

Monitoreo automático de estaciones de radio

J. A. Camarena Ibarrola
Facultad de Ingeniería Eléctrica
DES Ingenierías Arquitectura, UMSNH.

Resumen: A fin de detectar de manera automática la aparición de anuncios publicitarios en estaciones de radio se diseñó una firma digital de audio basada en la de-modulación del nivel de información de la señal. El esquema propuesto funcionó a la perfección a pesar de las deformaciones a las que se somete la señal que son inherentes a la transmisión de ondas de radio en Amplitud Modulada (AM). Se Monitorearon doce anuncios publicitarios en cinco diferentes estaciones de radio de AM a lo largo de 24 Horas de Transmisión. El sistema de Monitoreo detectó de manera precisa el instante en que ocurrieron todos los anuncios sujetos de monitoreo sin confundirlos con otros segmentos de audio.

Introducción

Las empresas que contratan publicidad en radio, cuestionan el cumplimiento riguroso tanto en número de anuncios publicitarios como en horario en que debieran aparecer los mismos. Para ello, es normal que la empresa que paga la publicidad contrate a una tercer empresa para que esta lleve a cabo el monitoreo de la publicidad contratada. Algunas empresas prefieren dedicar parte de su personal para realizar dicho monitoreo. En cualquier caso existe un costo a pagar para realizar la supervisión del cabal cumplimiento de la publicidad contratada.

Los Gobiernos también están interesados en hacer monitoreo de medios, por ejemplo, para dictaminar si un proceso electoral se puede calificar como justo, es común que se monitoreen los anuncios de los partidos políticos que participan en la elección.

Tanto empresas como gobiernos gastan mucho dinero en llevar a efecto el monitoreo de la señal de radio debido a que utilizan personas para llevarlo a cabo, como las estaciones de radio diseminadas por el país son muchas y utilizan una persona por cada estación que se desea monitorear, el gasto en nómina resulta excesivo.

Computadoras equipadas con tarjetas sintonizadoras de radio pueden utilizarse para realizar el monitoreo durante las 24 horas del día y los 7 días de la semana. Cada computadora podría ser capaz de monitorear varias estaciones de radio en forma simultánea.

Antecedentes: Existen básicamente dos esquemas para llevar a cabo el monitoreo automático de estaciones de radio, estos son el Monitoreo Basado en Marcas de Agua y el Monitoreo Basado en Firmas Digitales de Audio.

Monitoreo Basado en Marcas de Agua. Bajo este esquema, la estación transmisora envía junto con el audio, información inaudible denominada “Marca de Agua” en donde se almacena información que identifica al audio que esta transmitiendo [1]. La Marca de Agua se puede embeber en anuncios publicitarios; en canciones o en cualquier audio sujeto de monitoreo. El sistema de Monitoreo debe descubrir la Marca de Agua para identificar el audio. El programa que embebe la marca de agua funciona en colaboración con el que la descubre, en realidad, ambos programas forman parte de un sistema integral. Bajo el esquema de Monitoreo Basado en Marcas de Agua se asume que se tiene acceso al audio que es sujeto de monitoreo y que este puede ser modificado antes de que sea transmitido por la estación radiodifusora.

Monitoreo Basado en Firmas Digitales de Audio. Normalmente, un Sistema de Monitoreo de anuncios publicitarios no tiene acceso a la señal de audio antes de que sea transmitido por la estación de radio. Es necesario entonces que se identifique al audio extrayendo de él las características perceptuales que hacen posible a las personas reconocer una señal de audio que les resulta familiar. Una Firma digital de Audio, conocida también como Huella de Audio pretende ser precisamente una síntesis de las características salientes del sonido que permiten identificarlo. Decidir acerca de cuáles son las características perceptuales relevantes y que permanecerán en la señal aun si esta es deformada por ruido, compresión con pérdidas, conversión Analógica-Digital y Digital Analógica, ecualización y otras formas de degradación de la señal de audio es aún motivo de debate [2]. En 2002, Hellmuth *et al* [3] presentó un sistema de monitoreo bajo el esquema de Firmas Digitales de Audio basadas en una medida de lo plano que es el espectro que forma parte del estándar MPEG-7. En 2005, Oliveira *et al* [4] presentó un

sistema de monitoreo bajo el esquema de Firmas Digitales de Audio basadas en la energía del espectro.

