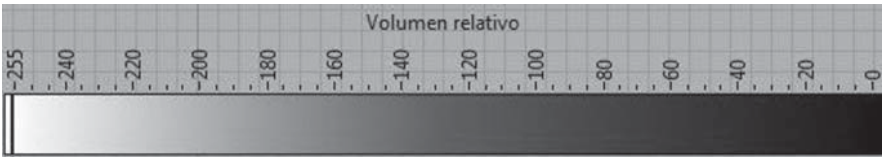
a. La información de color ha sido ignorada de tal manera que se utilizan solo imágenes en tonos de gris para la codificación final.
b. La resolución de tonos ha sido trabajada en codificación de 8 bits en la que, salvo con fines de ensayo, se ha optimizado el uso de los códigos disponibles mediante un procesamiento de imágenes ecualizando las imágenes.
c. Se ha trabajado en rangos de frecuencia flexibles de manera de adaptar dichos rangos para interferir solo parcialmente en rango de frecuencias típicamente usados por la voz humana.

5.1 *La codificación de los tonos de gris*

Los tonos de gris han sido codificados en función del volumen, así la ausencia de sonido implica oscuridad total mientras que el incremento del volumen significa tonos cada vez más claros. En todos los casos, esta codificación es relativa ya que la codificación se realiza modulando la amplitud de las ondas sonoras al margen del control del volumen hacia los parlantes que es controlado por una vía independiente. De esta forma los volúmenes relativos representan tonos de gris dentro de un rango de volúmenes controlable por el usuario. La ventaja de esta metodología radica en que el propio usuario ajustará el volumen en función de sus necesidades y de la sensibilidad que ha desarrollado por las necesidades obligadas por la invidencia.

Desde que las imágenes en tonos de gris se representan en matrices XY mediante valores numéricos, donde la ausencia de color (negro) se representa por el número 0 (cero) y el blanco por el mayor valor del rango (dependiendo de la resolución en la representación), dicha representación se traduce directamente en volumen relativo en nuestra codificación, según se muestra en el gráfico 1. Esta es parte de la codificación de un punto de imagen que solo codifica el tono de gris para un punto cualquiera de la imagen; para identificar la ubicación de dicho tono de gris se emiten sonidos de diversas frecuencias, de tal manera que el tono de gris se codifica en volumen y la frecuencia codifica la posición, como se verá en el inciso 5.3.

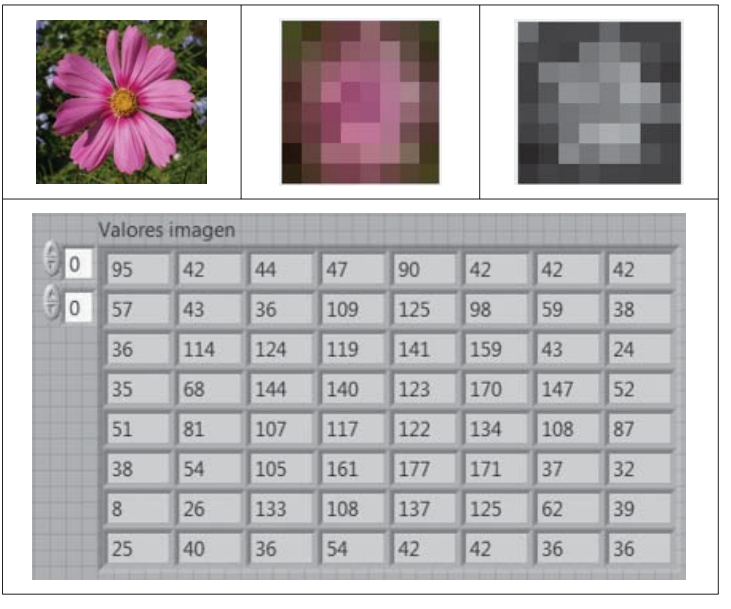
Gráfico 1
Relación del tono de gris con el volumen relativo



Elaboración propia.

En el gráfico 2 se muestra el procedimiento seguido para generar los valores de los tonos de gris que corresponden a una imagen a color. En la primera fila se presenta una imagen de buena resolución y su equivalente a color de muy baja resolución (8x8x3 bytes), así como su representación solo en tonos de gris de muy baja resolución (8x8 bytes), mientras que en la segunda columna se muestra la representación matricial de dicha imagen. Según esta codificación, dicha matriz también puede ser interpretada como la matriz de volúmenes relativos.

Gráfico 2
Codificación a 8x8 bytes de imagen



Elaboración propia.

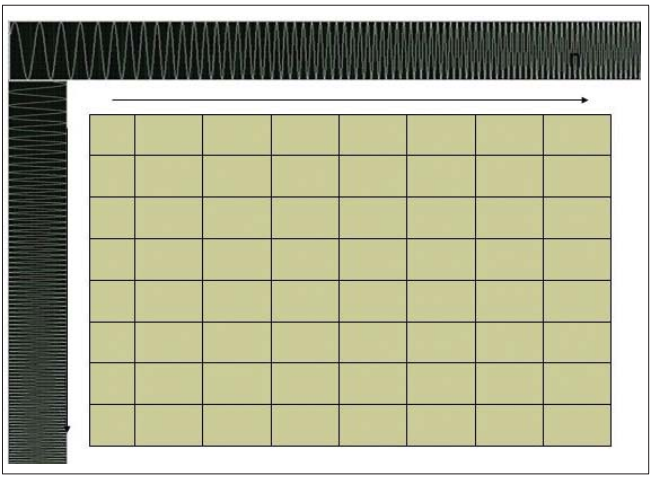
5.2 Codificación de la posición

La posición se codifica en todos los casos por variación de la frecuencia en una variedad de formas dependiendo de los criterios que se han seleccionado para cada una de las alternativas de barrido ensayadas, como son el barrido derecha-izquierda, el barrido manual de una imagen y el barrido aleatorio de una imagen.

