SISTEMA DE WATERMARKING DE AUDIO

Instituto de Ingeniería Eléctrica
Facultad de Ingeniería
Universidad de la República Oriental del Uruguay

Fecha de aprobación: 9 de Marzo de 2007
Carrera: Ingeniería Eléctrica
Plan: 97

Juan Enrique Artagaveytia
José Luis Barattini
Bernardo Brum

Tutor responsable: Juan Pechiar

Índice de contenido

1 - Introducción

El uso de “marcas de agua” como sistema de protección es casi tan antiguo como la fabricación de papel. Durante cientos de años, cualquiera que poseyera o fabricase un documento u obra de arte valiosa lo marcaba con un sello de identificación o marca de agua (visible o no), no sólo para establecer su propiedad, origen o autenticidad, sino para desalentar a aquellos que pudieran intentar robarlo.

Por lo general dicha marca debe ser imperceptible de manera que permanezca oculta. La ciencia que estudia cómo transmitir información secreta sin que sea detectada se denomina esteganografía (del griego escritura secreta). No es una ciencia nueva; pueden rastrearse antecedentes incluso entre los antiguos griegos.

Algunos ejemplos de las técnicas de esteganografía que han sido usados en la historia son:

- Mensajes ocultos en las tabletas de cera en la antigua Grecia, la gente escribía mensajes en una tabla de madera y después la cubrían con cera para que pareciera que no había sido usada.

- Mensajes secretos en papel, escritos con tintas invisibles entre líneas o en las partes en blanco de los mensajes.

- Durante la segunda guerra mundial, agentes de espionaje usaban micro-puntos para mandar información. Los puntos eran extremadamente pequeños comparados con los de una letra de maquina de escribir por lo que en un punto se podía incluir todo un mensaje.

- Mensajes escritos en un cinturón enrollado en un bastón, de forma que solo el diámetro adecuado revela el mensaje.

Las técnicas de marcas de agua son utilizadas para la autenticación (tanto del distribuidor o propietario legal, como de que el original no ha sido falsificado) de la información, así como para el seguimiento de copias, ya que permiten la identificación del autor, propietario, distribuidor y/o consumidor autorizado de un documento.

Como técnica empleada en documentos digitalizados surge inicialmente para imágenes a partir de los años ochenta. Posteriormente a partir de la década del noventa comienzan a surgir para audio digital.

Clasificación de las técnicas de watermarking

Las técnicas de watermarking se clasifican en:

1. Non Blind watermarking
2. Blind watermarking

En las técnicas non blind watermarking el receptor necesita procesar el audio original (sin la marca) por lo que son teóricamente interesantes, pero no tan útiles en la práctica, ya que requiere el doble de memoria y el doble de ancho de banda para la detección de la marca de agua. Como ejemplo podemos encontrar el algoritmo de watermarking de fase.

Mediante las técnicas blind watermarking se pueden detectar y extraer la marca de agua sin utilizar la señal de audio original. Por lo tanto, requiere solamente la mitad de memoria comparado con Non Blind Watermarking. Estos algoritmos son los más utilizados y por lo tanto le vamos a dar más importancia. Los ejemplos son: LSB, Spread Spectrum, Patchwork y Echo Hiding

Características de las marcas de agua

Una característica muy importante de la marca de agua es que esté incluida en la misma señal de audio. No en el encabezado del archivo de audio, ni en un flujo de bits independiente, ni en otro archivo.

Además es necesario que la marca se repita a lo largo de todo el fragmento de audio, de manera que si alguna marca no se llegara a detectar, existen otras marcas que si podrán ser detectadas. Esto mejora la redundancia y la robustez del sistema.

Imperceptibilidad

La *imperceptibilidad* o *transparencia* de la marca tiene como base el comportamiento del sistema perceptual humano. Una marca de agua es imperceptible, si la degradación que causa en el audio donde se ha insertado es muy difícil de apreciar. Para conseguir este objetivo, la mayoría de los sistemas utilizan una etapa de análisis psicoacústico. Los modelos psicoacústicos son popularmente conocidos por la compresión perceptual del audio (por ejemplo el sistema de compresión “mp3”).

Robustez

Un sistema absolutamente transparente es de poca utilidad si la marca es altamente volátil y no permanece luego de realizar manipulaciones sobre el audio marcado (agregado de ruido blanco, conversión D/A – A/D, compresión mp3, conversión de la frecuencia de muestreo, etc.)

Una marca de agua se considera robusta si perdura luego de realizadas dichas manipulaciones. En el caso que la marca fuera removida, el sistema sería considerado robusto si dicha eliminación disminuyera apreciablemente la calidad del audio.

Largo de la marca de agua

La cantidad de la información a enviar está fuertemente vinculada con la aplicación que se va a utilizar. Típicamente la mayoría de las aplicaciones no requieren grandes largos de marca. Por ejemplo, para almacenar el código de país, código de propietario, el año de la grabación, y un número de serie alcanzarían con 60 bits.

Tasa de bits

Como en la mayoría de los sistemas de comunicación, un sistema de watermarking de audio destinará parte de los bits totales transmitidos para técnicas de detección o corrección de errores. El resto de los bits serán la carga útil. La velocidad de transmisión de los bits totales es del orden de los 100 bps.

Compromiso entre los parámetros más importantes

Los requerimientos más importantes del watermarking son imperceptibilidad, robustez y la taza de bits. Una forma fácil de ver visualmente estos parámetros es a través del esquema de la figura 1.1:

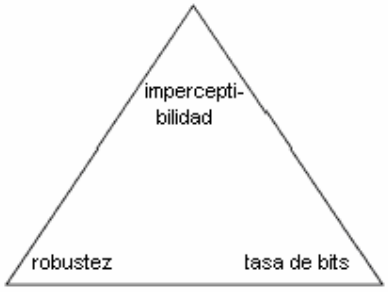


Figura 1.1 – Compromiso entre robustez, imperceptibilidad, y tasa de bits.

La figura 1.1 representa el compromiso entre la taza de datos y la robustez a ciertos ataques, mientras que conserva la imperceptibilidad en la marca de agua en un nivel aceptable.

A continuación se detallan las distintas aplicaciones para el caso del audio digital.

1.1 - Aplicaciones de watermarking

Transmisión oculta de información

A partir de lo comentado anteriormente, una de las posibles aplicaciones es la transmisión de información de manera oculta en el audio digital. Dado que la marca de audio no es perceptible, no puede ser escuchada. Sin embargo, el receptor del audio con el método adecuado puede recuperar la marca.

Identificación de Contenido

Muchos problemas que van desde la prueba de propiedad (proof of ownership), y copyright, hasta el monitoreo de uso y seguimiento de regalías (pago que se realiza al titular de derechos de autor, patentes o marcas registradas, a cambio del derecho de usarlos) utilizan marcas de agua.

Tales marcas de agua serían insertadas en el audio en la última etapa de producción, asegurando de esta manera que todas las copias siguientes queden marcadas con la misma identificación.

El monitoreo del uso de una canción (ya sea para propósitos de regalías u otros) se basará típicamente en el despliegue de estaciones de monitoreo. Estas estaciones deberían poder extraer los identificadores de contenido de todo el audio marcado, independientemente de sus fuentes. Por consiguiente el formato apropiado para este caso, para las marcas de agua, no debería ser propietario.

Identificación de Transacción o de Número de Serie

Las marcas de agua pueden también servir para la trazabilidad de una cadena de distribución específica. Por ejemplo, un distribuidor de contenido puede elegir ponerle un número de serie a copias con idéntico contenido pero que van dirigidas a distintos clientes, de manera de identificar la fuente (distribuidor) y el destinatario de esa copia específica. Este procedimiento no evitaría la posibilidad de que el cliente pudiera realizar copias ilegales del contenido, sin embargo serviría como una herramienta que podría rastrear todas las copias ilegales hasta el cliente original.

No hay un incentivo para que los distribuidores de contenido se pongan de acuerdo en una única estructura para dichas marcas de agua. En realidad, muchos de ellos desean mantener un completo control sobre el formato de la marca. Es por esta razón que los formatos propietarios de marcas de agua son los preferidos para este tipo de aplicación.

Control de Uso

Cuando el poder disuasivo de herramientas que actúan luego de cometido el ilícito no es suficiente, las marcas de agua pueden también servir para soluciones de seguridad más activas. La idea es usar el canal de datos de la marca de agua como un medio para transmitir reglas de uso para componentes y dispositivos (reproductores, grabadores, o cualquier otro componente diseñado para trabajar con audio), permitiendo o no la reproducción o grabación de determinado contenido.

Las marcas de agua para el control de uso deberían poder ser extraídas por una cantidad potencialmente grande de dispositivos compatibles. Es por este motivo que dichas marcas de agua no pueden ser propietarias. El agregado de las marcas puede llevarse a cabo en varias etapas de la cadena de distribución.

Identificadores de audio

Mediante esta aplicación las señales de audio de difusión se transforman en medios a través de los cuales se obtiene información para realizar transacciones. Mediante dispositivos capaces de extraer información de la marca insertada por un distribuidor, el dispositivo puede descifrar un identificador con el cual, a través de consultas en una base de datos, obtiene información adicional. Esta información puede consistir en: nombre del grupo, sitio del grupo en la web, próximos conciertos, posibilidad de adquirir el tema en alta fidelidad, etc.

SMDI

La Iniciativa de Música Digital Segura (SMDI por sus siglas en inglés) fue un foro creado en 1998, en el cual intervinieron más de 200 empresas, con el propósito de desarrollar especificaciones tecnológicas que protegieran la reproducción, almacenamiento, y distribución de música digital.

SDMI fue principalmente conocido por el desafío público SMDI que fue propuesto en el año 2000, para ver si alguien podía vulnerar el sistema. Los resultados del desafío fueron que los sistemas de protección fueron vulnerados ampliamente. La iniciativa SDMI ha estado inactiva desde mayo de 2001.

DRM

La Gestión de Derechos Digitales (DRM) es un conjunto tecnologías orientadas a ejercer restricciones sobre los usuarios de un sistema informático, o a forzar los derechos digitales permitidos por comisión de los poseedores de derechos de autor. Si bien dicha tecnología incluye varios tipos de contenido digital, dentro de ellos se encuentra el audio digital.

2 - Técnicas de watermarking de audio

En este capitulo se analizan de forma esquemática distintos métodos de watermarking de audio. En dichos análisis se consideran las ventajas y desventajas de cada uno, con el fin de seleccionar el que se implementará en una segunda etapa.

2.1 – Estudio de las técnicas existentes

2.1.1 - Método de codificación del LSB

Unas de las primeras técnicas estudiadas en el campo del watermarking de audio digital es la codificación del bit menos significativo (LSB). La idea aplicada a las secuencias de audio digital es la introducción de datos de marca por medio de la alteración de muestras individuales de la señal de audio original.

El codificador de este sistema de watermarking utiliza un subconjunto de todas las muestras de audio, x, disponibles elegidas por medio de una clave secreta. A las muestras de este subconjunto se le aplica la operación de sustitución $x_j[i] \rightarrow m[i]$ en los LSB (donde m es el conjunto de bits del mensaje de watermarking). El proceso de extracción recupera la marca leyendo el valor de estos bits conociendo qué muestras fueron utilizadas para la codificación. Generalmente el tamaño de este subconjunto es mucho menor al de toda la señal de audio por lo que la robustez de este método se puede mejorar repitiendo la marca en otros subconjuntos.

La modificación de los LSB de las muestras usadas para la introducción de la marca le agrega a la señal original un ruido que se puede modelar como ruido blanco uniforme aditivo (AWUN) de baja potencia. Teniendo en cuenta que el sistema auditivo humano (HAS) es muy sensible a este AWUN y que este método generalmente no usa modelos psicoacústicos para evitar que el ruido introducido en reemplazo de los LSB sea percibido, existe una limitación importante en el número de LSB que pueden ser modificados de forma imperceptible.

La principal desventaja de este método es su extremadamente baja robustez debido a que cambios aleatorios en los LSB destruyen la marca codificada por lo que no se puede aplicar a señales que serán recodificadas (transformadas a analógicas y luego nuevamente a digital). Por otro lado, dado que la transformación de la señal original no requiere demasiado procesamiento computacional, este algoritmo tiene muy bajo retardo, lo que lo hace apto para aplicaciones en tiempo real.

2.1.2 - Watermarking en la fase de la señal original

Los algoritmos que realizan watermarking modificando la fase de la señal sacan provecho de otra característica del HAS que es la insensibilidad a

corrimientos relativos constantes en la fase de una señal de audio estacionaria. Existen dos métodos usados en el watermarking de fase de una señal de audio: codificación de fase y modulación de fase.

La codificación de fase divide toda la señal de audio en bloques e introduce toda la marca en la fase del espectro del primer bloque. La señal original c_0 se divide en $M = [l(c_0)/N]$ bloques $c_{0j}, \quad 0 \le j \le M-1$ con $N = 2l(m)$ muestras. Se define l(m) largo de la marca en muestras y l(c_0) largo de la señal original en muestras.

1. A cada bloque de c_0 se le aplica la transformada de Fourier $C_{0j} = F\{c_{0j}\}, \forall j$. A partir de esto se construye una matriz con las fases $\Phi_{0j}[w_k]$ y las magnitudes $|A_{0j}[w_k]|, 0 \le k \le N/2-1$

2. Se calcula una nueva matriz con las diferencias de fase entre los M bloques vecinos: $\Delta\Phi_{0j+1}[w_k] = \Phi_{0j+1}[w_k] - \Phi_{0j}[w_k], \forall j,k$

3. La marca es codificada en la fase del espectro del primer bloque:
   $\Phi_{w0}[w_k] = (-1)^{m[k]+1}\,\pi/2$, con $m[k] \in \{0,1\}, 0 \le k \le N/2-1$

4. Con el objetivo de garantizar la inaudibilidad de los cambios de fase entre los bloques, las diferencias de fase de cada bloque deberán ser ajustadas.
   $\Phi_{wj+1}[w_k] = \Phi_{wj}[w_k] - \Delta\Phi_{0j+1}[w_k], \forall j,k$

5. Las magnitudes originales $|A_0|$ y la fase modificada del espectro Φ_w de los bloques se utilizan para obtener la señal marcada en el tiempo
   $c_{wj} = F^{-1}\{C_{wj}\}, \forall j$

Para la decodificación de la marca es necesario un pre-procesamiento de la señal para sincronizarse con el comienzo de la primera secuencia. Para esto es necesario tener en el receptor el conocimiento del largo de la marca l(m). El procedimiento para la detección consistirá en:

1. Sincronización con el primer bloque c_{w0},

2. Transformación del bloque: $C_{w0} = F\{c_{w0}\}$,

3. Lectura de los bits de la marca mediante la información de la fase del primer bloque $\Phi_{w0}[w_k], para \le k \le N-1$.

Una desventaja del método de codificación de fase es que lleva una muy baja información agregada (watermark) dado que sólo el primer bloque es usado para introducir la marca. Otro inconveniente es que dado que la marca no se encuentra distribuida sobre toda la señal de audio, sino que se encuentra localizada en un bloque específico, ésta puede ser más fácilmente removida.

Otra forma de incluir la información de la marca en la fase es realizando modulaciones de fase multibanda independiente. Modificaciones de fase inaudibles se utilizan en este algoritmo por medio de alteraciones de fase multibanda controladas de la señal original. El audio original c_o se segmenta en M bloques, $0 \le m \le M-1$ con $M = \frac{l(c_0)-N}{2N}$, usando ventanas solapadas. Siendo *N* el número de muestras de cada bloque y $l(c_0)$ la cantidad de muestras del audio original. La función utilizada para el enventanado es:

$$win[\mathrm{n}] = \sin\left(\frac{\pi(2n+1)}{2N}\right), \quad 0 \le n \le N-1 \qquad (2.1)$$

Dos bloques adyacentes consisten en un bloque original y uno marcado. El k-ésimo bloque marcado (*k=2m*) lleva la k-ésima secuencia de la marca.

Para asegurar la inaudibilidad se introducen solo pequeños cambios en la envolvente, la modulación de fase se realiza cumpliendo la siguiente restricción:

$$\left|\frac{\Delta\phi(z)}{\Delta z}\right| < 30^{\circ} \qquad (2.2)$$

Siendo ϕ la fase de la señal y z la frecuencia en la escala de Bark (ver sección 3.2.2). Usando un bloque de gran tamaño ($N=2^{14}$) se consiguen cambios de fase lentos en el tiempo. Para el marcado de los bloques se procede de la siguiente manera:

1. Cada bloque se transforma al dominio de la frecuencia obteniéndose los coeficientes de Fourier Cok[f].

$$C_{0k} = F\{c_{0k}\}, \quad k = 2m, \quad 1 \le m \le \frac{M-1}{2} \qquad (2.3)$$

2. El siguiente paso es construir la función de modulación de fase ϕ_k(b). Cada unidad en la escala de Bark transporta un bit de la marca. Cada bit del mensaje es representado mediante una función ventana de fase, centrada al final de la correspondiente banda de Bark, extendiéndose 2 bandas de Bark.

$$\phi(z) = \sin^2\left(\frac{\pi(z+1)}{2}\right), \quad -1 \le z < 1 \qquad (2.4)$$

El signo de la función ventana de fase $a_k[j]$, se determina por el j-ésimo bit del mensaje (m_k[j]) de la k-ésima secuencia. La modulación de fase total es obtenida por la combinación lineal de las funciones de fase solapadas.

$$\Phi_k(z) = \sum_{j=1}^{J} a_k[j]\,\phi(z-j), \quad 0 \le z < J \qquad (2.5)$$

3. Usando $\Phi_k(z)$, los bits se insertan en la fase del k-ésimo bloque, multiplicando los coeficientes de Fourier con la función de modulación de fase:

$$A_{wk}[f] = A_{ok}[f] \times e^{i\Phi_k[f]} \qquad (2.6)$$

4. La señal de marca se obtiene al antitransformar los coeficientes modificados de Fourier ($A_{wk}[f]$) de los bloques individuales. Todos los bloques son enventanados y sumados solapadamente para obtener la señal de marca.

La robustez se puede mejorar aumentando el valor de nz (cantidad de bandas críticas necesarias para codificar un bit) (nz >1) con la consiguiente disminución en la tasa de bits .La tasa de bits (bit rate) de este método se puede calcular como el número de bits por bloque multiplicado por el número de bloques por segundo:

$$\frac{N_B}{n_z} \times \frac{f_s}{N} \qquad (2.7)$$

Siendo f_S la frecuencia de muestreo (44100Hz), N_B la cantidad de bandas críticas (24 para f_S =44.1 kHz), N la cantidad de muestras del audio original, y n_z es la cantidad bandas críticas para codificar un bit.

La extracción de la marca requiere un procedimiento de sincronización muy preciso para realizar el alineamiento de cada bloque marcado, usando la señal original como referencia (método non blind). Dicho alineamiento será posible si la señal marcada no tiene grandes distorsiones.

2.1.3 - Método réplica

La señal original es usada como una marca de audio. Un buen ejemplo de esto es el ocultamiento del eco (echo hiding). La modulación de replica introduce parte de la señal original en el dominio de la frecuencia como una marca de agua, de ahí el nombre modulación réplica (introduce una réplica). El detector genera la réplica a partir de la señal marcada de audio y calcular la correlación. La ventaja más significativa de este método es su alta inmunidad al ataque de sincronización. El método de la modulación de la réplica es un esquema que introduce una réplica (versión modificada de la señal original).

Existen tres formas de realizar la réplica:

- esquemas de corrimiento en frecuencia
- de corrimiento en fase
- de corrimiento en amplitud.

El método de corrimiento en frecuencia transforma la señal de audio original s(n) en el dominio de la frecuencia y copia una fracción de los componentes de baja frecuencia (p. ej. el rango de 1kHz – 4kHz) y lo modula (p. ej. moviendo 20Hz, con un factor de escala apropiado) y se lo agrega a los componentes originales (cubriendo el rango entre 1020Hz y 4020Hz) y dicha señal se pasa al dominio del

tiempo obteniendo así la señal marca w(n). De esta manera se obtiene la señal marcada: x(n)=s(n)+αw(n).

La réplica en el dominio de la frecuencia w'(n) puede ser generada a partir de la señal marcada x(n) siguiendo el proceso inverso por el cual se introdujo la marca. Luego se calcula la correlación entre x(n) y w'(n) de la siguiente forma para decidir si la marca está o no presente:

$$c = \frac{1}{N}\sum_{i=1}^{N} s(i)w'(i) + \frac{1}{N}\sum_{i=1}^{N} \alpha w(i)w'(i) \qquad (2.8)$$

Dado que w'(n) sólo tiene componentes de frecuencia en un pequeño rango de frecuencias comparado con s(n), la correlación entre ambos será muy pequeña en la ecuación. Por otro lado el espectro del producto w(n)*w'(n) tiene un fuerte componente de DC y por lo tanto c contiene un termino con el valor medio de w(n)*w'(n), esto es, contiene la señal auxiliar escalada por a en el último termino de la ecuación.

2.1.4 - Echo hiding

Un gran número de algoritmos de watermarking de audio desarrollados se basan en el método de ocultamiento de eco. Estos esquemas introducen la marca en la señal original agregando ecos para producir la señal marcada; de esta manera se evita el problema de la sensibilidad del HAS al ruido agregado. Luego de que fue agregado el eco en la señal original, la señal marcada mantiene la misma estadística y sus características de percepción.

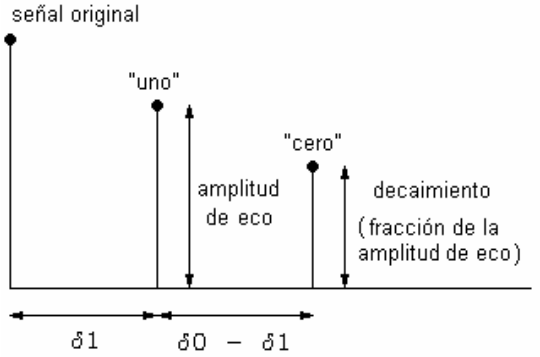


Figura 2.1 - Señal original y eco

El retardo entre la señal original y la marcada es lo suficientemente pequeño para que el eco sea percibido por el HAS como una resonancia. Los parámetros básicos de este método estan presentados en la figura y son las amplitudes de los ecos y los retardos del eco "1" y del eco "0". El proceso de agregado de la marca se puede ver como un sistema que aplica una de dos funciones posibles, estas funciones vistas en el dominio de la frecuencia son exponenciales discretas que difieren en el retardo que introducen.

La señal original es dividida en pequeños bloques para poder codificar más de un bit, a cada uno de estos bloques se le aplica una de las exponenciales discretas

para agregarle el eco correspondiente al bit que se quiera codificar. La señal marcada final (conteniendo varios bits) es una composición de todos los bloques codificados independientemente. Como siguiente paso se deben ajustar las transiciones entre los bloques codificados para prevenir los cambios abruptos en la resonancia de la señal marcada.

$$x_i(n) = \begin{cases} s_i(n) + \alpha s_i(n - \partial 1) & w = 1 \\ s_i(n) + \beta s_i(n - \partial 0) & w = 0 \end{cases} \qquad (2.9)$$

Teniendo en cuenta que la información es insertada en la señal original agregando eco con 2 posibles retardos, la extracción de la información añadida consiste en detectar el espaciamiento entre los ecos. La magnitud de la autocorrelación del cepstrum (el cepstrum es la transformada de Fourier en decibeles del espectro de una señal) de la señal codificada, ecuación 2.4.

$$F^{-1}\left\{\log\left(\left|F(x)\right|^2\right)\right\} \qquad (2.10)$$

, donde F representa la transformada de Fourier y F^{-1} la transformada inversa de Fourier, puede ser examinada en 2 regiones correspondientes a los retardos de las funciones “uno” y “cero”. Si el autocepstrum es mayor en d1 que en d0, el bit introducido será decodificado como un “uno” y viceversa.

Existen muchas modificaciones que se le pueden aplicar al algoritmo básico de echo-hiding, pero en general tienen como desventaja una alta complejidad en el cómputo del cepstrum o el autocepstrum durante la detección. Otra desventaja es que cualquiera puede detectar el eco sin previo conocimiento, lo que lo hace sensible a ataques.

2.1.5 - Método de los dos conjuntos (patchwork)

El método de los dos conjuntos surge originalmente como un método de watermarking de imágenes. La idea principal consiste en dividir pseudo-aleatoriamente al conjunto de muestras original en dos conjuntos disjuntos A y B. A uno de esos conjuntos, por ejemplo al A, se le suma un valor constante. A las muestras del conjunto B se le resta es mismo valor.

patch A	patch B
$a_i^* = a_i + d$	$b_i^* = b_i - d$

Al realizar el valor esperado de (a*-b*) obtenemos:

$$E\left(a^* - b^*\right) = E\left(a + d - \left(b - d\right)\right) = E\left(a - b\right) + 2d \qquad (2.11)$$

Como E(a-b) es igual a E(a)-E(b) y tanto A como B son conjuntos disjuntos elegidos al azar, es razonable esperar que E(a-b) sea aproximadamente cero. Por lo tanto el valor esperado de un bloque al cual se le aplicó la marca de agua va a ser 2d.

En el caso que la marca de agua no esté presente el valor esperado, será aproximadamente cero. Cuanto mayor sea el valor del parámetro d, mayor será la robustez del algoritmo. Sin embargo también se verá afectada la inaudibilidad.

El proceso de detección comienza con la resta de los valores de las muestras entre los dos conjuntos. En función del valor esperado de la resta se decide si está presente o no la marca de agua.