El resto del documento está organizado de la siguiente manera: En la siguiente sección se describirá el sistema de monitoreo implementado, enseguida se describirán los experimentos realizados así como los resultados obtenidos, finalmente se aportarán conclusiones y se discutirán trabajos futuros.

Descripción del sistema propuesto

La señal de audio se digitaliza en monoaural a una precisión de 16 bits por muestra, a una frecuencia de muestreo de 16 Kilohertz. La señal se divide en marcos de 250 milisegundos con un traslape de 7/8 de manera que se produce un vector de características cada 31.25 milisegundos. Para la frecuencia de muestreo establecida, esto significa utilizar marcos de 4096 muestras procesando un vector de características cada 512 muestras. A cada marco se le aplica la ventana de Hann [5] y luego se le aplica la Transformada Discreta de Fourier. El oído humano es mas sensible a las frecuencias bajas que a las altas, por eso prácticamente todos los sistemas de identificación de audio dividen el espectro de frecuencias en bandas distribuidas logarítmicamente, algunos utilizan la escala de Mel [6], otros utilizan la escala de Bark [7] y otros usan las octavas musicales. En este trabajo se optó por utilizar la escala de Bark aprovechando los estudios hechos por Zwicker respecto a la sensibilidad del oído humano a las diferentes frecuencias audibles [7]. La Ecuación (1) se puede utilizar para convertir Hertz a Barks.

$$z = 13 \arctg\left(\frac{0.76f}{1000}\right) + 3.5 \arctg\left(\frac{f}{7500}\right)^2 \quad (1)$$

Donde f es la frecuencia en Hertz. Las Bandas críticas definidas en honor a Barkhausen tienen precisamente un ancho de banda de un Bark y cada una corresponde a

una sección de la cóclea de 1.3 milímetros. La Tabla 1 muestra algunas de las 25 bandas críticas de la famosa escala de Bark.

Tabla 1: Algunas Bandas de la escala de Bark

Banda crítica	Inicio (Hz)	Fin (Hz)
1	20	100
2	100	200
⋮	⋮	⋮
10	1080	1270
⋮	⋮	⋮
20	5300	6400
21	6400	7700
⋮	⋮	⋮
25	15500	20000

En este trabajo se utilizaron las primeras 21 Bandas críticas de Bark, es decir, se utilizan solo las frecuencias inferiores a 7700 Hertz. Por cada banda, se determina la entropía de Shannon asumiendo que las partes real e imaginaria del espectro de una señal de audio se comportan como variables aleatorias independientes que siguen una distribución gaussiana como lo demostró Rainier en [8], se determina la entropía H por cada banda utilizando (2).

$$H = \ln(2\pi) + \frac{1}{2} \ln(\sigma_{xx}\sigma_{yy} - \sigma_{xy}^2) \quad (2)$$

Donde σ_{xx} y σ_{yy} son las varianzas de la parte real e imaginaria del espectro respectivamente y $\sigma_{xy} = \sigma_{yx}$ es la co-varianza entre la parte real y la parte imaginaria del espectro.

Por cada 31.25 ms obtenemos entonces un vector de 21 entropías, un valor por cada banda crítica. La secuencia de tales vectores de entropías forma una matriz con 21 renglones y un número de columnas que depende de la duración de la señal de audio.

Se obtiene ahora la diferencia de cada vector de entropías con el vector del marco anterior, el vector resultante puede considerarse como el vector con las derivadas respecto al tiempo de las entropías para las primeras 21 bandas críticas. Por cada elemento de este vector de

diferencias si su valor es positivo se agrega un bit “encendido” y “apagado” en caso contrario. Precisando, el bit para la banda b del n-ésimo marco de la Firma Digital de Audio se calcula de acuerdo a (3)

$$F(n,b) = 1 \quad \forall \quad H(n,b) - H(n-1,b) > 0 \\ = 0 \quad \forall \quad H(n,b) - H(n-1,b) \leq 0 \quad (3)$$

Donde F es la Matriz de bits resultante que constituye la Firma digital de Audio y H es la matriz de flotantes con las entropías de Shannon recién calculadas.

