Técnica de watermarking para legalización de audio

Darío A. Villarreal, Martín Kinen y Matías A. Eberhardt
Trabajo práctico final de Procesamiento Digital de Señales, II-FICH-UNL.

***Resumen*— Las técnicas de watermarking pueden ser usadas para agregar información extra a las señales, de audio, imágenes, etc. Existen numerosas técnicas en función de como se implemente, cada una de éstas con determinadas características. Sus aplicaciones van desde copyright hasta legalización de audio judicial. Se presenta un método de watermarking para legalización de audio judicial, basado en que el oído humano no puede detectar ecos de muy corta duración. Se codificará en la watermark la fecha de la grabación y el ID del dispositivo usado, para además de reconocerse la marca al realizar la detección, se pueda recuperar esta relevante información.**

***Palabras clave*— watermarking inaudible, legalización de audio, codificación de información**

I. Introducción

HOY en día el ámbito judicial esta regido por muchas leyes aplicadas de tal forma que no deben existir lugar a dudas sobre las pruebas y declaraciones utilizadas, más aún cuando se trata de juicios relacionados a crímenes de gravedad en países donde está vigente la pena de muerte.

Las características deseadas que persigue el diseño de las distintas técnicas son:

- *Seguridad:* Consta de dos niveles. En el primero, un usuario no autorizado no puede leer o descifrar una marca de agua ni puede detectar si una señal contiene una marca de agua. En el segundo, se permite saber que la señal contiene una watermark, pero ésta no puede ser descifrada sin las correspondientes claves.
- *Robustez:* se refiere a que la watermark debe ser resistente a las distorsiones introducidas por diferentes procesamientos comunes de la señal (compresión, edición, cambio de formato, etc) y a ataques a la misma.
- *Imperceptibilidad:* El embebido de la marca no debe agregar componentes perceptibles al oído humano. Por otro lado, para una alta robustez es necesario que la amplitud de la watermark sea el máximo posible. Por lo tanto, en el diseño del método hay que tener en cuenta que el aumento de una característica reduce necesariamente a la otra.
- *Recuperación de la marca con o sin los datos originales:* La recuperación de una watermark es más robusta si se cuenta con la señal original, lo cual no es posible en todos los casos. La mayoría de las técnicas de watermarking realizan algún tipo de modulación en la cual se introduce una distorsión en la señal. Si esta distorsión es conocida o puede ser modelada en el proceso de recuperación, se puede realizar la extracción de la marca de agua sin el conocimiento de la señal original. A estas se las llama *Técnicas ciegas de watermarking*.

Las técnicas de watermarking de audio que podemos usar se pueden dividir en distintos grupos:

- Las basadas en agregar a la señal ruido blanco, de modo tal que el cambio resulte inaudible. Las watermarks generadas de esta manera son frágiles a la mayoría de ataques de procesamiento o compresión de la señal.
- Las que explotan la insensibilidad del oído humano ante ecos de muy corta duración. Estas watermarks son dependientes de la señal y de los retardos de los ecos elegidos, y resolver esto torna al sistema demasiado complejo.
- Las que embeben la marca en el dominio frecuencial. El problema que presentan es poca robustez en señales con pocos componentes en el dominio transformado.
- Las que emplean el concepto de dispersión en el espectro de frecuencias. Presentan una alta complejidad computacional.

En este caso, se busca usar una técnica que permita asegurar que una grabación, por ejemplo de una declaración jurada, sea legítima. En otras palabras, debemos asegurar que no ha sufrido modificaciones, tales como tomar palabras fuera de contexto, suprimir partes, etc. Por lo tanto, no nos centraremos en crear una watermark robusta, dado que con sólo reconocer que la señal fue modificada, nos basta para declararla legalmente inválida.

Este trabajo está organizado de la siguiente manera: Sección II, detalles y características de la watermark a utilizar; Sección III, Embebido de la marca en la señal; Sección IV, decodificación de la watermark; Sección V, técnicas suplementarias, código Hamming de autocorrección y codificación NRZ. Finalmente en la Sección VI se presentan las conclusiones del trabajo.

II. Watermarking

La técnica de watermarking que utilizaremos se basa en ocultar datos en el dominio del tiempo incrustando ecos en la señal original mediante la convolución en el tiempo con un impulso determinado.

