

Codification of the contour of the visual image to acoustics patterns and its decoding by blind

José Rodolfo Martínez y Cárdenas¹, Felipe Rolando Menchaca García², Virginia Berrón Lara³

¹Centro Interdisciplinario de Investigación para el Desarrollo Integral Regional (CIIDIR), Instituto Politécnico Nacional (IPN), Calle Hornos N° 1003, C.P. 71230.

Sta. Cruz Xoxocotlán, Oax., México. Telf. 01(951)5170400 ext. 82739. rodolfo_mc@yahoo.com

²Centro de Investigación de Cómputo (CIC), IPN, México D.F. fmenchac@gmail.com

³Universidad Tecnológica de la Mixteca. Huajuapán de León, Oax., México.

berron@mixteco.utm.mx

Abstract

In this paper is proposed the development's first phase of a conceptual model for a non-invasive prosthesis for helping blind people to obtain information about the shape of objects. The prosthesis has been called "Acoustic Viewer". It is based on an encoding of digital image contours into an acoustic pattern which is conceptually simpler than those are found in previous papers. The digital imaging processing algorithms and the method for transforming image into sound using *CSound* are described here. In order to evaluate coding's suitability were chosen 33 (basic) figures, which are formed of straight lines and curves. A training course in their decoding was designed and an experiment with two groups of people, the first consisting only of blind people and the second consisting of people with normal vision, who were trained and finally they participated in a random trial to assess their acquired ability to decoding these patterns. Our results show that both groups developed the same ability for decoding the proposed patterns. This paper shows "Acoustic Viewer" could help blind people.

Key words: acoustic patterns and blindness, non-invasive prosthesis for blind people, sensorial substitution, acoustic vision.

Codificación del contorno de la imagen visual a patrón acústico y su decodificación realizada por invidentes

Resumen

En este trabajo se propone la primera etapa del desarrollo de un modelo conceptual de una prótesis no invasiva para ayudar a los invidentes a obtener información sobre la forma que poseen los objetos. Se ha llamado a esta prótesis "Visor Acústico". Se basa en una codificación del contorno de la imagen digital en un patrón acústico conceptualmente más sencilla que las presentadas en trabajos anteriores; se describen los algoritmos de procesamiento digital de la imagen utilizados para la extracción del contorno y los de su posterior conversión a sonido mediante el uso de *CSound*. Para evaluar la idoneidad de la codificación se propuso la transformación de 33 contornos (básicos) formados por líneas rectas y curvas, un curso de entrenamiento en su decodificación y un experimento en el que a dos grupos de personas, uno de invidentes y el otro de sujetos con visión normal, se les entrenó, y después, sometió a una prueba aleatoria de estímulos con 8 de los patrones propuestos con la finalidad de valorar su capacidad de decodificación ad-

quirida. Los resultados muestran que ambos grupos desarrollaron la misma habilidad para decodificar los patrones. Este trabajo demuestra que el "Visor Acústico" podrá ayudar a los invidentes.

Palabras clave: patrones acústicos y ceguera, prótesis no invasivas para personas ciegas, sustitución sensorial, visión acústica.

Introducción

El sentido de la vista en el humano juega un papel importante en nuestra vida diaria; su carencia nos enfrenta a retos muy difíciles, sobre todo en la generación de conceptos sobre las cosas que nos circundan y en la navegación por las ciudades. La Organización Mundial de la Salud [1] estimó, en mayo de 2009, que la comunidad de invidentes y discapacitados visualmente ascendía a 314 millones de personas en el mundo, siendo 45 millones de ellas ciegas, lo que resulta un problema social y educativo a nivel mundial [2].

A consecuencia del desarrollo de propuestas de sistemas de ayuda para este grupo social, en la década de los 70's se introdujo el concepto de sustitución sensorial no invasiva [3], con la idea de "sustituir" parcialmente una modalidad sensorial por el uso de otra, y el concepto de la codificación de una imagen a sonido [4]. Así mismo, se estableció el concepto de sonificación el cual plantea la transformación de datos en sonidos, con la intención de transmitir información [5].

Hasta ahora los usuarios de los sistemas que convierten la imagen en sonido han manifestado insatisfacción debido a la complejidad del patrón acústico al momento de llevar a cabo su lectura [6]. Por esta razón, aquí se plantea una alternativa de codificación del contorno de la imagen digital a patrón acústico buscando que sea lo menos compleja posible.

