

IDENTIFICACION AUTOMÁTICA DE PAUTAS PUBLICITARIAS EN SEÑALES DE AUDIO

*José A. Camarena⁽¹⁾, Edgar Chávez⁽²⁾

(1) Universidad Michoacana de San Nicolás de Hidalgo, camarena@umich.mx

(2) Universidad Michoacana de San Nicolás de Hidalgo, elchavez@umich.mx

Resumen

Identificar el contenido de señales de audio transmitido por estaciones de radio y televisión es de interés para instancias gubernamentales interesadas en monitoreo de medios y para empresas del sector productivo interesadas en corroborar el cumplimiento de sus contratos de publicidad.

El monitoreo automático por computadora se puede llevar a cabo usando "*Marcas de agua de audio*" lo que implica alterar la señal de audio antes de ser transmitida (gran inconveniente). Alternativamente, el monitoreo se puede llevar a cabo usando "*Firmas de Audio*". En este artículo presentamos una solución basada en una firma de audio diseñada por nosotros basada en entropía espectral multi-banda que ha demostrado ser más robusta que las existentes. La robustez de una firma de audio indica cuanto ruido puede estar presente en una señal de audio o cuanta deformación puede haber sufrido (Ej. Por ecualización o por conversión Digital/Analógica) de manera que aún pueda ser identificada.

En nuestros experimentos utilizamos 95 horas de audio grabado de cinco diferentes estaciones de radio FM y comprimidas al formato mp3@64kbps, en estos archivos de audio se realizó un barrido secuencial buscando ocurrencias de 13 anuncios publicitarios de interés encontrando el cien por ciento sin falsos positivos. El barrido de las 95 horas de audio se realizó en solo una hora. Nuestra técnica permite monitorear hasta 10 estaciones de radio con una sola computadora de escritorio y de manera más confiable que las del estado del arte, se incluyen resultados comparativos.

Palabras clave: Identificación, Señal, Firma, Entropía

1. Introducción.

Considere una empresa del sector productivo que ha contratado publicidad en diversas estaciones de radio para cierto producto. Dado los altos costos de la publicidad, es de interés de la empresa el corroborar si se está cumpliendo el contrato de publicidad por parte de las estaciones de radio, para ello, puede distraer de sus labores a algunos de sus empleados o bien contratar los servicios de otra empresa para auditar los contenidos de las estaciones donde se supone aparecerá la publicidad, de cualquier manera esto incrementará los gastos de la empresa. Las autoridades con la obligación de arbitrar los procesos electorales también están interesadas en monitorear las estaciones de radio y televisión con el objeto de llevar estadística de los anuncios de los partidos políticos

participantes. Existen mas razones para llevar a cabo monitoreo de señales de audio como lo es el medir el éxito de piezas musicales por su frecuencia de aparición en el aire o el control de calidad de la misma radiodifusora.

2. Antecedentes.

Hay básicamente dos maneras de llevar a cabo la identificación de pautas publicitarias, utilizando "*Marcas de Agua de Audio*" o bien mediante "*Huellas de Audio*" también conocidas como "*Firmas de Audio*".

2.1 Marcas de Agua de Audio.

Es posible insertar información oculta en la señal de audio, a esta información se le denomina *marca*. La "marca de agua" insertada no debe producir distorsión perceptible en la señal de audio, adicionalmente, la *marca* debe ser reconocible a pesar de que la señal de audio sea sujeta a degradaciones propias de la transmisión por radio frecuencia. Tanto el programa que lee la marca como el que la inserta comparten la información relacionada a como se ocultó la marca y como descubrirla, en realidad, ambos programas forman parte de un sistema integral, este tipo de sistemas son adecuados cuando la estación transmisora acepta modificar la señal que va a transmitir [14], [7].

2.2 Huella de Audio.

Llamamos huella o firma de audio a la representación compacta de una señal de audio específica, es usada para fines de identificación. Una huella de audio debe ser robusta a degradaciones tales como contaminación por ruido, ecualización, conversión digital/analógica, etc. Una huella de audio se determina extrayendo características preceptuales relevantes de la señal de audio. En contraste con sistemas de audio basados en marcas de agua, los sistemas basados en huellas de audio no requieren que la señal de audio sea alterada previo a la transmisión de la misma por la estación de radio lo cual es muy conveniente para nuestro propósito. Los coeficientes cepstrales de frecuencia de Mel (MFCC por sus siglas en inglés) [11],[6]; la frecuencia de modulación y frecuencia acústica conjunta [12] [13]; la rugosidad espectral [5]; los centroides de sub-bandas espectrales [10]; la tonalidad [3] y los valores cromáticos [9] son características que han sido utilizadas en el diseño de huellas de audio exitosas [1].