5.3 Codificación de un punto

La forma general de codificación de un punto de imagen es asignando una frecuencia al eje X y otra al eje Y, de tal manera que cada punto queda identificado por un par de coordenadas (fx, fy). En el caso de que se usen dos rangos de frecuencia diferenciados (un rango para X y otro para Y) cada punto queda identificado sin ambigüedades.

Gráfico 3
Codificación de posición (x, y) como frecuencia



Elaboración propia.

A este punto, la codificación completa de la imagen como patrón de sonido implica la equivalencia (x, y, tono de gris) a (fx, fy, volumen).

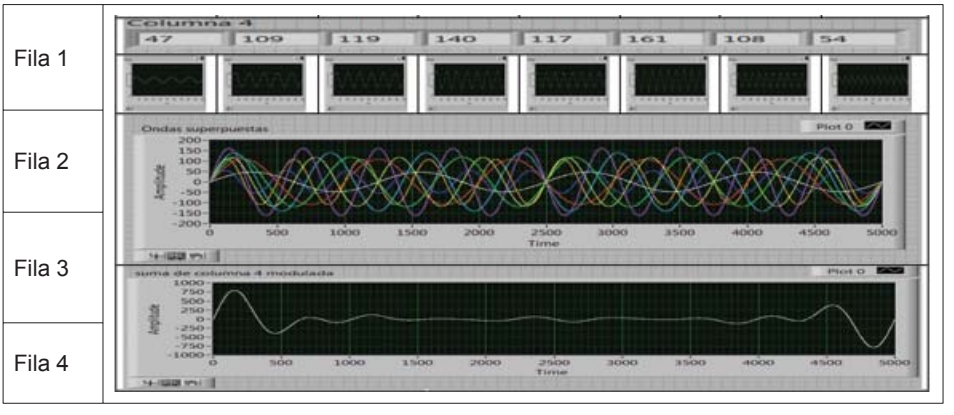
5.4 Barrido horizontal

En el barrido horizontal se ha hecho una variación de la forma de codificar en la que la posición del eje x se representa como instantes sucesivos en el tiempo de barrido, de tal manera que en cada instante sucesivo se codifica solo una línea vertical de la imagen, empezando por la primera columna y terminando en la última. En este caso, el entrenamiento permitirá al invidente ubicar la posición horizontal de la línea cuyo sonido se está ejecutando en cada instante.

5.4.1 Codificación de línea de imagen y composición

Para cada línea de imagen hay tantos puntos en la vertical como resolución tiene la imagen en dicho eje, por lo que para cada posición X se tiene un conjunto de frecuencias que corresponde a cada punto de la imagen en dicha línea; debido a que no se cuenta con un sistema multiparlante, se deben disponer en un solo sonido todos los componentes presentes en dicha línea.

Gráfico 4
Secuencia de generación de patrón de sonido para una línea de imagen de muy baja resolución (8X8)



Elaboración propia.