Es importante hacer notar que si bien todo lo expuesto hasta ahora es para el dominio del tiempo, en la práctica el algoritmo casi siempre se realiza en el dominio de la frecuencia. De esta manera, un cambio en algunas componentes de frecuencia se traduce en un cambio en todas las muestras en el dominio del tiempo, haciendo más robusto al algoritmo.

Modificaciones del algoritmo

Para hacer más robusto el algoritmo se realizan las siguientes modificaciones:

1) Además de calcular los valores esperados, también se calcula la varianza conjunta.

2) Se supone distribución normal para el conjunto de muestras elegidas al azar. En el algoritmo original se usaba distribución uniforme.

3) Se utiliza un valor d que se adapta para cada bloque. Esto es muy importante porque permite esconder la información de watermarking de acuerdo a las características del audio. De esta manera se aumenta considerablemente la robustez y la inaudibilidad.

En este método se calculan tanto los valores esperados como la desviación estándar conjunta:

$$S = \sqrt{\frac{\sum_{i=1}^{n}\left(a_i - \bar{a}\right)^2 + \sum_{i=1}^{n}\left(b_i - \bar{b}\right)^2}{n(n-1)}} \qquad (2.12)$$

$$a_i^* = a_i + sg\left(E(a) - E(b)\right)\sqrt{C}\,\frac{S}{2} \qquad (2.13)$$

$$b_i^* = b_i - sg\left(E(a) - E(b)\right)\sqrt{C}\,\frac{S}{2} \qquad (2.14)$$

siendo C una constante adaptada para cada bloque;

$$d = \sqrt{C}\,\frac{S}{2} \qquad (2.15)$$

Se aplican estos cambios y luego se realiza la transformada inversa.

Detección de la marca de agua

Para la detección de la marca de agua, se calcula el siguiente estadístico:

$$T^2 = \frac{\left(E(a) - E(b)\right)^2}{S^2} \qquad (2.16)$$

Se compara el estadístico con un umbral τ que se determinará.

En caso de superarlo significa que está presente la marca de agua, en caso contrario no está presente. Este procedimiento se realiza en distintos bloques de la señal de audio logrando un mensaje de varios bits.

Desventajas

Algunas desventajas del método patchwork aquí presentado son:

1) No incorpora información sobre el sistema auditivo humano
2) Es vulnerable a un ataque de sincronización
3) Tiene una baja tasa de transferencia de información (un bit por bloque)

2.1.6 - Método de Spread-Spectrum

El esquema de Spread Spectrum se basa en el cálculo de la correlación entre secuencias. Para eso introduce a la señal origen una secuencia pseudo-aleatoria y luego detecta la marca calculando la correlación entre la secuencia de ruido pseudo aleatorio y la señal de audio marcada. Este método es uno de los más estudiados y es relativamente fácil de implementar. Tiene como extra la necesidad de utilizar modelos psicoacústicos que permitan que el ruido agregado a la señal sea inaudible (el uso de estos modelos no está limitado a la técnica de Spread Spectrum).

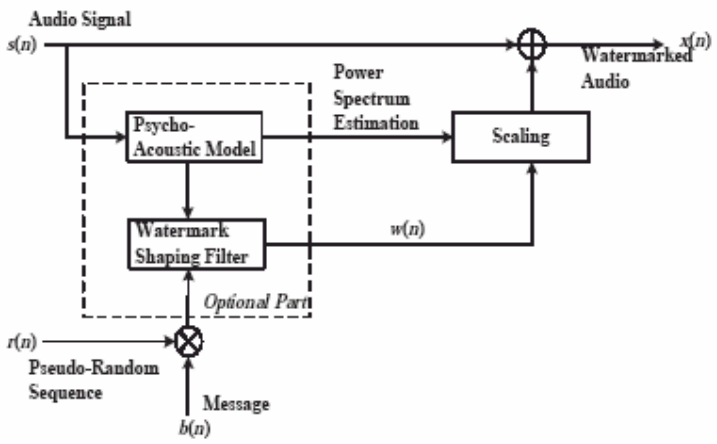


Figura 2.2 - Esquema de introducción de la marca en el método Spread Spectrum

Idea básica

Como ya se mencionó el método consiste en esparcir una secuencia pseudo aleatoria a lo largo de toda la señal de audio. Este ruido puede ser agregado tanto en la señal en el dominio del tiempo como en un dominio transformado sin importar qué transformada sea utilizada; las más usadas son DCT (Discrete Cosine Transform), DFT (Discrete Fourier Transform) y DWT (Discrete Wavelet Transform). El mensaje de marca (v = {0,1} , o su equivalente bipolar b = {1,-1}) es modulado por una secuencia pseudo aleatoria r(n) binaria de media nula, generada por medio de una clave secreta. Luego la marca modulada, w (n) = b . r(n), se atenúa según un factor α de acuerdo a la energía de la señal original s(n); este factor α se encarga de mantener un compromiso entre la robustez y la inaudibilidad de la marca agregada. Esta señal modulada y escalada adecuadamente se agrega a la señal original para generar la señal marcada x(n).

$$x(n) = s(n) + \alpha.w(n) \quad , w(n) = b.r(n) \qquad (2.17)$$

Para la detección se utiliza un esquema de correlación lineal. Dado que la secuencia pseudo-aleatoria r(n) se conoce y puede ser regenerada por la clave secreta, la marca se detecta calculando la correlación entre x(n) y r(n)

$$c = \frac{1}{N}\sum_{i=N}^{1} x(i)\cdot r(i) \qquad (2.18)$$

donde N es el largo de la señal. La ecuación (2.18) se puede descomponer como la suma de dos términos a partir de la ecuación (2.17)

$$c = \frac{1}{N}\sum_{i=1}^{N} s(i)\cdot r(i) + \frac{1}{N}\sum_{i=1}^{N} \alpha\cdot b\cdot r^{2}(i) \qquad (2.19)$$

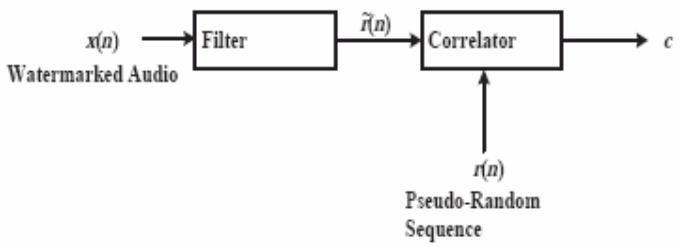


Figura 2.3 - Diagrama del proceso de detección en el método Spread Spectrum

Asumiendo que s(n) y r(n) son independientes el primer término de la ecuación (2.19) debería ser muy pequeño; muchas veces esto no se cumple por lo tanto la señal marcada es preprocesada como muestra la figura 2.3 para que esto se cumpla. Existen varios métodos de prefiltrado, una posible solución podría ser filtrar s(n) de x(n) por ejemplo incluyendo un filtrado pasa alto.

Estos métodos de preprocesamiento permiten que el segundo término de la ecuación (2.13) sea mucho mayor en magnitud y que el primero prácticamente sea nulo. Si esto no se cumpliera (o sea el primer termino fuera del orden o mayor que el segundo) la detección resultaría errónea. Basados en el test de hipótesis; usando el valor de la correlación (c) y un umbral predefinido k, la salida del detector será:

$$m = \begin{cases} 1 & , c > k \\ 0 & , c \le k \end{cases} \qquad (2.20)$$

Típicamente el valor de k es 0. El umbral de detección tiene efecto directo sobre las probabilidades de falsos positivos y falsos negativos. Falso positivo es un error en el detector en el que incorrectamente determina que la marca esta presente en una señal no marcada, de forma análoga un falso negativo es un error en el cual el detector determina que no hay marca en una señal marcada.

Secuencias pseudo-aleatorias

Las secuencias pseudo-aleatorias tienen propiedades estadísticas similares a las que realmente son señales aleatorias, pero pueden ser regeneradas exactamente con el conocimiento de información privilegiada. Las señales pseudo-

aleatorias tienen idealmente buenas propiedades de correlación lo que significa que el valor de la correlación cruzada entre ellas es muy bajo mientras que el valor de autocorrelación es moderadamente alto.

Procesamiento de la señal marca

Cuando se agrega una secuencia pseudo-aleatoria o ruido a la señal de audio original se debe considerar que puede causar sonidos audibles e indeseados cualquiera sea el esquema de watermarking que se use. Esto no se soluciona simplemente bajando el valor de atenuación de la secuencia pseudo-aleatoria dado que el oído humano es muy sensible especialmente cuando la energía del sonido es muy baja un pequeño ruido con bajo valor de α puede ser percibido. Bajar el valor de α hace al método de Spread Spectrum perder robustez. Una solución para asegurar inaudibilidad de la marca es procesar la señal pseudo-aleatoria mediante un modelo psicoacústico el cual tiene en cuenta las características del HAS. Esto permite incrementar la robustez al máximo dado que establece un límite bajo el cual la señal se hará imperceptible por lo que se podrá aumentar el valor de α mientras se mantenga por debajo del margen establecido.

Los modelos psicoacústicos usados en la compresión de audio utilizan las propiedades del HAS de enmascaramiento en el tiempo y en frecuencia y el procesamiento del ruido se hace de acuerdo a los umbrales de enmascaramiento. Estos modelos consideran el HAS como un analizador de frecuencias con un conjunto de filtros pasabanda, de esta manera se obtiene la curva de umbral de audición que es la intensidad requerida de un sonido simple (expresada en decibeles) para que éste sea escuchado en ausencia de otros sonidos.

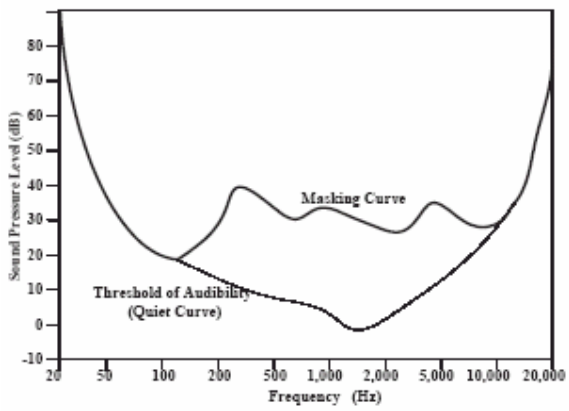


Figura 2.4 - Umbral de quietud

La figura 2.4 muestra la curva del umbral de audición en silencio, en este caso esta curva es igual a la curva de umbral de enmascaramiento (los sonidos por debajo de esta curva son imperceptibles). En presencia de otros sonidos la curva de enmascaramiento se mueve hacia arriba. El enmascaramiento se produce cuando un sonido cercano (en el tiempo o en frecuencia) a otro, afecta sus características, el primero se llama sonido enmascarador y el segundo sonido enmascarado. Los modelos analizan la señal de entrada, s(n), para luego calcular los umbrales de enmascaramiento.

La señal de marca audible se transforma en una señal inaudible mediante el proceso basado en estos modelos psicoacústicos (ver figura 2.5).

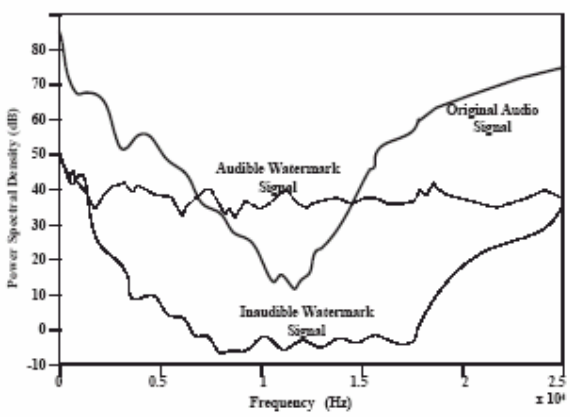


Figura 2.5 - Marca de agua audible e inaudible

El procedimiento para el enmascaramiento en frecuencia se puede resumir en los siguientes pasos: 1) Calcular la potencia espectral, 2) Localizar los componentes de los tonos, 3) Reducir el número de enmascaradores para eliminar los enmascaramientos irrelevantes, 4) Computar los umbrales de enmascaramiento individuales, 5) Determinar el umbral mínimo de enmascaramiento en cada sub-banda. Este umbral mínimo define la respuesta del filtro que procesara la señal de marca. Luego la señal filtrada es escalada para que el ruido de la marca este debajo del umbral de enmascaramiento.

La tarea de procesamiento de la señal marca consume mucho tiempo especialmente cuando se trata de explotar las propiedades de enmascaramiento dado que se debe calcular los coeficientes del filtro con el que se transformara la señal de marca. Otra posibilidad podría usar la curva del umbral de audición en silencio, lo que disminuiría el procesamiento. Utilizar el mínimo nivel de ruido no es óptimo en términos de la potencia de la señal de marca. Esto resulta en una fuerte reducción de la robustez del método. Otra forma de aumentar la robustez, en vez de maximizar el umbral de enmascaramiento, es incrementar el largo de la secuencia pseudo-aleatoria aunque esto reduce la capacidad de mensaje que puede ser introducido en la señal original.

2.2 - Justificación de la técnica seleccionada

En esta sección analizamos algunas de las diferencias entre los métodos, para luego poder elegir el más adecuado para nuestros propósitos.

Por ejemplo, con el método LSB se obtiene el mayor bitrate, sin embargo es muy poco robusto. Por el solo hecho de realizar una conversión digital-analógica y luego una conversión analógica-digital se pierde la marca de agua. Es por eso que el método LSB se aplica principalmente para la transmisión digital de audio (en la cual no se pasa por el dominio analógico).

Según la complejidad computacional, el método LSB es de los más sencillos de implementar. Mientras que los métodos que trabajen en el dominio de la frecuencia requerirán tanto la realización de la transformada como de la transformada inversa.

Como vimos existe una clasificación de los métodos según sean blind watermarking o no. Esta característica se refiere a si se necesita o no la señal original del audio sin marcar para el proceso de detección de la marca de agua. Los métodos más populares son los que no utilizan esa información (blind watermarking) dado que tienen mayor espectro de aplicación. Los métodos que modifican la fase del audio original son del tipo non-blind, en esta categoría entra el método de watermarking de fase por lo que no será considerado en este proyecto.

Los métodos de Spread Spectrum, de los dos conjuntos, y echo-hiding son los más usados. El método de echo-hiding es mejor en términos de imperceptibilidad pero como se mostró anteriormente este método requiere mucho procesamiento computacional en el cálculo del cepstrum. Cuando introducimos marcas mediante los métodos de Spread-Spectrum o patchwork se debe considerar algún modelo psicoacústico para que la marca agregada no se perciba, esto es porque estos métodos agregan ruido aleatorio que pueden producir sonidos audibles indeseables.

En la investigación de los distintos métodos se encontró que el método más estudiado y desarrollado por distintos autores es el de Spread Spectrum, dentro de este método existen distintas variantes que pueden ser utilizadas; esto es lo que nos hace decidirnos por este método.

En el desarrollo del algoritmo se focalizará en las restricciones planteadas en la especificación del proyecto (robustez e imperceptibilidad) y no tanto en la cantidad de información que se introducirá (bitrate). Un punto muy importante en el sistema de watermarking es la redundancia de la marca introducida, esto si bien disminuye la tasa de información aumenta la robustez. Dicha redundancia nos permitirá disminuir la tasa de error en la detección.

Una vez elegido el método se debe estudiar un modelo psicoacústico, el cual se desarrolla a continuación, y un modelo del sistema de comunicación basado en Spread Spectrum que usaremos como base teórica para la implementación del algoritmo.

3 - Conceptos de psicoacústica

La psicoacústica se encarga fundamentalmente del estudio de los aspectos psico-físicos de la audición; incluyendo la estructura física del oído así como características sobre el proceso en que el oído transforma el sonido a señales en el cerebro. La percepción del audio, si bien es un área dentro de la psicoacústica, no es de las centrales dado que generalmente la psicoacústica busca eliminar cualquier juicio subjetivo en sus temas de estudio.

La percepción intenta cuantificar, en la medida de lo posible, las características importantes, en lo que a reproducción de audio se refiere, de lo que el humano interpreta y como estas características influyen en el análisis del juicio acerca de la calidad del sonido.

Algunas tecnologías que manejan audio digital, como el desarrollo del MPEG (algoritmo de compresión), se basan en un conocimiento detallado del HAS que provee la psicoacústica. A partir de estos conocimientos se desarrolla un modelo del HAS que luego es utilizado en el procesamiento de las señales de audio.

El desarrollo de modelos sofisticados del HAS es basado en gran parte en el uso de datos experimentales lo cual necesariamente implica cierta subjetividad; este aspecto puede ser compensado en parte manejando un conjunto grande de datos. Físicamente el sonido es descrito por la variación en el tiempo de la presión p(t). La entrada al HAS son las variaciones temporales de la presión sonora. Luego del procesamiento en el HAS se obtiene una salida que contiene información de las características temporales y espectrales del sonido así como también la localización de la fuente sonora.

En el procesamiento dinámico realizado en los diferentes modelos del HAS se consideran distintas características de éste, dependiendo de las aplicaciones de cada algoritmo. Para el sistema que se desarrollará, se utilizará un modelo en el cual se contemplan las siguientes características

- El oído enmascara las componentes de señal pequeñas en presencia de componentes de mayor amplitud. Este efecto es más evidente entre señales cercanas en frecuencia.

- El oído tiene una discriminación frecuencial finita de señales complejas, por lo que puede ser aproximado como un banco de filtros pasabanda. Los anchos de cada filtro se conocen como las bandas críticas. Los anchos de las bandas críticas varían con la frecuencia.

- El oído 'junta' señales complejas pertenecientes a una misma banda crítica interpretándolas como un tono simple.

- El oído, y por lo tanto la percepción, tienen un comportamiento no lineal.

- El oído es capaz de procesar los sonidos con una ventana temporal de 25ms como un único evento.

En las siguientes secciones se profundizará en los conceptos de enmascaramiento sonoro y bandas críticas de manera de poder comprender luego el algoritmo que implementa el modelo psicoacústico utilizado en el sistema de watermarking desarrollado.

3.1 - Enmascaramiento sonoro

3.1.1 - Definición

El enmascaramiento sonoro puede definirse como el proceso en el cual el umbral de audibilidad correspondiente a un sonido se eleva, debido a la presencia de otro sonido. El umbral de audibilidad representa la sensibilidad del aparato auditivo, es decir, el valor mínimo de presión sonora que debe tener un tono para que éste sea apenas perceptible en ausencia de otros sonidos. El umbral de audibilidad depende de la frecuencia de la señal sonora.

Dependiendo de la ubicación temporal de la señal de prueba (P) con respecto a la señal enmascarante (E), se pueden distinguir tres situaciones:

1) Enmascaramiento frecuencial o simultáneo: E y P se presentan solapados en el tiempo. E esta presente durante toda la duración de P.

2) Enmascaramiento previo: E se presenta después de P.

3) Enmascaramiento posterior: E se presenta antes de P.

Cada tipo de enmascaramiento tiene características distintas. A continuación se analizará con detalle las características del enmascaramiento frecuencial, cuyos conceptos serán considerados en el modelo que implementaremos.

3.1.2 - Enmascaramiento frecuencial

Para entender el proceso de enmascaramiento frecuencial se hará un breve repaso por el proceso de audición del HAS.

El sonido se introduce a través del oído y choca con el tímpano haciéndolo vibrar. La vibración es recibida por tres huesos articulados en cadena y controlados por dos pequeños músculos que trasmiten el movimiento al estribo, ver figura 3.1, que en su extremo se une con la ventana oval. La ventana oval es el lugar por donde penetra el sonido (oído interno) a la cóclea o caracol. Los movimientos del estribo producen desplazamientos del líquido en el oído interno que estimulan las terminaciones nerviosas o células ciliadas, lugar donde realmente comienza el proceso auditivo. Las células nerviosas estimuladas, envían la señal por el nervio auditivo hasta los centros del cerebro, donde el estimulo eléctrico es procesado.

En la cóclea se encuentra la membrana basilar. Si estiramos esta membrana, se puede observar que cada punto tiene una frecuencia de resonancia diferente, la frecuencia de resonancia va disminuyendo a medida que avanzamos por la membrana basilar desde la ventana oval. Es decir, nuestro oído funciona como un analizador de espectros, cada frecuencia excita un nervio determinado (figura 3.1).

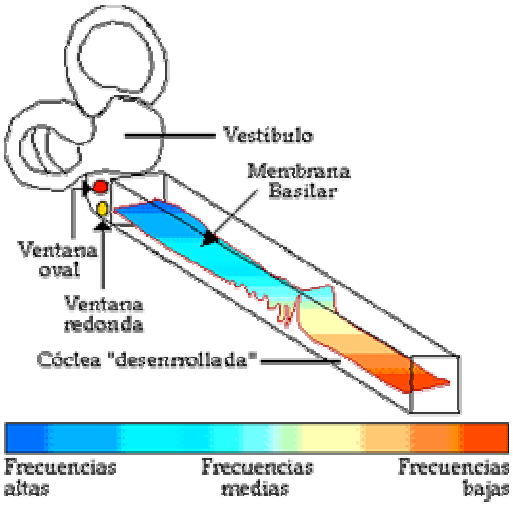
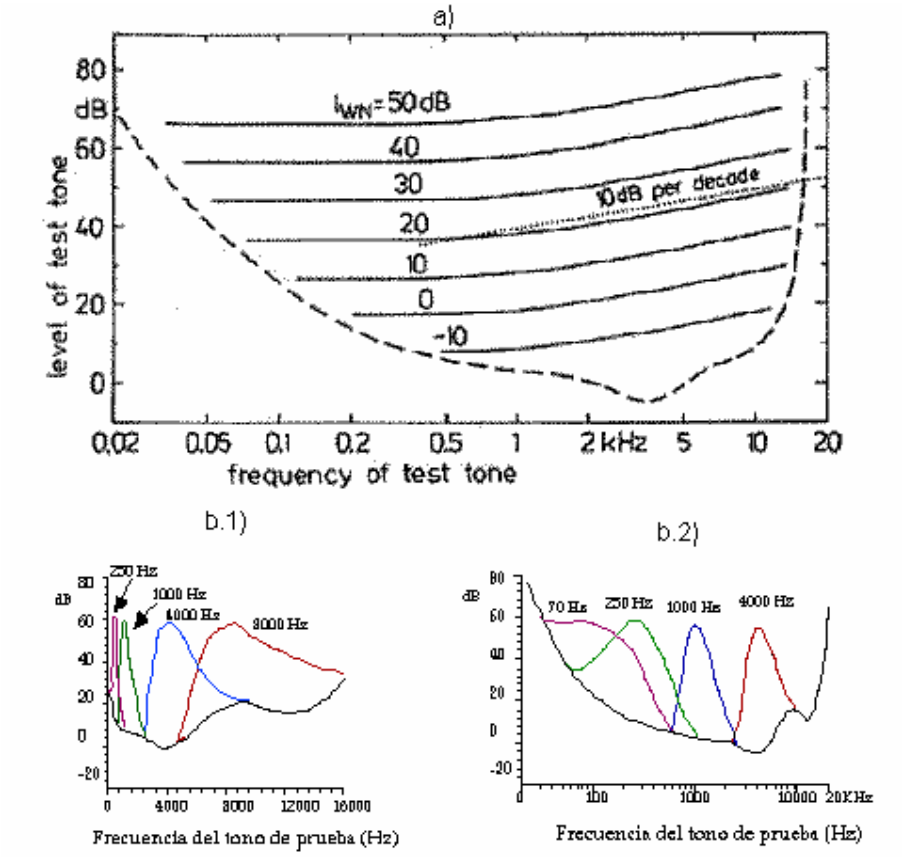


Figura 3.1 – Estructura del oído interno

Esto no es rigurosamente cierto, ya que si penetra en nuestro oído un tono puro no sólo se excitan los nervios correspondientes a esa frecuencia sino también, aunque con menor intensidad, los nervios adyacentes que se corresponden con frecuencias mayores y menores. La amplitud de la excitación a lo largo de la membrana basilar, cuando oímos un tono puro, define lo que llamamos curvas de enmascaramiento para esa frecuencia. De esta forma, cada componente en frecuencia de cada señal sonora ocasiona un nivel de actividad neuronal (excitación) en diversas zonas de la membrana basilar, lo que altera la detectabilidad de otras componentes.

Las curvas de enmascaramiento frecuencial varían dependiendo del tipo de enmascarador (tono o ruido) del rango de frecuencia y de la amplitud del enmascarador. La figura 3.2 muestra el patrón de enmascaramiento generado por ruido blanco (espectro plano entre 20 Hz y 20 kHz) y por tonos en función de la frecuencia y la amplitud.



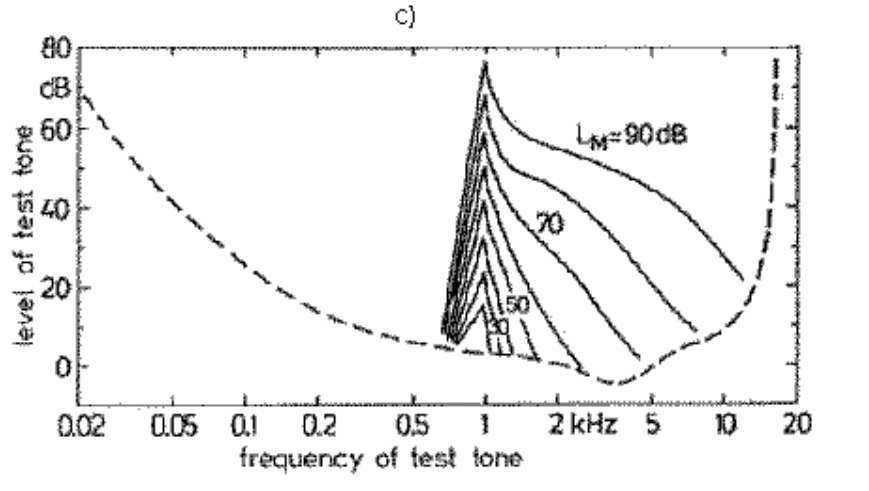


Figura 3.2 – Curvas de enmascaramiento

Se observa que:

1) El umbral de enmascaramiento correspondiente al ruido blanco (a) es prácticamente constante en el rango de 20 a 500Hz y con una leve pendiente por encima de los 500Hz.