2.2.1 Monitoreo de Radio basado en Huellas de Audio.

Un sistema de monitoreo automático basado en huellas de audio fue propuesto en [8], en ese trabajo extraen una huella de audio basada en un análisis del contenido de energía tanto en el tiempo como en la frecuencia, los autores reportan un régimen de identificación correcta de 95.4% y 1% de régimen de falsos positivos. El monitoreo del contenido de estaciones de radio presentado en [4] utiliza una huella de audio basada en rugosidad espectral tal y como se definió en el estándar de MPEG-7 [2].

2.3 Monitoreo en-línea y monitoreo fuera de línea.

La detección de incidencias de pautas publicitarias en señales de audio puede hacerse de dos maneras a saber en línea y fuera de línea, es decir en tiempo real o bien grabando la señal y realizando el monitoreo posteriormente haciendo uso de dichas grabaciones. Para la mayoría de las aplicaciones, realizar el monitoreo fuera de línea resulta mas conveniente ya que al final del día se cuenta con la evidencia (las grabaciones) de lo que el sistema de monitoreo reporta. Incluso las empresas que monitorean estaciones de radio empleando personas para tal efecto han descubierto que monitorear fuera de línea les resulta mas productivo, de hecho, normalmente, un empleado normalmente revisa 24 horas de grabación en su turno de 8 horas mientras que si trabajaran en línea les tomaria 24 horas revisar justamente 24 horas. La alternativa de monitoreo automático en línea es la elección solo en demostraciones o expo-eventos donde se presente un sistema de este tipo.

3. Nuestro Sistema de monitoreo

La huella de audio de cada pauta publicitaria de la que se desea llevar la cuenta de sus ocurrencias debe ser extraída previamente. La huella de la señal de audio sujeta a monitoreo debe también ser extraída

3.1 Extracción de la Huella

La huella de audio que utilizamos la denominamos: "Firma de Audio basada en entropía espectral multi-banda o MBSES por sus siglas en inglés (Multi-Band Spectral Entropy Signature), su proceso de extracción es el siguiente:

- a). La señal es procesada en marcos de tiempo de 256 ms, esto asegura un soporte de tiempo adecuado para la estimación de la entropía. Los Marcos se encuentran traslapados un 87.5 por ciento (7/8), esto implica que un vector de características se determinará por cada 32 ms de audio
- b). La señal correspondiente a un marco de tiempo se multiplica por la ventana de Hann y después se aplica la transformada discreta de Fourier.
- c). La entropía de Shannon se determina para las primeras 21 bandas críticas de acuerdo a la escala de Bark (Frecuencias en el rango de 20 Hz a 7700 Hz). Para calcular la entropía de Shannon utilizamos (1) donde: σ_{xx} , σ_{yy} y σ_{xy} son los elementos de la matriz de co-varianzas de los coeficientes espectrales que es de 2x2 debido a que consideramos la parte real y la parte imaginaria como dos variables aleatorias

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

- d). Por cada banda crítica decidir si la entropía está aumentando o disminuyendo respecto al marco de tiempo anterior. La Ecuación (2) establece como el bit

correspondiente a la banda b y marco n se determina usando los valores de entropía de los frames n y n-1 para la banda b.

$$F(n,b) = \begin{cases} 1 & \text{si } [H_b(n) - H_b(n-1)] > 0 \\ 0 & \text{en caso contrario} \end{cases} \quad (2)$$

3.2 El Procedimiento de Monitoreo

Para detectar pautas publicitarias en una señal de audio calculamos la distancia de Hamming entre el MBSES de cada pauta de interés y el MBSES del segmento de señal mas recientemente leído que es del mismo tamaño que la pauta de interés en turno. Recuerde que el MBSES de una señal de audio es una matriz binaria, por ende, puede representarse gráficamente con puntos negros y blancos, en la Fig. 1, el MBSES de una pauta publicitaria (magnificada) concuerda con un pedazo del MBSES el audio monitoreado.