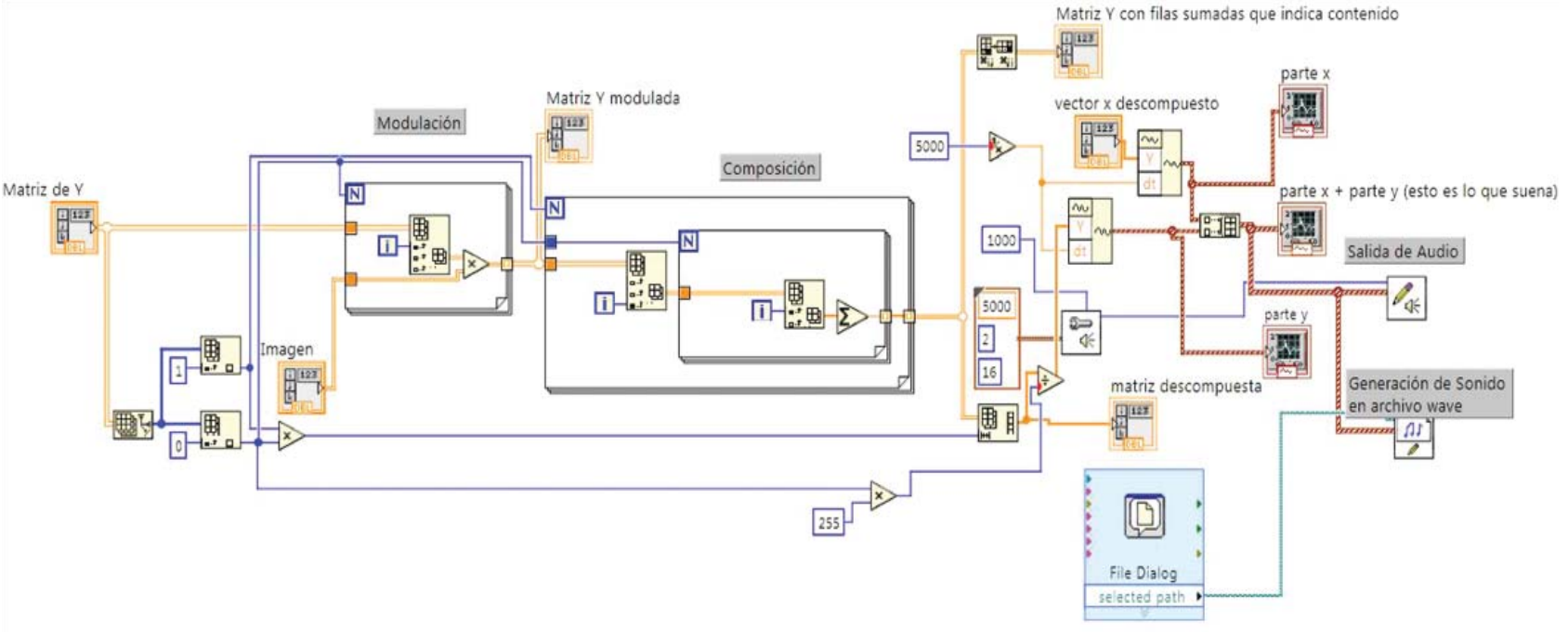
En el gráfico 4, la primera fila muestra los tonos de gris correspondientes a la columna 4 de la imagen de 8X8 mostrada en el gráfico 2, en la segunda fila están representadas las frecuencias que corresponden a dicha vertical modulada con los tonos de gris de tal manera que cada onda mostrada en esta fila codifica tanto la posición vertical como el tono de gris para dicha posición. En la tercera fila se han incluido las 8 componentes de esta línea en la que se pueden apreciar las diferentes amplitudes de las ondas representando el tono de gris, mientras que en la última fila se muestra la suma de todas las componentes que codifica toda la línea de imagen.

Este proceso se realiza con cada una de las líneas verticales de la imagen, generando la onda sonora y concatenando secuencialmente las ondas compuestas para ser emitidas por el parlante.

El entrenamiento en este caso se orientará a fortalecer la relación baja frecuencia = arriba y alta frecuencia = abajo, ya que las frecuencias se incrementan de arriba hacia abajo. Los primeros instantes de la emisión corresponden al lado izquierdo mientras que los subsiguientes implican las posiciones del barrido hacia la derecha.

El *software* ha sido desarrollado usando LabVIEW ™; por ello mostramos, a manera de información general, el diagrama base de la aplicación en el diagrama 1. Aquí el código fuente muestra el procedimiento para la generación del patrón de sonido en el barrido regular. Se parte de una matriz de frecuencias que identifica cada punto de la imagen, y se modula con la imagen para tener una en la que cada punto tenga una modulación en frecuencia (FM) y en amplitud (AM). A este punto, la imagen es bidimensional; para hacer el barrido se debe convertir en unidimensional, por ello se suman las frecuencias línea (vertical) por línea, generando la secuencia de sonidos que representa a la imagen en la estructura denominada composición. Luego se incorpora la información temporal para acondicionar los datos tanto para generar el archivo con extensión .wav como para emitirla por el parlante.

Diagrama 1
Código fuente LabVIEW



Elaboración propia.

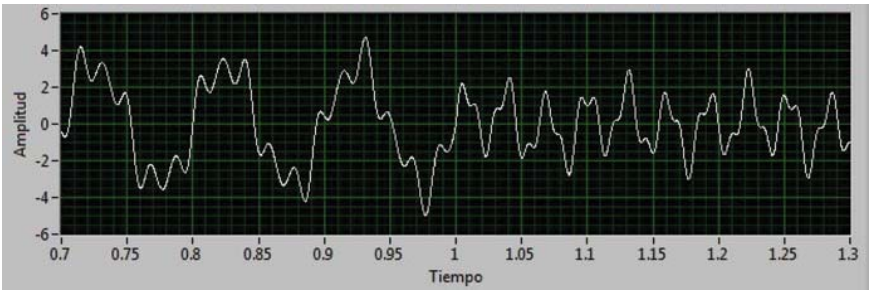
5.5 *Barrido manual*

El concepto del barrido manual consiste en asociar el tacto sobre una superficie rectangular (como la imagen), con el sonido que caracterice la estructura de la imagen en el lugar donde se está «tocando» virtualmente la imagen. En este caso, el usuario escucha el patrón de sonido de la posición en la cual está moviendo el dedo sobre la imagen proveniente de la cámara.