Imagínese a una imagen como si estuviera embebida dentro de una pantalla de televisión blanco y negro formada por una multitud de líneas horizontales, en donde a cada una se le asocia una frecuencia correspondiente a una tecla del piano. Estas frecuencias se ordenarán en forma ascendente (de abajo hacia arriba de la imagen), de modo que las frecuencias más altas queden en la parte superior (Figura 1).

Se tiene una línea vertical t , llamada línea de barrido, inicialmente colocada en el extremo izquierdo de la imagen, la cual se desplaza hacia la derecha con una rapidez uniforme explorando

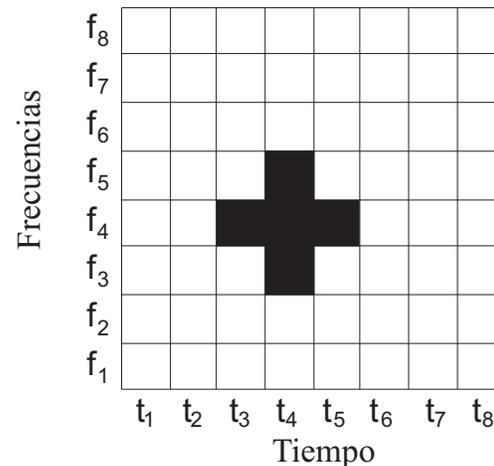


Figura 1. Aquí se muestra como cada línea se asocia a una frecuencia.

durante su trayectoria a la imagen en su totalidad.

La línea vertical hará sonar las frecuencias de los píxeles negros (contorno) que vaya encontrando en la imagen al pasar sobre ésta. En caso de que haya varios píxeles en el mismo tiempo t_i la salida del sistema en ese momento estará compuesta por la ejecución simultánea de las frecuencias correspondientes. Bajo este principio, la codificación de la imagen digital consiste en mapear el contorno de imagen representada en un espacio ortogonal bidimensional, siendo el eje horizontal el tiempo y el eje vertical el de las frecuencias. Como fuente generadora de frecuencias de sonido se escogió al piano, uno de los instrumentos musicales más completos en lo que respecta a la generación de frecuencias audibles y armoniosas. De las 88 teclas del piano se seleccionaron sólo 50 de ellas con la finalidad de simplificar lo más posible el patrón acústico (Tabla 1).

Parte experimental

Este trabajo estuvo constituido por cinco etapas:

Tabla 1
Las 50 frecuencias correspondientes a las 50 teclas del piano escogidas

Tecla del piano	Nota del piano*	Frecuencia en Hz	Tecla del piano	Nota del piano*	Frecuencia en Hz
1	A ₀	25.5	26	E ₄	329.63
2	B ₀	30.868	27	F ₄	349.23
3	C ₁	32.703	28	G ₄	392.0
4	D ₁	36.708	29	A ₄	440.0
5	E ₁	41.203	30	B ₄	493.88
6	F ₁	43.654	31	C ₅	523.25
7	G ₁	48.999	32	D ₅	587.33
8	A ₁	55.00	33	E ₅	659.26
9	B ₁	61.735	34	F ₅	698.46
10	C ₂	65.406	35	G ₅	783.99
11	D ₂	73.416	36	A ₅	880.00
12	E ₂	82.407	37	B ₅	987.77
13	F ₂	87.307	38	C ₆	1046.5
14	G ₂	97.999	39	D ₆	1174.7
15	A ₂	110.0	40	E ₆	1318.5
16	B ₂	123.47	41	F ₆	1396.9
17	C ₃	130.81	42	G ₆	1568.0
18	D ₃	146.83	43	A ₆	1760.0
19	E ₃	164.81	44	B ₆	1975.5
20	F ₃	174.61	45	C ₇	2093.0
21	G ₃	196.00	46	D ₇	2349.3
22	A ₃	220.00	47	E ₇	2637.0
23	B ₃	246.94	48	F ₇	2793.8
24	C ₄	261.63	49	G ₇	3136.0
25	D ₄	293.66	50	A ₇	3520.0

*Sistema de notación musical inglesa, es una manera abreviada de representar las notas musicales con letras del alfabeto.

