

Sonificación de Imágenes 2D

Antonio S. MONTEMAYOR, Alberto L. CORRALES, Ángel SÁNCHEZ, Juan José PANTRIGO
Dpto. Informática, Estadística y Telemática, Universidad Rey Juan Carlos
Móstoles, Madrid 28933, España

RESUMEN

La transformación de información de un medio a otro ha sido, y sigue siendo, un tema de investigación activo. Dentro de este campo podemos incluir la sonificación, que se define como la transformación de relaciones de datos en relaciones acústicas con el propósito de facilitar la comunicación y la interpretación. En los últimos años, y gracias a las facilidades multimedia que han aportado las nuevas tecnologías, la sonificación de datos se ha potenciado en gran medida. El acceso a la información de las personas con discapacidades visuales se ha visto favorecido por este motivo, aunque aún persisten grandes barreras. Actualmente, existen numerosas aplicaciones de lectura de textos en documentos o páginas web basados en sistemas OCR, así como del paso inverso, conversión de voz a texto. Sin embargo, aún no hay buenas soluciones para la percepción y el entendimiento de las imágenes. El optófono es el instrumento que permite sonificar una imagen o escena. En este trabajo se presenta un prototipo sobre el que se han experimentado diversas y muy conocidas técnicas de segmentación de imágenes. Por último, las técnicas aplicadas a la sonificación de imágenes en escala de grises se han extendido de forma preliminar para imágenes en color.

Palabras Clave: Sonificación, procesamiento de imágenes, segmentación, optófono, tiflotecnía.

1. INTRODUCCIÓN

La transformación de información de un medio a otro puede resultar de especial interés en casos en los que es difícil asimilar una determinada cantidad o ciertos tipos de datos. El uso de sonido no hablado (*nonspeech audio*) para tratar o canalizar información o, más específicamente, la transformación de relaciones de datos en relaciones acústicas con el propósito de facilitar la comunicación y la interpretación, se denomina sonificación [9]. Bajo este nombre, se engloban muchas técnicas multidisciplinares, que integran conceptos de la percepción humana, la acústica, las artes o el diseño, por citar algunos, y en la que tiene cabida a la acción combinada de profesionales de la informática, la física, la música o la psicología, entre otros. Gran parte de la investigación centrada en la sonificación se ha dirigido a establecer las bases teóricas de los umbrales auditivos, las escalas psicofísicas y los modelos de percepción auditivos. En particular, se ha insistido en el estudio de los parámetros involucrados, como la intensidad, la frecuencia y la discriminación temporal de sonidos estáticos [5, 11, 12], los determinantes de la tonalidad y volumen, los efectos de enmascaramiento [4], y propiedades de localización auditiva [1]. El tratamiento de grandes cantidades de información mediante la incorporación del sonido es un campo de investigación de innegable utilidad. El sonido es esencialmente una onda de presión propagada sobre un medio material. Las características del sonido son el volumen (directamente

relacionado con la amplitud de la onda), el tono (relacionado logarítmicamente con la frecuencia de la onda), el timbre, la localización, la duración y el ritmo [9]. En el contexto de la visualización de datos, el sonido puede ser utilizado para facilitar la interpretación de los mismos en función de las relaciones que se puedan establecer entre ellos [6].

Dentro de la sonificación podemos incluir una interesante aplicación: el optófono. El optófono es el instrumento que permite sonificar una imagen basando su sonido en características de ésta. El uso de este instrumento puede tener especial relevancia en personas con alguna discapacidad visual pues, en algunos casos, la percepción del entorno puede verse limitada a canales de recepción como el oído y el tacto.

Para el presente trabajo hemos desarrollado en software un optófono similar al del investigador Peter Meijer [10], que tiene en cuenta la intensidad de los niveles de gris y la posición de éstos en la imagen en escala de grises para realizar la sonificación. La finalidad de este prototipo es la de experimentar con diferentes métodos de segmentación de imágenes que resulten más apropiados para la sonificación. Asimismo, nos ha servido para poner a punto la metodología necesaria para la sonificación de imágenes a color, como explicaremos en la última sección. Es necesario tener en cuenta que una de las limitaciones más acusadas de la transferencia de información a través del sonido es la restricción de memoria que tiene una persona frente a un conjunto denso o continuo de información de este tipo. La sonificación de una imagen compuesta por muchos detalles no permitirá al usuario discernir lo superfluo de lo trascendente, restando funcionalidad al instrumento. Por este motivo, es de vital importancia la simplificación de la escena a sonificar.