2) Se puede apreciar cómo aumenta la amplitud de las curvas de enmascaramiento al aumentar la densidad espectral en el caso de ruido blanco o el nivel de amplitud de los tonos (a y c).

3) Para el caso de las señales tonales (b) se observa que el efecto de enmascaramiento se extiende en un rango más amplio hacia las altas frecuencias que hacia las bajas. Comparando las figuras (b1) y (b2) se ve que en una escala logarítmica de frecuencias, la forma y las pendientes de las curvas correspondientes a las frecuencias 1000 Hz y 4000 Hz son muy similares mientras que en una escala lineal las curvas correspondientes a frecuencias por debajo a 500Hz son similares entre si.

4) La pendiente de la curva descripta por el umbral de enmascaramiento de señales tonales (b y c) es más pronunciada hacia las bajas frecuencias, respecto a la frecuencia del tono, (de lo que deduce que frecuencias cercanas más altas que la enmascaradora pueden ser más fácilmente enmascarables).

A partir de los puntos 1 y 3 podría pensarse que el efecto del enmascaramiento depende de la frecuencia en forma lineal por debajo de los 500Hz y en forma logarítmica por encima de los 500Hz.

3.2 - Bandas críticas

Como se vio en la sección anterior, el comportamiento de la cóclea como analizador de frecuencia puede resumirse en dos características: las componentes espectrales de una señal sonora son procesadas por más de un receptor auditivo; análogamente, cada receptor auditivo procesa diversas componentes del espectro de la señal. Esto muestra que la selectividad en frecuencia del sistema auditivo tiene sus limitaciones.

3.2.1 - Definición de banda crítica

El concepto de ancho de banda crítico puede interpretarse como una medida de la selectividad frecuencial del oído. El ancho de banda crítico permite explicar por qué, dado un tono de una cierta frecuencia, una banda de ruido estrecha centrada en dicha frecuencia produce la misma cantidad de enmascaramiento sobre el tono que una banda ancha de ruido, aún cuando el nivel de densidad espectral de ambos ruidos sea igual y, por ende, la energía del ruido de banda estrecha sea menor.

Numerosos experimentos psicoacústicos indican que las respuestas de los sujetos ante distintos fenómenos perceptibles cambian abruptamente cuando los estímulos sobrepasan un cierto ancho de banda.

Así pues, se define una banda crítica como un intervalo de frecuencia que representa la máxima resolución frecuencial del sistema auditivo en diversos experimentos psicoacústicos. Adicionalmente, puede decirse que una banda crítica constituye el intervalo de frecuencia en el cual el oído interno efectúa una integración espectral de la intensidad de la señal sonora: la banda crítica es el intervalo en el cual se "suma" la energía de las distintas componentes espectrales de la señal.

Este concepto resulta importante dado que implica que la porción de una señal compleja que pertenece a una banda crítica puede ser tratada como una composición de señales simples. De esta manera una señal compleja puede ser analizada como la combinación, debida a las distintas bandas críticas, de señales simples.

3.2.2 - Escala de bandas críticas

El ancho de banda de las bandas críticas tiene un valor constante de 100Hz hasta la frecuencia central de 500Hz. A partir de esa frecuencia los anchos de banda van creciendo hasta cubrir todo el espectro audible. Esta subdivisión continua del rango de frecuencias audibles en intervalos solapados entre sí de una banda crítica de ancho, lleva a la necesidad de obtener para cada frecuencia fo, un valor que represente el número (no necesariamente entero) de bandas críticas adyacentes y no solapadas contenidas en el intervalo de 0 a fo Hz. Los valores así obtenidos constituyen la denominada tasa de bandas críticas, también llamada escala de bandas críticas, de gran utilidad para el análisis del procesamiento realizado por el HAS.

Para los valores de la escala de bandas críticas, se ha definido como unidad el "Bark": un intervalo de frecuencia de 1 Bark es, por definición, un intervalo de una banda crítica de ancho en cualquier punto del rango de frecuencias audibles. Así la primera banda abarca el intervalo de 0 a 1 Bark, la segunda banda crítica el intervalo de 1 a 2 Barks, y así sucesivamente. La relación entre la tasa de bandas críticas y la frecuencia puede ser expresada mediante la ecuación 3.1, la cual permite calcular la tasa de bandas críticas (en Barks), z, correspondiente a la frecuencia en Hz, f, con un error inferior a ±0,2 Barks:

$$z = 13\tan^{-1}\left[\frac{0.76 * f}{1000}\right] + 3.5\tan^{-1}\left[\left(\frac{f}{7500}\right)^2\right] \qquad (3.1)$$

La relación entre la escala de bandas críticas y parámetros tales como la posición en la membrana basilar (escala fisiológica), el número de incrementos de frecuencia apenas perceptibles y el cociente de frecuencias subjetivas o "alturas del sonido" (escalas psicoacústicas) resulta prácticamente lineal. Por todo lo expuesto, la escala de bandas críticas no sólo está asociada con una medida de la selectividad en frecuencia, como lo son las bandas críticas, sino que además constituye una escala más natural y conveniente que la escala de frecuencias para representar gráficamente e interpretar fenómenos perceptuales.

3.3 - Modelos psicoacústicos

Con lo que se ha mostrado acerca de la psicoacústica, se concluye que no todos los sonidos tienen la misma relevancia. Las propiedades analizadas son usadas por los mecanismos de compresión de audio para disminuir la cantidad de datos necesarios para representar un sonido, basándose en un modelo psicoacústico que simula el comportamiento del oído humano. Considerando las limitaciones del oído para percibir todas las componentes en una onda de audio compleja, los mecanismos de compresión calculan lo que se oirá de un sonido particular, descartando el material indetectable o codificándolo con menos precisión, idealmente sin cambiar la calidad del sonido percibido.

Si bien el objetivo de nuestro sistema no es la compresión de audio, utilizaremos los resultados obtenidos por el modelo psicoacústico del algoritmo de compresión MPEG1 en nuestro sistema de watermarking. Durante la compresión MPEG se debe calcular la relación señal-mascara, para esto es necesario previamente calcular el mínimo umbral de enmascaramiento. Dicho umbral puede ser utilizado, aplicado al watermarking, para introducir información por debajo de este umbral logrando que sea imperceptible al combinarlo con el audio. A continuación se detalla el algoritmo del modelo psicoacústico 1 del estándar ISO/IEC 11172 (MPEG-1) [14] válido para capas 1 y 2 que luego será implementado y adaptado para nuestro sistema de watermarking.

3.3.1 - Modelo psicoacústico 1 de la norma ISO/IEC 11172-3:1993

Este modelo calcula la asignación de bits de las 32 sub-bandas en base de la relación señal máscara de todas las sub-bandas. Por lo tanto, es necesario determinar para cada sub-banda, el máximo nivel de señal y el mínimo umbral de enmascaramiento

El cálculo de la relación señal máscara está basado en los siguientes pasos:

1. *Cálculo de la transformada rápida de Fourier para la conversión del dominio tiempo al dominio frecuencia.*
2. *Determinación del nivel de presión sonora para cada sub-banda.*
3. *Determinación del umbral de silencio (umbral absoluto).*
4. *Encontrar las componentes del tipo tonal y no tonal de la señal de audio.*
5. *Diezmado de los enmascaradores, para obtener solamente aquellos relevantes.*
6. *Cálculo de los umbrales de enmascaramiento individuales.*
7. *Determinación del umbral de enmascaramiento global.*
8. *Determinación del umbral de enmascaramiento mínimo para cada sub-banda.*
9. *Cálculo de la relación señal máscara para cada sub-banda.*

*1. Transformada rápida de Fourier***.**

El umbral de enmascaramiento se obtiene estimando la densidad espectral de potencia que se calcula a través de la transformada rápida de Fourier de orden 512 para la capa I (layer I) y de orden 1024 para la capa II (layer II). La transformada

rápida de Fourier se calcula directamente desde la entrada PCM enmarcada con una ventana del tipo Hanning.

Para que haya coincidencia en el tiempo entre la asignación de bits y las muestras de sub-banda correspondientes, la señal PCM entrante que ingresa a la transformada rápida de Fourier, debe retardarse.

- El retardo del filtro sub-banda es de 256 muestras, lo que corresponde a 5,3 milisegundos de audio a 48 KHz. de tasa de muestreo. Se requiere un desplazamiento de ventana de 256 muestras para compensar la demora en el banco de filtros de análisis de sub-banda.

- La ventana Hanning debe coincidir con las muestras por sub-banda de la trama. Para Layer I esto se adiciona a un desplazamiento de ventana de 64 muestras. Para Layer II se requiere un desplazamiento de ventana de – 64 muestras.

Datos técnicos de la Transformada Rápida de Fourier (FFT):

	Layer I	**Layer II**
Longitud de la transformada	512 muestras	1024 muestras
Tamaño de la ventana (fs=48KHz)	10,67 mseg	21,3 mseg
Tamaño de la ventana (fs=44.1Khz)	11,6 mseg	23,2 mseg
Tamaño de la ventana (fs=32kHz)	16 mseg	32 mseg
Resolución de frecuencia	Frec_muestreo/512	Frec_muestreo/1024

*Ventana de Hanning**:***

$h(i)= (8/3)^{1/2} * 0.5 * [1 – \cos(2\pi i /N)]$ $\quad i = 0... N-1$

Densidad espectral de potencia:

$$X_{(k)} = 10 * \log_{10}\left[1/N * \sum_{l=0}^{N-1} h_{(l)} * s_{(l)} * e^{[-j.k.l.2\pi/N]} \right]^2 dB \qquad k = 0,....,N/2$$

Donde $s_{(l)}$ es la señal de entrada.

Se ha hecho una normalización al nivel de referencia de 96 dB SPL (Sound Pressure Level), de forma tal que el máximo valor posible corresponda a 96 dB.

2. Determinación del Nivel de Presión Sonora (Sound Pressure Level) (volumen)

El nivel de presión sonora (SPL) L_{sb} en la subbanda n se calcula como:

$L_{sb}(n) = MAX [X_{(k)}, \ 20*\log(scf_{máx}(n)* 32768) – 10]$ dB $X_{(k)}$ en la sub-banda n

Donde:

- $X_{(k)}$ es el SPL de la línea espectral con un índice k de la FFT con el máximo de amplitud en el rango de frecuencia correspondiente a la sub-banda n.

- $Scf_{(máx)}$ es en Layer I el SCF(factor de escala [14]) y en Layer II el máximo de los 3 SCF de la subbanda n de la trama.
- El término -10 dB corrige la diferencia entre el nivel de pico y RMS. ($V_p/\sqrt{2}=V_{rms}$ $20.\log_2(1/\sqrt{2}) = -10dB$)
- L_{sb} se calcula para cada sub-banda n.

3. Consideración del Umbral en silencio

El umbral en silencio $LT_q(k)$, también denominado umbral absoluto (UA) se detalla en las tablas D1 a,b,c para layer I y D1 d,e,f para layer II. Estas tablas dependen de la frecuencia de muestreo de la señal PCM de entrada. Los valores están disponibles para cada muestra en el dominio de la frecuencia donde se calcula el umbral de enmascaramiento.

Un offset dependiente de la tasa total de bits se utiliza para el umbral absoluto. Este offset es de –12 dB para tasa de bits mayores o iguales a 96 Kbps y 0 dB para tasas de bits menores a 96 Kbps.

5. Localización de las componentes tonales y no tonales

La tonalidad de una componente de enmascaramiento tiene influencia en el Umbral de Enmascaramiento (UE). Por este motivo, resulta importante discriminar entre las componentes del tipo Tonal y no Tonal**.** Para el cálculo del umbral de enmascaramiento global es necesario discriminar ambas componentes del espectro de la FFT.

La operación comienza con la determinación del máximo local, luego se extrae la componente tonal (tipo sinusoide) y se calcula la intensidad de las componentes no tonales dentro de un ancho de banda comprendido en una banda crítica. Los límites de las bandas críticas se dan en las tablas D2 a,b y c para layer I y D2 d,e y f para layer II).

El ancho de banda de las bandas críticas varía con la frecuencia, con un ancho de banda de solamente 0.1 KHz para bajas frecuencias hasta un ancho de banda de aproximadamente 4 KHz en altas frecuencias. Estos límites se determinaron en experimentos psicoacústicos donde se comprobó que el oído humano tiene mejor resolución absoluta en bajas frecuencias que en la región de frecuencias altas.

Para determinar si el máximo local puede ser una componente del tipo tonal se examina un rango de frecuencia (df) alrededor del máximo. Este rango de frecuencia está dado por:

Sampling rate: 48 kHz

Layer I:	**df = 187,5 Hz**	**0 kHz < f <= 6,0 kHz**
	df = 281,25 Hz	**6,0 kHz < f <= 12,0 kHz**
	df = 562,50 Hz	**12,0 kHz < f <= 24,0 kHz**
Layer II:	**df = 93,750 Hz**	**0 kHz < f <= 3,0 kHz**
	df = 140,63 Hz	**3,0 kHz < f <= 6,0 kHz**
	df = 281,25 Hz	**6,0 kHz < f <= 12,0 kHz**
	df = 562,50 Hz	**12,0 kHz < f <= 24,0 kHz**

Sampling rate: 32 kHz						
Layer I:	**df = 125 Hz**	**0 kHz**	**<**	**f**	**<=**	**4,0kHz**
	df = 187,5 Hz	**4,0 kHz**	**<**	**f**	**<=**	**8,0 kHz**
	df = 375 Hz	**8,0 kHz**	**<**	**f**	**<=**	**15,0kHz**
Layer II:	**df = 62,5 Hz**	**0 kHz**	**<**	**f**	**<=**	**3,0 kHz**
	df = 93,75 Hz	**3,0 kHz**	**<**	**f**	**<=**	**6,0 kHz**
	df = 187,5 Hz	**6,0 kHz**	**<**	**f**	**<=**	**12,0 kHz**
	df = 375 Hz	**12,0 kHz**	**<**	**f**	**<=**	**24,0 kHz**
Sampling rate: 44,1kHz						
Layer I:	**df = 172,266 Hz**	**0 kHz**	**<**	**f**	**<=**	**5,512kHz**
	df = 281,25 Hz	**5,512 kHz**	**<**	**f**	**<=**	**11,024 kHz**
	df = 562,50 Hz	**11,024 kHz**	**<**	**f**	**<=**	**19,982kHz**
Layer II:	**df = 86,133 Hz**	**0 kHz**	**<**	**f**	**<=**	**2,756 kHz**
	df = 129,199 Hz	**2,756 kHz**	**<**	**f**	**<=**	**5,512kHz**
	df = 258,398 Hz	**5,512 kHz**	**<**	**f**	**<=**	**11,024 kHz**
	df = 516,797 Hz	**11,024 kHz**	**<**	**f**	**<=**	**19,982kHz**

Para listar las líneas espectrales $X_{(k)}$ que son componentes tonales y no tonales se deben realizar las siguientes 3 operaciones:

a) Etiquetado del máximo local**.**

La línea espectral $X_{(K)}$ se etiqueta como un máximo local si:
$X_{(K)} > X_{(K-1)}$ *y* $X_{(K)} >= X_{(K+1)}$

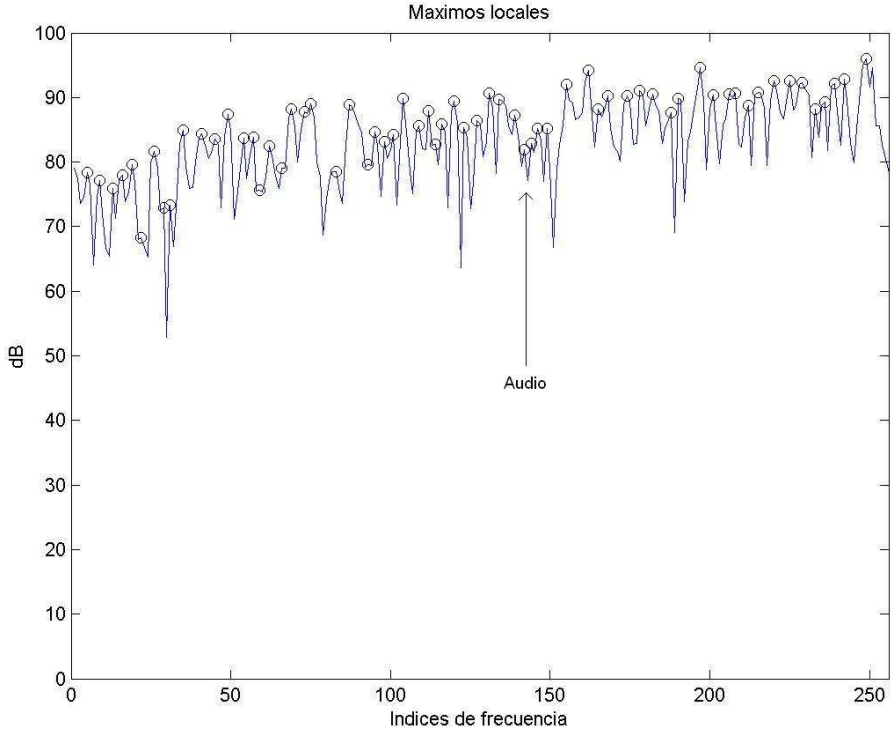


Figura 3.3 – Máximos locales

b) Listado de los componentes tonales y cálculo de SPL

Un máximo local se coloca en la lista de componentes tonales si:
$X_{(K)} - X_{(K+j)} >= 7dB$

Donde j es elegido de acuerdo a:

Layer I:
J= -2,+2 para 2<k<63
J= -3,-2,+2,+3 para 63<=k<127
J= -6,...,-2,+2,...,+6 para 127<=k<=250

Layer II
J= -2,+2 para 2<k<63
J= -3,-2,+2,+3 para 63<=k<127
J= -6,...,-2,+2,...,+6 para 127<=k<255
J= -12,...,-2,+2,...,+12 para 255<=k<500

Si resulta que $X_{(K)}$ es un componente tonal, entonces se listan los siguientes parámetros:

- El índice k de la línea espectral
- SPL $X_{tm(K)} = 10 * \log_{10}[10^{X(k-1)/10} + 10^{X(k)/10} + 10^{X(k+1)/10}]$, en dB
- La bandera tonal (tonal flag)

Luego, todas las líneas espectrales dentro de las frecuencias analizadas se ajustan a $-\infty$

c) Listado de las componentes no tonales y cálculo de la energía

Las componentes no tonales (asimilables a ruido) se calculan a partir de las líneas espectrales remanentes. Para calcular las componentes no tonales a partir de estas líneas espectrales $X_{(K)}$**,** se determinan las bandas críticas $z_{(k)}$ usando las tablas D2 a,b y c para Layer I y D2 d,e y f para Layer II [14].

En Layer I:

- 23 bandas críticas son usadas para las tasas de muestreo de 32 KHz
- 24 bandas críticas son usadas para las tasas de muestreo de 44.1 KHz
- 25 bandas críticas son usadas para las tasas de muestreo de 48 KHz

En Layer II:

- 24 bandas críticas son usadas para las tasas de muestreo de 32 KHz
- 26 bandas críticas son usadas para las tasas de muestreo de 44.1 KHz y 48 KHz

Dentro de cada banda crítica, la energía de las líneas espectrales remanentes luego que las componentes tonales se establecieran en cero, son sumadas para formar el SPL de la nueva componente no tonal $X_{nm(k)}$ correspondiente a esa banda crítica.

Los siguientes parámetros son listados:

- El índice K de la línea espectral más cerca de la media geométrica de la banda crítica
- SPL $X_{nm(k)}$ en dB
- La bandera no tonal (non tonal flag)

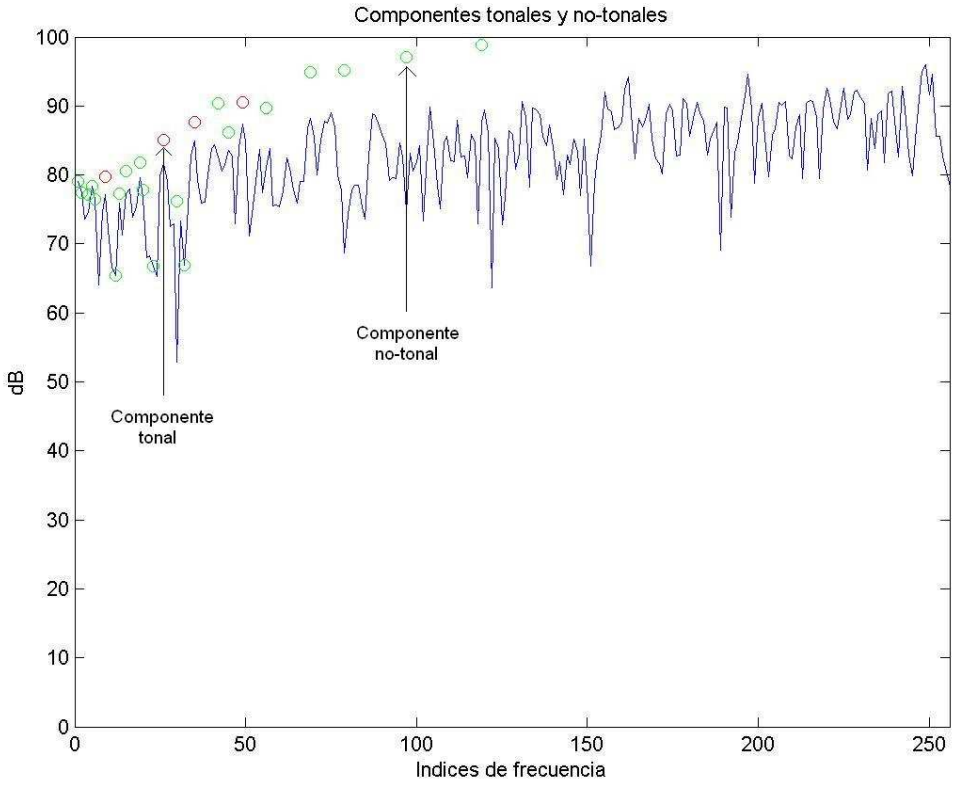


Figura 3.4 - Componentes tonales y no tonales

6. Diezmado de las componentes tonales y no tonales.

El diezmado es un procedimiento que se utiliza para reducir el número de enmascaradores (maskers) que son utilizados para el cálculo del umbral de enmascaramiento global.

a.- Los componentes tonales $X_{tm(k)}$ o NO tonales $X_{nm(k)}$ son considerados para el cálculo del umbral de enmascaramiento solamente si:
$X_{tm(k)} >= LT_{q(k)}$ o $X_{nm(k)} >= LT_{q(k)}$

En la expresión, LTq(k) es el umbral absoluto (o umbral en silencio) a la frecuencia del índice k. Estos valores se dan en la tabla D1 a,b, y c para Layer I y D1 d,e, y f para layer II.

b.- El diezmado de dos o más componentes tonales dentro de una distancia de 0.5 Bark: dejar la componente con la energía más alta y remover las

componentes menores de la lista de componentes tonales. Para esta operación, se utiliza una ventana que se desplaza en la banda crítica con ancho de 0.5 Bark.

En adelante, el índice j se utilizará para indicar las componentes tonales y no tonales relevantes de la lista combinada diezmada.

7. Cálculo de los umbrales de enmascaramiento individuales.

De las N/2 muestras originales en el dominio de la frecuencia, indexadas en *k*, solamente un subconjunto de muestras, indexadas en ***i*** son consideradas para el cálculo del umbral de enmascaramiento global. Las muestras están en las tablas D1 a,b y c para layer I y D1 d,e y f para layer II [14].

Layer I:

- Para las muestras de frecuencia correspondientes a la región de frecuencia cubierta por las primeras 6 sub-bandas, no se utiliza submuestreo.
- Para las muestras de frecuencia correspondientes a la región de frecuencia cubierta por las siguientes 3 sub-bandas se considera cada segunda línea espectral.
- Finalmente, en el caso de 44.1 KHz y 48 KHz de tasa de muestreo, en la región de frecuencias correspondientes a las sub-bandas remanentes, son consideradas cada 4 líneas espectrales, hasta 20 KHz.
- En el caso de 32 KHz de frecuencia de muestreo, en la región de frecuencias correspondientes a las sub-bandas restantes se consideran cada 4 líneas espectrales, hasta 15 KHz.

Layer II:

- Para las muestras de frecuencia correspondientes a la región de frecuencia la cual es cubierta por las primeras 3 sub-bandas, no se utiliza submuestreo.
- Para las muestras de frecuencia correspondientes a la región de frecuencia la cual es cubierta por las siguientes 3 sub-bandas se considera cada segunda línea espectral.
- Para las muestras de frecuencia correspondientes a la región de frecuencia la cual es cubierta por las siguientes 6 sub-bandas se considera cada 4 líneas espectrales.
- Finalmente, en el caso de 44.1 KHz y 48 KHz de tasa de muestreo, en la región de frecuencias correspondientes a las sub-bandas remanentes, cada 8 líneas espectrales son consideradas hasta 20 KHz.
- En el caso de 32 KHz de frecuencia de muestreo, en la región de frecuencias correspondientes a las sub-bandas restantes, cada 8 líneas espectrales se consideran hasta 15 KHz.