Primera etapa

Se *planteó* el modelo para obtener el patrón acústico, a partir del cual se obtendría la codificación del contorno de la imagen digital a sonido. Se partió del hecho de que dicho contorno es extraído de una imagen digital con formato BMP almacenada en la memoria de la computadora, la

que previamente habría sido capturada por un dispositivo digital [7]. Esta imagen mostraría ya sea un obstáculo u objeto que se encuentra virtualmente frente al invidente, sin especificar por ahora su distancia con respecto al sujeto. Ésta es transformada a imagen binaria (Figura 2), en donde el contorno de la imagen se representa por

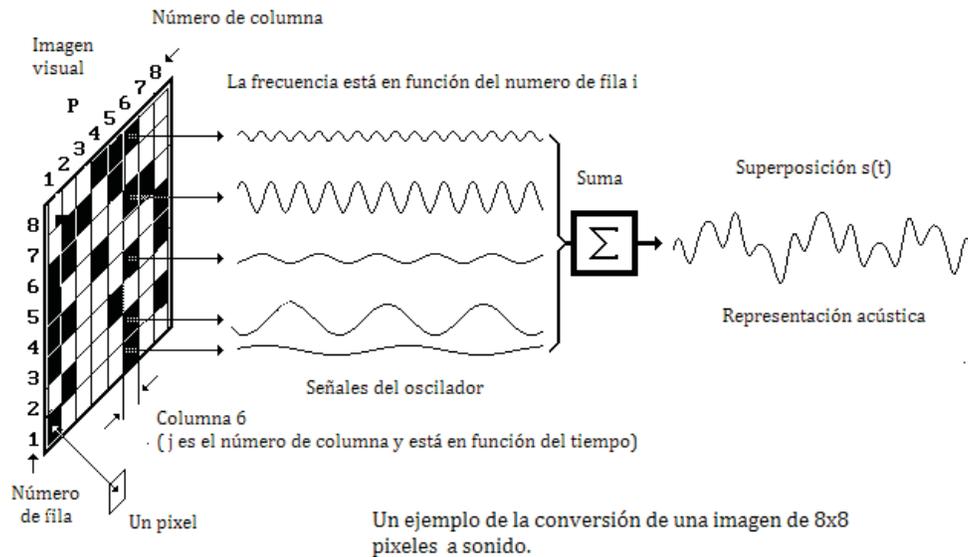


Figura 2. Se muestra el principio propuesto de barrido de la imagen visual y su transformación a patrón acústico del sistema Visor Acústico.

medio de píxeles negros correspondiéndole a cada pixel una frecuencia del piano preestablecida.

En este modelo (Figura 2) a cada línea horizontal se le asoció la frecuencia de una nota musical correspondiente a una tecla del piano. Se ordenaron las frecuencias de estas notas en forma ascendente (de tonos graves a agudos) sobre el eje de las frecuencias (eje y), de tal modo que los tonos más graves correspondieran a la parte inferior del contorno de la imagen y los tonos agudos a la parte superior del contorno, esparciéndose las 50 notas a lo largo del eje vertical. Todos los contornos de las imágenes quedaron insertados en el rango de frecuencias comprendido entre 4 octavas del piano iniciando con la frecuencia de Do2 hasta llegar a Do6.

Segunda etapa

Se llevó a cabo la *determinación* de los algoritmos de procesamiento digital de la imagen necesarios para prepararla para su codificación en sonido. Se implementaron los programas en un lenguaje de alto nivel permitiendo primero cargar imágenes en colores en la memoria de la computadora para luego procesarlas.

Para preparar la imagen se utilizaron los siguientes tres bloques de algoritmos: *transformaciones elementales*, donde se introducen las ope-

raciones más simples de tratamiento de imágenes; *transformaciones de vecindad*, donde se introducen aquellas operaciones que tienen en cuenta la vecindad de cada píxel para llevar a cabo la transformación, considerándose especialmente las de suavizado; el último grupo de algoritmos es el que corresponde al *histograma y ecualización de la imagen*, se estudian las propiedades del histograma, su modificación y el realzado de imágenes mediante variaciones del histograma (Figura 3).

Estos tres bloques sumaron un total de 11 algoritmos, los que se combinaron unos con otros obteniéndose diferentes resultados para cada combinación. A continuación se procedió a hacer una sistematización ordenada de la experimentación de las etapas y sus opciones, con la finalidad de no pasar por alto ninguna posible transformación importante.