A continuación, se describe la estructura del resto del trabajo. En la sección 2 se detalla la problemática de la sonificación de imágenes, y en concreto se presenta el prototipo de optófono implementado. En la sección 3, se discuten los aspectos derivados de la simplificación de las imágenes, describiendo los métodos de simplificación utilizados en este trabajo. En la sección 4 se presentan algunos resultados experimentales, y finalmente, en la sección 5 se muestran las conclusiones del presente trabajo y se reflexiona sobre futuros avances.

2. SONIFICACIÓN DE IMÁGENES

Los optófonos actuales generan sonido a partir de imágenes digitales en escala de grises basándose en las características visuales de la misma. Para ser más precisos, la imagen se sonifica por columnas, creando una onda sonora para cada uno de los píxeles que contiene. Para la sonificación de imágenes se debe establecer una estrecha correspondencia entre las características de la imagen y las propias del sonido. Una forma intuitiva de relacionar ambas es a través de lo que se conoce Transformación Piano que hace uso del volumen y el tono del sonido para representar los píxeles de la imagen. En esta

transformación se representa el nivel de intensidad de cada píxel a través del volumen y la posición espacial en la dirección vertical por el tono (frecuencia). De este modo, la intensidad de los píxeles es proporcional a la amplitud (volumen) del sonido que se genera. Por otra parte, los píxeles inferiores en altura generan ondas de frecuencia baja y viceversa.

Para cada columna se suman las ondas generadas por cada uno de sus píxeles y se hace un barrido de izquierda a derecha sonificando así todas las columnas y por tanto toda la imagen (Fig. 1).

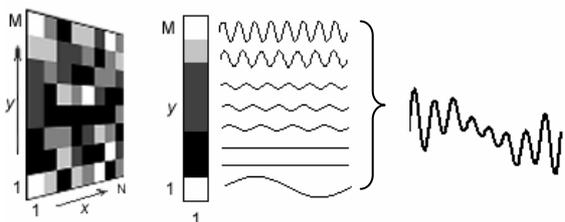


Fig. 1: Sonificación de la primera columna de una imagen sintética a partir de sus píxeles constituyentes. La suma de los sonidos generados para cada píxel es el sonido resultante para la columna.

La función sinusoidal utilizada para la generación de la onda sonora asociada a cada píxel viene dada por la Ec. (1).

$$S(t) = A(I(y)) \cdot \text{sen}(\omega(y) \cdot t) \quad (1)$$

siendo S el valor de la onda en cada instante t , A su amplitud, dependiente del nivel de intensidad del píxel, $I(y)$, y ω su frecuencia, dependiente de la altura del píxel en la columna, y .

Como ejemplo, en la figura 1 se observa que las ondas correspondientes al píxel blanco de las filas superior e inferior tienen la misma amplitud, sin embargo ambas difieren en su frecuencia. El sonido correspondiente a la primera columna se forma sumando la contribución de cada uno de sus píxeles. Para sonificar la imagen completa, se barren sucesivamente las columnas. La contribución a la onda de la columna de los dos píxeles blancos del ejemplo anterior viene expresada respectivamente por:

$$S_1(t) = A(255) \cdot \text{sen}(\omega(8) \cdot t) \quad (2)$$

$$S_2(t) = A(255) \cdot \text{sen}(\omega(1) \cdot t) \quad (3)$$

siendo 255 el nivel de gris correspondiente al color blanco y 1 y 8 las alturas respectivas de ambos píxeles. También es interesante observar en la misma figura que los píxeles negros generan ondas de amplitud igual a cero, sean cuales sean sus alturas, y por tanto sus frecuencias.

La función resultante para una columna x de la imagen se expresa a través de la suma de las contribuciones dadas por la

Ec. (1) de cada uno de los píxeles que contiene, llegando así a la Ec. (4).

$$S_T(x,t) = \sum_{y=1}^M A(I_{x,y}) \cdot \text{sen}(\omega_y \cdot t) \quad (4)$$

siendo $x=1..N$ la posición de la columna en la imagen de dimensiones $N \times M$, $y=1..M$ la posición del píxel en la columna, $I_{x,y}$ el nivel de gris del píxel situado en la posición (x, y) de la imagen, y ω_y la frecuencia asociada a la altura y . La adhesión sucesiva de todas las funciones suma correspondientes a las columnas formará finalmente el sonido de la imagen completa. De esta manera, la duración del sonido debido a la sonificación de toda la imagen será suma de las duraciones parciales de las ondas de todas las columnas.

En la práctica, el acoplamiento de las funciones correspondientes a las columnas para crear la sonificación de toda la imagen produce sonidos extremadamente complejos y sin aparente sentido. Esta complejidad se incrementa notablemente con imágenes grandes, compuestas por multitud de niveles de gris y con detalles finos. De aquí surge la necesidad de preprocesar la imagen original y simplificarla basándose en las características generadoras del sonido, es decir, el número de niveles de gris de sus píxeles y el tamaño de la misma. En la siguiente sección se abordarán los diversos tratamientos que se han elegido para realizar los experimentos.