La Firma de Audio descrita solo requiere 21 bits por cada 31.25 milisegundos de audio. Lo que se esta almacenando realmente es una indicación por cada Banda Crítica que nos dice si el nivel de información (entropía) está incrementándose o no. Esta Huella de Audio resultó ser extremadamente resistente a ruido, ecualización, conversión A/D y D/A, compresión con pérdidas y otras formas de degradación.

Experimentos realizados

Para probar qué tan seguro es el sistema de monitoreo propuesto en este trabajo se llevaron a efecto los siguientes pasos:

- Se conectó la salida RCA de un sintonizador de radio AM a la entrada auxiliar de la tarjeta de sonido de la computadora

- Se eligieron 12 anuncios publicitarios de 5 diferentes estaciones transmisoras, se grabaron en archivos de audio con formato MP3 con un régimen de bits de 128 kbps. Los 12 Anuncios Elegidos conforman la lista de anuncios de interés.

- Se determinó la Firma Digital descrita en la sección anterior para cada uno de los doce anuncios publicitarios de interés. Las Firmas de Audio de estos 12 anuncios de interés se almacenaron en otros doce archivos binarios.

Las Firmas Digitales de los anuncios publicitarios son matrices de bits con 21 renglones, y un número de columnas que depende de su duración, por ejemplo, una pauta publicitaria de 20 segundos tiene una Firma de 615 columnas, es entonces un arreglo de 21 x 615 bits. Por otra parte, la Firma Digital del Audio de la estación de radio monitoreada consiste de un flujo interminable de vectores de bits (cada uno con 21 bits).

Para detectar la ocurrencia de un anuncio publicitario se compara su Firma Digital con el arreglo de bits del mismo tamaño formado con los vectores de bits mas recientes extraídos de la señal de audio de la estación de radio que está siendo monitoreada, esto se hace para cada uno de los anuncios publicitarios de interés. Para comparar la Firma de Audio de un anuncio publicitario con la Firma del segmento de audio mas reciente que tiene la misma duración que el anuncio se utilizó la Distancia de Hamming que reporta el número de bits diferentes entre los dos arreglos de bits.

El monitoreo consiste entonces en determinar la distancia de Hamming de la Firma de la pauta con la primer sub-matriz del mismo tamaño que es parte de la Firma del Audio de la estación monitoreada, La Firma de la pauta se va recorriendo un marco de audio (columna de 21 bits en la Firma) a la vez y se recalcula la distancia de Hamming. Cuando la Distancia de Hamming cae por debajo de un valor umbral se reporta que ha ocurrido el anuncio publicitario y se guarda el instante en que este ocurrió. En la Figura 1, la Firma de Audio del anuncio publicitario se representa por un rectángulo de pequeños cuadros blancos y negros que corresponde con el arreglo de bits de dicha Firma, este rectángulo se recorre a lo largo de una interminable secuencia de vectores de bits (de nuevo puntos blancos y negros) que constituye la Firma de Audio de la estación que se está monitoreando.

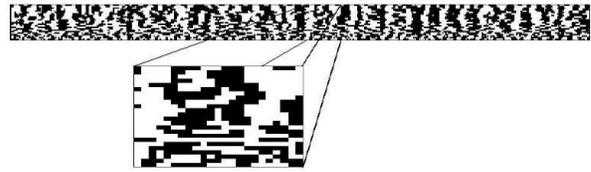


Fig. 1 Firma de Anuncio Publicitario Localizada dentro de la Firma del audio de una estación transmisora

En la Figura 2, se graficó la distancia de Hamming entre la Firma de un anuncio publicitario y la Firma de la señal de audio de una estación a lo largo de 50 minutos, se ve claramente como cae la distancia de Hamming al aparecer el anuncio publicitario el cual ocurre en el minuto 2 y en el minuto 40. Por la forma tan brusca en que cae la distancia de Hamming al ocurrir el anuncio se aprecia lo simple que es establecer el valor umbral, sin embargo, la distancia de Hamming debe normalizarse para poder establecer dicho valor umbral de manera segura.