El número de muestras, n, en el dominio de las frecuencias submuestreadas es diferente dependiendo de las tasas de muestreo y de las capas.

Tasa de muestreo (KHz)	Layer I (n=)	Layer II (n=)
32	108	132
44.1	106	130
48	102	126

Cada componente tonal y NO tonal se le asigna el valor del índice ***i*** que más cerca se corresponda a la frecuencia de la línea espectral original $X_{(k)}$. Este índice ***i*** se da en las tablas D1 a,b y c para layer I y D1 d,e y f para Layer II.

Los umbrales de enmascaramiento individuales para tanto las componentes tonales y NO tonales se dan en la siguiente expresión:

$LT_{tm}[z(j),z(i)] = X_{tm}[z(j)] + av_{tm}[z(j)] + vf[z(j),z(i)]$ dB
$LT_{nm}[z(j),z(i)] = X_{nm}[z(j)] + av_{nm}[z(j)] + vf[z(j),z(i)]$ dB

En las fórmulas, LT_{tm} y LT_{nm} son el valor de los umbrales de enmascaramiento individuales en la frecuencia z(i) en Bark, debida al enmascarador ubicado en la frecuencia z(j). Los valores en dB pueden ser tanto positivos como negativos.

El término $X_{tm}[z(j)]$ es el SPL de la componente de enmascaramiento con el número índice j en la banda crítica correspondiente z(j).

El término av se denomina el índice de enmascaramiento y vf es la función de enmascaramiento de la componente de enmascaramiento $X_{tm}[z(j)]$. El índice de enmascaramiento av es diferente para los enmascaradores (maskers) tonales y no tonales (av_{tm} y av_{nm}).

Para los enmascaradores tonales está dado por:

$av_{tm} = -1{,}525 - 0{,}275 * z(j) - 4{,}5$ dB

Para los enmascaradores no tonales está dado por:

$av_{nm} = -1{,}525 - 0{,}175 * z(j) - 0{,}5$ dB

La función de enmascaramiento vf de un enmascarador se caracteriza por diferentes pendientes inferiores y superiores, las cuales dependen de la distancia (en Bark) dz **=** z(i)-z(j) al enmascarador. En la expresión anterior, i es el índice de la línea espectral a la cual se calcula la función de enmascarado. (y ***j*** es la del enmascarador).

Las bandas críticas z(j) y z(i) se encuentran en las tablas D1 a,b y c para Layer I y D1 d,e y f para Layer II. La función de enmascarado, que es la misma para los enmascaradores tonales y NO tonales está dada por:

$vf = 17*(dz+1) - (0{,}4 * X[z(j)] + 6)$	dB	para -3 <=dz<-1 Bark
$vf = (0{,}4 * X[z(j)] + 6) * dz$	dB	para -1 <=dz< 0 Bark
$vf = -17 * dz$	dB	para 0 <=dz< 1 Bark
$vf = -(dz-1) * (17 - 0{,}15 * X[z(j)]) - 17$	dB	para 1 <=dz< 8 Bark

En estas expresiones X[z(j)] es el SPL de la j-ésima componente de enmascarado (en dB). Por razones de complejidad de implementación, el enmascaramiento no se considera más allá, si dz < -3 Bark o dz >= 8 Bark (LT_{tm} y LT_{nm} se ajustan en $-\infty$ dB)

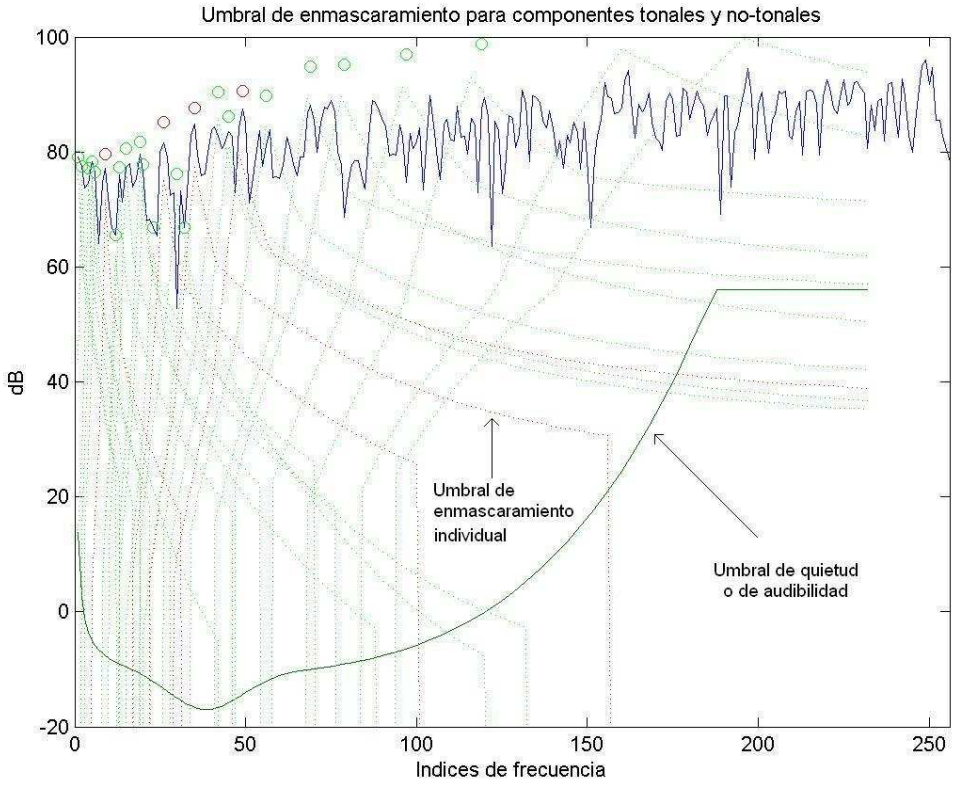


Figura 3.5 – Umbrales de enmascaramiento individual

7. Cálculos del Umbral de enmascaramiento global (LTg)

El umbral de enmascaramiento global LT_g (i) en la i-ésima muestra de frecuencia se deriva a partir de las pendientes superiores e inferiores de los umbrales de enmascaramiento individuales de cada uno de los enmascaradores tonales y no tonales "j" y a partir del umbral en silencio LT_q(i). Este umbral se da en las tablas D1.a, D1.b, D1.c para el Layer I y de las tablas D1.d, D1.e y D1.f para Layer II. El umbral de enmascaramiento global se obtiene sumando las potencias correspondientes la los umbrales de enmascaramiento individuales en dB y el umbral de enmascaramiento en silencio.

El número total de enmascaradores tonales está dado por *m* y el número total de enmascaradores no tonales por *n*. Dada *i*, el rango de *j* puede ser reducido para abarcar esas componentes de enmascaramiento que se hallan entre –8 y +3 Bark desde *i*. Fuera de este rango, LT_{tm} y LT_{nm} se consideran $-\infty$ dB.

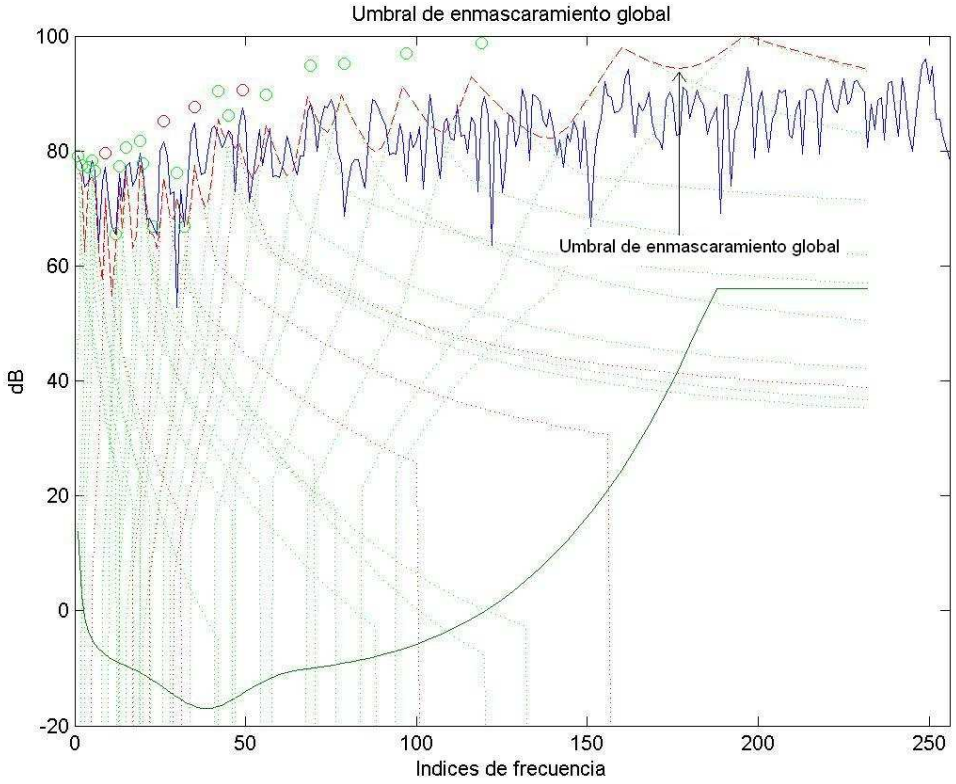


Figura 3.6 – Umbral de enmascaramiento global

*8. Determinación del umbral de enmascaramiento mínimo***.**

El mínimo nivel de enmascaramiento $LT_{mín}(n)$ en la subbanda *n* se determina a partir de la siguiente expresión:

$LT_{mín}(n)=MIN\ \{LT_g(i)\}$ [dB]

f(i) en la sub-banda n

Donde f(i) es la frecuencia de la i-ésima muestra frecuencial. Las f(i) están tabuladas en las tablas D1.a, D1.b, D1.c para el Layer I y de las tablas D1.d, D1.e y D1.f para Layer II.

Un nivel mínimo de enmascaramiento ($LT_{mín}(n)$) se computa para cada sub-banda.

9. Cálculo de la relación Señal-Máscara (SMR)

La relación señal-máscara se computa para cada sub-banda (n) de la siguiente manera:

$SMR_{sb}(n) = L_{sb}(n) – LT_{mín}(n)$ en dB

Habiendo analizado detalladamente el algoritmo que implementa el estándar MPEG1, estamos en condiciones de comprender el funcionamiento del sistema desarrollado de watermarking de audio, que se detalla en el siguiente capitulo.

.

4 - Implementación del sistema de watermarking

El algoritmo de watermarking desarrollado en esta sección combina la técnica de comunicación Spread Spectrum con el modelo psicoacústico explicado en el capítulo anterior para lograr el objetivo planteado. El sistema fue desarrollado mediante software con el lenguaje de programación de Matlab para audio en formato wav, monoaural y calidad de CD (16 bits por muestra, frecuencia de muestreo 44100 Hz) basándose en [7],[15].

Considerando el sistema como un sistema de transmisión de datos se separó en 2 grandes bloques: el transmisor, encargado de generar e introducir los datos en el audio y el receptor, encargado de la recuperación de dichos datos. El detalle de implementación de estos bloques se realizará por separado a continuación.

4.1 - Detalles de implementación del Transmisor

El proceso de generación e introducción de la marca se puede ver en el diagrama de bloques de la figura 4.1. La secuencia de bits de información (w), que se quiere agregar al audio, se modula para generar una señal de audio mediante un conjunto de parámetros para controlar su esparcimiento en el espectro. Por otro lado el audio original es analizado por el modelo psicoacústico; de este análisis se obtiene información sobre el umbral de enmascaramiento, usado para filtrar la marca. La salida del sistema transmisor es la versión marcada del audio que puede ser almacenada o transmitida.

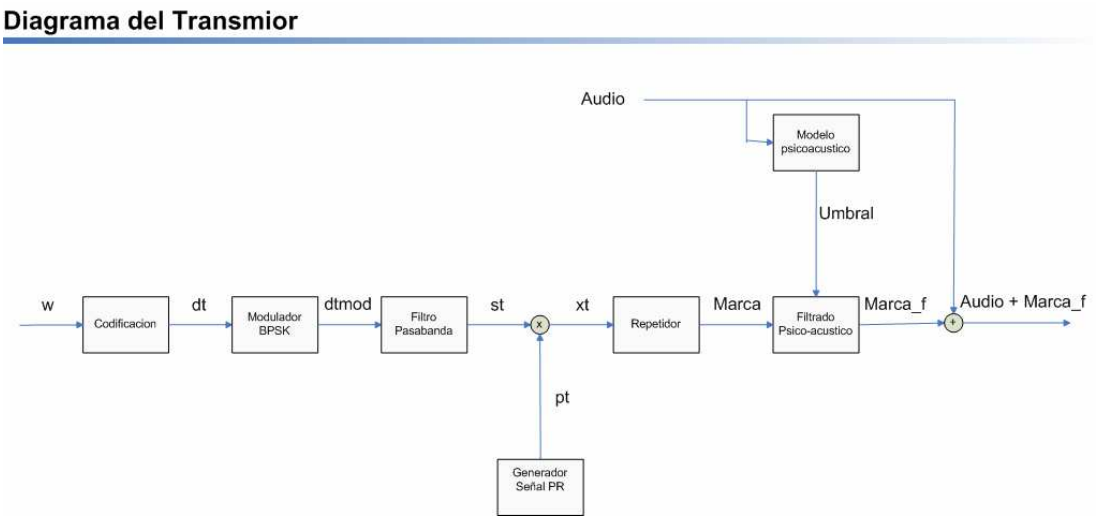


Figura 4.1 – Diagrama del transmisor

4.1.1 - Generación de la marca

El objetivo de esta primera etapa del transmisor es generar una señal de audio que contenga los datos que quieren ser transmitidos. Esta señal a la que llamaremos 'marca' luego será acondicionada para su introducción en el audio que se desea marcar.

La secuencia de datos que formarán la marca es organizada en tramas, las cuales a su vez están compuestas por tres campos: el encabezado, los datos y un campo para la de detección de errores. El encabezado, cuyo objetivo es el de localización del inicio de trama en el receptor, esta compuesto por una secuencia fija de 16 bits (1 1 1 1 1 1 1 1 0 1 0 1 0 1 0 1). El campo de datos está compuesto por los bits del mensaje que quiere ser transmitido, el sistema fue diseñado para transmitir palabras de 32 bits. Finalmente la trama cuenta con un campo para la detección de errores de 8 bits generados mediante un código de redundancia cíclica (CRC).

El bloque codificador se encarga de realizar el armado de la trama. Para esto primero toma la secuencia de datos de 32 bits (w) recibida como entrada del sistema y la reordena utilizando una matriz interleaver. La utilización de un interleaver en el transmisor para reordenar los bits y un de-interleaver en el receptor para volver a la secuencia original, se realiza para que la interferencia que pueda afectar a un bit de datos sea independiente de los otros. Para llevar a cabo esta operación, el codificador utiliza una matriz I x H (I=8, H=4) en la que se colocan los bits de datos de arriba hacia abajo llenando cada una de las I columnas, para luego ser leídos de izquierda a derecha comenzando por la primer fila, ver figura 4.2. La secuencia de datos w = {X1, X2, X3, X4,....} se coloca de la siguiente manera:

X1	X9	X17	X25
X2	X10	X17	X25
X3	X11	X19	X27
X4	X12	X20	X28
X5	X13	X21	X29
X6	X14	X22	X30
X7	X15	X23	X31
X8	X16	X24	X32

Figura 4.2 – Matriz interleaver

obteniendo la secuencia w_R = {X1, X9, X17, X25, X2, X10, X18, X26, etc}. Seguidamente el codificador concatena el encabezado con la secuencia obtenida del interleaver obteniendo una nueva secuencia d = {encabezado}+{w_R}. Por último el codificador añade a esta secuencia 8 bits de redundancia de forma que el polinomio resultante sea divisible por el polinomio generador $P(X) = X^8+X^2+X+1$ (utilizado en las celdas de las redes de datos ATM). El receptor verificará si el polinomio recibido es divisible por P(X), si no lo es, habrá un error en la transmisión y la trama será descartada. De esta manera el codificador arma una trama de 56 bits (dt) la cual deberá ser modulada para obtener a partir de ella una señal de audio.

Para modular la trama recibida del codificador se usa la técnica de Spread Spectrum DS (Direct Sequence) junto con modulación BPSK. Esta elección se basa en la característica de imperceptibilidad que deberá tener la marca final. Desde ese punto de vista la señal de audio puede ser considerada como el principal obstáculo para el algoritmo de watermarking. Para nuestro sistema, el audio será visto como 'interferencia', teniendo éste mucha más potencia que la marca. Cuando a una señal se la expande sobre el espectro, su potencia espectral se incrementa, esto hace que la potencia transmitida aumente sobre un ancho de banda más extenso, dificultando la detección de forma normal (es decir, sin la utilización de ninguna secuencia pseudoaleatoria). Este hecho también implica una reducción de las interferencias, con lo cual, el espectro ensanchado puede sobrevivir en un medio adverso y coexistir con otras señales en la misma banda de frecuencia.

El modulador BPSK recibe esta trama (dt) y realiza una conversión de la secuencia de bits con el fin de efectuar modulación BPSK diferencial. En esta variante la información no va en la fase sino en las transiciones de fase. Esto es, en el caso de que el bit a enviar sea un uno, se codifica como una transición de fase. En el caso de que el bit enviado sea un cero, no hay transición en la fase de la portadora. Dado que el receptor se sincroniza con el primer máximo de la portadora para la demodulación, en el caso de utilizar modulación BPSK normal, éste no sabría diferenciar si corresponde a un seno o a un -seno, pudiendo recuperar una trama invertida. Una vez adaptada la secuencia a transiciones se procede a modular los bits con una portadora muestreada a 44.1 kHz, frecuencia fc=18.2 kHz y tiempo de bit Tb= 234 / 18200 s; de manera de tener un número entero de ciclos de portadora en cada bit, evitando las discontinuidades entre bits.

De este proceso se obtiene una señal muestreada a 44.1 kHz de duración Tb*57 segundos (dado que en el proceso de pasar a transiciones se agrego un bit más a la trama) que será filtrada para obtener una señal limitada en banda. Esto es

realizado por un filtro pasabajos que mantiene las componentes de frecuencia de los lóbulos primario y secundario donde se encuentra la mayor parte de la potencia de la señal modulada. Así se obtiene la señal s(t).

El siguiente paso es expandir la señal s(t) en el espectro mediante la técnica Spread Spectrum. Para esto se debe generar una señal binaria pseudo aleatoria, p(t), con una tasa de bits Tc, (Tc = Tb/N) N veces más rápida que la señal BPSK, y por lo tanto con una densidad espectral de potencia N veces más ancha. Ver apéndice A. Para nuestro sistema N = 21.

La señal p(t) es generada por el bloque generador señal PR. La secuencia pseudo aleatoria generada debe ser lo suficientemente larga para poder hacer la dispersión de una trama completa sin repetir ninguna porción de ella. Para esto el bloque implementa un registro de desplazamiento con retroalimentación lineal de 11 bits, logrando una secuencia pseudo aleatoria de de período 2047 bits (2^{11}-1). El proceso de generación de la secuencia se puede ver en la figura 4.3

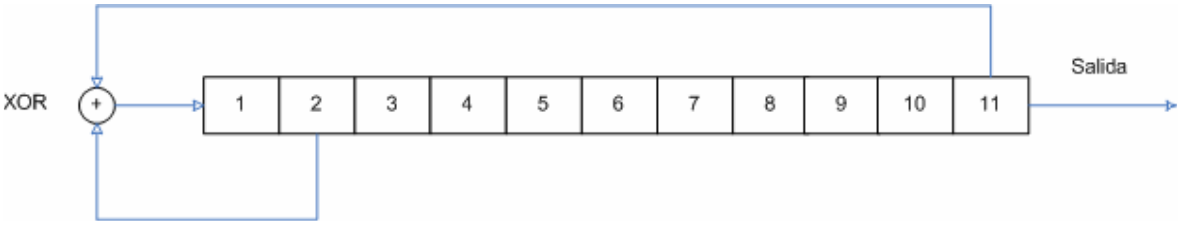


Figura 4.3 – Generación de secuencia pseudo-aleatoria

El registro de 11 bits se inicializa con una secuencia predeterminada llamada clave de generación fija para nuestro sistema, con la restricción que al menos uno de los bits sea distinto de cero. La secuencia se obtiene a partir de los valores que toma el bit n° 11 del registro, dicho valor se actu aliza ejecutando iterativamente los siguientes pasos:

- Se calcula A = XOR (bit 2, bit 11); bits que producirán periodo máximo de secuencia.

- Se realiza un corrimiento hacia la derecha del registro donde el bit 1 tomará el valor A.

Este procedimiento se ejecuta hasta obtener un periodo de la secuencia completo. A partir de esta secuencia de bits se procede a conformar una onda cuadrada polar con tiempo de bit Tc y frecuencia de muestreo 44.1 kHz. De esta señal se toman las primeras muestras obteniendo la señal p(t) del mismo largo que s(t).

La modulación Spread Spectrum se realiza multiplicando en el tiempo la salida filtrada del modulador BPSK, s(t), con la onda binaria pseudo aleatoria p(t), obteniendo la señal Spread Spectrum x(t), x(t) = s(t). p(t).

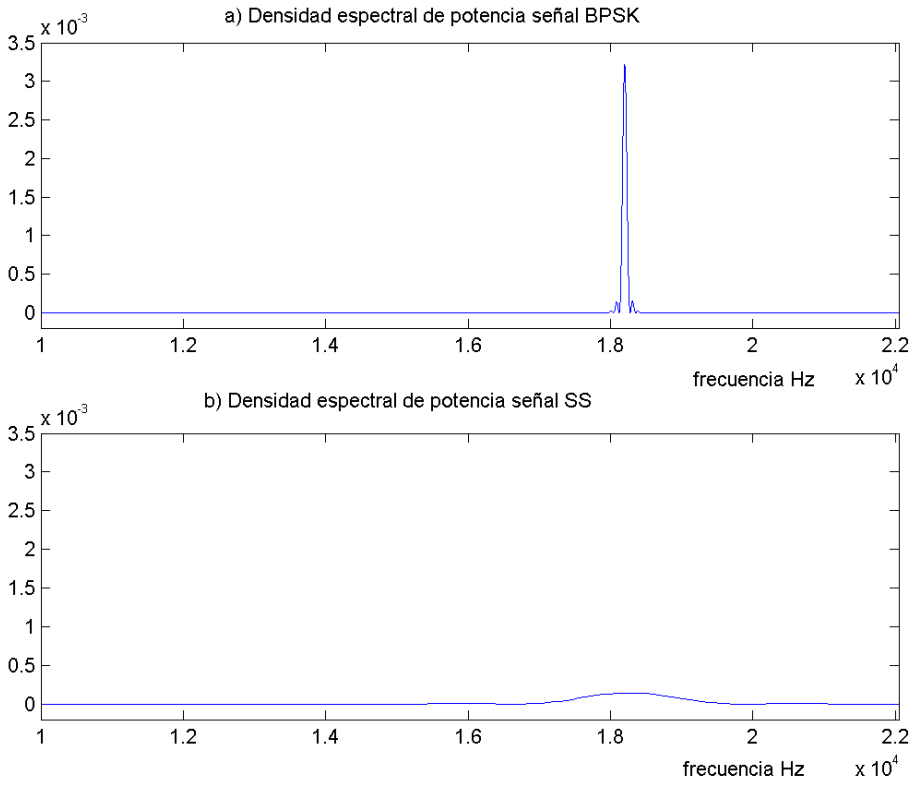


Figura 4.4 – Comparación de densidades espectrales

En este punto del transmisor contamos con una señal de audio, x(t), correspondiente a una trama modulada cuya duración es aproximadamente 0.73 segundos (Tb*57). El bloque repetidor se encarga simplemente de concatenar esta señal para obtener la señal marca del mismo largo que el audio que se quiere marcar, logrando que la información esté distribuida sobre todo el audio. Esta redundancia permite recuperar los datos aún cuando se degrade la señal en algún tramo, aumentando la robustez del sistema.

Esta señal de audio contiene la información que se quiere transmitir. El siguiente paso es adaptar esta señal al audio que se desea marcar con el fin de que no sea percibida y luego combinarla con el audio para su transmisión o almacenamiento.

4.1.2 - Filtrado y combinación con el audio

Para realizar la adaptación de la marca, se procede al análisis de la señal de audio original mediante el bloque modelo psicoacústico. Como se vió en el capítulo anterior, el modelo del sistema de compresión MPEG1 calcula la relación señal máscara (SMR) necesaria para la asignación de bits en cada sub-banda con la que realiza una codificación perceptual. Si bien el SMR no es de utilidad para el caso de nuestra aplicación de watermarking, sí lo es el umbral de enmascaramiento calculado en un paso intermedio. Por este motivo se realiza una adaptación del algoritmo planteado en el capítulo anterior efectuando los siguientes pasos:

1. Cálculo de la transformada rápida de Fourier para la conversión del dominio tiempo al dominio frecuencia.
2. Encontrar las componentes del tipo tonal y no tonal de la señal de audio.
3. Diezmado de los enmascaradores, para obtener solamente aquellos relevantes.
4. Cálculo de los umbrales de enmascaramiento individuales.
5. Determinación del umbral de enmascaramiento global.
6. Determinación del umbral de enmascaramiento mínimo para cada sub-banda.