Se seleccionó de estas combinaciones aquellas que generaron una imagen binarizada considerablemente limpia. Sin embargo, estas imágenes aun presentaron objetos (agujeros) por afuera y por dentro del contorno, mismos que se eliminaron de la imagen tomando en consideración su tamaño. Para esto, se determinó una 4-adyacencia para los píxeles blancos y una 8-adyacencia para los negros, para poder aplicar el algoritmo de etiquetación de componentes (Fi-

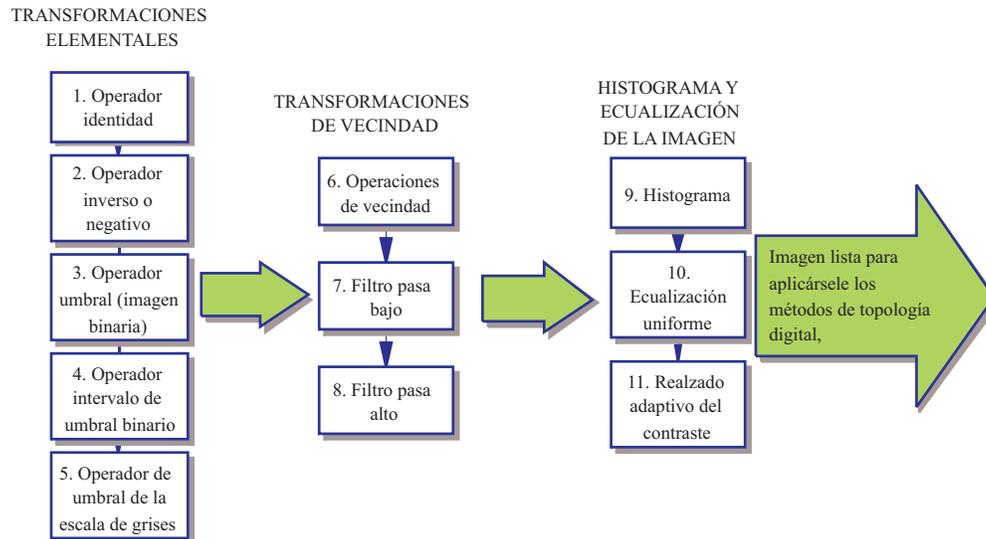


Figura 3. Representación de los algoritmos que fueron usados en el tratamiento de la imagen y en su transformación antes de llevarse a cabo la topología digital.

gura 4); se hizo un barrido de la imagen pixel a pixel de izquierda a derecha y de arriba hacia abajo. Las etiquetas asignadas fueron valores mayores que 1, el valor 0 representó un punto blanco y el valor 1 un punto negro. Este algoritmo fue utilizado tanto para etiquetar los píxeles que eran 8-adyacentes (puntos negros), como para etiquetar los píxeles que eran 4-adyacentes. Después etiquetar las componentes se eliminaron aquellas que no cumplieron con el tamaño definido.

Tercera etapa

Uno de los problemas que se presentaron al llevar a cabo la *transformación a un patrón acústico* fue que al hacer el barrido de la imagen de izquierda a derecha a veces era necesaria la ejecución de más de una frecuencia en un instante dado de tiempo. Esta situación fue resuelta utilizando el lenguaje de programación musical CSound, con el cual se crearon los archivos de los patrones acústicos con extensión *.wav. Dado que el contorno de la imagen está representado por una matriz de datos, se desarrolló un programa en lenguaje de alto nivel para el barrido de dicha matriz y detección de los puntos correspondientes al contorno de la imagen para transportarlo a una base de datos. Posteriormente, esta base de datos fue requerida por CSound con el objetivo de generar las frecuencias correspondientes al patrón acústico de cada una de las

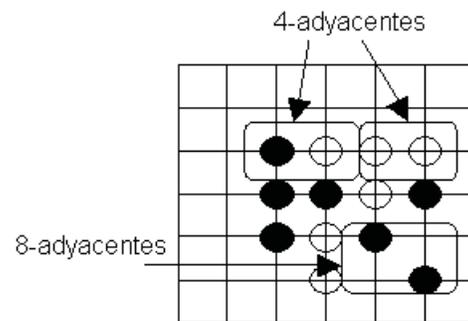


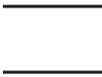
Figura 4. Adyacencia entre puntos.

imágenes digitales. Una vez hecho esto, se reprodujo el resultado del archivo de sonido a través del convertidor digital analógico (DAC) de la tarjeta madre o bien por medio del DAC de la tarjeta de sonido de la computadora personal.