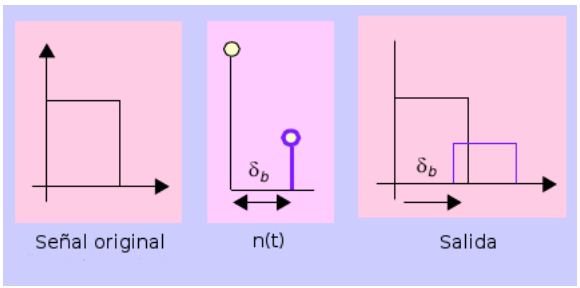

Fig. 1. Convolución señal-eco

Tenemos para esto dos funciones núcleo $n1$ y $n0$ que codificarán 1 y 0 respectivamente, definidas por los siguientes parámetros: amplitud, offset y tasa de decaimiento.

Las funciones núcleo utilizan dos tiempos de retardo, uno para representar al cero (offset) y otro para representar al uno (offset + delta). Ambos retardos fueron determinados empíricamente utilizando diferentes oyentes, resultando indetectables al oído humano.

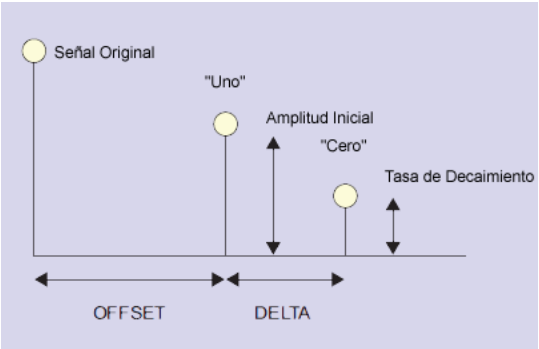

Fig. 2. Núcleos

III. Embebido

Utilizando las funciones núcleo anteriores procederemos a embeber la watermark en la señal. Para esto dividimos la señal original en pequeñas ventanas (tantas como bits se necesite codificar en ella) y convolucionamos éstas con dichas funciones núcleo. De esta forma, cada ventana codifica un 1 o un 0 de manera independiente. Para simplificar esta realización se puede generar una señal aplicándole a la señal original las funciones núcleo correspondientes al 0 ($S0$) y otra con la correspondiente al 1 ($S1$), donde S_k es la ventana k-ésima de la señal.

$$S0_k(i) = \sum_{i=1}^{N} S_k(j) * n0(j-i)$$

$$S1_k(i) = \sum_{i=1}^{N} S_k(j) * n1(j-i)$$

Generamos también dos señales llamadas $Mixer1$ y $Mixer0$ complementarias, compuestas por ceros y unos correspondientes a la información binaria a codificar. Las transiciones entre unos y ceros son rampas y la suma entre los mixers es siempre uno. Esto nos da una transición suave entre las porciones codificadas con distintos bits y evita los cambios bruscos en la resonancia de la señal resultante.

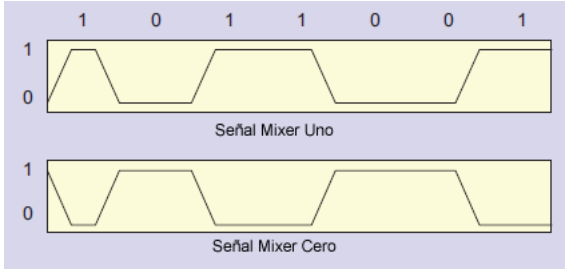

Fig. 3. Señales Mixer 0 y Mixer 1

Luego, al multiplicar la señales $Mixer0$ por $S0$ y $Mixer1$ por $S1$ y sumando los resultados obtenemos la señal codificada. Esto se puede apreciar gráficamente en la Figura 4.

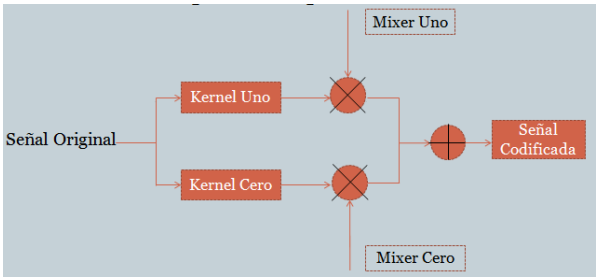

Fig. 4. Proceso de codificación

IV. Decodificación

La extracción de la información embebida se logra detectando la separación de los ecos. Para ello dividimos la señal codificada (Sw) en ventanas iguales a las utilizadas en la codificación y calculamos el *cepstrum* de las mismas:

$$Cs_k = F^{-1}(\log(F(Sw_k))^2)$$

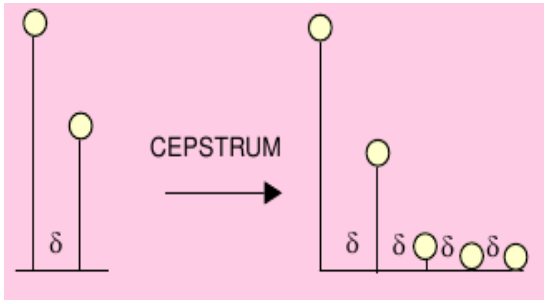

Fig. 5. Cepstrum del kernel

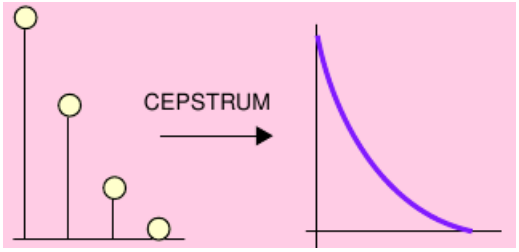

Fig. 6. Cepstrum de la señal original

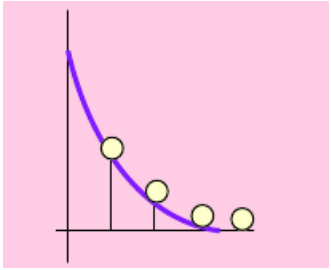

Fig. 7. Cepstrum de la señal codificada

Con esto logramos ver el espacio entre el eco y la señal original. El *cepstrum* duplica el eco cada δ segundos y la magnitud de los impulsos que representan los ecos son pequeños con respecto a la señal original, lo que dificulta su detección. Para solucionar esto, se toma la autocorrelación del *cepstrum*.

Luego examinamos los valores de la autocorrelación del *cepstrum* en búsqueda de los picos del mismo. Éstos, si la señal tiene nuestra watermark y no fue modificada se encontrarán a la distancia de offset u offset+delta. Dependiendo del lugar en que se encuentre el pico (offset u offset+delta) el valor codificado será un cero o un uno.

A continuación podemos apreciar dos ventanas con distintos símbolos codificados, a las cuales se les calculó el *cepstrum* y la autocorrelación del *cepstrum* (*autocepstrum*).

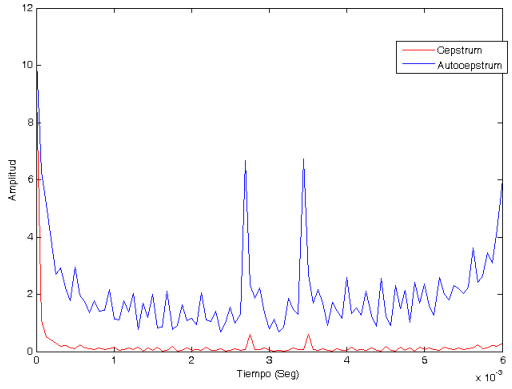

Fig. 8. Cepstrum y autocepstrum de una ventana con un 0 codificado

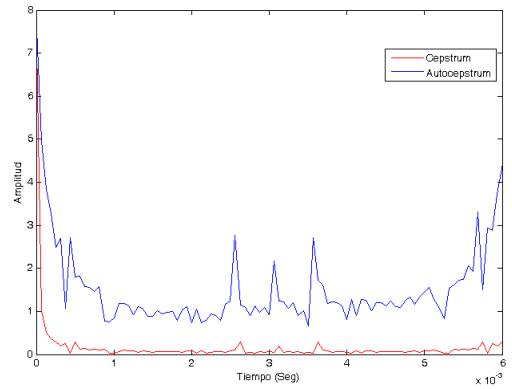

Fig. 9. Cepstrum y autocesptrum de una ventana con un 1 codificado

Una vez procesadas todas las ventanas, recuperamos la información binaria que codificamos en el embebido. Analizando la información recuperada llegamos a la conclusión de que, si la información es correcta, la señal no fue modificada. En cambio si lo recuperado no coincide con lo codificado concluiremos que no es la señal no es la original y la declararemos inválida.

V. Técnicas suplementarias

A. Código de autocorrección de Hamming

Es bien conocido que en la transmisión, podemos tener errores causados por ruido, ya sean internos o externos (internos, impulsivos, etc), también en el almacenamiento o hasta en el procesamiento de nuestra señal. Sin que estos errores impliquen que la señal a sido modificada sino que en algunos casos solo es debido al umbral utilizado en la detección y con valores muy cercanos a este o redondeos y truncamientos con los números usados. Por ello podemos agregarle más bits para aumentar así la distancia entre dos palabras, utilizaremos el código de Hamming 7,4 que cada 7 bits agrega 4, los cuales permiten detectar y corregir errores de un bit.