El bloque modelo psicoacústico ejecuta estos pasos siguiendo el estándar para capa 1 (Layer I) y frecuencia de muestreo 44.1 kHz. Dicho procesamiento se realiza mediante el método de ventanas deslizantes (sliding window) que permite procesar con mayor eficiencia señales de gran duración como lo es una señal de audio. De esta manera se calcula el umbral de enmascaramiento para un bloque de 512 muestras de audio original. Esta información la toma el bloque filtrado psicoacústico y filtra el bloque de 512 muestras de la señal marca que corresponderá al mismo intervalo de tiempo. El proceso se repite recorriendo el audio original con bloques solapados de 512 muestras. Así, se obtienen bloques filtrados de marca que serán combinados mediante ventanas de Hanning (dado que se solapan), para obtener la señal marca filtrada.

El umbral de enmascaramiento, que recibe el bloque de filtrado, es interpretado como el máximo valor de potencia absoluta por sub-banda que debe tener una señal para que sea imperceptible al ser combinada con el audio original. A partir de esto se expresa dicho umbral en índices de frecuencia (0 – 511) manteniendo la potencia por sub-banda. Por otro lado el bloque de filtrado cuenta con una tabla donde se expresa para cada índice de frecuencia (0 - 511) la densidad espectral de potencia de la señal marca. Haciendo una división elemento a elemento entre el umbral y la densidad espectral de potencia de la marca, se obtiene el filtro con el que se logra ubicar la marca por debajo del umbral, haciéndola imperceptible, ver figura 4.5.

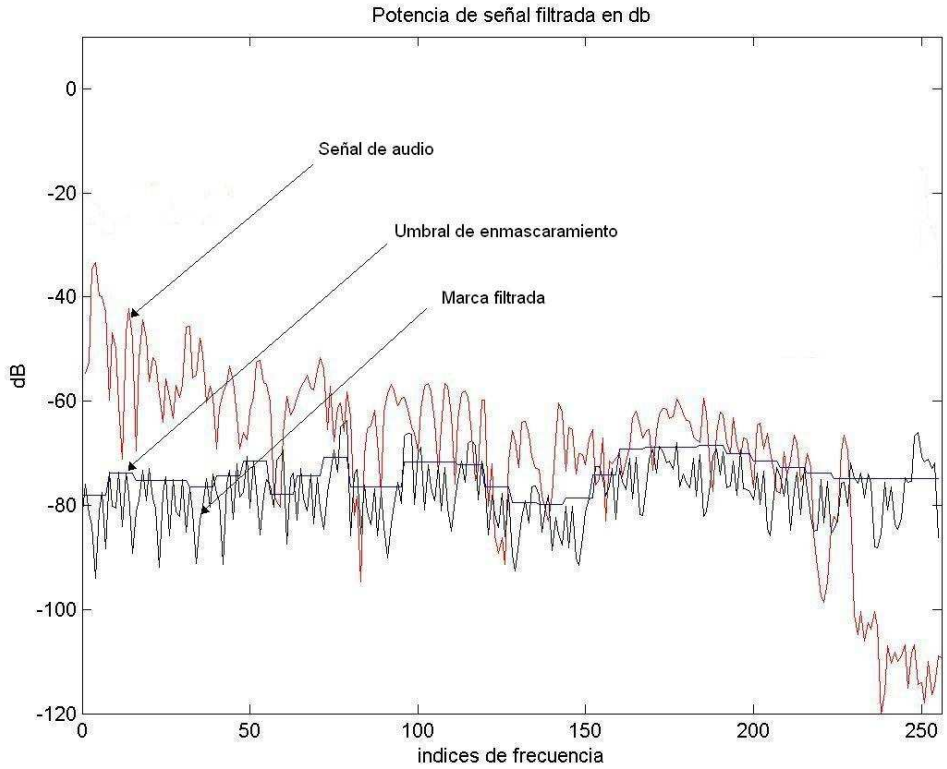


Figura 4.5 – Potencia de la señal filtrada

Este filtrado permite obtener la máxima potencia de nuestra señal de datos con la restricción de imperceptibilidad. De todos modos el sistema desarrollado permite atenuar la señal marca filtrada para aumentar la imperceptibilidad de la información resignando robustez en el sistema. La combinación de las dos señales se realiza sumando en el tiempo dichas señales, obteniéndose la señal marcada lista para transmitir o almacenar.

La figura 4.6 resume los pasos que se realizan en el transmisor. Aquí se muestra la señal marca antes del filtrado donde se distingue su estructura redundante de concatenación de tramas (b) y cómo esta señal se adapta a la forma del audio para hacerse imperceptible (c).

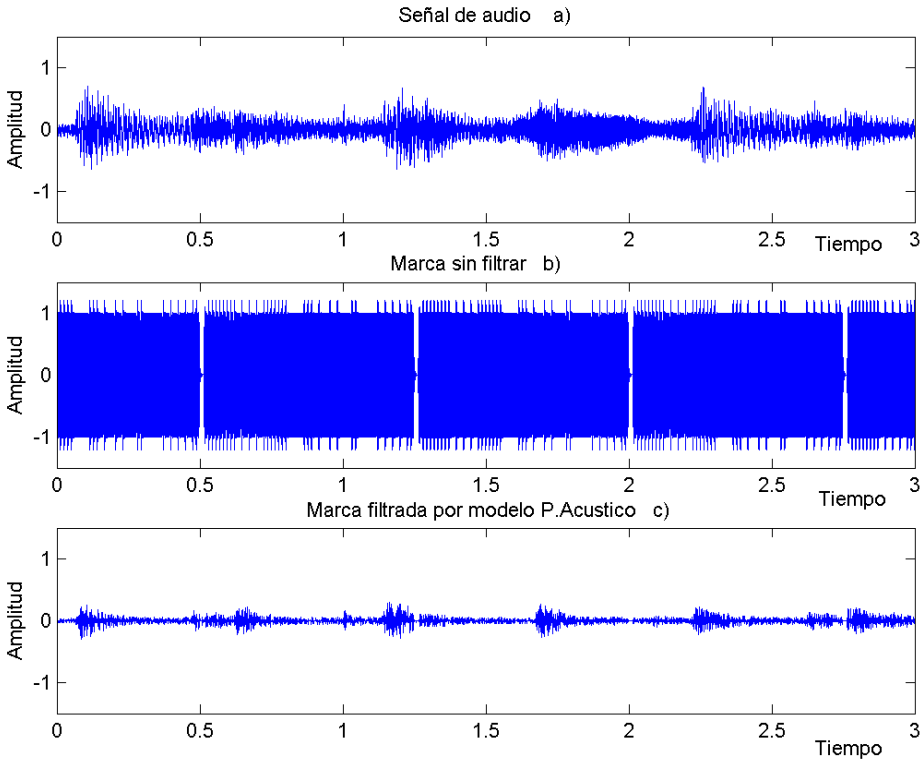


Figura 4.6 - Comparación de la marca antes y después del filtrado

4.2 - Detalles de implementación del Receptor

El proceso de recuperación de la marca se puede ver en el diagrama de bloques de la figura 4.7. La entrada es el audio marcado luego de la transmisión. El mismo modelo psicoacústico usado en el transmisor es aplicado a la entrada para ecualizar la marca. Conociendo los parámetros utilizados en el transmisor se genera la señal encabezado de la trama. Con esto y mediante un filtro de correlación se procesa la marca ecualizada buscando la ocurrencia de un encabezado o lo que es lo mismo el inicio de una posible trama. Luego de esto se realiza la demodulación de las tramas encontradas para recuperar los datos.

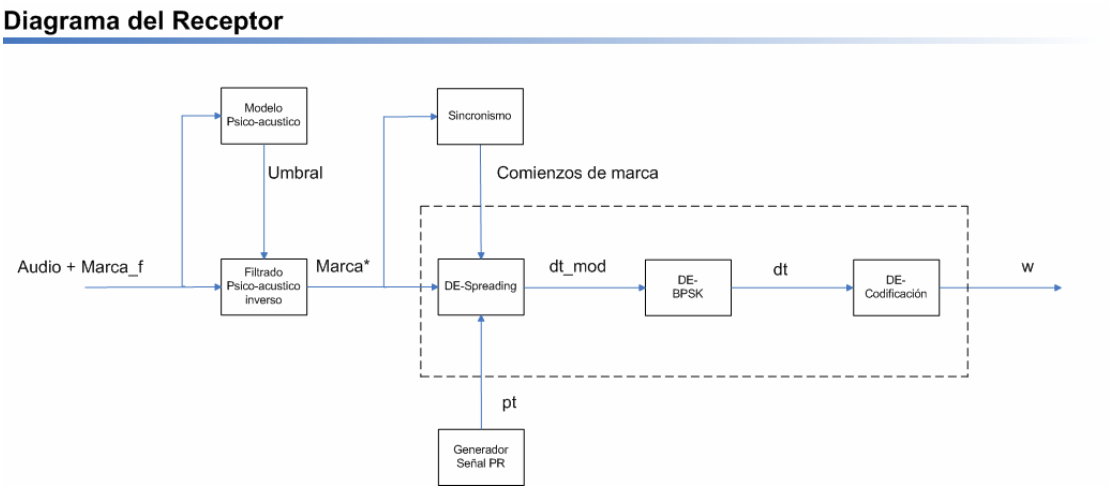


Figura 4.7 – Diagrama del receptor

4.2.1 - Filtrado psicoacústico inverso

La primera etapa del receptor se ocupa de recuperar la marca de la distorsión sufrida en el transmisor debida al filtrado psicoacústico. Este proceso lo llevan a cabo en conjunto los bloques modelo psicoacústico y filtrado psicoacústico inverso. El bloque modelo psicoacústico calcula el umbral de enmascaramiento de la misma manera que en el transmisor con la diferencia que para este calculo no dispone del audio original sino del audio marcado. Esto si bien no producirá exactamente el mismo umbral se puede suponer que será muy parecido, dado que la potencia de la marca es mucho menor en comparación a la del audio original.

El bloque filtrado psicoacústico inverso recibe el umbral, con el que calcula un filtro psicoacústico de la misma manera que en el transmisor. A partir de este filtro, calcula el filtro inverso con el que filtrará la señal de entrada. Este proceso se realiza, igual que en el transmisor, en conjunto entre los dos bloques antes mencionados utilizando la técnica de ventanas deslizantes.

El efecto de este filtrado, suponiendo que la señal de entrada corresponde a audio con marca, provoca por un lado que la componente de la señal correspondiente a la marca sufra el proceso inverso que en transmisión por lo que se puede suponer que contamos con una señal similar a la señal marca del transmisor (figura 4.6 b). Por otro lado la componente de audio original se verá distorsionada lo cual no es de importancia dado que el objetivo del receptor es recuperar la información de la marca.

El siguiente paso será ubicar en dicha señal, marca'(t) en el diagrama de bloques, la posición exacta del inicio de cada trama para luego demodularla y obtener la información de la marca.

4.2.2 - Ubicación y demodulación de la trama

La búsqueda de las tramas dentro de la señal marca'(t) la realiza el bloque sincronismo. Para poder realizar dicho proceso el bloque dispone de una señal fija header(t), interna al receptor, igual al fragmento que corresponde al encabezado de

la trama modulada Spread Spectrum en el transmisor. A partir de la señal header(t), el bloque sincronismo aplica un filtro de correlación a la señal de entrada, marca'(t).

La correlación es un producto discriminador de similitud entre dos funciones. El producto de correlación entre las funciones z(t) e y(t) se define como:

$$c(t) = z(t) \otimes y(t) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} z(t') * y^*(t' + t) dt' \qquad (4.1)$$

La función de correlación c(t) describe topográficamente la similitud entre z(t) e y(t) en términos energéticos. La similitud entre las dos funciones se manifiesta como elevaciones de energía localizadas (máximos de correlación), cuya altura proporciona una medida del grado de similitud entre ellas.

El bloque sincronismo calcula c(t), para las funciones marca'(t) y header(t). Los máximos de esta señal indican en que posición exacta dentro de la señal x(t) están los inicios de cada trama. El proceso se realiza nuevamente mediante la técnica de ventanas deslizantes. De esta manera, se divide la señal marca'(t) en bloques solapados del mismo largo que la señal header(t). La operación de correlación se aplica a cada bloque, siguiendo la ecuación 4.2 equivalente a la 4.1:

$$C(t) = \mathrm{Re}\left(IFFT(B * HEADER^*)\right) \qquad (4.2)$$

donde B corresponde a la FFT del bloque procesado de la señal marca'[nT] y HEADER* al conjugado de la FFT de la señal header[nT]. Luego los bloques de señal C(t) que se obtienen son combinados para obtener la señal correlación(t) del mismo largo que la señal marca'(t). La figura 4.8 muestra un fragmento de dicha señal.

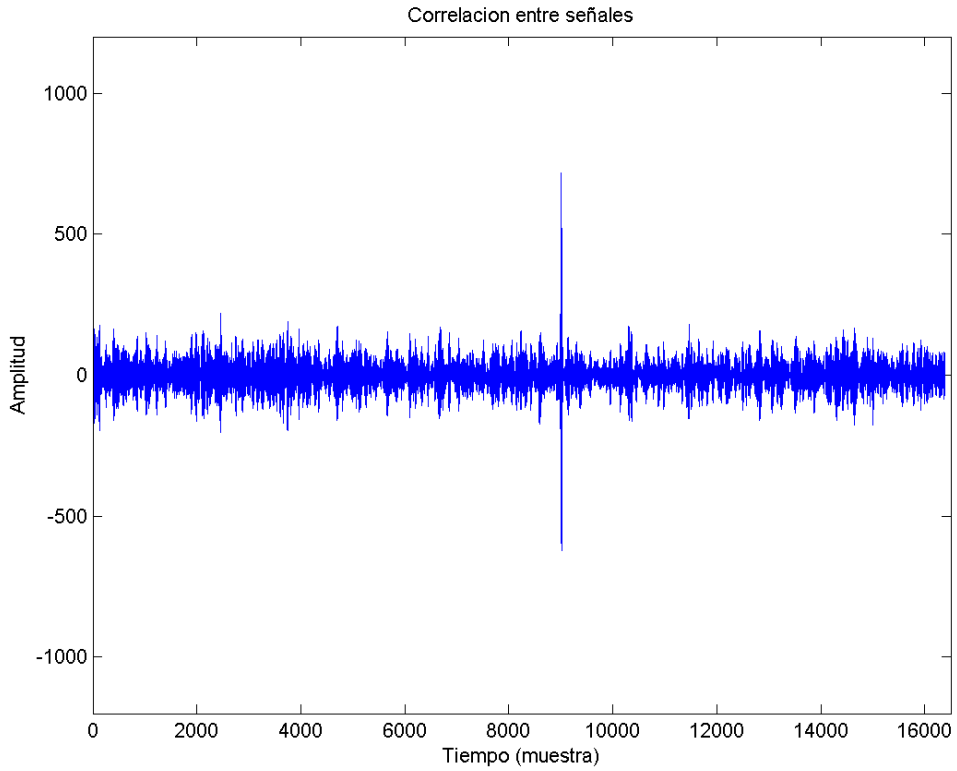


Figura 4.8 – Señal de correlación

A partir de esta señal, el bloque sincronismo selecciona qué máximos locales serán considerados como inicios de trama. Para esto utiliza una variable llamada UMBRAL con la cual compara los máximos de correlación(t). Si un máximo es mayor que $UMBRAL * \overline{correlacion(t)}$ entonces será considerado como un inicio de trama. Esta variable se ajusta automáticamente (dentro de un intervalo predefinido) para obtener un número razonable de posibles inicios de trama.

La salida del bloque sincronismo es un vector que contiene las posiciones de los inicios de trama en la señal marca'(t). Los bloques que se encuentran dentro del recuadro punteado de la figura 4.7 realizan la demodulación de las tramas encontradas por el bloque sincronismo. De no encontrar ningún inicio de trama el proceso de recepción culmina, concluyendo que la señal de audio no estaba marcada.

De aquí en más se realiza un proceso iterativo para cada inicio de trama encontrado comenzando con la demodulación Spread Spectrum. Como primer paso se extrae de la señal marca'(t) el fragmento que corresponde a la trama conociendo su inicio y su largo. La señal que se obtiene es demodulada mediante la onda cuadrada p(t) generada de la misma manera que en transmisión, esto es posible dado que el receptor cuenta con una copia de la clave utilizada para inicializar el registro con el que se genera la secuencia pseudo aleatoria. Seguidamente el bloque DE-Spreading realiza una multiplicación en el tiempo entre las dos señales, recuperando la señal BPSK. Con esto se logra aumentar la potencia del mensaje con relación a la del residuo de la señal de audio original en la banda donde está

ubicada la señal BPSK. Luego se procede a filtrar dicha señal mediante un filtro pasabanda, para quedarnos solamente con las componentes del espectro donde está la señal de interés.

El siguiente paso es realizar la demodulación BPSK de la trama, para esto el bloque DE-BPSK genera una señal sinusoidal con la misma frecuencia que la portadora utilizada para modular la trama en el transmisor. A continuación busca un máximo en la señal modulada BPSK y se sincroniza con éste. Con la señal sinusoidal sincronizada se procede a multiplicar dichas señales para obtener la trama en banda base. La señal obtenida es filtra para eliminar componentes fuera del rango de frecuencias de la señal de interés, ver figura 4.9.

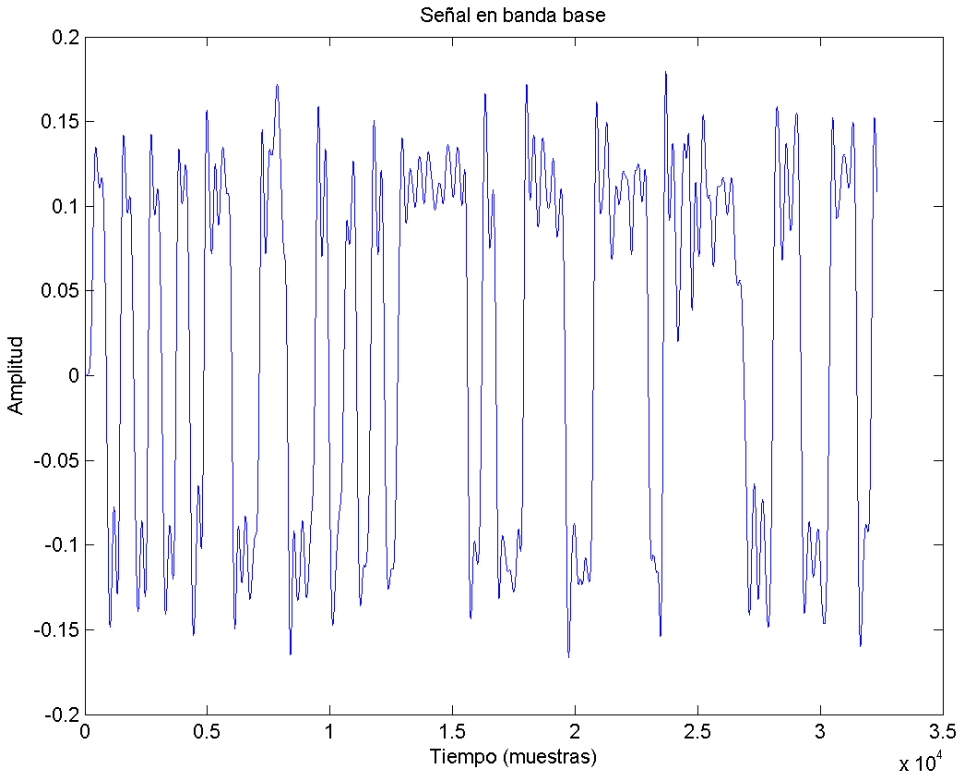


Figura 4.9 – Salida del demodulador BPSK

Finalmente el bloque DE-codificador devuelve la información relevante a partir de la señal en banda base que recibe. Para esto realiza los siguientes pasos intermedios: determinación del instante de muestreo a partir de diagrama de ojo, conversión de transiciones a bits, detección de errores, eliminación del encabezado y de-interleaver.

El proceso que lleva a cabo el diagrama de ojo a partir de la señal en banda base es el de obtener una secuencia de datos binarios. Como la señal recibida no es una onda cuadrada perfecta se debe determinar el instante de muestreo óptimo dentro de cada pulso para obtener correctamente la secuencia de bits. Para esto todos los símbolos recibidos se superponen en un único período de símbolo. En la figura 4.10 se puede apreciar el diagrama correspondiente a una trama. La región

interior del ojo se denomina apertura del ojo y su forma va a condicionar la calidad del sistema.

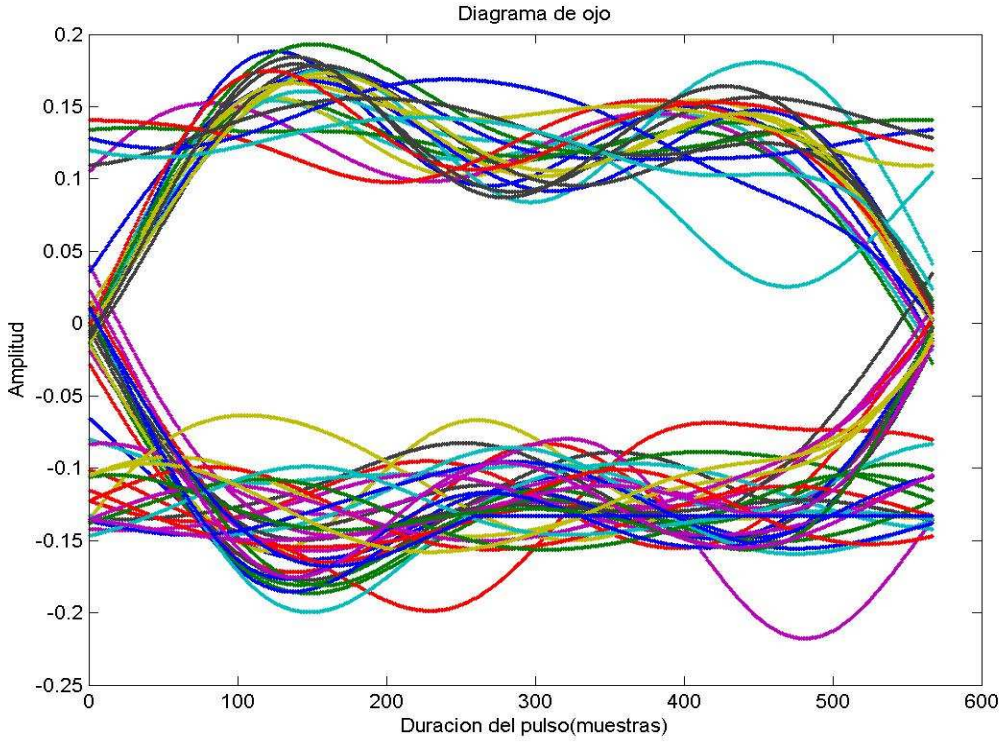


Figura 4.10 – Diagrama de ojo de una trama

El ancho de apertura del ojo indica el intervalo de tiempo en el que se puede muestrear sin error. Como es evidente, el mejor instante de muestreo corresponderá a aquel instante temporal en que la apertura del ojo es mayor. Este cálculo es realizado dentro del DE-codificador para obtener luego del muestreo una secuencia binaria de 57 bits.

La secuencia de 57 bits obtenida del muestreo deberá ser convertida de transiciones a bits, proceso inverso al realizado por el transmisor. Esta conversión se efectúa analizando la secuencia de la siguiente manera: en caso de haber una transición de un bit al siguiente se interpreta como un uno, en caso contrario se interpreta como cero. Luego de esta conversión se obtiene una nueva secuencia de 56 bits de los cuales los primeros 16 bits corresponden al encabezado de la trama, los siguientes 32 bits corresponden al mensaje y los últimos 8 bits corresponden al control de errores.

El proceso de detección de errores se lleva a cabo a continuación. Para esto se utiliza el mismo polinomio, $P(X) = X^8+X^2+X+1$, usado para generar los bits de paridad en transmisión. A partir de este polinomio y los 48 primeros bits de la secuencia, que corresponden al encabezado y los datos, se calculan 8 bits de paridad los cuales serán comparados con los últimos 8 bits de la trama recibida. Si la trama recibida no sufrió errores se obtendrá el mismo resultado dado que se efectúa la misma operación que en la transmisión. En caso que los 8 bits calculados coincidan con los recibidos, la trama recibida se considerará correcta y se seguirá

con su decodificación de lo contrario se considerará que sufrió errores y será descartada.

El siguiente paso que se realizará a la secuencia en caso de no detectarse errores será la eliminación de los bits de encabezado y de control de errores para quedarse con los 32 bits que contienen el mensaje. Estos bits deberán ser reordenados de forma inversa a la efectuada por el interleaver del transmisor. Para esto los bits w' = {Y1, Y2, Y3, Y4,....} son colocados en una matriz de-interleaver como muestra la figura 4.11:

Y1	Y2	Y3	Y4
Y5	Y6	Y7	Y8
Y9	Y10	Y11	Y12
Y13	Y14	Y15	Y16
Y17	Y18	Y19	Y20
Y21	Y22	Y23	Y24
Y25	Y26	Y27	Y28
Y29	Y30	Y31	Y32

Figura 4.11- Matriz de-interleaver

la nueva secuencia de bits será leída a lo largo de las columnas obteniendo finalmente la secuencia w = { Y1, Y5, Y9, Y13, etc....} que representa el mensaje recibido.