Cuarta etapa

En esta etapa quedó establecida la capacidad de poder convertir cualquier imagen digital a patrón acústico. Sin embargo, esto no garantizó por sí solo que los invidentes pudieran decodificar la imagen, ya que se comprobó que al escuchar los patrones, estos no les proporcionaban ningún significado real o concepto. Así que nació la necesidad de *establecer un código* basado en figuras geométricas (Tabla 2) elementales y que sirvieran a su vez como un facilitador para

Tabla 2
Código propuesto de 33 figuras geométricas, el cual representó los patrones acústicos utilizados en nuestros experimentos

Patrones con una línea recta							
							
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Patrones con dos líneas rectas							
							
(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
							
(17)	(18)	(19)	(20)				
Patrones con tres líneas rectas							
							
(21)	(22)	(23)	(24)	(25)	(26)		
Patrones con cuatro líneas rectas							
							
(27)	(28)	(29)	(30)				
Patrones con líneas curvas							
							
(31)	(32)	(33)					

representar segmentos de contornos más complejos.

Se consideró que todos los contornos en la naturaleza están contruidos a partir de líneas y curvas cuyas combinaciones generan variaciones de todas las formas físicas de la naturaleza; con base en esto, se propuso el siguiente código constituido por 33 figuras geométricas (Tabla 2).

Quinta etapa

Como un primer paso para lograr el *aprendizaje de la lectura* de los patrones acústicos se utilizó el aprendizaje táctil: se dibujó y se pegó el contorno de cada imagen del código propuesto en relieve sobre tarjetas de cartón. Se le fue dando una a una cada tarjeta a cada uno de los invidentes, quienes exploraron con la yema de los dedos los contornos. Posteriormente, el instructor del curso les proporcionó a los sujetos el nombre particular de la imagen. Con esta información adquirida por el oyente de la forma táctil del objeto y de su nombre particular se logró construir un concepto mental primario de la forma física del objeto en cuestión. Como siguiente etapa, se inició el entrenamiento acústico, el cual consistió de tres sesiones, en cada una de las cuales se expuso cada uno de los patrones acústicos durante tres ocasiones, reforzando el aprendizaje con las tarjetas con los relieves y repitiendo el nombre de la imagen por parte del instructor. A partir de estos tres eventos, el táctil, el de nombre particular y el acústico, el oyente construyó un conocimiento del significado cognoscitivo del patrón acústico y su relación intrínseca con la forma del contorno del objeto.

Para evaluar la capacidad de decodificación de los patrones acústicos por parte de los invidentes se utilizó un modelo de experimentación basado en el uso de los Cuadrados Latinos, para lo cual se formaron dos grupos de ocho personas, uno de control con sujetos con visión normal y el otro formado por invidentes, cabe aclarar que ambos grupos llevaron el mismo taller de entrenamiento.

El primer Cuadrado Latino estuvo constituido por 8 individuos con visión "normal", cuatro hombres y cuatro mujeres; en el segundo Cuadrado de 8 invidentes, cuatro hombres y los restantes, mujeres. Se seleccionaron 8 de las 33

figuras del código propuestas por medio del método de números aleatorios, formando el grupo de estímulos acústicos.

Resultados

En la Figura 5 se muestra una de las secuencias de las imágenes a las que se les aplicó la mejor combinación de los 11 algoritmos de procesamiento digital de la imagen propuestos. Esta secuencia correspondió a la serie de algoritmos etiquetados como 5, 8, 9 y 3, la cual proporcionó una imagen con pocos objetos de diferentes tamaños (Figura 3). Para quitarle estos elementos indeseables fue sometida a las técnicas de topología digital. Finalmente se obtuvo una imagen etiquetada como 14_1, la cual presentó el contorno más nítido. De esta misma manera se procesaron diferentes imágenes digitales llegando al mismo resultado. Al obtenerse un solo contorno se procedió a para hacer la transformación a patrón acústico.

Imágenes como la 14_1 pueden ser fáciles de interpretar por sujetos con visión normal, no así por los invidentes, debido a la falta de entrenamiento adecuado y a que muchos de ellos no tienen claro el concepto de formas (geométricas) en su mente. Por esto, en la parte experimental se decidió usar sólo figuras geométricas. En la Tabla 3, se presentan los resultados obtenidos al compararse los datos proporcionados por los dos grupos de 8 sujetos cada uno, el de los de visión normal y el de invidentes. Estos datos fueron obtenidos después de ser sometidos ambos grupos a una evaluación que constó de 8 de los 33 patrones acústicos seleccionados en forma aleatoria para formar el grupo de sonidos utilizados en los Cuadrados Latinos.