5.6 *Barrido aleatorio*

El barrido aleatorio es una alternativa análoga de lo que sucede con la visión, en la que los conos y bastones de la retina captan fotones individuales y los mapean a la zona del cerebro de donde las toman las redes neuronales para interpretar la imagen. Esta se percibe como continua debido a la cantidad de fotones que llegan simultáneamente a dichas células y al tiempo de extinción de los fotorreceptores. Se trata de definir un equivalente al fotón al que podríamos llamar «sonón» (aun cuando ya existen otros usos de dicho término), que viene a ser un sonido instantáneo de duración, que conserva la frecuencia que identifica su posición, sea vertical, horizontal o combinada, de tal modo que un punto individual de imagen se emite como un sonido de frecuencia que identifica la posición horizontal por el audífono derecho y la vertical por el izquierdo, simultáneamente, con un volumen proporcional al tono de gris de dicho punto. De las muchas alternativas existentes para su definición, se ha optado por una onda sinodal de duración arbitraria. El método requiere que aleatoriamente se elijan uno o más puntos, y que se emita los sonidos que los identifican, de tal manera que se deja al cerebro del usuario la tarea de sincronizar los sonidos del lado derecho y del izquierdo para que, conjugando con los sonidos sucesivos, se logre interpretar una especie de patrón complejo de una imagen formada con un porcentaje razonable de los pixeles que la componen. En el gráfico 5 se muestra la emisión sucesiva de dos «sononeS», en este caso cada sonón codifica las dos frecuencias que identifican sin ambigüedad cada posición. Aquí, por la amplitud, el primer punto es más claro que el segundo y está (por la frecuencia) más a la derecha y más arriba que el segundo.

Gráfico 5
Sonidos consecutivos de dos puntos de imagen



Elaboración propia.

Si se discrimina en sonido estéreo se puede usar el mismo rango de frecuencias para el eje 'x' como para el eje 'y', con la ventaja de la simetría, lo cual se logra utilizando el audífono o parlante derecho para un eje y el izquierdo para el otro. La dificultad por resolver radica en que para tener un patrón de suficiente resolución se requiere combinar un número grande de puntos simultáneamente, sin embargo, a medida que se aumenta el número de puntos emitidos al mismo tiempo, aumenta la probabilidad de ubicar ambiguamente puntos que aparecen simultáneamente en la horizontal o en la vertical, por lo que es más conveniente utilizar un rango de frecuencias diferente para la horizontal que para la vertical.

5.7 *Resumen de metodologías*

En las metodologías desarrolladas o ensayadas, se han mantenido los criterios en dos aspectos: el tono de gris de cualquier punto siempre se codifica con el volumen y la posición con la frecuencia, por lo que las metodologías analizadas representan variantes en la forma de presentar las imágenes al usuario.

En la tabla 1 se resumen las metodologías, donde la simbología ∑Fy.Iy implica que se suma cada onda de frecuencia que identifica la posición vertical modulada por el valor del pixel en dicha posición en toda la vertical. Fx es una onda de frecuencia que identifica la posición horizontal, Fx*Fy es una onda superpuesta de frecuencias que identifi-

can la posición (x, y). Ixy es el valor del pixel en una posición arbitraria (x,y). En los casos donde hay alternativa monoaural o estéreo, se han separado las simbologías con comas.

Tabla 1
Resumen de modalidades de codificación propuestos

Método de barrido	Codificación			Forma
	Tono de gris	Eje x	Eje y	
Regular	Volumen	Tiempo	∑Fy.Iy	Columna por columna
Indexado estéreo	Volumen	Fx, derecha	∑ Fy.Iy, izquierda	Columna por columna
Indexado monoaural	Volumen	Tiempo	∑(Fx*Fy).Ixy	Columna por columna
Manual	Volumen	Manual+ Fx.Ix, (Fx*Fy).Ixy	Manual+Fy.Iy, (Fx*Fy).Ixy	Punto por punto
Aleatorio	Volumen	(Fx) , (Fx*Fy)	(Fy), (Fx*Fy)	Punto(s)
Patrón complejo	Volumen	∑(∑(Fx*Fy).Ixy)		Imagen completa

Elaboración propia.

5.8 *Hardware* para aplicación eficiente

Para tener sistemas funcionales se requiere que sean al mismo tiempo portátiles, de bajo consumo de energía y de alta velocidad; en este sentido, las alternativas posibles incluyen el uso de programación incorporada (*embebed*) basada en compuertas programables en el campo (FPGA) o en dispositivos personales (iPod, Tablet, iPhone o similares), que están aumentando rápidamente su desempeño, disminuyendo su costo, están diseñados para un uso eficiente de las facilidades multimedia de audio y video.

6. *RESULTADOS*

Los resultados de los ensayos se pueden presentar en forma expresable impresa o en archivo de texto. Algunas de las metodologías han quedado plasmadas en programas de computadora cuyos resultados son audiovisuales; por ello la presentación de dichos resultados en esta sección es incompleta o inexistente.

6.1 *Resultados de barrido horizontal*

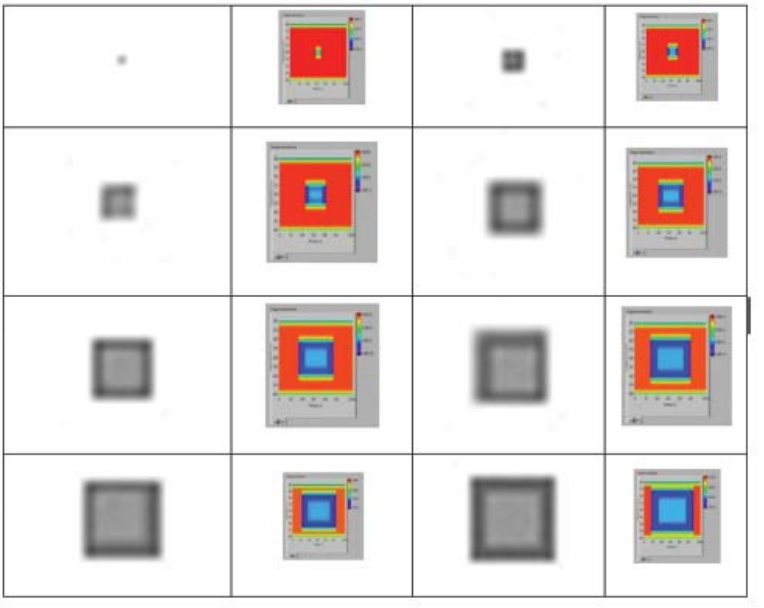
Con la finalidad de probar el concepto, se han hecho ensayos utilizando varias resoluciones de imagen y aplicando metodologías de transformadas rápidas de Fourier (FFT) así como de transformada rápida de Fourier segmentada en el tiempo (TSFFT).