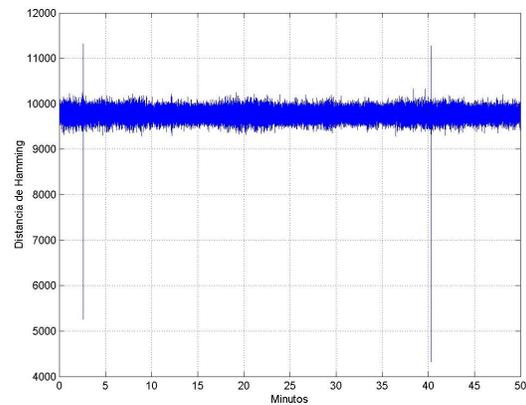


Fig. 2 Distancias de Hamming entre la Firma de un anuncio publicitario y todas las posibles sub-matrices del mismo tamaño de la Firma de Audio de una estación transmisora a lo largo de 50 minutos. Se observa que el anuncio ocurre en el minuto 2 y en el minuto 40

Análisis de sensibilidad. En forma simultánea al proceso de monitoreo se grababa las señal de audio para efectos de hacer un estudio sobre la sensibilidad del sistema.

Cuando la Distancia de Hamming es menor al valor umbral definido para tal efecto, el sistema reporta un “positivo”, en caso contrario se considera un “negativo” y no hay nada que reportar. Los anuncios detectados por el sistema fueron corroborados uno por uno utilizando las grabaciones del audio de las estaciones para tal efecto, esto fue simple puesto que el sistema nos reporta exactamente en que instante ocurre el anuncio publicitario, al validar el anuncio publicitario, se puede decidir si se trata de un “positivo verdadero” o bien de un “positivo falso”, un positivo falso es entonces un reportes erróneo de ocurrencia de anuncio por parte del sistema. Por otra parte, se levantó un “inventario” de anuncios publicitarios incluidos en las grabaciones, para corroborar la ausencia de “negativos falsos”, es decir anuncios no detectados por el sistema. La Tabla 2 resume las definiciones necesarias para el análisis de sensibilidad.

Tabla 2: Definiciones para el análisis de sensibilidad

	dist < umbral (El Sistema reporta una ocurrencia)	dist > umbral (El Sistema no reporta nada)
El Anuncio sí ocurrió en realidad	POSITIVO VERDADERO (PV)	NEGATIVO FALSO (NF)
El Anuncio no ocurrió en realidad	POSITIVO FALSO (PF)	NEGATIVO VERDADERO (NV)

Se define el Régimen de Positivos Verdaderos (RPV) mediante (4) y el Régimen de Positivos Falsos (RPF) mediante (5).

$$RPV = \frac{PV}{PV + NF} \quad (4)$$

$$RPF = \frac{PF}{PF + NV} \quad (5)$$

Las curvas ROC sirven para evaluar el desempeño en cuanto a la precisión de un sistema de identificación como el que se presenta en este trabajo, para generarlas se debe ir variando el valor umbral con el que se decide si ocurre o no un anuncio. Para cada valor umbral se calcula RPV y RPF, de esta manera se obtiene un punto de la curva ROC. Al variar el valor umbral se genera la curva ROC, ver [9]. En la Figura 3 se muestra la curva ROC obtenida al utilizar la Firma de Audio Basada en Entropía Espectral Multi-banda y la curva ROC obtenida al utilizar la Firma de Audio Basada en una medida de lo plano del espectro que forma parte del estándar de MPEG-7. El hecho de que el área bajo la curva ROC que corresponde a la Firma basada en entropía espectral sea mayor es un claro indicativo del mejor desempeño de esta Firma para efectos de monitoreo de estaciones de radio AM.

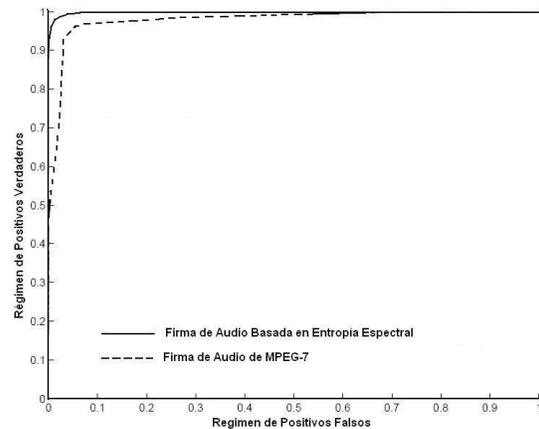


Fig. 3 Curvas ROC para el Sistema basado en la Firma de Audio Basada en Entropía Espectral Multi-banda y el Basado en la Firma de Audio de MPEG-7

Conclusiones y Trabajos Futuros

Los anuncios publicitarios son en su mayoría invitaciones verbales para promover un producto o servicio, ello explica el por qué la resolución que finalmente resultó adecuada para la Huella de Audio que se diseñó en este

trabajo para fines de monitoreo de Radio se acerca a la que se usa en reconocimiento de voz, sin embargo, al estar basada en la demodulación del nivel de información contenida en la señal de audio, la Huella aprovecha muy bien cualquier contenido de música que pueda tener la señal de audio para fines de identificación.