Se puede observar que para ambos grupos de sujetos les es difícil identificar el círculo, una curva, y el patrón de líneas ascendentes y descendentes. Llegado a este punto fue pertinente hacer una evaluación cuantitativa para ver si existía o no una diferencia significativa entre los dos grupos. Así que se aplicó la *t-student* con una alfa de 0.01 a los datos de la Tabla 3. Concluyéndose finalmente que no existía ninguna diferencia significativa real ente ambos grupos en términos de la capacidad para hacer la decodificación de los patrones acústicos. Es importante hacer

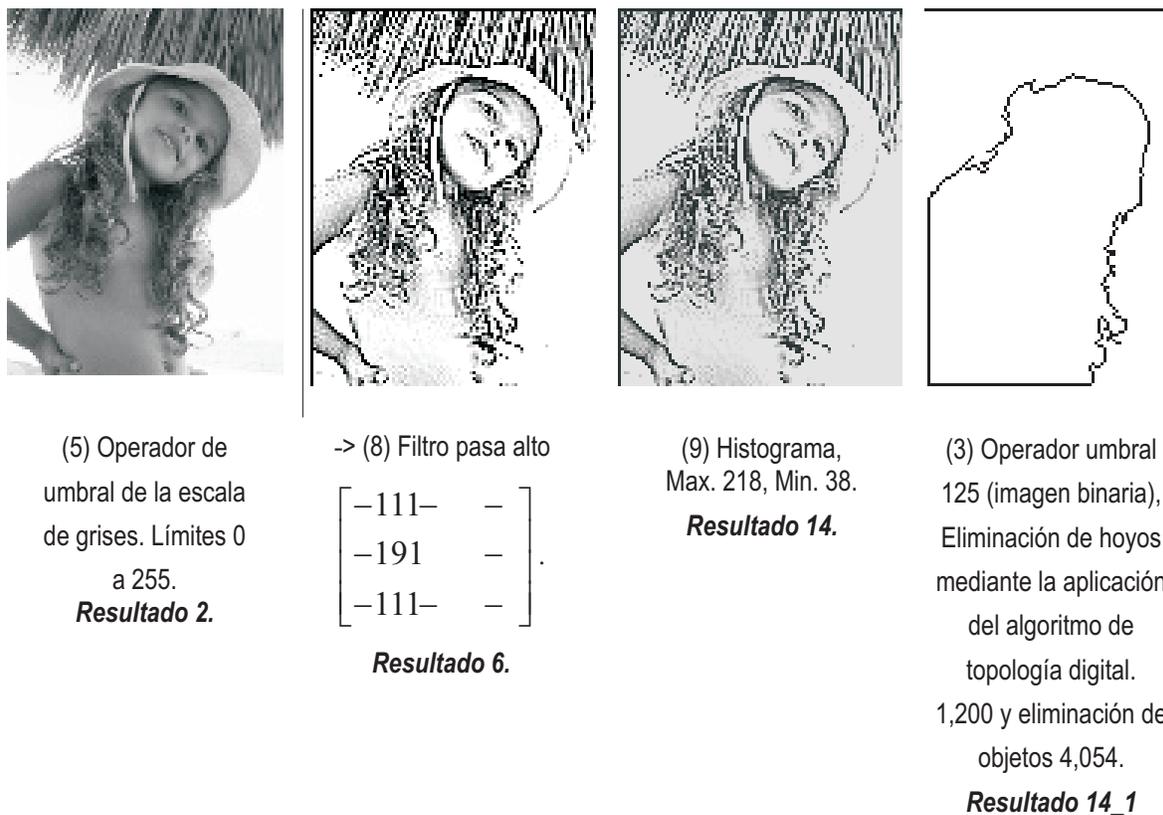


Figura 5. El mejor resultado de la combinación de los algoritmos aplicados.