En el gráfico 6 se ve la secuencia de imágenes de 64X64 bytes que representa patrones cuadrados con un fondo claro (blanco de máxima intensidad) en el que hay un cuadrado gris claro dentro de un cuadrado gris oscuro.

Los cuadros a color representan la reconstrucción de las imágenes a partir de los datos generados por el programa que codifica la imagen como patrón de sonido. Toda vez que las posiciones en el eje 'y' son codificadas en frecuencia mientras que en el eje 'x' son codificadas en tiempo, en el método de barrido regular, al aplicar la transformada de Fourier segmentada en el tiempo se obtiene directamente la imagen codificada. Para resaltar la estructura de la imagen se ha utilizado una paleta de colores denominada arcoíris, debido a que los tonos de gris aparecen en falso color y los tonos oscuros se representan por los azules, los intermedios por los verdes y los claros por los rojos.

Teniendo en cuenta que las precisiones esperadas en la interpretación del patrón de sonido serán menores para un sentido no especializado, debemos buscar una mayor resolución y un mejor contraste.

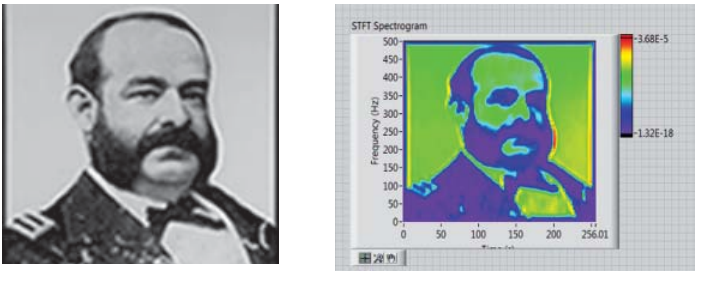
Gráfico 6
Patrones de 64X64 en gris. Imagen recuperada en falso color



Elaboración propia.

En el gráfico 7 se muestra la reconstrucción de la imagen a partir del sonido para un caso de fácil interpretación: mejor resolución, buen contraste y una imagen fácilmente identificable (solo para invidentes que perdieron la visión en la adolescencia o adultez). Los resultados gráficos evidencian este hecho.

Gráfico 7
Imagen original de 128x128 pixeles y su reconstrucción a partir del patrón de sonido generado

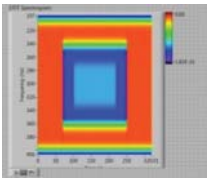
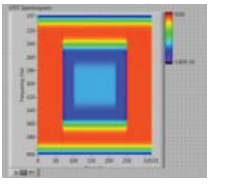
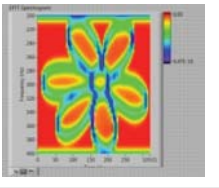
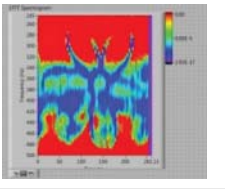
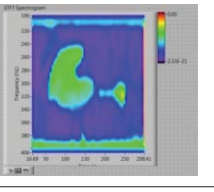
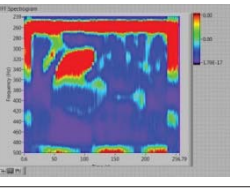
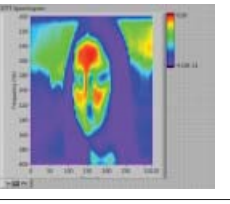
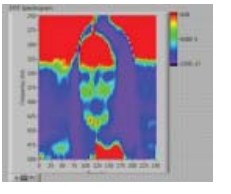


Elaboración propia.

La reconstrucción matemática de la imagen a partir del patrón de sonido es buena bajo la suposición de que no existe distorsión del sonido al ser emitido por el parlante para ponerlo a disposición del invidente, que hará el trabajo equivalente a la herramienta matemática de recuperación de la imagen.

Para una evaluación más completa de la reconstrucción de las imágenes posteriores a la emisión del patrón de sonido se ha elaborado el gráfico 8, que muestra las imágenes originales, la reconstrucción a partir del sonido puro (sonido a emitir) y la reconstrucción a partir del patrón de sonido emitido por los parlantes y grabado en un equipo de sonido convencional.

Gráfico 8
Varias imágenes con su reconstruccion a partir del patrón de sonido sin y con distorsión de los parlantes

Original	Reconstruida de sonido generado	Reconstruida de grabación externa
		
		
		
		

Elaboración propia.