El sonido creado tras el barrido de la imagen completa representará de una manera sonora la propia imagen y llevará toda la información que en principio pudiera contener en su representación habitual. Por este motivo, y aunque “escuchar” una imagen no sea una tarea sencilla, se puede llegar a entenderla de una forma grosera. Esta aproximación al significado de la escena puede tener especial importancia en personas con discapacidades visuales, y en concreto en casos de cegueras parciales o totales.

3. PROCESAMIENTO DE IMÁGENES

En la sección anterior se ha justificado que las imágenes con exceso de detalle, múltiples niveles de gris o de gran tamaño generan sumas de ondas cuya sonificación es potencialmente compleja. Por tanto, es necesaria una labor de simplificación que se lleva a cabo en una etapa de preprocesamiento. Esta sección describe algunos de los métodos de simplificación de imágenes utilizados.

Umbralización

Una manera de disminuir drásticamente el número de niveles de gris de una imagen consiste en aplicar una umbralización. Mediante una umbralización simple y global dividimos el histograma de la imagen usando un umbral único. Los píxeles cuyo nivel de intensidad superan este umbral se transforman a blanco y el resto a negro. Este proceso produce la binarización de la imagen. En la práctica, este método sólo será aplicable para imágenes tomadas en entornos altamente controlados [4]. En efecto, aplicar la umbralización puede provocar un exceso de píxeles blancos o negros dependiendo de los niveles de gris originarios en la imagen y el umbral escogido (Fig. 2). Por esta razón, es preferible utilizar la umbralización junto con algún método añadido de segmentación.

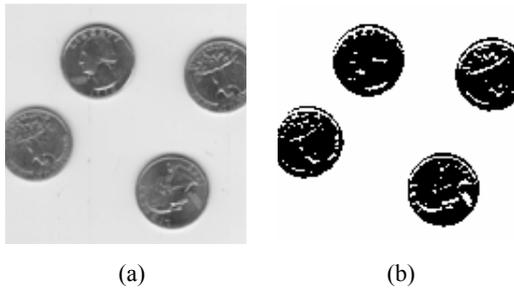


Fig. 2: a) imagen original; b) imagen umbralizada

Detección de Bordes

Una cualidad que puede interesar a la hora de sonificar una imagen es la distinción de las siluetas de los objetos presentes en la escena. Esta interpretación se puede conseguir mediante una detección de bordes. La detección de bordes en imágenes en escala de grises es una de las técnicas de segmentación más utilizada y, con mucho, la aproximación más común para detectar discontinuidades importantes [3]. Consiste en destacar los píxeles correspondientes a las fronteras entre regiones de niveles de gris diferentes. Por esta razón, se consigue una reducción en el número de píxeles principales, pues pasan a ser tan sólo los ubicados en las fronteras de regiones homogéneas (Fig. 3).

Así pues, la idea subyacente en la mayoría de las técnicas de detección de bordes es el cálculo de un operador de derivada local. En muchos casos se adoptan aproximaciones al operador gradiente a través de convoluciones con determinadas máscaras de convolución. Ejemplos de estas máscaras son, entre otras, las de Sobel, de Prewitt o de Roberts. En este trabajo se ha escogido la de Sobel por ser ampliamente utilizada en la bibliografía.

Para evitar una sobredetección causada por el ruido o imperfecciones de la imagen se han incluido filtros de suavizado. La aplicación de estos filtros es opcional aunque recomendada y, en concreto, se han implementado los filtros de media local y gaussiana.

La manera más habitual de proceder tras la detección de bordes consiste en umbralizar la imagen. De esta manera, evitamos introducir niveles de amplitud muy diferentes acotándolos a tan solo dos.

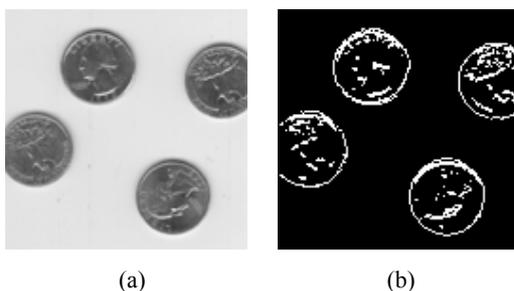


Fig. 3: a) Imagen original; b) imagen de bordes correspondiente

Reducción de Resolución Espacial

Otra forma directa y efectiva de disminuir la cantidad de información a procesar consiste en reducir espacialmente de la imagen original. De este modo, el número de píxeles disminuye y por tanto el número de sonidos a crear también. El factor clave a tener en cuenta en la reducción del tamaño de una imagen es el método de interpolación que se utiliza para que la imagen no pierda contenido semántico. Los métodos de interpolación más simples y utilizados son los de interpolación bilineal, bicúbica y por vecindad. Los dos primeros ponderan los niveles de intensidad de los píxeles vecinos y producen un nuevo valor de gris, diferente en general de los originales del entorno. La interpolación por vecindad es más simple y no añade nuevos valores de intensidad, sino que utiliza los pertenecientes a la propia imagen. Para la sonificación de imágenes es importante no mezclar multitud de sonidos de características diferentes y por esta razón es preferible la elección de una interpolación por vecindad.