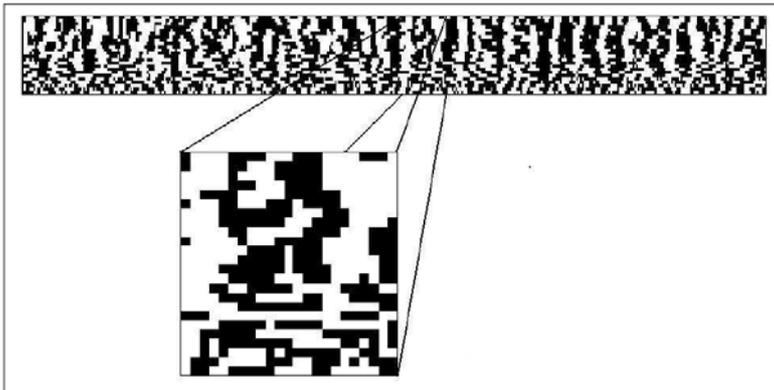


Figura 1. La firma de una pauta publicitaria es una matriz binaria que se presenta como una retícula de cuadros blancos y negros. La firma concuerda en algún punto que corresponde a la aparición del anuncio publicitario.

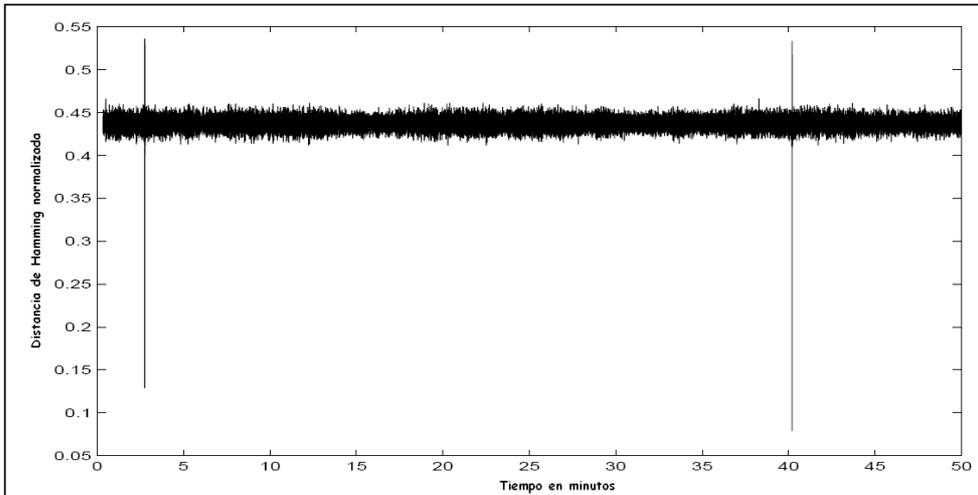


Figura 2. Aparición de una pauta publicitaria encontrada en el minuto 3 y en el minuto 41 donde la distancia de Hamming cae drásticamente.

La distancia de Hamming entre dos firmas es igual al número de bits que son diferentes entre ellas. Si la distancia de Hamming entre el MBSES de la m -ésima pauta y el del último segmento de audio recibido cae por debajo de un umbral preestablecido se declara que acaba de ocurrir la pauta publicitaria etiquetada por el identificador m . En la Fig. 2 una pauta fue detectada en el minuto 3 y vuelta a detectar en el minuto 41 pues como se aprecia la distancia cae drásticamente en esos instantes.

3.2.1 Búsqueda simultánea de varias pautas publicitarias

Por cada unidad de tiempo (dependiendo de la precisión deseada en cuanto al tiempo de aparición que aparecerá en el reporte de ocurrencias de las pautas monitoreadas) se utiliza una ventana de tamaño variable, el tamaño de la ventana se ajusta al tamaño de cada pauta de interés. En cada posición r , la distancia de Hamming entre el MBSES de cada pauta de interés y el MBSES que corresponde a la ventana de longitud variable se determina, la menor de estas distancias se compara con el umbral pre-establecido, si es menor se declara una ocurrencia en el instante r de la pauta con la menor distancia. Las distancias de Hamming son normalizadas de manera que su valor no dependa de la duración del anuncio publicitario. Una distancia normalizada tiene valores en el rango de cero a uno.

4 Experimentos.

Para nuestros experimentos utilizamos 19 Horas de grabaciones continuas de cinco diferentes estaciones de radio, estas grabaciones fueron proporcionadas por la empresa *Contacto Media Research Mexico SA de CV* en el formato mp3@64kbps repartidos en 95 archivos de una hora de duración cada uno. Utilizamos también 13 grabaciones de anuncios publicitarios y los resultados del monitoreo manual de estas estaciones realizado por empleados entrenados para