Tabla 3
Porcentaje de identificación de los patrones acústicos

Número de Patrón Acústico referido a la Tabla 1	Descripción de la imagen	Porcentaje de identificación de sujetos con visión normal	Porcentaje de identificación de sujetos invidentes
15		100	100
31		50	50
26		62.5	100
5		100	100
27		75	100
13		100	100
1		100	87.5
17		87.5	50

notar que aunque existe una gran diferencia en lo que respecta a experiencia sensorial entre ambos grupos a lo largo de sus vidas, esto no influye en la decodificación de los patrones acústicos.

Como una última prueba (subjetiva) se les pidió a algunos invidentes que dibujaran los patrones acústicos que en forma aleatoria escuchaban, con el fin de demostrar su capacidad para decodificarlos y poder constatar de esta forma qué era exactamente lo que construían en sus mentes (Tabla 4). Esta prueba representó para los invidentes un reto, ya que tuvieron que construir primero una imagen mental de la forma y después plasmarla en una hoja de papel.

Discusión de resultados

Con este trabajo se concluye que el modelo propuesto para la obtención de un patrón acústico menos complejo es viable y representativo del contorno de una imagen digital. Las 33 figuras geométricas fueron fácilmente transformadas a patrones acústicos, y se demostró la capacidad desarrollada por los sujetos para poderlas decodificarlas.

Esta idea forma parte de un modelo conceptual de lo que será en un futuro una prótesis no invasiva para la rehabilitación parcial de la ceguera, a la que hemos llamado "Visor Acústico".

Se demostró que se puede codificar cualquier contorno representativo de una imagen digital a patrón acústico mediante el procedimiento propuesto. Los participantes hicieron ver la necesidad de implementar más talleres para capacitar a los sujetos en la lectura de los patrones acústicos de imágenes de objetos comunes a nuestro entorno. En base a esto afirmamos que existe todavía un camino largo por recorrer entre la decodificación de las 33 imágenes propuestas y las imágenes digitales que puedan representar objetos comunes. Son necesarias propuestas de nuevas estrategias de aprendizaje para facilitar la decodificación rápida y eficaz de imágenes digitales complejas.

Los dibujos realizados por los invidentes sustentan la capacidad que tienen para decodificar los patrones acústicos y la construcción mental de conceptos geométricos y de nuevo conocimiento.

Tabla 4
Dibujos realizados por un invidente interpretando los patrones acústicos

Imagen del patrón acústico según Tabla 1	Dibujo a mano alzada realizada por invidente
 (1)	
 (16)	
 (15)	
 (13)	
 (10)	
 (12)	
 (23)	
 (24)	
 (26)	
 (25)	
 (31)	
 (27)	
 (28)	
 (21)	

Conclusiones

Aun con estos resultados alentadores, existe un camino largo por descubrir, ya que surge la necesidad de experimentar más con base al código propuesto e incursionar con contornos más complejos, de tal manera que poco a poco nos vayamos acercando a la decodificación de imágenes reales. Este proceso de aprendizaje es semejante al aprendizaje de otros códigos, como por ejemplo: el código Braille y el código Morse. Este trabajo permite demostrar las ventajas que tendrá el Visor Acústico como una prótesis visual no invasiva para los invidentes.

Referencias bibliográficas

1. Organización Mundial de la Salud, Centro de Prensa: "Ceguera y Discapacidad Visual". Nota Descriptiva No. 282, Mayo (2009).
2. Aponte Y.: "Secundarias para ciegos". La Gaceta de la Univ. Guadalajara, Año 7, Edición 570, Lunes 9 de febrero (2009) 16.
3. Matta S., Kumar D. K., Yu X. and Burry M.: "An Approach For Image Sonification". First International Symposium on Control, Communications and Signal Processing, (2004) 431-434.
4. Meijer P.: "Vision Technology for the totally blind". Seeing with Sound. The vOICe. 2007. Disponible en (14/8/08): <http://www.seeingwithsound.com/>
5. Jihong L. and Xiaoye S.: "A Survey of Vision Aids for the Blind". Proceedings of the 6th World Congress on Intelligent Control and Automation, Dalian, China, June 21-23 (2006) 4312- 4316.
6. Cimarolli V., Boerner K. and Wang S.: "Life goals in vision rehabilitation: Are they addressed and how?" Journal of Visual Impairment & Blindness, June (2006) 343-352.
7. Pajares G., De la Cruz M. J., Molina M. J., Cuadrado J. y López A.: "Imágenes Digitales, procesamiento práctico con JAVA". Ra-Ma, España 2004.

Recibido el 23 de Febrero de 2010

En forma revisada el 4 de Abril de 2011