Cuantización de Niveles de Intensidad

Las pequeñas diferencias en amplitud de un sonido se perciben peor que las diferencias en otras características como por ejemplo las del tono [9]. La reducción del número de posibles amplitudes hace aumentar las diferencias entre ellas y posibilita discernirlas mejor. De la Ec. (1) se observa que el número de posibles amplitudes está íntimamente ligado con el número de niveles de gris en la imagen. Se deduce fácilmente que la reducción del número de niveles de gris de la imagen implica una simplificación deseable en el resultado final de la sonificación.

Para que la reducción de niveles de intensidad tenga algún efecto notable en el sonido generado es deseable que el número de niveles de gris de la imagen final no sea mayor de 16, incluso menos si ésta se compone de finos detalles (Fig. 4).

4. RESULTADOS EXPERIMENTALES

Ya se ha mencionado que para sonificar una imagen se toma una correspondencia entre la altura del píxel y la frecuencia de la onda que se genera. Esta correspondencia se basa en una escala de frecuencias que se debe tomar. En nuestro prototipo (Fig. 5) se han implementado varias escalas ya utilizadas en el trabajo del profesor Peter Meijer, como la escala lineal, exponencial, la de Mel o la de Bark [10]. Sin embargo, hemos creído oportuno añadir entre las opciones de sonificación la escala musical natural. Creemos que la escala natural musical puede expresar de modo más intuitivo la geometría expuesta por los píxeles de la imagen. Nuestra hipótesis es que una persona occidental tiene el oído culturalmente educado de tal manera que dos notas musicales consecutivas que no guardan la misma distancia en frecuencia que otro par consecutivo, son interpretadas como equidistantes. Este hecho es fácilmente demostrable. Definimos un salto unitario como el intervalo entre una nota musical y la siguiente. De esta forma, a pesar de que los intervalos Mi4-Fa4 y Fa4-Sol4 son de un semitono y un tono respectivamente, según nuestra definición corresponderían ambos a un salto unitario (Fig 6b). Para este ejemplo, las distancias en frecuencia de ambos intervalos son de 20 y 43 Hz respectivamente, lo que contrasta con las demás escalas de frecuencias utilizadas, principalmente la lineal (Fig. 6a) pues queda patente que la escala natural es definitivamente no lineal.

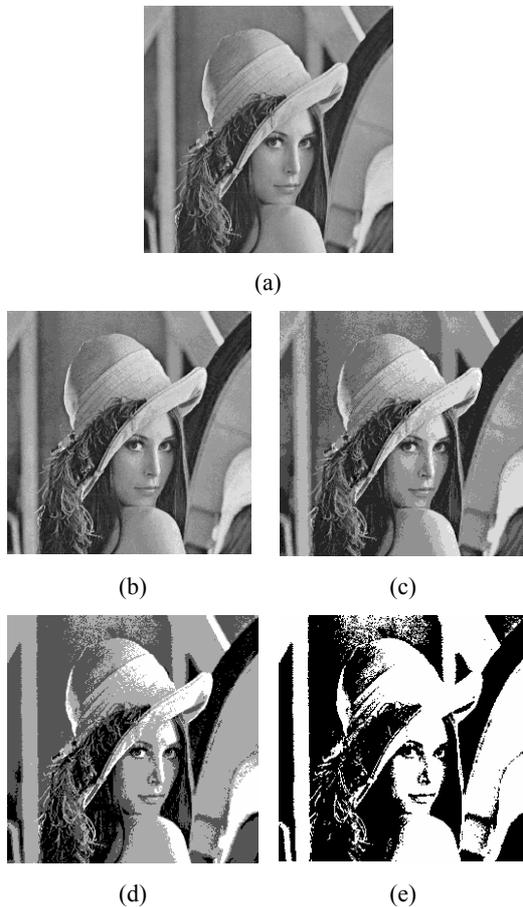


Fig. 4: a) Imagen original, 256 niveles de gris; b) imagen cuantizada a 16 niveles de gris; c) a 8 niveles de gris; d) a 4 niveles de gris; e) a 2 niveles de gris