El sistema despliega, luego de procesar todas las tramas encontradas en el bloque sincronismo, datos de la recepción como la duración de la señal procesada, el valor de la variable UMBRAL (que da una idea de con cuánta claridad se encontraron tramas), la cantidad de inicios de trama encontrados, la cantidad de tramas detectadas correctamente (sin error), y el mensaje recibido en caso de detectar alguna trama sin error. De lo contrario desplegará el mensaje "No se detecto ninguna marca".

4.3 - Problemas surgidos durante la implementación

En esta sección se analizarán las principales modificaciones que se le fueron realizando al sistema hasta lograr un correcto funcionamiento del mismo. El sistema analizado en las secciones 4.1 y 4.2 es el resultado final de un proceso de correcciones al sistema realizadas con el fin de lograr el objetivo planteado.

El principal problema planteado fue lograr que el receptor lograra detectar alguna marca. Originalmente el receptor no contaba con el filtro psicoacústico inverso. Este filtro fue necesario para lograr una recuperación de la marca ya que el filtro basado en el modelo psicoacústico del transmisor atenúa y distorsiona mucho la marca, con el fin de hacerla imperceptible. El filtro inverso se implementó intentando ser un bloque inverso al filtro del trasmisor; aunque dadas las características del sistema no es posible implementar dicho filtro debido a que no se dispone del audio original. Para evaluar los resultados de dicho filtro se le aplicó como entrada solo la señal marca filtrada (sin audio) tanto al receptor original (sin el filtro) como al nuevo receptor. Para el receptor original no se lograba detectar

prácticamente ninguna trama mientras que con el nuevo sistema se aumentó notoriamente el porcentaje de tramas detectadas.

Sin embargo al intentar recibir la marca dentro de la señal de audio (que es como esta pensado el sistema) no se lograba la detección de ninguna trama. Esto se atribuyó a las diferencias de potencia apreciable que había entre la señal marca y el audio. Analizando el proceso de filtrado se intentó ubicar la marca en otra región del espectro menos sensible al oído y por lo tanto donde sufra menos atenuación. Así fue que se paso de una portadora BPSK de 3.5 kHz a 18.2 kHz pasando a una región mucho menos sensible. Lamentablemente con este cambio de frecuencia en la portadora se resigna robustez debido que la marca podrá eventualmente ser removida más fácilmente sin alterar la calidad del audio que estando localizada en 3.5 kHz. Sin embargo este cambio fue necesario dado que a partir de este se comenzaron a detectar tramas. Junto con este cambio se aumentó la frecuencia de la onda pseudo aleatoria utilizada para la modulación Spread Spectrum (pasando de Tc = Tb/9 a Tc = Tb/21), con el objetivo de aumentar más la potencia de la marca en recepción.

Otros cambios realizados ayudaron a mejorar el funcionamiento del sistema. Dentro de estos se destacan:

- Control de errores
- Eliminación de filtros para un correcto DE-Spreading
- Modulación BPSK diferencial
- Velocidad de procesamiento

En un comienzo el polinomio generador era calculado por una función de biblioteca del Matlab. Este polinomio no nos dio buenos resultados para la detección de errores ya que con distintas secuencias de bits este polinomio nos generaba los mismos bits de redundancia. Debido a esto, marcas detectadas con algún bit de error el sistema las consideraba como correctas. Para solucionar este problema creamos nuestro propio polinomio generador. Como ya vimos, usamos el polinomio de las celdas ATM. Éste nos dio muy buenos resultados.

Se realizó la eliminación de filtros utilizados para acotar en frecuencia la señal de marca entre el Spreading y el DE-Spreading. Estos filtros distorsionaban la señal pseudoaleatoria con la que modulábamos nuestros datos. Esto impedía la realización de un correcto DE-Spreading ya que para esto se debe cumplir que $c^2(t)=1\ \forall t$, siendo c(t) la secuencia pseudoalatoria.

Otro de los cambios realizados fue pasar de BPSK a BPSK diferencial. En el primer caso se transmiten bits, en el segundo se transmiten transiciones de bits. Este cambio nos ayudó en la detección BPSK porque nos independizamos de los bits transmitidos. Esto se debe a la forma de efectuar la demodulación. Dado que el demodulador se sincroniza con el primer máximo encontrado en la señal BPSK y a partir de éste multiplica por una señal sinusoidal en fase, el demodulador no sabrá si este primer máximo corresponde a un uno o a un menos uno. Esto puede provocar que se recupere invertida la señal. Dado que al usar BPSK diferencial nos independizamos de cuál es el primer bit que se transmitió, decidimos adoptar esta variante.

Por último vamos a mencionar un cambio que es exclusivamente de la forma de implantación y que está relacionado con el software utilizado. El cambio se realizó en la implementación del modelo psicoacústico que como ya vimos en el modelo trabaja con ventanas corredizas. Al comienzo el vector de salida del filtro psicoacústico le íbamos agregando valores al mismo tiempo que el modelo calculaba los umbrales. El manejo de vectores largos y variables hacía que el programa Matlab trabajara con una velocidad de procesamiento cada vez más lenta a medida que el vector aumentaba su tamaño. La solución fue crear el vector salida del largo correspondiente contiendo solamente ceros. Luego a medida que modelo calculaba los umbrales se actualizaba el contenido del vector con los valores correspondientes. Como resultado se paso de demorar casi 2 horas para procesar 30 segundos se audio a 8 minutos, logrando un cambio sustantivo en su funcionamiento.

5 - Evaluación del sistema

Se llevaron a cabo una serie de evaluaciones del sistema para dar una idea aproximada de su funcionamiento y estudiar su alcance. En este capítulo se describen las evaluaciones realizadas y se presentan los resultados.

5.1 - Prueba de robustez

Con el fin de evaluar el funcionamiento del sistema se diseñaron una serie de pruebas. Estas pruebas consistieron principalmente en hacer diferentes tipos de modificaciones a la salida del transmisor para ver cómo varían las detecciones de las tramas en el receptor.

Para comprobar la robustez del sistema se realizaron diferentes pruebas detalladas a continuación:

- Géneros musicales
- Banco de señales
- Codificación a MP3
- Agregado de ruido blanco
- Recodificación

Para realizar las pruebas se utilizaron seis estilos de audio diferentes en formato .wav, monoaural y muestreada a 44.1 kHz de duración treinta segundos. Las seis canciones seleccionadas fueron:

1	Coldplay - A Rush of Blood to the Head
2	Coldplay - In My Place
3	Frank sinatra - New York, New York
4	Iron Maiden - Aces High
5	Laura Pausini – Vivimi
6	Mozart – Eine kleine nachtmusik

Tabla 5.1

Como ya se vio, la salida del transmisor corresponde a la suma del audio con la marca filtrada. Con el objetivo de evaluar el sistema con diferentes compromisos entre robustez e imperceptibilidad atenuamos la marca filtrada a distintos decibeles. Se utilizó las etiquetas w0, w2, y w4 para definir los diferentes atenuaciones. La etiqueta w0 representa que la marca filtrada no está atenuada, w2 que está atenuada 2dB y w4 4dB. Cuanto más atenuada está la marca el sistema tiene más imperceptibilidad y menos robustez. A continuación se muestra un ejemplo de análisis:

Ejemplo de prueba:

Fragmento	Prueba	Duración (seg)	Máximos	Umbral	Tramas sin error	Mensaje	Repeticiones
1	sm	30,2771	0	8	0	no	0
1	w0	30,2771	32	32	22	ok	22
1	w2	30,2271	25	32	17	ok	17
1	w4	30,2271	23	30	16	ok	16

En cada prueba se rellenaron nueve campos con la información correspondiente.

Estos campos son:

- Fragmento: Representa la canción según la Tabla 5.1.
- Prueba: Representa el tipo de prueba.
  - w0 sin atenuación de la marca;
  - w2 con 2 dB de atenuación de la marca;
  - w4 con 4 dB de atenuación de la marca;
  - sm sin marca.
- Duración: La duración del fragmento en segundos.
- Máximos: Son la cantidad de máximos encontrados por la función sincronismo. Cada máximo es un posible comienzo de trama en el fragmento. Para más información ir a la función sincronismo en el capitulo 4.
- Umbral: Es un valor que define cuáles son máximos y cuáles no. Para más información ir a la función sincronismo en el capítulo 4.
- Tramas sin error: Es la cantidad de tramas en las cuales no se detectó error según el CRC.
- Mensaje: Indica si detectó alguna trama correctamente.
- Repeticiones: Cantidad de tramas iguales. (Puede pasar que el CRC no detecte un error).

5.1.1 - Géneros Musicales

Evaluamos el sistema con seis fragmentos de diferentes estilos musicales. Los seis fragmentos con los cuales fue realizada la prueba son los fragmentos de la Tabla 5.1. También realizamos la recepción con estos fragmentos sin insertar la marca para comprobar que no exista detección.

Resultados obtenidos:

Fragmento	Prueba	Duración (seg)	Máximos	Umbral	Tramas sin error	Mensaje	Repeticiones
1	sm	30,2771	0	8	0	no	0
1	w0	30,2771	32	32	22	ok	22
1	w2	30,2271	25	32	17	ok	17
1	w4	30,2271	23	30	16	ok	16
2	sm	30,2076	1	8	0	no	0
2	w0	30,2076	47	32	39	ok	39
2	w2	30,2076	43	32	34	ok	34
2	w4	30,2076	31	32	21	ok	21
3	sm	30,1206	4	8	0	no	0
3	w0	30,1206	45	32	30	ok	30
3	w2	30,1206	42	32	27	ok	27
3	w4	30,1206	25	32	17	ok	17
4	sm	30,0262	0	8	0	no	0
4	w0	30,0262	42	28	22	ok	22
4	w2	30,0262	29	28	16	ok	16
4	w4	30,0262	27	26	14	ok	14
5	sm	30,1013	0	8	0	no	0
5	w0	30,1013	24	32	16	ok	16
5	w2	30,1013	22	30	11	ok	11
5	w4	30,1013	24	26	7	ok	7
6	sm	30,2559	0	8	0	no	0
6	w0	30,2559	25	28	19	ok	19
6	w2	30,2559	21	26	15	ok	14
6	w4	30,2559	24	22	13	ok	12

Conclusión:

El sistema detecta la marca para diferentes géneros musicales. Además podemos observar que en los fragmentos sin marcas la detección es nula.

5.1.2 - Bancos de señales

Esta prueba consiste en utilizar diferentes señales en lugar de los fragmentos vistos en la Tabla 5.1. Las señales utilizadas la podemos dividir en dos categorías. La primera consiste en una señal formada por la suma de tonos a diferentes frecuencias. La otra categoría consiste en ruido pasabanda.

Resultados obtenidos:

- *Suma de tonos:*

Para esta prueba generamos señales con cinco tonos. Los primeros cuatro tonos corresponden a las frecuencias 1500Hz, 1700Hz, 1800Hz, 2000Hz. El quinto tono lo fuimos variando para ver los diferentes resultados. En la tabla podemos ver la frecuencia del quinto tono en el campo llamado Suma de tonos.

Para esta prueba no utilizamos atenuación de la marca.

Suma de tonos	Prueba	Duración (seg)	Máximos	Umbral	Tramas sin error	Mensaje	Repeticiones
Con tono en f=10000	w0	10,0000	13	32	13	ok	13
Con tono en f=15000	w0	10,0000	13	32	11	ok	11
Con tono en f=17000	w0	10,0000	13	28	12	ok	12
Con tono en f=17500	w0	10,0000	13	16	0	no	0
Con tono en f=18000	w0	10,0000	8	10	0	no	0
Con tono en f=18200	w0	10,0000	0	8	0	no	0

- *Ruido pasabanda:*

El ruido pasabanda que utilizamos tiene una ancho de banda de 4KHz. Realizamos dos pruebas. La primera con el ruido centrado en la frecuencia 10KHz y la segunda con el ruido centrado en 18.2KHz. Esta última es la frecuencia de BPSK utilizada en el sistema.

Ruido pasabanda	Prueba	Duración (seg)	Máximos	Umbral	Tramas sin error	Mensaje	Repeticiones
f=10000	w0	10,0000	13	32	13	ok	13
f=18200	w0	10,0000	11	20	4	ok	4

Conclusión:

En la primera prueba comprobamos que con tonos cercanos a la frecuencia de BPSK la marca no se detecta.

En la prueba con ruido pasabanda la marca la detectamos cuando el espectro del ruido estaba fuera del espectro de la marca como también cuando estaba dentro. Sin embargo, notamos que la cantidad de tramas detectadas disminuyó en una cantidad considerable cuando el espectro del ruido esta centrado en la frecuencia 18200Hz.

5.1.4 - Codificación MP3

Evaluamos el sistema convirtiendo los fragmentos marcados a formato MP3 y volviendo a convertirlos a formato wav para hacer la detección. La conversión MP3 la hicimos para diferentes Kbps. La prueba se realizó sólo para los fragmentos 1, 2 y 3 de la Tabla 5.1.

Los parámetros de MP3 que utilizamos en esta prueba fueron los siguientes:

- Mode: Mono
- Quality: Low (frecuency cut-off 16KHz)
- VBR(variable bit rate): None
- Encoder: Lame MP3 encoder (versión 1.2)

Resultados obtenidos:

- *A 128 Kbps:*

Fragmento	Prueba	Duración (seg)	Máximos	Umbral	Tramas sin error	Mensaje	Repeticiones
1	w0	30,3020	24	26	13	ok	13
1	w2	30,3020	33	22	14	ok	14
1	w4	30,3020	27	20	10	ok	10
3	w0	30,1453	25	28	15	ok	15
3	w2	30,1453	25	26	11	ok	11
3	w4	30,1453	21	24	10	ok	10
6	w0	30,2759	24	24	11	ok	11
6	w2	30,2759	30	20	17	ok	17
6	w4	30,2759	25	18	7	ok	6

- *A 96 Kbps:*

Fragmento	Prueba	Duración (seg)	Máximos	Umbral	Tramas sin error	Mensaje	Repeticiones
1	w0	30,3020	21	22	6	ok	6
1	w2	30,3020	22	20	2	ok	2
1	w4	30,3020	22	16	0	no	0
3	w0	30,1453	21	24	7	ok	7
3	w2	30,1453	26	20	2	ok	2
3	w4	30,1453	27	16	2	ok	2
6	w0	30,2759	29	20	11	ok	11
6	w2	30,2759	23	18	8	ok	8
6	w4	30,2759	31	14	4	ok	4

Conclusión:

El sistema soporta la codificación MP3. Observando las columnas de mensajes y repeticiones se concluye que lo más restrictivo que soporta el sistema sin perder la detección de la marca es MP3 a 96 kbps, a menor tasa de bits que 96 kbps, la marca no es detectada.

5.1.5 - Agregado de ruido blanco

A los fragmentos se les agregó ruido blanco con diferentes amplitudes con respecto al fragmento. Ruido blanco con un K=0.5 significa que el ruido tiene la mitad de la amplitud del fragmento correspondiente. Si el fragmento tiene una amplitud A, entonces el ruido tiene una amplitud de K*A.

El ruido utilizado es el resultado de multiplicar una señal con una distribución normal N(0,1) con la constante K*A. Por lo que la potencia del ruido será $K^{2*}A^2$. Para estimar la potencia de la señal tomamos como referencia la potencia de una sinusoide de amplitud A. Esta sinusoide tiene una potencia de A^2*0.5 y se considera

como la máxima potencia que puede llegar a tener un fragmento de audio. Se estima que en promedio el audio tiene una potencia de A^2*0.1.

Entonces tenemos: $SNR = 10 * \log\left(\frac{A^2 * 0.1}{K^2 * A^2}\right) = 10 * \log\left(\frac{0.1}{K^2}\right)$

Resultados obtenidos:

- *Con K=0.04 (SNR=17.9dB)*

Fragmento	Prueba	Duración (seg)	Máximos	Umbral	Tramas sin error	Mensaje	Repeticiones
1	w0	30,2771	21	14	0	no	0
2	w0	30,2076	25	24	15	ok	14
2	w2	30,2076	31	20	13	ok	13
2	w4	30,2076	29	18	6	ok	6
3	w0	30,1206	24	14	3	ok	3
3	w2	30,1206	21	12	1	ok	1
3	w4	30,1206	24	10	1	ok	1
4	w0	30,0262	36	18	4	ok	4
4	w2	30,0262	29	16	3	ok	3
4	w4	30,0262	33	12	0	no	0
5	w0	30,1013	23	22	10	ok	10
5	w2	30,1013	23	20	2	ok	2
5	w4	30,1013	26	16	4	ok	4
6	w0	30,2559	29	10	0	no	0

- *Con K=0.1 (SNR=10dB)*

Fragmento	Prueba	Duración (seg)	Máximos	Umbral	Tramas sin error	Mensaje	Repeticiones
1	w0	30,2771	11	8	0	no	0
2	w0	30,2076	29	12	2	ok	2
2	w2	30,2076	36	10	0	no	0
3	w0	30,1206	22	8	0	no	0
4	w0	30,0262	40	8	0	no	0
5	w0	30,1013	25	14	0	no	0
6	w0	30,2559	11	8	0	no	0

Conclusión:

Las tramas se empiezan a detectar con más facilidad con un ruido blanco de amplitud al 4% de la amplitud del fragmento (K=0.04). Con un ruido blanco a mayor amplitud, la detección es muy inferior como podemos observar en el segundo recuadro. En éste mostramos una prueba con ruido blanco al 1% y sólo se detectó una trama en el fragmento.

5.1.5 - Recodificación

La prueba consiste en aumentar la frecuencia de muestreo y volver a la frecuencia original para la detección (remuestreo). Esto simula una conversión DA/AD ideal. La frecuencia de muestreo de los fragmentos es de 44100Hz y se aumentó a una frecuencia de 48000Hz.

La prueba se realizó para los fragmentos 1, 3 y 6 de la Tabla 5.1.

Resultados obtenidos:

Fragmento	Prueba	Duración (seg)	Máximos	Umbral	Tramas sin error	Mensaje	Repeticiones
1	w0	30,2771	22	20	7	ok	7
1	w2	30,2771	21	18	4	ok	4
1	w4	30,2771	28	14	2	ok	2
3	w0	30,1206	26	20	7	ok	7
3	w2	30,1206	26	28	4	ok	4
3	w4	30,1206	26	16	4	ok	4
6	w0	30,2559	29	18	8	ok	8
6	w2	30,2559	28	16	4	ok	4
6	w4	30,2559	24	14	2	ok	2

Conclusión:

El sistema logra hacer una detección correcta en todos los fragmentos. Pero las tramas correctas disminuyeron considerablemente.

5.2 - Prueba de imperceptibilidad

Uno de los requerimientos planteados para el sistema de watermarking es mantener la calidad con la que se percibe la señal de audio al insertarle la marca; referido como imperceptibilidad de la marca. La imperceptibilidad de nuestro algoritmo de watermarking fue probada usando 4 señales de audio marcadas (W+2, W0, W-2, W-4), correspondientes a audio con la marca amplificada 2dB, sin amplificar, -2dB y -4dB respecto a la potencia de la señal obtenida por el filtrado psicoacústico. De esta manera podremos calibrar, si fuera necesario, el nivel de la marca que usará el sistema.

Como mecanismo de prueba de audición, utilizaremos el método ABX que consiste en hacer escuchar a cada oyente un fragmento de audio A (correspondiente a audio sin marca), un fragmento de audio B (correspondiente al mismo fragmento pero con la marca presente) y un fragmento X (que puede ser de forma aleatoria igual a A o igual a B). El oyente deberá decidir luego si X corresponde al fragmento A o al B. De acuerdo a las respuestas correctas se decidirá si el sistema cumple con el requisito de imperceptibilidad.

Para realizar esta prueba se preparó un CD de audio conteniendo 16 pistas, cada una contiene 4 tests ABX correspondientes a W+2, W0, W-2, W-4. Los fragmentos A, B y X de cada test duran cada uno 12 segundos y están separados por 2 segundos de silencio. Los tests de cada pista están separados con 8 segundos de silencio. El fragmento de audio utilizado es el mismo dentro de cada test (W+2, W0, W-2, W-4), pero distintos entre sí. Para el test W+2 se usó el fragmento 6, para W0 el 2, para W-2 el 5 y para W-4 el 4 (en la tabla 5.1 estan las referencias de los números de los fragmentos); con esto probamos el sistema con distintos estilos musicales y evitamos el cansancio en el oyente.

Con las 16 pistas se cubren todas las combinaciones posibles de X y mediante el generador de números aleatorios de una calculadora se decide qué pista se le

hará escuchar a cada oyente de forma de garantizar la aleatoriedad de X.

Para definir el número de oyentes de la prueba así como para el análisis de los resultados se utiliza la teoría explicada en el apéndice B, que analiza los resultados del test ABX mediante un test de hipótesis. En esta publicación se presentan fórmulas basadas en la distribución normal como una aproximación a la binomial; distribución más apropiada para modelar estadísticamente una población con características dicotómicas.

Dichas formulas relacionan parámetros utilizados en el test de hipótesis dentro de los cuales se destacan:

- n (Muestras de la población, corresponde al número de oyentes que realizará el test).

- Ho (Hipótesis nula. Las respuestas correctas obtenidas son debidas al azar. Bajo esta hipótesis es de esperar obtener 50% de respuestas correctas, de obtener un porcentaje superior será atribuido al azar).

- H1 (Hipótesis alternativa. Los resultados obtenidos son debidos a un factor distinto del azar. Bajo H1 un porcentaje mayor del 50% de respuestas correctas es atribuido a diferencias audibles).

- c (Número de eventos con una determinada característica, en el test ABX corresponde al número de respuestas correctas).

- p1 (Proporción de eventos en la población si se esta bajo Ho, p1 es 0.5).

- p2 (Proporción hipotética de eventos en la población si se está bajo H1. En el caso del test ABX debe ser mayor a 0.5, siendo aceptable un valor de 0.7).

- a (Riesgo de error de tipo 1, es la probabilidad de rechazar Ho cuando Ho es verdadero. Equivale a la probabilidad de que se produzcan c o más eventos en n muestras de la población debido al azar.)

- b (Riesgo de error de tipo 2, es la probabilidad de aceptar Ho cuando Ho es falso).

- c’ (Valor crítico de c. Para valores de c superiores a c' se rechaza Ho con a<a' donde a' es una cota preestablecida de a).

Como criterio de diseño del test se definen a'=0.1 y b'=0.1 como cotas de a y b, a partir de estos valores y de p1 y p2 se calcula el número de muestras n de la población. Según las fórmulas de aproximación se obtiene que n debe ser mayor a 37, por lo que se toma n = 42 para estar cubiertos (a cada oyente de los 42 se le hará escuchar una pista conteniendo los 4 test, W+2, W0, W-2, W-4). Teniendo el número de muestras se halla c'= 27. Con estos valores se obtiene a=0.08 y b=0.09 valores que verifican las restricciones planteadas (a<a' y b<b').

Resultados:

Test	Numero de muestras (n)	Identificaciones correctas (c)
W+2	42	34
W0	42	20
W-2	42	27
W-4	42	26

Conclusiones:

Comparando la cantidad de identificaciones correctas con el umbral calculado anteriormente (c'= 27) se observa que para los test W0, W-2 y W-4 estamos por debajo de dicho umbral, por lo que se acepta la hipótesis Ho; lo que equivale a concluir que las identificaciones correctas obtenidas fueron debidas al azar y no a una diferencia apreciable entre los fragmentos.

En cambio, para el test W+2 las identificaciones correctas superan ampliamente el umbral, por lo tanto se concluye que la diferencia entre el audio con y sin marca se podía apreciar.

Con estos resultados vemos que el modelo psicoacústico utilizado está correctamente calibrado. De todas maneras hay cierta incoherencia dado que se obtuvieron más identificaciones correctas en los test W-2 y W-4 que en el W0. Esta diferencia se puede atribuir: a que los distintos géneros musicales sean más o menos propicios para la percepción de la marca, o simplemente al azar.

5.3 - Comparación con otro sistema

El EyM audio watermarking es un software basado en Spread Spectrum que se encuentra a la venta en la página web: http://www.metois.com/Eymwatermark/eymawm.htm. En esta página web se hacen algunos comentarios sobre los resultados que obtuvieron con pruebas de imperceptibilidad y persistencia. Para probar la imperceptibilidad de la marca en el audio utilizaron el test ABX igual que nosotros. La conclusión a la que llegaron es que el sistema no es extremadamente imperceptible, un oído entrenado puede notar diferencias entre el audio marcado y el original.

Si se quiere ver el contenido completo de estas pruebas ir a la página web: http://www.metois.com/Eymwatermark/eymawmperf.htm.

Nos pareció interesante poder hacer algunas comparaciones con nuestro sistema. Para esto, hicimos algunas pruebas similares para comparar algunos resultados utilizando la versión demo que se encuentra disponible de forma gratuita en la página.

Resultados:

Fragmentos	Ganancia	Packets T	Packets R	MP3-96	MP3-64	RB(0,04)	RB(0,1)	REC
1	0	14	0	0	0	0	0	0
1	1	14	10	11	11	1	0	12
1	2	14	11	12	12	8	3	12
3	0	14	0	0	0	0	0	0
3	1	14	10	10	9	0	0	11
3	2	14	11	13	11	9	5	12
6	0	14	0	0	0	0	0	0
6	1	14	13	12	11	3	0	11
6	2	14	13	12	12	12	6	13

Las pruebas son las mismas realizadas en el capítulo de robustez. Para ver detalles sobre las mismas dirigirse a la sección mencionada.