tal efecto por dicha empresa. Determinamos las firmas de audio (MBSES) de cada uno de los 95 archivos de aproximadamente una hora así como de los 13 archivos de duración variable (la duración de las pautas publicitarias). El proceso de detección de ocurrencias de las pautas publicitarias en las 95 horas de grabación se realizó aproximadamente en una hora. Nuestro sistema detectó 272 ocurrencias, todos ellos fueron verificados en las grabaciones que corresponden a la estación e instante reportado por el sistema de manera que no existieron falsas alarmas (positivos falsos). Adicionalmente, todas las ocurrencias de pautas detectadas por los empleados de *Contact Media Research* fueron también detectadas por nuestro sistema de manera que la tasa de reconocimiento (el régimen de positivos verdaderos) resultó del 100%. La tasa de reconocimiento reportado por Hellmuth *et al* en [14] en experimentos similares es de 99.8%, su monitoreo también es fuera de línea a partir de archivos en formato de compresión con pérdida, en su caso mp3@64kbps y buscando segmentos de audio de 20 segundos (el tamaño promedio de un anuncio publicitario). La Tabla 1 presenta un comparativo donde se incluyen los resultados reportados por Oliveira *et al* [13].

Tabla 1. Comparación con resultados reportados en investigaciones similares

Sistema	Régimen de positivos verdaderos (Tasa de reconocimiento)	Régimen de Falsos positivos (Tasa de errores)
Nuestro sistema propuesto	100 %	0 %
Hellmuth <i>et al</i> [14]	99.8 %	-
Oliveira <i>etal</i> [13]	95.4 %	1 %

5. Conclusiones y Trabajo Futuro.

Corroboramos que nuestra Firma de Audio basada en entropía espectral multi-banda (MBSES) es adecuada para el monitoreo automático de estaciones de radio. La resolución de la firma fué adaptada para el problema particular de detectar anuncios publicitarios que suelen tener alto contenido de voz. En el futuro haremos uso de un índice de proximidad para acelerar la búsqueda de anuncios publicitarios cuando en lugar de ser solo 13 como en los experimentos presentados aquí sean miles de ellos.

Referencias

- [1]. P. Cano, E. Battle, T. Kalker, J. Haitsma. (2002) A review of algorithms for audio Fingerprinting. *Multimedia Signal Processing, IEEE Workshop.* :169-167.
- [2]. A. Group. (2001). Text of ISO/IEC Final Draft International Standar 15938-4. *Information Technology - Multimedia Content Description Interface - Part 4 Audio.* :
- [3]. P. Hellman. (1972) Asymmetry of masking between noise and tone. *Perception and Psychophysics.* :241-246.

- [4]. O. Hellmuth, E. Allamanche, M. Cremer, T. Kastner, C. Neubauer, S. Schmidt, F. Siebenhaar. (2001). Content-based broadcast monitoring using mpeg-7 audio-Fingerprints. *International Symposium on Music Information Retrieval*. : .
- [5]. J. Herre, E. Allamanche, O. Hellmuth. (2001) Robust matching of audio signals using spectral Flatness features. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. :127-130.
- [6]. B. Logan. (2000) Mel frequency cepstral coefficients for music modeling. *International Symposium on Music Information Retrieval (ISMIR)*. : .
- [7]. T. Nakamura, R. Tachibana, S. Kobayashi. (2002) Automatic music monitoring and boundary detection for broadcast using audio watermarking. *SPIE*. : .
- [8]. B. Oliveira, A. Crivellaro, M. César. (2005) Audio-based radio and tv broadcast Monitoring. *Proceedings of the 11th Brazilian Symposium on Multimedia and the web (Webmedia '05) ACM*. :1-3.
- [9]. S. Pauws. (2004) Musical key extraction from audio. *International Symposium on Music Information Retrieval (ISMIR)*. : .
- [10]. S. Seo, M. Jin, S. Lee, D. Jang, D. Yoo. (2005) Audio Fingerprinting based on normalized spectral subband centroids. *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. : .
- [11]. S. Sigurdsson, K. Petersen, T. Lehn-Schioler. (2006) Mel frequency cepstral coefficients: An evaluation of robustness of mp3 encoded music; *International Symposium on Music Information Retrieval (ISMIR)*. : .
- [12]. S. Sukittanon, E. Atlas. (2002). Modulation frequency features for audio Fingerprinting. *IEEE, International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. **2**: 1773-1776.
- [13]. S. Sukittanon, J. Pitton, K. Filali. (2005). Improved modulation spectrum through multi-scale modulation frequency decomposition. *IEEE, International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. **4**: 517-520.
- [14]. M. Veen, F. Bruekers, T. Kalker. (2000). Audio watermarking for monitoring and copy protection; *ACM Multimedia*. :119-122.