6.1.1 Desempeño temporal

Para ser útiles, las aplicaciones propuestas tienen que cumplir con algunas exigencias en cuanto al tiempo para preparar la información, para emitirla y para asimilarla. Como en el desarrollo de este proyecto se ha utilizado lenguaje de alto nivel, se espera mejorar substancialmente el desempeño cuando la metodología desarrollada sea implantada para el uso de invidentes.

De todas las metodologías ensayadas para mejorar el desempeño, se optó por la generación de las ondas base normalizadas una sola vez para almacenarlas en una matriz de datos relativamente grande (dependiendo de la resolución de la imagen y la resolución temporal que se desee), que luego de compiladas quedan disponibles para la generación de patrones de sonido de todas las imágenes capturadas de la cámara.

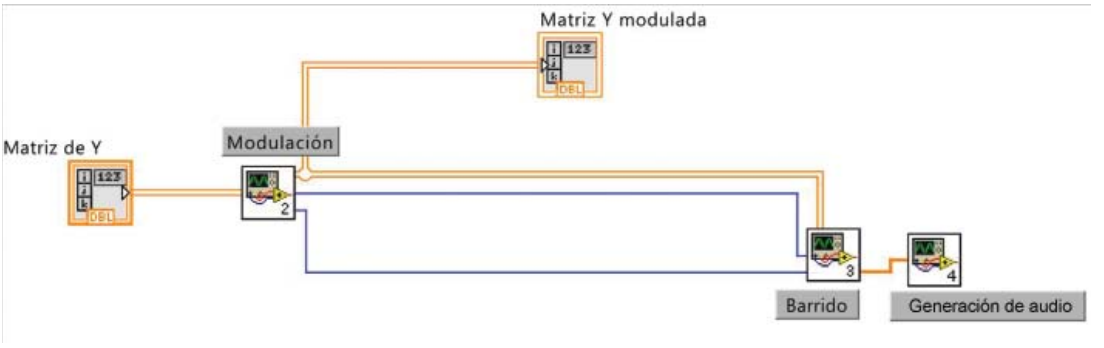
El proceso total incluye los siguientes pasos:

1. Generación de la matriz de ondas base normalizada
2. Lectura o captura de la imagen
3. Modulación de la matriz de ondas con la imagen
4. Barrido columna por columna sumando las componentes de cada columna y su normalización
5. Generación del vector o archivo de sonido
6. Aplicación de la transformada de Fourier para recuperar la imagen.

Los pasos 1) y 6) se hacen con programas separados, de tal manera que la aplicación haga solo los pasos 2) al 5).

El diagrama 2 muestra el diagrama de flujo de datos LabVIEW del programa principal, donde se han segmentado las funciones que hacen los pasos mencionados.

Diagrama 2
Diagrama de bloques del programa de barrido horizontal



Elaboración propia.

La matriz 'Y' contiene las ondas base de frecuencias para cada punto de la imagen, la subrutina de modulación modula las ondas con el contenido de la imagen; solo con fines de experimentación se muestra el resultado como matriz modulada. La subrutina Barrido compone el patrón de sonido y la subrutina de generación de audio genera y emite el sonido por el parlante.

En la tabla 2 están los tiempos promedio de generación de la matriz de frecuencias base para tres tamaños de imagen. Los resultados de las mediciones evidencian que los tiempos dependen de la cantidad de información generada.

Tabla 2
Tiempos de generación de matriz de frecuencias base

Tamaño de matriz	Número de puntos	Rango de frecuencias	Tiempo (ms)
32x32	5.000	1.000-2.000	471,9
64x64	5.000	1.000-2001	1.957
128x128	2.000	1.000-2.002	3.006

Elaboración propia.

Los tiempos importantes desde el punto de vista de la aplicación se muestran en la tabla 3. Es posible generar el patrón de sonido en alrededor de 1.5 s utilizando un computador personal con las características que se muestran en la tabla 4.

Tabla 3
Tiempos de ejecución (ms)

Comparación	128x128	64x64	32x32
Parte 1 Modulación	936	439,4	105,3
Parte 2 Barrido	436,8	309.4	92,3
Parte 3 Generación de audio	15,6	20.8	6,5
Otros procesos de aplicación principal	390	130	79,3
Reconstrucción	1.107,6	1.962	998
Total	2.886	2.861	1.282

Elaboración propia.

La parte correspondiente a otros procesos de la aplicación principal muestra tiempos que en la práctica pueden ser ignorados, ya que las funciones usadas allí fueron solo para mostrar resultados con propósitos de pruebas que no serán necesarias para el usuario. La subrutina de generación de audio tiene incluida una parte que se utilizó para generar un archivo que luego pueda ser utilizado en cualquier equipo de sonido, por lo que puede excluirse, en gran medida, del análisis de los tiempos.

Tabla 4
Características del computador

Procesador	Intel® Core™ i7 CPU Q720 @ 1,60GHz
RAM	6,00 GB
Tipo sistema	64 bit
Sistema operativo	Windows 7 Home Premium Service Pack 1

Elaboración propia.

Esta metodología es apropiada utilizando computadores personales, y se espera que con el uso de programación de bajo nivel se reduzca considerablemente el tiempo de generación del patrón de sonido en el barrido regular.