Interpretación de la tabla:

- Fragmentos: Corresponden a los fragmentos de la Tabla 5.1
- Ganancia: Parámetro del sistema EyM. Con más ganancia se tiene más robustez pero menos imperceptibilidad y con menos ganancia lo contrario.
- Paquetes T: Los paquetes transmitidos en el audio. Paquetes es la forma con que denominan a las tramas en el sistema EyM.
- Paquetes R: Son los paquetes recuperados por el receptor.
- MP3-96: Prueba de Mp3 a 96Kbps
- MP3-64: Igual que la anterior pero con MP3 a 64Kbps.
- RB(0.04): Prueba de ruido blanco con K=0.04.
- RB(0.1): Misma prueba con K=0.1.
- REC: Prueba de recodificación

Para obtener una noción sobre la imperceptibilidad del sistema hicimos una pequeña prueba basándonos en el test ABX con los integrantes del proyecto. En la prueba utilizamos fragmentos de audio con marcas de ganancia uno y dos. Como conclusión de la prueba realizada se puede decir que se llegó a apreciar algunas diferencias que se hacían más perceptibles al aumentar la ganancia de la marca. Obviamente estas pruebas no tienen ningún rigor científico.

Conclusión:

No hay mucha información sobre cómo esta hecho el sistema EyM. Pero basándonos en la página web encontramos algunas similitudes con nuestro sistema.

Estas similitudes son el uso de Spread Spectrum y la inserción de varias tramas en el audio, también usaron el mismo test (ABX) para las pruebas de imperceptibilidad.

Con respecto a las pruebas realizadas el sistema EyM tiene una velocidad de procesamiento mucho mayor que la nuestra. En las pruebas de compresión MP3 y recodificación obtuvimos excelentes resultados. Sin embargo en la prueba con ruido blanco encontramos una fuerte baja en la detección de los paquetes. Sin poderlo demostrar científicamente, llegamos a la conclusión que el audio marcado con ganancia igual (y mayor que uno), tiene ciertas diferencias perceptibles respecto al audio original.

6 - Conclusiones

En el marco de este proyecto se ha realizado un estudio del watermarking aplicado al audio digital. Dicho estudio se llevó a cabo con el fin de seleccionar una técnica que nos permitiera la posterior implementación de un sistema. A partir de dichos estudios consideramos que la técnica más apropiada para dicho objetivo era el método Spread Spectrum aplicado conjuntamente con un modelado psicoacústico. Una vez seleccionada dicha técnica, se llevó a cabo la implementación del sistema y con el fin de evaluar este sistema se realizaron pruebas tanto de robustez como de imperceptibilidad.

El objetivo del proyecto consistía en implementar un sistema de watermarking que cumpliera ciertos niveles de robustez e imperceptibilidad. Podemos afirmar que este objetivo lo hemos cumplido ya que logramos implementar un sistema, descrito en este documento, que se desempeñara con los requerimientos exigidos. En una primera instancia se pensó que se llegaría a una solución más directa y eficiente, sin embargo, a lo largo de la implementación fueron surgiendo dificultades que nos mostraron los problemas que existen a la hora de desarrollar un sistema.

Como hemos mencionado, el objetivo del proyecto fue cumplido pero existen varias áreas, en las cuales se puede profundizar para perfeccionar el sistema. Dentro de éstas se destaca la generalización del algoritmo para distintos formatos y parámetros del audio, como lo son la frecuencia de muestreo, la cantidad de canales, etc. Otra área de investigación es la búsqueda e implementación de modelos psicoacústicos más precisos que posibiliten mejorar el compromiso entre robustez e imperceptibilidad. Dentro del aspecto robustez se puede estudiar la forma de reubicar la marca en un rango del espectro donde el oído sea más sensible, con el objetivo de dificultar su eliminación sin la perdida de calidad en el audio. Para que el sistema pueda ser utilizado comercialmente existen dos aspectos que son necesarios perfeccionar. Estos son el aumento de la velocidad de procesamiento del sistema y el manejo de tramas de mayor tamaño. Una de las posibilidades para mejorar la velocidad de procesamiento es el uso de lenguajes de programación más eficientes como por ejemplo lenguaje C. También se podría implementar dicho sistema en un DSP con el que se podría lograr una ejecución en tiempo real.

Como observación final cabe destacar que el estudio de watermarking aplicado al audio no tiene precedentes en el IIE. Debido a esto y a las necesidades de investigación expuestas en el párrafo anterior, el sistema desarrollado constituye una primera aproximación al tema en cuestión. Con lo cual se espera que los conceptos estudiados sirvan como base para futuras investigaciones.

Apéndice

A - Teoría de Spread Spectrum

El Spread Spectrum (también llamado espectro ensanchado, espectro esparcido, espectro disperso, o SS) es una técnica por la cual la señal transmitida se ensancha a lo largo de una banda de frecuencias, mucho más amplia que el ancho de banda mínimo requerido para transmitir la información que se quiere enviar. No se puede decir que las comunicaciones mediante Spread Spectrum son medios eficientes de utilización del ancho de banda. Sin embargo, rinden al máximo cuando se los combina con sistemas existentes que hacen uso de la frecuencia. La señal de Spread Spectrum, una vez ensanchada puede coexistir con señales en banda estrecha, ya que sólo les aportan un pequeño incremento en el ruido. En lo que se refiere al receptor de Spread Spectrum, él no ve las señales de banda estrecha, ya que está escuchando un ancho de banda mucho más amplio gracias a una secuencia de código preestablecido.

Podemos concluir diciendo que todos los sistemas de Spread Spectrum satisfacen dos criterios:

- El ancho de banda de la señal que se va a transmitir es mucho mayor que el ancho de banda de la señal original.

- El ancho de banda transmitido se determina mediante alguna función independiente del mensaje y conocida por el receptor.

A continuación, se presentan cinco técnicas de Spread Spectrum:

Sistemas de secuencia directa

La *secuencia directa* es quizás uno de los sistemas de Spread Spectrum más ampliamente conocido, utilizado y relativamente sencillo de implementar. Una portadora en banda estrecha se modula mediante una secuencia pseudo-aleatoria. Para la secuencia directa, el incremento de ensanchado depende de la tasa de bits de la secuencia pseudo-aleatoria por bit de información. En el receptor, la información se recupera al multiplicar la señal con una réplica generada localmente de la secuencia de código.

Sistemas de salto de frecuencia

En los sistemas de *salto de frecuencia*, la frecuencia portadora del transmisor cambia (o salta) abruptamente de acuerdo con una secuencia pseudo-aleatoria. El orden de las frecuencias seleccionadas por el transmisor viene dictado por la secuencia de código. El receptor realiza los mismos saltos en frecuencia que el transmisor para detectar la señal.

Sistemas de salto temporal

Un sistema de *salto temporal* es un sistema de Spread Spectrum en el que el período y el ciclo de trabajo de una portadora se varían de forma pseudo-aleatoria bajo el control de una secuencia pseudo-aleatoria. El salto temporal se usa a menudo junto con el salto en frecuencia para formar un sistema híbrido de Spread Spectrum mediante acceso múltiple por división de tiempo (TDMA).

Sistemas de frecuencia modulada pulsada (o *Chirping*)

Se trata de una técnica de modulación en Spread Spectrum menos común que las anteriores, en la que se emplea un pulso que barre todas las frecuencias, llamado chirp, para expandir la señal espectral. El *chirping*, como también es conocido, suele usarse más en aplicaciones con radares que en la comunicación de datos.

Sistemas híbridos

Los *sistemas híbridos* usan una combinación de métodos de Spread Spectrum para beneficiarse de las propiedades más ventajosas de los sistemas utilizados. Dos combinaciones comunes son secuencia directa y salto de frecuencia. La ventaja de combinar estos dos métodos está en que adopta las características que no están disponibles en cada método por separado.

A.1 - Watermarking con SS/secuencia directa usando una señal modulada en BPSK

Modelo y Parámetros Fundamentales

En el análisis siguiente, el proceso de generar una marca que sea insertada en una señal de audio se expresa en terminología de Spread Spectrum. La señal audio original será llamada "ruido" y la corriente de bits que conforma la marca será la señal de datos. Este proceso de agregar ruido a un canal o a una señal se llama "jamming". El objetivo de un jammer en un sistema de comunicación es degradar la performance de la transmisión, utilizando el conocimiento del sistema de comunicación. En el algoritmo de marcado la señal audio (es decir la música) se considera el jammer, y tiene mucho más energía que la corriente de bits transmitida (marca).

Un sistema básico se muestra en la Figura A.1, con los siguientes parámetros:

Wss = Ancho de banda disponible de la señal Spread Spectrum
Rb = Data rate (bits/segundo)
S = Energía de la señal (en la entrada del receptor)
J = Energía del Jammer (en la entrada del receptor)

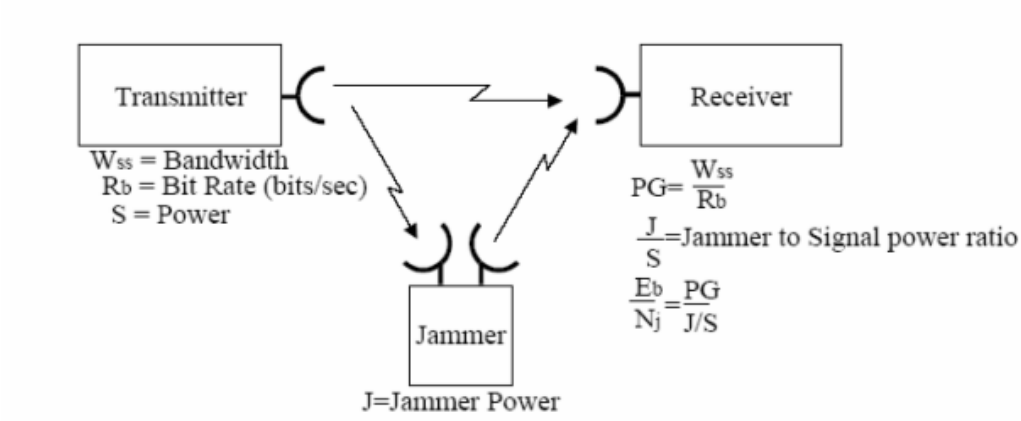


Figura A.1 – Sistema de comunicación básico de Spread Spectrum

Se define Wss como el ancho de banda disponible del Spread Spectrum que puede utilizar el transmisor. El Rb es el bit data rate sin codificar usado durante la transmisión. La energía de la señal S y la energía del jammer J son la energía promedio en el receptor.

Esparcimento BPSK de la secuencia directa

El esparcimiento con modulación Binary Phase-Shift-Keying de la secuencia-directa se conoce como DS/BPSK. Puede ser explicado con un ejemplo simple. Las señales de BPSK se expresan a menudo como:

$$s(t) = \sqrt{2S}\, sen\left[w_0 t + \frac{d_n \pi}{2} \right] \qquad (A.1)$$

$$nT_b \le t < (n+1)T_{b,} \qquad n = entero$$

Donde T_b es el tiempo de bit: $\left(\frac{1}{R_b} \right)$

d_n es la secuencia de bits de datos, con los valores posibles de 1 o -1; y probabilidad igual de ocurrencia.

Ec. A.1 se puede expresar cómo:

$$s(t) = d_n \sqrt{2S} \cos(w_0 t) \qquad (A.2)$$

$$nT_b \le t < (n+1)T_b, \qquad n = entero$$

BPSK se puede ver como modulación de fase en Ec.A.1 o modulación de amplitud en Ec.A.2. El espectro de una señal de BPSK está generalmente de la forma mostrada en la Figura A.2.

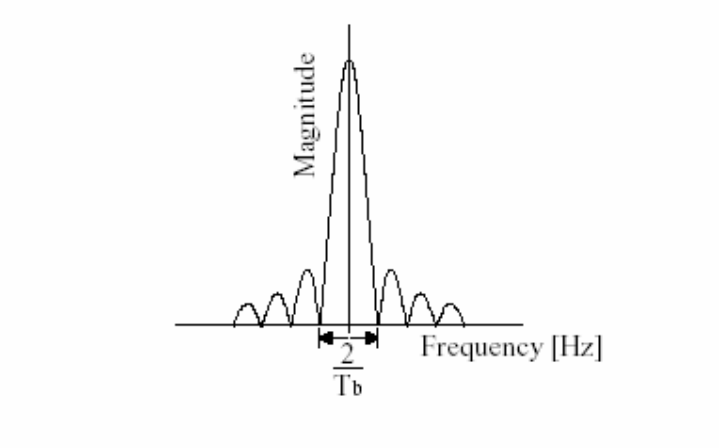


Figura A.2 – Espectro de la señal BPSK

Esto es una función como $(\sin^2 x)/x^2$ y el primer cruce por cero es en 1/Tb. Esto muestra la mínima banda que se necesitaría para transmitir la señal s(t) y recuperarla en el receptor.

La teoría de Spread Spectrum requiere que la señal sea esparcida en un espectro más grande que el mínimo necesario para la transmisión. El esparcimiento de la secuencia directa se hace usando una secuencia binaria pseudo-aleatoria (PN) llamada {c}. Los valores de esta secuencia son 1 o -1 y su velocidad es N veces más rápida que los datos {d}. El tiempo Tc, de cada bit de la secuencia PN esta dado como:

$$T_c = \frac{T_b}{N} \qquad (7.3)$$

El esparcimiento directo de la secuencia de la señal Spread Spectrum tiene la forma:

$$x(t) = \sqrt{2S}\, sen\left[w_0 t - d_n c_{nN+k}\, \pi/2\right] = d_n c_{nN+k} \sqrt{2S}\cos(w_0 t) \qquad (A.4)$$

$$nT_b + kT_c \le t < nT_b + (k+1)T_c$$
$$k = 0,1,2...N-1$$
$$n = entero$$

La señal es muy similar al BPSK común, excepto que el bit rate es N veces más rápido y la densidad espectral de potencia es N veces más larga, según se muestra en la Figura A.3.

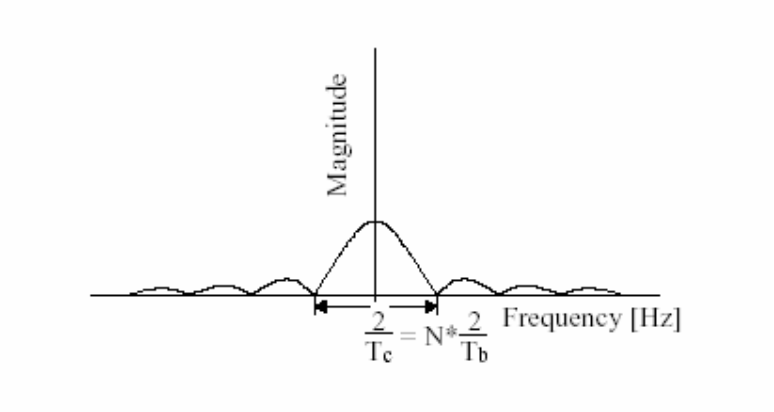


Figura A.3 – Espectro de la señal BPSK después del Spreading

La ganancia esta dada por:

$$PG = \frac{W_{SS}}{R_b} = N \qquad (A.5)$$

Wss es el ancho de banda de la secuencia directa de Spread Spectrum:

$$\frac{1}{T_c} = N\frac{1}{T_b}$$

Si se define la función de datos como:

$$d(t) = d_n, \qquad nT_b \le t < (n+1)T_b \qquad (A.6)$$
$$n = entero$$

y la secuencia PN es:

$$c(t) = c_k, \qquad KT \le t < (k+1)T_c \qquad (A.7)$$
$$k = entero$$

Ec. A.4 se puede ampliar como:

$$x(t) = \sqrt{2S}\,sen\left[w_0 t + c(t)d(t)\,\pi/2\right] = c(t)d(t)\sqrt{2S}\cos(w_0 t) \qquad (A.8)$$

La Figura A.4 muestra el diagrama de bloques de la modulación normal de DS/BPSK.

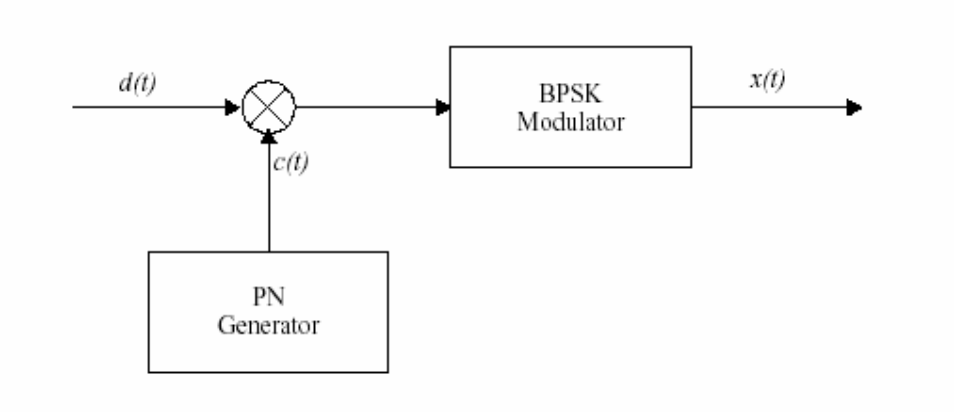


Figura A.4 – DS/BPSK modulación

Y la Figura A.5 muestra un modelo equivalente usado en el paso siguiente del análisis.

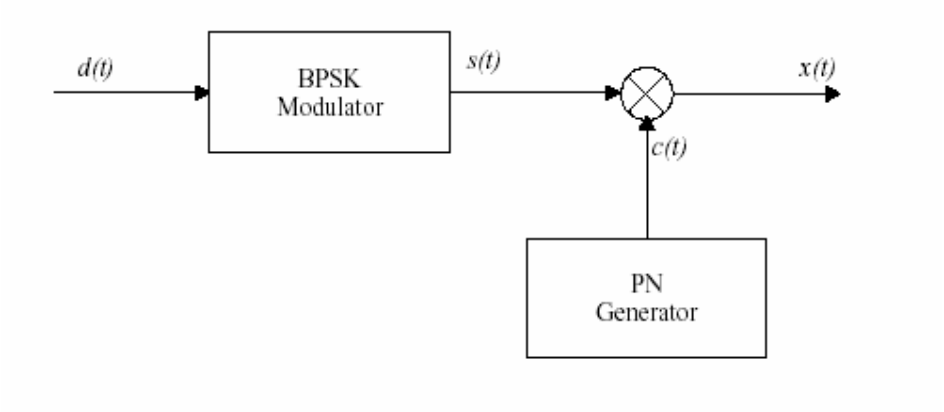


Figura A.5 – DS/BPSK modificado

La Figura A.6 muestra la señal d(t) y c(t):

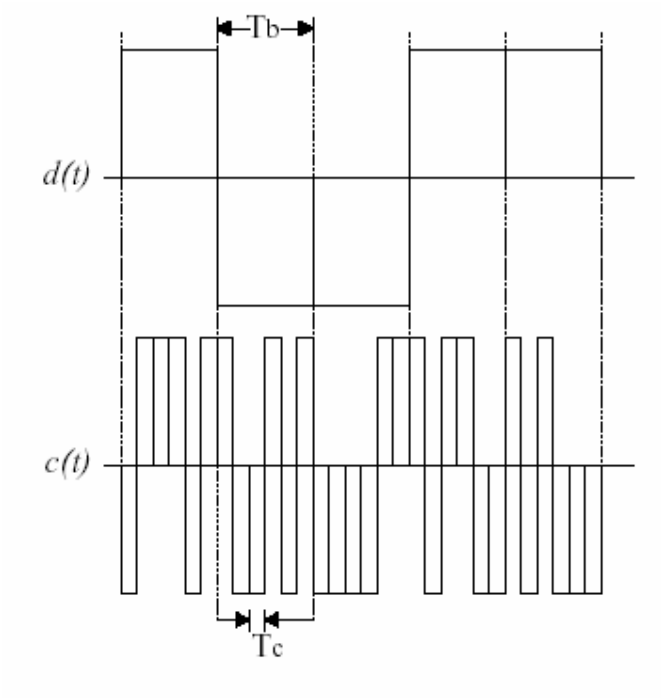


Figura A.6 – Señales antes del Spreading

Y la Figura A.7 muestra la señal c(t)d(t) con N=6:

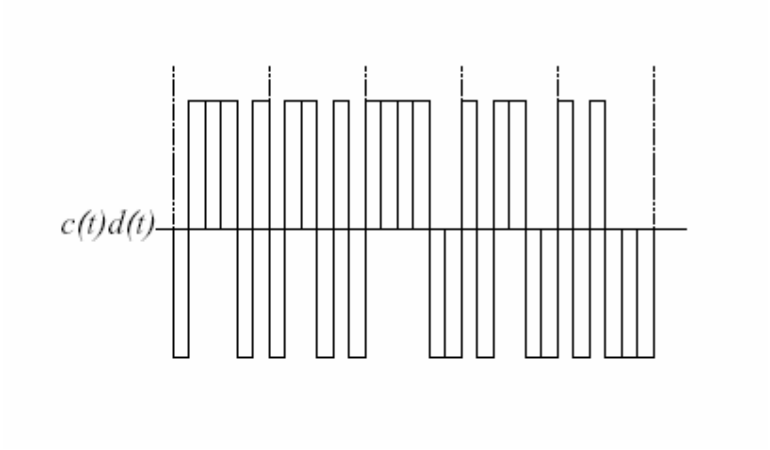


Figura A.7 – Señal después del Spreading

De la Figura A.5, la forma equivalente de x(t) esta dada por:

$$x(t) = c(t)s(t) \qquad (A.9)$$

Donde

$$s(t) = d(t)\sqrt{2S}\cos(w_0 t) \qquad (A.10)$$

Ésta es la señal original de BPSK. La característica:

$$c^2(t) = 1 \qquad \forall t \qquad (A.11)$$

Este es el punto clave usado para recuperar la señal original de BPSK:

$$c(t)x(t) = s(t) \qquad (A.12)$$

Si el receptor posee una copia de la secuencia PN y puede sincronizar la copia local con la señal recibida x(t), es posible de-Spread la señal y recupera los datos transmitidos.

Jammer con ruido de potencia constante sobre toda la banda de frecuencia

Un jammer, J(t), con la potencia constante J se muestra en la Figura A.8:

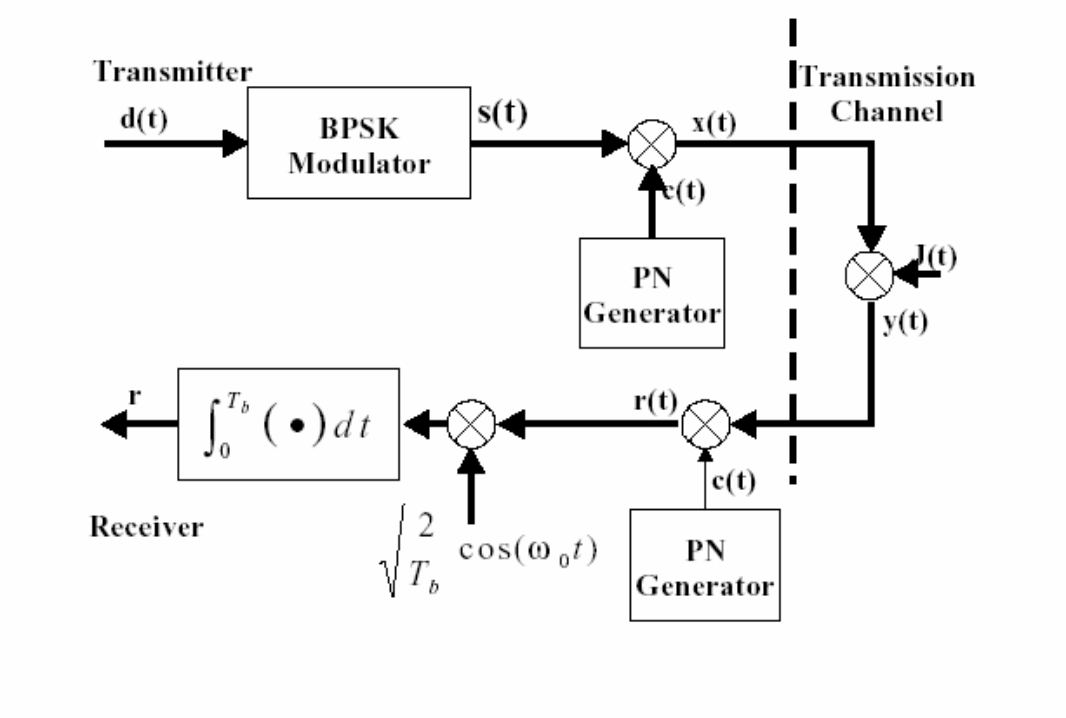


Figura A.8 – DS/BPSK

El sistema también asume que no tiene ningún ruido en el canal de transmisión. Se asume un demodulador ideal de BPSK después de que la señal recibida y(t) sea multiplicada por la secuencia PN. La salida del canal es:

$$y(t) = x(t) + J(t) \qquad (A.13)$$

Ésta es multiplicada por la secuencia PN c(t):

$$r(t) = c(t)y(t) = c(t)x(t) + c(t)J(t) = s(t) + c(t)J(t) \qquad (A.14)$$

Este término muestra la señal original de BPSK más el ruido dado por c(t)J(t). La salida del detector convencional de BPSK es entonces:

$$r = d\sqrt{E_b} + n \qquad (A.15)$$

Donde:

d es el bit de datos para el intervalo actual Tb.
Eb=STb es la energía del bit.
n es la componente equivalente del ruido.

n se define como:

$$n = \sqrt{\frac{2}{T_b}} \int_0^{T_b} c(t) J(t) \cos(w_0 t) dt \qquad (A.16)$$

La regla de decisión para BPSK es:

$$\hat{d} = \begin{cases} 1, & si \quad r > 0 \\ -1, & si \quad r \leq 0 \end{cases} \qquad (A.17)$$

A.2 - Ventajas y desventajas

El Spread Spectrum tiene muchas propiedades únicas y diferentes que no se pueden encontrar en ninguna otra técnica de modulación. Para verlo mejor, se listan debajo algunas ventajas y desventajas que existen en los sistemas típicos de espectro ensanchado:

Ventajas

- Resiste todo tipo de interferencias, tanto las no intencionadas como las malintencionadas (más conocidas con el nombre de jamming), siendo más efectivo con las de banda estrecha.