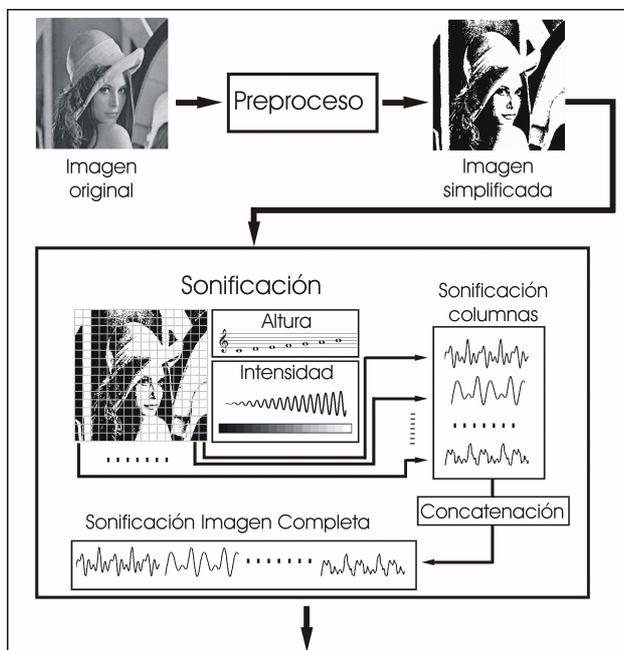


Fig. 5: Prototipo de sonificador.

La escala natural la empleamos fundamentalmente para imágenes de dimensiones pequeñas pues tratamos de no utilizar más notas que las que contenga una octava. Sin embargo, otra escala que podemos aplicar es la cromática para la que nuestro salto unitario se corresponde con un semitono. Esta escala es útil para imágenes algo más grandes, pues en una única octava tenemos hasta 12 sonidos diferentes (Fig 6c).

En primera aproximación y para comprobar la bondad de la sonificación se han hecho pruebas con nuestro prototipo y con imágenes sintéticas en escala de grises generadas externamente. En la Fig. 7 se muestran dos imágenes que resultan especialmente útiles para experimentar con las escalas de frecuencias y con la dependencia de la amplitud de las ondas con los niveles de gris de los píxeles. Una diagonal cuyos píxeles son de igual nivel de intensidad en un fondo negro interesará para discernir posibles diferencias entre las escalas de frecuencias utilizadas para sonificar cada uno de sus píxeles (Fig. 6a). Una recta horizontal cuyos píxeles tengan niveles de gris variables, de oscuros a claros, nos mostrará la dependencia de las amplitudes de las ondas generadas con los niveles de intensidad de sus píxeles manteniendo fijas las frecuencias (Fig. 7b).

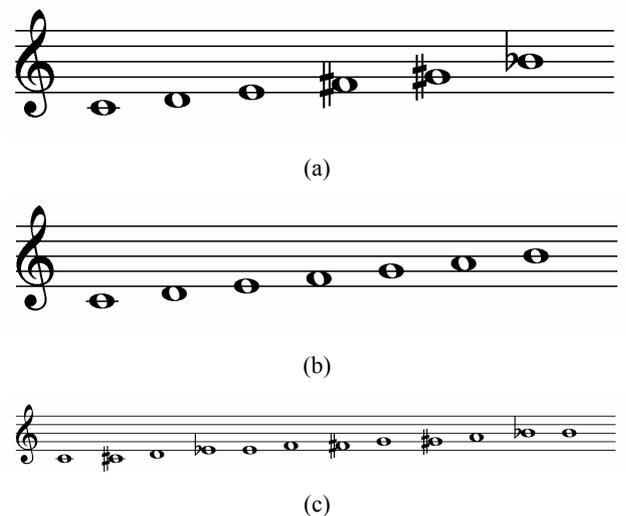


Fig. 6: a) Escala de tonos completos; b) escala natural; c) escala cromática

Tras las pruebas con las imágenes sintéticas se han realizado experimentos con imágenes más realistas aún en escala de grises. Para ellas se aplica un preprocesamiento con los métodos descritos en la sección anterior. En concreto, la detección de bordes es muy determinante en la extracción de las propiedades geométricas de los objetos presentes en la escena. Se puede aplicar la técnica de Sobel sobre la imagen directamente (Fig. 8a), aunque se obtienen mejores resultados si antes se suaviza con algún método de filtrado como el gaussiano (Fig. 8b). Incluso después de la simplificación obtenida tras este proceso, si la imagen final es suficientemente grande, se reducirá su tamaño a uno más práctico. El tamaño final deberá guardar un compromiso entre su definición y el sonido resultante del proceso de sonificación. En nuestros experimentos este tamaño se ha limitado a 32x32 o 64x64 píxeles para imágenes cuadradas, aunque a priori no está restringido (Fig. 8c, 8d).