6.2 *Resultado del método de barrido manual*

El programa captura una imagen y con una rutina de ecualización (procesamiento numérico de imágenes) mejora el contraste. El programa ha sido acondicionado para que este proceso de captura se realice a voluntad del usuario cada vez que se presione el botón del ratón (*mouse*). Tan pronto la imagen es ecualizada es posible, usando el *mouse pad* o la pantalla sensible al tacto (*touch screen*) deslizando el dedo sobre la superficie, «escuchar» cualquier parte de la imagen. El programa está diseñado para emitir sonido solo cuando se ingresa a la región de la pantalla donde está la imagen capturada.

7. DISCUSIÓN

Todas las alternativas de codificación de imágenes, provenientes de captura de video en tiempo real, se sustentan en lo que se conoce —y ha sido extensivamente estudiado— como «plasticidad de la mente»; en este caso, se puede reemplazar la función de la vista adaptando el sentido del oído para interpretar información visual mediante un entrenamiento adicional de las redes neuronales del cerebro.

7.1 *Multidimensionalidad*

La gran ventaja del sentido de la visión es la magnitud de información que es captada por la retina y procesada por el cerebro. La retina es capaz de captar información con más de 5 millones de conos y más de 100 millones de bastones; no es posible codificar tal resolución con frecuencias disponibles en el rango útil de la audición, ya que si suponemos un rango de 10 KHz implicaría discriminar 0,01 Hz en este rango de frecuencias. Por otro lado, la visión estereoscópica implica duplicar el número de puntos de imagen y procesarla en paralelo para generar la tercera dimensión, que en el caso del oído también existe, pero que es utilizada solo para identificar la dirección de la fuente sonora.

Al captar la imagen con una cámara perdemos una buena parte de la resolución espacial y una dimensión espacial por lo que ahora tenemos información bidimensional de baja resolución (comparada con la de la retina).

Al generar los patrones de sonido que codifican la imagen, la resolución de las imágenes han sido reducidas drásticamente; en los ensayos hemos usado resoluciones de alrededor de 16 K pixeles. Estas resoluciones las podemos catalogar como de muy baja resolución.

En los métodos de barrido la imagen bidimensional es emitida como unidimensional; debido a ello el proceso de aprendizaje deberá compensar este hecho para poder reconstruir una imagen bidimensional que requiere entrenar intensamente la memoria, para retener y ordenar 128 líneas verticales de 128 puntos cada una.

En el método de sonido aleatorio, la bidimensionalidad se mantiene a costa de perder resolución o hacer más complicado el proceso de memorización, ya que en este caso se deben superponer patrones aleatorios de baja resolución para generar un patrón bidimensional de mayor resolución, que se aproximará, en teoría, tanto como se quiera pero sin llegar a igualar la resolución de la imagen capturada.

7.2 *Temporalidad*

Los tiempos de captura, procesamiento y generación del patrón de sonido, debido a la cantidad de información que se debe procesar, requieren de un lapso durante el cual la imagen no debe cambiar, por lo que se perderán eventos que ocurren en ese tiempo, así como los que ocurran durante la emisión del patrón de sonido. Esta desventaja puede ser parcialmente mejorada con emisión de las líneas actuales de video durante el barrido; en este caso, cada línea emitida es en tiempo real, pero la imagen es la composición de varios instantes del barrido, que implicaría artefactos extraños al poner como contiguas líneas que pueden no serlo, y, en el caso del barrido aleatorio, complicaría mucho la interpretación del patrón de sonido. Queda claro que la resolución temporal también será baja.

7.3 *Interpretación*

El entrenamiento, en general, para los métodos descritos tiene dos componentes fundamentales: la posición y la luminosidad. En el caso de los que no son invidentes de nacimiento es muy probable que tales conceptos, desde el punto de vista visual, ya existan. En el caso

de los invidentes de nacimiento, el concepto de posición es parte de su experiencia de invidente y tienen asimilados con mucho énfasis los conceptos de derecha/izquierda así como los de arriba/abajo, por lo que el entrenamiento de la asociación de la frecuencia con la posición es asimilable en función de conceptos previamente adquiridos.

El concepto de claro/oscuro será mucho más difícil de enseñar, pues tienen relacionadas las formas con el tacto mas no con la iluminación. Desde este punto de vista, la aplicación desarrollada de barrido manual será una herramienta que puede usarse por sí sola como ayuda al invidente, pero también se convierte en un instrumento de entrenamiento en cualquiera de las otras metodologías propuestas.

7.4 *Ventajas y desventajas*

La mayor ventaja del método de barrido regular es que permite interpretar línea por línea de un lado a otro la imagen, lo que da tiempo al usuario de reconstruir la imagen a partir de los claros y oscuros que interpreta ubicados en la vertical. En el caso de usar un sistema monoaural la desventaja es que la posición horizontal tiene que ser calculada en función de la identificación del inicio del barrido y la duración típica de un barrido completo. Esta deficiencia se subsana en el estéreo, donde se puede dedicar un audífono a la sola misión de identificar la posición. Las desventajas son probablemente la asimetría, ya que el efectuar un barrido de derecha a izquierda o viceversa es arbitrario.

Las ventajas del método aleatorio radican en que el concepto intenta enviar un patrón simultáneo en todas las direcciones, lo cual es concordante con la forma como la mente analiza las imágenes, aun cuando desde el punto de vista práctico esta emisión se realice en instantes sucesivos, tratando de complementar en cada instante sucesivo información faltante del patrón de pocos puntos que se emiten en cada instante.