- Tiene la habilidad de eliminar o aliviar el efecto de las interferencias multisenda.

- Se puede compartir la misma banda de frecuencia con otros usuarios.

- Confidencialidad de la información transmitida gracias a los códigos pseudo-aleatorios (multiplexación por división de código).

Desventajas

- Ineficiencia del ancho de banda.

- La implementación de los circuitos es en algunos casos muy compleja.

Approximation Formulas for Error Risk and Sample Size in ABX Testing*

HERMAN BURSTEIN

New College of Hofstra, Hempstead, NY 11550, USA

When sampling from a dichotomous population with an assumed proportion p of events having a defined characteristic, the binomial distribution is the appropriate statistical model for accurately determining: type 1 error risk α; type 2 error risk β; sample size n based on specified α and β and assumptions about p; and critical c (minimum number of events to satisfy a specified α). Table 3 in [1] presents such data for a limited number of sample sizes and p values. To extend the scope of Table 3 to most n and p, we present approximation formulas of substantial accuracy, based on the normal distribution as an approximation of the binomial.

0 INTRODUCTION

This paper is principally an extension of Table 3 in the paper by Leventhal [1], which correctly stresses the frequent importance of considering type 2 error risk as well as type 1 in testing a sample from a dichotomous population. His paper is particularly relevant, but not limited, to the popular ABX listening tests, where subjects are asked to distinguish between two similar audio components, such as power amplifiers.

Leventhal's Table 3, based on the binomial distribution as the appropriate statistical model, provides the following for 16 selected sample sizes from 10 to 150: type 1 error risk α based on varying numbers of correct responses r and on the assumption that the proportion of correct responses p in the population is 0.5, due to chance alone (when the listening test compares two components, as in the ABX case); and type 2 error risk β based on r and on five p values ranging from 0.6 to 0.9—namely, hypothesized values of p ascribed to audible differences between audio components. These are one-tail risks.

Our purpose is to provide approximation formulas that extend the scope of Leventhal's Table 3 to most values of sample size n and proportion p. These formulas are used with Table A to determine error risk, sample size, and critical c (minimum number of correct responses to satisfy a specified α).

* Manuscript received 1988 January 5; revised 1988 September 1.

The formulas are based on the normal distribution as an approximation of the binomial distribution, with a continuity correction for the normal deviate [2]. We use the symbol c for the number of correct responses (instead of r, used by Leventhal). The continuity correction replaces c by $c + 0.5$ or by $c - 0.5$, as appropriate. The standard error is $\sqrt{np(1-p)}$. Thus the standard normal deviate z is $(np - c - 0.5)/\sqrt{np(1-p)}$. Table A provides values of z and the corresponding error risk (one tail).

For sample sizes of 15 or more, and p in the neighborhood of 0.5, the normal approximation supplies quite accurate indications of error risk. Accuracy remains good enough for practical purposes if p is not under 0.1 and not over 0.9. Calculated error risk then is generally inaccurate by no more than one unit in the second decimal place. Accuracy increases with sample size and also as the value of error risk increases, so that a calculated error risk of 0.05 or more tends to be sufficiently accurate for the purpose at hand. Examples accompanying the formulas compare approximations with exact values derived from the binomial distribution.

It is desirable at the outset to define symbols and terms. To an extent they differ from those used by Leventhal; when necessary, fill in omissions.

0.1 Definitions

H_0 = null hypothesis. Results of a test based on a sample are due to chance alone. In the ABX case, on average a correct response rate of 0.5 is expected due to chance. Under H_0, a

rate over 0.5 is attributed to chance.

H_1 = alternative hypothesis. Results of a test based on a sample are due to a factor other than chance. In the ABX case, under H_1, a rate over 0.5 is attributed to audible differences. To support H_1 it is necessary to discredit H_0. H_0 is discredited if significance level α is "sufficiently low." The term "sufficiently low" is judgmental and defined by whatever criterion of significance α' is arbitrarily chosen to evaluate the test results. To illustrate, if the chosen criterion is 0.10 and if α is 0.08, H_0 is discredited and H_1 supported. (See below for a discussion of α and α'.)

n = sample size.

c = number of events with a defined characteristic in a sample. In the ABX case, number of correct responses.

c/n = proportion of events in sample. In the ABX case, correct-response rate. See Note 1 below.

p = proportion of events in population (which is infinite or vastly larger than n). p is a general term; see p_1 and p_2 below.

p_1 = proportion of events in population if chance alone is operating. In the ABX case, p_1 is 0.5 inasmuch as two components are compared; for three, p_1 is 0.333; and so on.

p_2 = effect size. Hypothesized proportion of events in population if a factor other than chance is operating. In the ABX case, p_2 exceeds 0.5 (for example, 0.6, 0.7, etc.). p_2 is a factor in determining sample size. For specified type 1 and type 2 error risks (see below), required n increases as the difference between p_2 and p_1 decreases. p_2 may be chosen as the smallest p for which type 2 error risk should be satisfactorily low. To illustrate, type 2 risk may be unimportant, and therefore permissibly high, if p is under 0.7; but it may be deemed important, and therefore requisitely low, if p is 0.7 or more. Then p_2 would be 0.7.

α = significance level. Probability of c or more events in a sample of size n due to chance alone. In the ABX case, chance corresponds to $p_1 = 0.5$.

α' = criterion of significance, also called type 1 error risk. It is a specified value of α for the following purposes. As the criterion of significance, it serves to decide whether to accept or reject H_0. If $\alpha \leq \alpha'$, H_0 is rejected (and H_1 accepted). Otherwise H_0 is accepted (and H_1 rejected). Type 1 error signifies rejecting H_0 when H_0 is actually true. Type 1 error risk is the probability of doing so, and is equal to the value chosen as α'. To illustrate, if 0.1 is selected as α', the probability of committing type 1 error is 0.1. α' is a factor in determining sample size, along with β' (see below), p_1, and p_2. As α' decreases, n increases.

β = type 2 error risk. Type 2 error signifies accepting H_0 when H_0 is actually false. Type 2 error risk is the probability of doing so. For a given p_2 it is the probability of fewer than c' events (see below) in a sample of size n. $1 - \beta$ is the power of the test. For a given p_2 it is the probability of rejecting H_0 when H_0 is false. In the ABX case, power is the probability of c' or more correct responses in n trials for a specified p_2.

β' = *specified* type 2 error risk. β' is a factor in determining a sample size sufficient to keep actual type 2 error risk at or below the specified value. As β' decreases, n increases. $1 - \beta'$ is the specified power; for a given p_2 it is the probability of rejecting H_0 when H_0 is false.

c' = critical c. Minimum value of c which, together with n and p_1, can produce significance level α equal to or less than the specified criterion of significance α'. c' is also the minimum value of c which, together with n and p_2, can produce type 2 error risk β no greater than specified β', and, by the same token, power at least as great as specified $1 - \beta'$.

Table A. Values of error risk and corresponding z, based on normal distribution.*

Error Risk	z	Error Risk	z
.0010	3.09	.10	1.28
.0015	2.97	.11	1.23
.0020	2.88	.12	1.18
.0025	2.81	.13	1.13
.0030	2.75	.14	1.08
.0040	2.65	.15	1.04
.0050	2.58	.16	.99
.0060	2.51	.18	.92
.0070	2.46	.20	.84
.0080	2.41	.22	.77
.0090	2.37	.24	.71
.0100	2.33	.26	.64
.0125	2.24	.28	.58
.0150	2.17	.30	.52
.0175	2.11	.32	.47
.020	2.05	.34	.41
.025	1.96	.36	.36
.030	1.88	.38	.31
.035	1.81	.40	.25
.040	1.75	.42	.20
.045	1.70	.44	.15
.05	1.65	.46	.10
.06	1.56	.48	.05
.07	1.48	.50	0.00
.08	1.41		
.09	1.34		

* Error risk is Q, tail probability of the normal distribution. For positive values of z, z is the standard normal deviate. For an approximation of the cumulative binomial distribution, we employ $z = (|np - c| - .05)/\sqrt{np(1-p)}$, where c is the number of events in a sample, n is the sample size, and p is the proportion of events in the population. When z is negative, error risk is $1 - Q$. To illustrate, $z = -0.84$ corresponds to error risk of $1 - 0.20 = 0.80$. Similarly, to find z for error risk Q above 0.50, we obtain z for $1 - Q$ and sign it negative. To illustrate, if error risk is 0.80, we find that z for $1 - 0.80 = 0.84$, and we sign it -0.84.

The approximation formulas are as follows

1) Significance level α, based on c, n, and p_1.
2) Type 2 error risk β, based on c', n, and p_2.
3) Critical c (c'), based on n, p_1, and α'.
4) Implied effect size p_2, based on c', n, and β'.
5) Sample size n, based on p_1, p_2, α', and β'.
6) Sample size n, disregarding β'.
7) Sample size n, disregarding β' and p_2.

Note 1: c/n is an estimate of the population correct-response rate. But it should be distinguished from the known-response rate k/n, which estimates the proportion of trials where correct responses are based solely on knowledge and not on guessing. k is the number of trials where a correct response is due to knowledge. In the remaining trials $n - k$, half would result in correct responses by chance, as in the ABX case. Altogether, $c = k + 0.5(n - k)$ in the ABX case. Solving for k and dividing by n, we obtain for a two-component listening test

$$k/n = 2c/n - 1, \qquad \text{but} \geq 0 . \tag{A}$$

To illustrate, if a subject achieves 10 correct responses in 16 ABX trials, $c/n = 0.625$, whereas $k/n = (2 \times 10/16) - 1 = 0.25$. The estimated known-response rate of 0.25 is in sharp contrast with the estimated correct-response rate of 0.625.

Note 2: Some believe it is more accurate to conclude "H_0 is not accepted" rather than "H_0 is rejected." However, this issue is unimportant here because the approximation formulas apply equally well to either interpretation.

1 APPROXIMATION OF SIGNIFICANCE LEVEL α

Compute

$$z = \frac{c - 0.5 - np_1}{\sqrt{np_1(1 - p_1)}} \tag{1}$$

where

c = number of correct responses
n = sample size
p_1 = proportion of correct responses in population due to chance alone. In the ABX case, 0.5.

Refer z to Table A to find the corresponding α.

Example: $n = 16$, $c = 12$, and $p_1 = 0.5$,

$$z = \frac{12 - 0.5 - 8}{\sqrt{16 \times 0.5 \times 0.5}} = 1.75.$$

Referring z to Table A, the corresponding significance level α is 0.040.

Comments: The exact value of α, derived from the binomial distribution is 0.038. (That is, for $p = 0.5$, the binomial probability of 12 or more events in a sample of size 16 is 0.038.) Leventhal's Table 3 shows $\alpha = 0.0384$.

Inasmuch as p_1 is always 0.5 in ABX tests, Eq. (1) can be simplified by substituting 0.5 for p_1, reducing the formula to $z = (2c - n - 1)/\sqrt{n}$. However, Eq. (1) is the general case, permitting use of values other than $p_1 = 0.5$ when circumstances require.

2 APPROXIMATION OF TYPE 2 ERROR RISK β

Compute

$$z = \frac{np_2 - c' + 0.5}{\sqrt{np_2(1 - p_2)}} \tag{2}$$

where

c' = critical c; minimum c to satisfy α'
n = sample size
p_2 = effect size; hypothesized proportion of correct responses in population due to audible differences.

Refer z to Table A to find the corresponding β.

Example: $n = 16$, $c' = 12$, and $p_2 = 0.8$,

$$z = \frac{16 \times 0.8 - 12 + 0.5}{\sqrt{16 \times 0.8 \times 0.2}} = 0.81.$$

Referring z to Table A, the corresponding type 2 error risk β is between 0.20 and 0.22, but slightly nearer to 0.20. If we interpolate between these two values, we obtain $\beta = 0.209$.

Comment: The exact value of β, derived from the binomial distribution, is 0.202. Power, or $1 - \beta$, is 0.798; that is, for $p_2 = 0.8$, the probability of fewer than 12 events in 16 trials is 0.202, while the probability of 12 or more is 0.798. Leventhal's Table 3 shows $\beta = 0.2018$.

3 APPROXIMATION OF CRITICAL c (c')

Compute

$$c' = z\sqrt{np_1(1 - p_1)} + 0.5 + np_1 \tag{3}$$

rounded up to nearest integer, where

n = sample size
p_1 = proportion of correct responses in population due to chance alone. In the ABX case, 0.5
z = Table A value corresponding to specified criterion of significance (type 1 error risk) α'.

Example: $n = 16$, $p_1 = 0.5$, and $\alpha' = 0.05$. Referring error risk of 0.05 to Table A, the corresponding value of z is 1.65.

$$c' = 1.65 \times \sqrt{16 \times 0.5 \times 0.5} + 0.5 + 16 \times 0.5 = 11.8.$$

Rounding up to the nearest integer, $c' = 12$.

Comment: 12 is the correct value for c'. When $c = 12$, $n = 16$, and $p_1 = 0.5$, the exact significance level α is 0.038, which matches the specification $\alpha' = 0.05$ as nearly as integers for c and n allow. If we try $c' = 11$ instead, we obtain $\alpha = 0.105$, which considerably exceeds the specified type 1 error risk of 0.05. Leventhal's Table 3 also shows $c' = 12$ (labeled r in his table).

4 APPROXIMATION OF IMPLIED EFFECT SIZE p_2

Compute

$$p_2 = \frac{b + \sqrt{b^2 - 4ad^2}}{2a} \qquad (4)$$

where

c' = critical c; minimum c to satisfy α'
n = sample size
β' = specified type 2 error risk
z = Table A value corresponding to β'
$a = z^2n + n^2$
$b = z^2n + 2dn$
$d = c' - 0.5$.

Example: $c' = 20$, $n = 30$, and $\beta' = 0.10$. Referring error risk of 0.10 to Table A, the corresponding value of z is 1.28,

$$d = 20 - 0.5 = 19.5$$
$$a = 1.28^2 \times 30 + 30^2 = 949.15$$
$$b = 1.28^2 \times 30 + 2 \times 19.5 \times 30 = 1219.15$$

and

$$p_2 = \frac{1219.15 + \sqrt{1219.25^2 - 4 \times 949.15 \times 19.5^2}}{2 \times 949.15} = 0.751 .$$

Comment: An implied effect size of 0.751 is quite accurate. The exact effect size, derived from the binomial distribution, is 0.7524. For $p_2 = 0.75$, the exact probability of fewer than 20 correct responses in 30 trials is 0.106, which very closely matches the specification $\beta' = 0.10$. Leventhal's Table 3 shows $\beta = 0.1057$ for an effect size of 0.75.

5 APPROXIMATION OF SAMPLE SIZE n, BASED ON p_1, p_2, α', AND β'

Compute

$$n = \left[\frac{z_1\sqrt{p_1(1 - p_1)} + z_2\sqrt{p_2(1 - p_2)}}{p_2 - p_1}\right]^2 , \qquad (5)$$

an integer, where

p_1 = proportion of correct responses in population due to chance alone. In the ABX case, 0.5
p_2 = effect size; hypothesized proportion of correct responses in population due to audible differences
z_1 = Table A value corresponding to specified criterion of significance (type 1 error risk) α'
z_2 = Table A value corresponding to specified type 2 error risk β'.

Example:
$p_1 = 0.5$, $p_2 = 0.75$, $\alpha' = 0.05$, and $\beta' = 0.10$. Referring error risk of 0.05 to Table A, the corresponding value of z is 1.65; referring 0.10 to Table A, z is 1.28,

$$n = \left[\frac{1.65 \times \sqrt{0.5 \times 0.5} + 1.28 \times \sqrt{0.75 \times 0.25}}{0.75 - 0.5}\right]^2 = 30.4 .$$

Rounding to the nearest integer, $n = 30$.

Comments: Applying Eq. (3) to $n = 30$, $p_1 = 0.5$, and $\alpha' = 0.05$, we obtain $c' = 20$. For $c' = 20$, $n = 30$, and $p_1 = 0.5$, the significance level is 0.050 from Eq. (1) or 0.049 from the binomial distribution. These closely match the specification $\alpha' = 0.05$. For $c' = 20$, $n = 30$, and $p_2 = 0.75$, the type 2 error risk is 0.103 from Eq. (2) or 0.106 from the binomial distribution. These closely match the specification $\beta' = 0.10$. In sum, $n = 30$ is verified as the appropriate sample size. Leventhal's Table 3 is in agreement.

Eq. (5) is the recommended method of determining sample size because it takes into account type 2 error risk as well as type 1. When type 2 error risk is of little or no consequence, alternative sample size formulas are Eqs. (6) and (7).

6 APPROXIMATION OF SAMPLE SIZE n, DISREGARDING β'

Compute

$$n = \frac{L + \sqrt{L^2 - M}}{2M} , \qquad (6)$$

an integer, where

z_1 = Table A value corresponding to specified criterion of significance (type 1 error risk) α'
p_1 = proportion of correct responses in population due to chance alone
p_2 = effect size; hypothesized proportion of correct responses in population due to audible differences
$L = z_1^2p_1(1 - p_1) + p_2 - p_1$

$$M = (p_2 - p_1)^2.$$

Example: $p_1 = 0.5$, $p_2 = 0.6$, and $\alpha' = 0.05$. Referring error risk of 0.05 to Table A, the corresponding value of z is 1.65,

$$L = 1.65^2 \times 0.5 \times 0.5 + 0.6 - 0.5 = 0.780625$$

$$M = (0.6 - 0.5)^2 = 0.01$$

and

$$n = \frac{0.780625}{2 \times 0.01} + \frac{\sqrt{0.780625^2 - 0.01}}{2 \times 0.01} = 77.7 .$$

Rounding to the nearest integer, $n = 78$.

Comments: When Eq. (6) is used, the relationships are such that, rounded up to the nearest integer, $c' = np_2$. Thus $c' = 78 \times 0.6 = 47$. For c', n, and p_1 in the example, the significance level is 0.045 from Eq. (1), matching the specification $\alpha' = 0.05$ as nearly as integers for c' and n allow.

Applying Eq. (2) to c', n, and p_2, type 2 error risk β is 0.47. If this is unacceptably high, Eq. (5) should be used to determine n. Alternatively, one could keep n at 78 and trade an increase in α' for a reduction in β. For example, if one experimentally chooses 44 as c', Eq. (1) shows that type 1 error risk becomes 0.15, while Eq. (2) shows that type 2 error risk becomes 0.22.

7 MINIMUM SAMPLE SIZE n, DISREGARDING β' AND p_2

Compute

$$n = \frac{\log \alpha'}{\log p_1} \qquad (7)$$

rounded up to the nearest integer, where

α' = specified criterion of significance (type 1 error risk)

p_1 = proportion of correct responses in population due to chance alone.

Example: $\alpha' = 0.05$ and $p_1 = 0.5$,

$$n = \frac{\log 0.05}{\log 0.5} = \frac{-1.30103}{-0.30103} = 4.32 .$$

Rounding up to the nearest integer, $n = 5$.

Comments: In the special case of minimum sample size, $c' = n$; that is, the correct-response rate must be 1 in order for α to satisfy the specified type 1 error risk α'. If $c = c'$, the exact significance level is p_1^n. In the example, if $c = 5$, $\alpha = 0.5^5 = 0.03125$, which satisfies $\alpha' = 0.05$. If we try $n = 4$ instead, and if $c = c' = 4$, $\alpha = 0.5^4 = 0.0625$, which fails to satisfy α'.

The procedure of Eq. (7) assumes that the investigator is willing to accept high type 2 error risk β. To illustrate, assume $n = 5$ and that the hypothesized effect size p_2 is 0.75. Exact $\beta = 1 - p_2^n = 1 - 0.75^5 = 0.763$.

8 ACKNOWLEDGMENT

The author is indebted to an anonymous reviewer for helpful comments and suggestions.

9 REFERENCES

[1] L. Leventhal, "Type 1 and Type 2 Errors in the Statistical Analysis of Listening Tests," *J. Audio Eng. Soc.*, vol. 34, pp. 437–453 (1986 June).

[2] T. Yamane, *Statistics, An Introductory Analysis*, 3rd ed. (Harper & Row, New York, 1973), pp. 712–716, 141–143.

APPENDIX A
PROGRAM FOR THE CUMULATIVE BINOMIAL PROBABILITY

For those with access to a personal computer or a pocket calculator programmable in BASIC, the following is a program for computing the cumulative binomial probability. Inputs are c, n, p, namely, the number of events in a sample, the sample size, and the proportion of events in the population. The output, labeled B in the program, is the probability of c or more events. To illustrate, given inputs of 12, 16, and 0.05, the program produces $B = 0.038406637206$. The probability of fewer than c events is $1 - B$. To illustrate, assume inputs of 12, 16, and 0.8 for c, n, and p. The output B is 0.7982454418; the probability of fewer than c is $1 - B = 0.2017545582$.

```
 5 INPUT "c = ",C: INPUT "n = ",N: INPUT
   "p = ",P
10 Q = 1 - P
15 A = N * LOG P
20 B = 10 ↑ A
25 IF C = N THEN 50
30 FOR I = 1 TO N - C
35 A = A + LOG (Q/P) + LOG (N - I + 1)
   - LOG I
40 B = 10 ↑ A + B
45 NEXT I
50 PRINT B
55 END
```

Note: 10 ↑ A signifies 10^A.

THE AUTHOR

Herman Burstein earned a Ph.D. degree in economics from New York University in 1945. From 1945 to 1967 he was an economist and statistician for a well-known accounting firm, and from 1967 to 1981 when he retired, he worked as a professor at New College of Hofstra University, where he taught courses in economics, statistics, and research methods.

He is the author of several books and articles in these areas, primarily in statistics. Maintaining a strong avocational interest in audio since the late '40s, he has written some 200 magazine articles and six books in the field of audio. A contributing editor of *Audio*, he writes the monthly Q & A "Tape Guide" column.

Bibliografía

[1] Hyoung Joong Kim, “Audio Watermarking Techniques”, Kangwon National University, Department of Control and Instrumentation Engineering, Korea, 2000

[2] Ken C. Pohlmann, Principles of Digital Audio – 4th ed. , McGraw-Hill, 2000

[3] Yuval Cassuto, Michael Lustig, Shay Mizrachy, “Real time Digital Watermarking System for Audio Signals Using Perceptual Masking”, Technion – Israel Institute of Technology, Department of Electrical Engineering www-sipl.technion.ac.il

[4] Andrew Tanenbaum, Computer Networks 3rd Edition, Prentice Hall, 1996

[5] I. K. Yeo, H. J. Kim, “Modified patchwork algorithm: The novel audio watermarking scheme,” IEEE Transactions on Speech and Audio Processing, vol. 11, no. 4, pp. 381-386, July 2003

[6] A. R. Sánchez Quiroz, A. Castellanos Giacinti, I. Zapata Wolff, “Marcas de agua digitales en señales de audio”, Proyecto de fin de carrera, TEC de Monterrey, México, 2001

[7] Ricardo A. García, “Digital watermarking of audio signals using a psychoacoustic auditory model and Spread Spectrum”, University of Miami, School of Music Engineering Technology, April 1999

[8] E. Zwicker, U. T. Zwicker, “Audio Engineering and Psychoacoustics: Matching Signals to the Final Receiver, the Human Auditory System,” J. Audio Eng. Soc., vol. 39, pp. 115-126, March 1999

[9] Herman Burstein, “Approximation Formulas for Error Risk and Sample Size in ABX Testing,” *Journal of the Audio Engineering Society*, vol. 36, pp. 879-883, November 1988

[10] Herman Burstein, “Transformed Binomial Confidence Limits for Listening Tests”, *Journal of the Audio Engineering Society*, vol. 37, p. 363, 1989

[11] A. V. Oppenheim, R. W. Schafer, *Discrete-Time Signal Processing*, Prentice Hall, 1999

[12] Michael Arnold, Martin Schmucker, Stephen D. Wolthusen, *Techniques and Applications of Digital Watermarking and Content Protection*, Artech House Inc. ,2003

[13] Nedeljko Cvejic, *Algorithms for Audio Watermarking and Steganography*, University of Oulu, Department of Electrical and Information Engineering, Information Processing Laboratory, 2004

[14] Information Technology – Coding of moving pictures and associated audio for digital storage media at up to 1,5 Mbits/s – Part3: audio. British Standard. BSI, London. October 1993. Implementation of ISO/IEC 11172-3:1993. BSI, London. First edition 1993-08-01

[15] A. Bruce Carlson, *An introduction to Signals and Noise in Electrical Communication*, Third Edition, McGraw-Hill, 1986

[16] Leon W. Couch II, *Sistemas de comunicación digitales y analógicos*, 5ª. ed., Prentice Hall, México, 1998

[17] Michael Arnold, *Audio Watermarking: Features, Applications, and Algorithms*, Fraunhofer-Institute for Computer Graphics, Department for Security Technology for Graphics and Communications Systems

[18] E. Zwicker, H. Fastl, *Psychoacoustics Facts and Models*, Springer-Verlag, Berlin, 1990