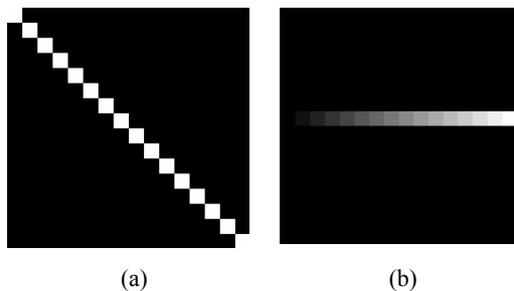


Fig. 7: a) Diagonal principal con píxeles blancos en fondo negro; b) recta horizontal con píxeles en escala de grises en fondo negro

La heurística adoptada consiste en suavizar en primer lugar la imagen inicial y posteriormente realizar la detección de bordes mediante el método ya descrito. La imagen de bordes resultante (Fig. 9a) se puede binarizar por umbralización. De este modo se obtiene una imagen con dos únicos niveles de intensidad: blanco o negro (Fig. 9b). Sin embargo, mediante esta práctica evitamos trabajar con múltiples niveles de gris, perdiendo uno de los dos grados de libertad de nuestro sonido estructurado, el volumen.

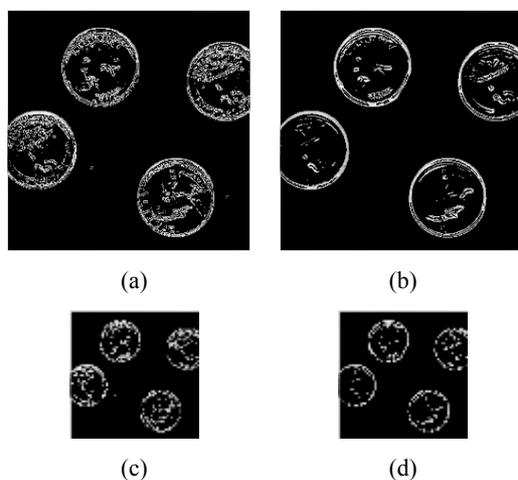


Fig. 8: a) Detección de bordes sin suavizado previo; b) detección de bordes con un suavizado gaussiano previo; c) y d) imágenes a) y b) respectivamente, redimensionadas.

Una posibilidad más integradora consiste en utilizar el umbral tan solo para transformar a negro aquellos píxeles cuyos niveles sean inferiores a dicho valor y dejar intactos los superiores, lo que hemos llamado detección de bordes umbralizada en lugar de binarizada (Fig. 9c). En este caso, la imagen resultante contendrá múltiples niveles de intensidad y, en general, tendrá la posibilidad de crear diversas amplitudes de onda. Este procesamiento admite mejoras utilizando una equalización del histograma, redistribuyendo los niveles de intensidad supervivientes en nuestro intervalo de valores permitidos y ampliando las diferencias en los niveles de gris, con la consiguiente ampliación de las diferencias en las amplitudes de las ondas. Por último, si además de esta equalización, contamos con una reducción de los niveles de gris resultantes, obtendremos lo que se ha llamado imagen de bordes umbralizada, equalizada y cuantizada (Fig. 9d). Esta técnica es

la que mejor mantiene el compromiso entre las dos características sonoras y la dificultad de interpretación de la sonificación.

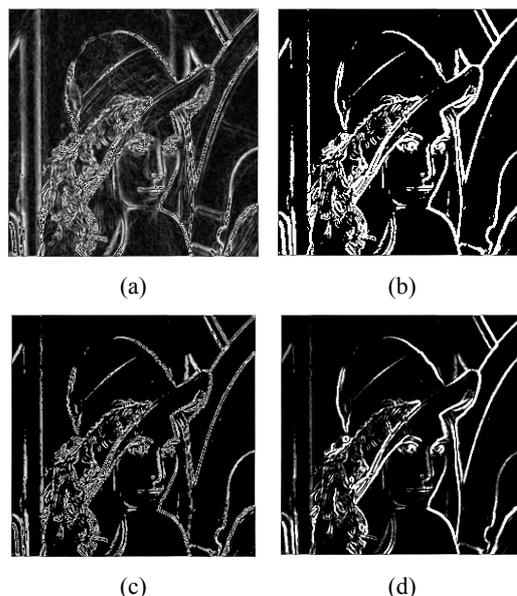


Fig. 9: a) Imagen de bordes; b) imagen de bordes binarizada; c) imagen de bordes umbralizada, equalizada y cuantizada

Para comprobar la bondad de la sonificación se han creado baterías de pruebas con imágenes sintéticas y se ha experimentado con 28 voluntarios sin discapacidad visual. El porcentaje de personas que consigue identificar más de un 50% de las imágenes presentadas en forma sonora llega hasta el 71% de los casos y un 21% de las personas entrevistadas identificaron más del 75% de los objetos sonificados. Teniendo en cuenta que los voluntarios no habían sido entrenados previamente consideramos que estos resultados son altamente satisfactorios. Sin embargo, nuestra hipótesis respecto al uso de las escalas musicales natural y cromática no ha podido ser contrastada estadísticamente hasta el momento de la escritura de este trabajo.