El uso de computadores es ventajoso por su permanente evolución y la tendencia es lograr más velocidad y capacidad de almacenamiento. Aun utilizando lenguajes de alto nivel, como en el presente caso, se pueden lograr tiempos razonablemente cortos como los mostrados en el capítulo de resultados. En el anexo 1 mostramos que una imagen de 64*64 pixeles con una excelente resolución temporal (5.000 valores por pixel) implica procesar un total de 20.480.000 valores haciendo varias

operaciones con cada uno de ellos. Las altas capacidades de almacenamiento de los computadores personales ha permitido también reducir drásticamente los tiempos de procesamiento, ya que empezamos leyendo una matriz muy grande de datos que no cambian y que representan la distribución espacial de frecuencias, que adicionalmente permiten precodificar algunos otros aspectos que son fundamentales desde el punto de vista de optimizar la interpretación de los patrones.

8. CONCLUSIONES

Se han ensayado metodologías de codificación de imágenes como patrones de sonido, demostrando que se pueden incluir las características de la imagen en patrones de sonido para ser emitidos con recursos disponibles en los computadores personales, y se pueden extrapolar a otros dispositivos portátiles cada vez más populares. Los computadores personales pueden ser utilizados como ayuda al invidente, usando lenguajes de alto nivel y recursos existentes en las actuales versiones portátiles con resoluciones aceptables y en tiempos razonables.

Las codificaciones en frecuencia y amplitud, en la mayoría de los casos, permiten verificar la factibilidad de la recuperación de imágenes, ya que las metodologías matemáticas que convierten el espacio temporal al de la frecuencia, realizan lo que el sentido del oído hace en forma natural al discriminar los componentes de frecuencia del sonido.

Desde que las imágenes convertidas a sonido son almacenadas como archivos de sonido convencionales, estos pueden ser posteriormente reproducidos en cualquier equipo de sonido capaz de leer tales archivos, por lo que los invidentes podrían guardar y reproducir fotografías del mismo modo que lo hacemos los no invidentes.

REFERENCIAS

Arno, P. et al. (2001). Auditory substitution of vision: Pattern recognition for the blind, *Appl. Cognit. Psychol. 15*, 509-519.

Blasch, B. B, Wiener, W. R., & Welsch, L. (1997). *Foundations of orientation and mobility*. (2.ª ed.). Nueva York: AFB Press.

Blogspot Ben Underwood, a celebration of life (2009). Recuperado en noviembre del 2011 de http://www.benunderwood.com

Caspi, A., Dorn, J. D., McClure, K. H., Humayun, M., Greenberg, R. J., & McMahon, M. J. (2009). Feasibility study of a retinal prosthesis, *Arch Ophthalmol, 127(*4), 398-401.

Cronly-Dillon, J., Persaud, K. C., & Blore, R. (2000). Blind subjects construct conscious mental images of visual scenes encoded in musical form, *Proc. R. Soc. Lond B,* 267, 2231-2238.

Maurice, P., Solvej, M., Albert, G. and Ron K. (2005). Cross-modal plasticity revealed by electrotactile stimulation of the tongue in the congenitally blind. *Brain, 128,* 606-614.

Meijer, P.B. L. (1992). An experimental system for auditory image representations, *IEEE Trans. on Biomed Eng*., *39*(2), 112-121.

Nakanishi, K. (2008). Image coding apparatus, image coding decoding apparatus, and image decoding method, *US patent Aplication 20080080779.*

Neville H. J., & Bavelier, D. (2000). Specificity and plasticity in neurocognitive development in humans. *The New Cognitive Neurosciences, 2,* 83-98.

Okada, S., Taketa, K., (2008). Image coding method and apparatus, and image decoding and apparatus, *US patent Application* 20080075373.

Portal World Access for the blind (2000). Recuperado en marzo del 2012, de http://www.worldaccessfortheblind.org/

Portal Reportajes (2008). *Ben Underwood, el joven ciego con visión sónica*. Recuperado en marzo del 2012 de http://www.reportajes.org/2008/07/23/ben-underwood-el-joven-ciego-con-vision-sonica/

Portal Zapp Internet (2009). *Ben Underwood - El chico que puede ver sin ojos*. Recuperado en marzo del 2012, de http://www.zappinternet.com/video/PaLlLorTuy/Ben-Underwood-El-chico-que-puede-ver-sin-ojos

Shouval, H. Z. (2011). What is the appropriate description level for synaptic plasticity? *PNAS. 108*, 19103-19104.

Szurman, P., Warga, M., Roters, S., Grisanti, S., Heimann, U., Aisenbrey, S., Rohrbach, J. M., Sellhaus, B., Ziemssen, F., Bartz-Schmidt, K. U., (2005). Experimental Implantation and

Long-term Testing of an Intraocular Vision Aid in Rabbits, *arch ophthalmol* 123, 964-969.

Tsuji, M., Suzuki, S., & Toyama, K. (2008). Sound encoding method and apparatus, *US patent Aplication 2008008325*.

Wilson, S. J., Lusher, D., Wan, C. Y., Dudgeon, P., & Reutens, D. C. (2009). The neurocognitive components of pitch processing: Insights from absolute pitch, *Cerebral Cortex*, 19, 724-732.

Zeki, S. (2005). The functional specialization of the brain in space and time, *Phil. Trans.* R. Soc. B 360, 1145-1183.