B. Codificación NRZ (No Retorno a Cero)

Si se codifica la señal de forma unipolar, se puede perder información al trasmitirla, ya que en algunos casos el receptor tendrá problemas para detectar si hubo o no transición en la señal enviada e interpretar 0 por 1 y viceversa. Para solucionar este inconveniente usamos la codificación NRZ, que codifica con un nivel de tensión distinto cada símbolo transmitido.

VI. Resultados

Con el proceso explicado anteriormente pudimos agregar fácilmente información relevante que puede corroborar la autenticidad de una grabación y a la vez brindarnos información de cuando fue realizada y con que dispositivo. Esto puede ser de mucho uso en el ámbito legal y judicial permitiendo a abogados y jueces basarse en pruebas verídicas y certificadas. Dicho proceso es tolerante a errores mínimos de procesamiento y codificación, esto permitido por el código de Hamming y a que el volumen de datos a querer introducir no es muy grande.

Al usar ecos muy pequeños, para que resulten imperceptibles al oído humano, el método presentó algunas fallas al recuperar la información codificada, aún usando el código autocorrector de Hamming. Estas fallas varían de acuerdo a la señal utilizada. Como ya fue comentado en la introducción, este método de watermarking es dependiente de la señal, y una solución a este problema es determinar los retardos y la magnitud del eco a utilizar adaptativamente, es decir, calcular los valores óptimos para cada ventana de una señal. Otra desventaja que presenta es que no es robusta ante ataques y cambios en la señal, esto en principio pudiese verse como un inconveniente, pero no lo es si nos basamos en nuestro objetivo inicial de sólo determinar si es original una grabación en cuestión.

Referencias

[1] J. A. W. O. Aweke Negash Lemma y L. van de Kerkhof, "A temporal domain audio watermarking technique." *IEEE Transactions on signal processing*, vol. 51, pp. 1088–1097, 2003.

[2] S. D. Larbi y M. Jaïdane-Saïdane, "Audio watermarking: A way to stationnarize audio signals," *IEEE Transactions on signal processing*, vol. 53, pp. 816–823, 2005.

[3] J. W. Seok y J. W. Hong, "Audio watermarking for copyrigth protection of digital audio data," *Electronics Letters*, vol. 37, pp. 60–61, 2001.

[4] R. B. Abdellatif Zaidi y P. Duhamel, "Audio watermarking under desynchronization and additive noise attacks," *IEEE Transactions on signal processing*, vol. 54, 2006.

[5] P. Moulin y R. Koetter, "Data-hiding codes," *Proceedings Of The IEEE*, vol. 93, 2005.

[6] S.-K. Lee y Y.-S. Ho, "Digital audio watermarking in cepstrum domain," *IEEE Transactions and Consumer Electronics*, vol. 46, 2000.

[7] C. I. Podilchuk y E. J. Delp, "Digital watermarking: Algorithms and applications," *IEEE Signal Processing Magazine*, 2001.

[8] S. K. Kumar y T. Sreenivas, "Increased watermark-to-host correlation of uniform random phase watermarks in audio signals," *Elsevier Signal Processing*, vol. 87, 2001.

[9] F. Hartung y M. Kutter, "Multimedia watermarking techniques," *Proceedings of the IEEE*, vol. 87, 1999.

[10] X. X. W. Li y P. Lu, "Robust audio watermarking based on rhythm region detection," *Electronics Letters*, vol. 41, 2005.

[11] I. P. Paraskevi Bassia y N. Nikolaidis, "Robust audio watermarking in the time domain," *IEEE Transactions on Multimedia*, vol. 03, 2001.

[12] N. K. K. Mohammad Ali Akhaee y F. Marvasti, "Robust audio and speech watermarking using gaussian and laplacian modeling," *Elsevier Signal Processing*, vol. 90, 2010.
[13] D. C. H. C. Hyun Wook Kim y T. Kim, "Selective correlation detector for additive spread spectrum watermarking in transform domain," *Elsevier Signal Processing*, vol. 90, 2010.
[14] E. Mandado y Y. Mandado, *Sistemas electrónicos digitales*. Marcombo, 2008.
[15] J. Proakis y D. Manolakis, "Tratamiento digital de señales," *Madrid [etc.]: Prentice-Hall*, 1997.