Por último, también se han realizado experimentos preliminares de sonificación de imágenes a color. Cada color de una imagen en el espacio RGB (red – green – blue) está compuesto por la suma de estos tres colores básicos (también llamados canales) en una determinada proporción. Para sonificar una imagen a color, se han asociado diferentes timbres sonoros (es decir, diferentes instrumentos) a cada uno de los canales. De esta forma, los colores rojo, verde y azul puro, producirán un sonido con un timbre puro. Los colores formados por la composición de estos tres colores básicos (blanco, amarillo, magenta, etc.), producen sonidos en los cuales están mezclados los timbres de los tres instrumentos puros en diferentes proporciones. Es importante notar que, si la tarea de interpretación de un sonido derivado del proceso de sonificación de imágenes en escala de grises es compleja, la correspondiente a la fusión de sonidos resultantes de varios instrumentos es aún mayor. Por esta razón, la simplificación de imágenes a color para su posterior sonificación, es aún más necesaria. En este sentido, se han llevado a cabo pruebas sobre son para imágenes simplificadas de banderas de pocos colores y sin estructuras internas, tales

como escudos o motivos geométricos. La simplificación consiste en tomar muestras representativas de cada uno de los colores que forman la bandera y componerlos sobre fondo negro (Fig. 10). Con esta transformación, se consigue simplificar el sonido resultante de la sonificación, debido a que el color negro no produce ningún sonido.

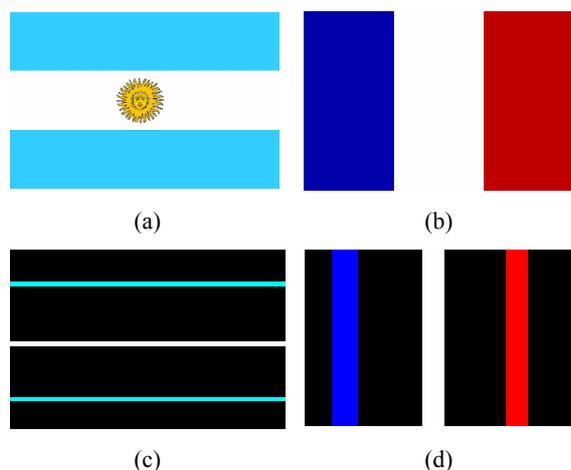


Fig. 10: a) y b) Imágenes originales de las banderas de Argentina y Francia; c) y d) Simplificación de a) y b) respectivamente.

5. CONCLUSIONES Y FUTUROS TRABAJOS

Este trabajo presenta un prototipo de optófono que permite sonificar imágenes digitales en niveles de gris utilizando los niveles de intensidad y la distribución espacial de sus píxeles. La mayor dificultad en la sonificación de imágenes es el exceso de información que contienen para describir una escena, fundamentalmente dado por el número y distribución de los píxeles. Por esta razón, aplicamos conocidos métodos de segmentación y simplificación de imágenes. Los resultados preliminares son satisfactorios y las pruebas realizadas sobre diferentes individuos indican que la asimilación de información sonora, a pesar de no estar exenta de dificultad, mantiene un grado de precisión acorde con las expectativas creadas. De este modo, el oyente consigue formarse una grosera idea de dicha escena, sin llegar a extraer finos detalles de la imagen.

No obstante, y aunque el presente trabajo sólo tiene en cuenta el volumen y el tono como características de sonido, se le ha dado una extensión natural al proceso de sonificación. Tradicionalmente, la sonificación sólo se ha aplicado a imágenes en escala de grises para, precisamente, simplificar las escenas. Sin embargo, hemos creído oportuno incorporar el color como característica a sonificar por la relevancia que tiene en el contexto de las imágenes. Para introducir el color en el proceso de sonificación se propone el uso del timbre del instrumento. El timbre es una característica musical determinante e identificativa del sonido de un instrumento, y es consecuencia directa de la forma, los materiales y el elemento que origina el sonido. Dado un sonido con una frecuencia principal, la distribución de intensidad de sus armónicos y parciales genera un timbre u otro [8]. De esta manera, se introduce en paralelo la información no sólo de la distribución espacial de los píxeles de la imagen sino también un contenido global a partir de su homogeneidad de color. Para la consecución de este objetivo se ha utilizado una librería MIDI