El esquema propuesto pasó las pruebas sobrepasando incluso las expectativas en cuanto a robustez, es decir la tolerancia a degradaciones de la señal debidas a ruido, conversión D/A y A/D, compresión con pérdidas, etc. Muy comunes en transmisión de radio AM.

En este trabajo se implementó y probó un sistema capaz de revisar la presencia en un flujo de audio de cualquiera de los anuncios de una lista previamente preparada de aquellos anuncios que son de interés. Sin embargo, la complejidad en el algoritmo de búsqueda de este método es $O(mn)$ donde m es el número de anuncios que forman la lista mencionada y n la longitud media de las Firmas de Audio de los anuncios. Quisiéramos implementar un algoritmo de búsqueda con complejidad lineal, es decir complejidad $O(n)$, esto quizá se logre utilizando técnicas modernas como las que se utilizan para buscar sub-secuencias en cadenas genéticas o para buscar virus de computadoras en archivos y que se discuten en [10]

Referencias

- [1] Seungwon Shin; Oanjin Kim; Jongweon Kim *et al.* "A Robust Audio Watermarking Algorithm Using Pitch Scaling", in Conference Proceedings *14th International Conference on Digital Signal Processing*, Vol 2, pp 701-704, 2002.
- [2] Cano, Pedro; Battle, E.; Kalker, T. *et al.*, "A Review of Algorithms for Audio FingerPrinting". *Multimedia Signal Processing*, IEEE , pp 169-167, 2002.
- [3] O. Hellmuth; E. Allamanche; M. Cremer; *et al* "Content-Based Broadcast Monitoring Using MPEG-7 Audio Fingerprints" in Conference Proceedings, *International Symposium on Music Information Retrieval(ISMIR)*, 2001.
- [4] Oliveira, Bruno; Crivellaro, Alexandre; and Cesar Jr, Roberto M. "Audio-Based Radio and TV Broadcast Monitoring" in Conference Proceedings *11th Brazilian Symposium on Multimedia and the Web*, pp 1-3, 2005.
- [5] Blackman, R. B. and Tukey, J. W. "Particular Pairs of Windows." In "The Measurement of Power Spectra From the Point of View of Communication Engineering" Dover Publications, 1959.
- [6] Stevens, Stanley Smith; Volkman; John; & Newman, Edwin. "A scale for the measurement of the psychological magnitude of pitch". *Journal of the Acoustical Society of America*, pp 185-190. 1937.
- [7] Zwicker, E. "Subdivision of the Audible Range into critical Bands". *Journal of the Acoustical Society of America*. 1961.
- [8] Rainer Martin, "Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics", *IEEE Transactions on Speech and Audio Processing*, Vol 9, Num 5, pp 504-512, 2001.
- [9] Fawcett, T. "Roc graphs: Notes and practical considerations for researchers". Tech report HPL2003-4, HP Labs Tech, 2003.
- [10] Gonzalo Navarro and Mathieu Raffinot. "Flexible Pattern Matching in Strings". Cambridge University Press. 2002.

José Antonio Camarena Ibarrola. Ingeniero Electricista egresado de la Universidad Michoacana en 1986. Obtuvo el grado de Maestría en Ciencias Computacionales en el Tecnológico de Toluca en 1996 y el grado de Doctor en Ciencias en Ingeniería Eléctrica Opción Sistemas Computacionales en 2008 en la Universidad Michoacana.

email: camarena@umich.mx

Dirección del autor: Edificio de la División de Estudios de Postgrado de la Facultad de Ingeniería Eléctrica, Laboratorio de Sistemas Computacionales PB, C.U. Morelia, Michoacán, México.