(*Musical Instrument Digital Interface*). La gran cantidad de información que poseen las imágenes en color realistas hace inviable la clara definición de los sonidos generados y pensamos que su uso sólo puede ser factible para entornos muy controlados, como la identificación de formas o motivos geométricos. Las imágenes sobre las que se ha probado este último prototipo son esencialmente imágenes sintéticas muy simplificadas como banderas o líneas de colores. Para los experimentos realizados se ha hecho una correspondencia entre cada uno de los canales de color (RGB) y tres instrumentos pertenecientes a la librería estándar MIDI (*piano, rock organ y tubular bells*). Creemos que esta idea puede ser mejorada en muchos aspectos y aunque por el momento es muy básica podría ser de gran utilidad para la visualización de gráficas de líneas donde se representan varios conjuntos de datos diferenciados en colores, especialmente para personas invidentes parciales o totales.

La correspondencia entre colores y timbres sonoros es una tarea no trivial y como ampliación se podría hacer uso de un sistema de inferencia borroso, pues el color como característica global de una imagen tiene un alto grado de indeterminación. Hay múltiples variantes para establecer esta correspondencia, entre ellas se pueden destacar la realización del proceso de sonificación con el instrumento cuyo color es mayoritario en la imagen o utilizar instrumentos más estridentes en función de las altas frecuencias de la imagen (mayores contrastes). El uso de combinación de notas musicales para la formación de acordes también podría explotarse en el proceso de sonificación.

Por otro lado, nos parece interesante explorar la sonificación de imágenes a color expresadas en espacios de color diferentes al RGB, como el HSI, YCrCb, etc. En estos espacios, se separa la componente relacionada con la Iluminancia de la imagen de las características del color, de forma que éste queda descrito únicamente por dos componentes (Saturación y Tono en el HSI o Crominancias en el YCrCb). Esperamos que la sonificación de la imagen representada en estos espacios sea más intuitiva y produzca menos estrés auditivo.

Otro punto clave que queda como futuro trabajo es la experimentación de estos métodos de sonificación sobre personas con discapacidades visuales, entrando en contacto con las instituciones pertinentes.

Por último, cabría destacar el futuro uso de este prototipo como sonificador de secuencias de imágenes. Debemos asumir que la sonificación periódica de imágenes generaría un excesivo estrés para el oyente mermando así las posibilidades prácticas de aplicación del optófono. Siendo conscientes de esta limitación se ve necesaria la adopción de más potentes técnicas de reducción de información.

REFERENCIAS

- [1] Blauert, J., **Spatial Hearing: The Psychophysics of Human Sound Location**. MIT Press, 1997.
- [2] Flowers, J.H., Buhman, D.C., Turnage, K. D. "Cross-modal Equivalence of Visual and Auditory Scatterplots for Exploring Bivariate Data Samples", **Human Factors**, 39, 1996, pp. 341-351.
- [3] Gonzalez, R. C., Woods, R. E., **Digital Image Processing**, Second Edition, Prentice-Hall, 2002.

- [4] Gulick, W. L., Gescheider, G. A., Frisina, R. D., **Hearing: Physiological Acoustics, Neural Coding, and Psychoacoustics**, Oxford University Press, 1989.
- [5] Hartmann, W.M., **Sounds, Signals, and Sensation: Modern Acoustics and Signal Processing**. Springer Verlag: New York, 1997.
- [6] Kramer, G. "Some Organizing Principles for Representing Data with Sound". **Proceeding of the First International Conference on Auditory Display (ICAD)**, 1994, pp. 185-221.
- [7] Kramer, G. "Auditory Display: Sonification, audification, and auditory interfaces". **Proceeding of the First International Conference on Auditory Display (ICAD)**, 1994.
- [8] López, M. R., **Ingeniería Acústica**, Ed. Paraninfo, 2000.
- [9] Madhyastha T. M., Reed, D. A., "Data Sonification: Do You See What I Hear?", **IEEE Software**, Vol. 12, No. 2, 1995, pp. 85-90.
- [10] Meijer, P.B.L, "An Experimental System for Auditory Image Representations", **IEEE Transactions on Biomedical Engineering**, Vol. 39, No. 2, 1992, pp. 112-121.
- [11] Moore, B. C. J., **Handbook of Perception and Cognition**, Vol 6, Hearing, Academic Press, 1995.
- [12] Moore, B. C. J., **An Introduction to the Psychology of Hearing**, 4th ed., Academic Press, 1997.