

De las matemáticas abstractas hasta el reconocimiento de música

► Dr. Alain Manzo Martínez, Dra. Graciela M. J. Ramírez Alonso, Dr. Luis Carlos González Gurrola, Dr. Fernando Martínez Reyes y el Dr. Rodolfo S. González Garza*

Facultad de Ingeniería, Universidad Autónoma de Chihuahua, Instituto Tecnológico de Morelia*
FINGUACH Año 4, Núm. 13, septiembre - noviembre 2017

Desde la antigüedad la música ha sido un aliciente para la vida del ser humano. En la actualidad la música y en general el audio tienen diferentes fines que van desde terapia hasta el puro placer de escucharla. Desde el punto de vista científico existen diferentes categorías que ayudan a estudiar el comportamiento del audio/música en ciertos entornos cotidianos. Una categoría es Recuperación de la Información de la Música (RIM) la cual es un área que crece rápidamente, especializándose en el reconocimiento, clasificación, indexación y recuperación de datos de la música, principalmente para aplicaciones en *internet*, bases de datos y librerías.

Con respecto a la categoría RIM, existe una subcategoría que hace referencia a los Sistemas de Identificación de Audio Basados por su Contenido. La mayoría de las investigaciones que tratan sobre estos sistemas proponen procedimientos o técnicas de extracción de la Huella de Audio (HA). Una HA se podría decir que es la firma con la que una pieza de audio se identifica y por lo tanto, es única. La HA se utiliza para enlazar "audio desconocido" a sus correspondientes metadatos (nombre de la pieza, intérprete y autor, entre otros) sin importar el formato del archivo de audio.

Para extraer la HA por lo general se utilizan diferentes herramientas matemáticas o de inteligencia artificial. Una herramienta matemática poco explorada, que es sencilla en su interpretación y que puede ser bastante robusta para extraer la HA de una señal de audio es el mapeo o proyecciones de triángulos sobre un círculo unitario en el plano complejo. Como se sabe, un triángulo en geometría es un polígono de tres lados determinado por tres segmentos de tres rectas que se cortan en denominados lados o aristas. Los tres puntos no alineados donde se cortan las rectas son llamados vértices. Dependiendo de las características de los triángulos, éstos se pueden clasificar por la longitud de sus lados o por la amplitud de sus ángulos.

Considerando las definiciones anteriores y tomando en cuenta que un triángulo está determinado por sus vértices en un plano cartesiano, si se consideran los vértices de un triángulo como una terna de números complejos, se puede asociar a cada triángulo con un punto sobre un círculo unitario que está ubicado en el plano complejo. Para mostrar lo antes mencionado se desarrolló un programa en MATLAB donde se generan diferentes tipos de triángulos y se realizan las proyecciones sobre el círculo unitario. En la Figura 1 se muestra la ubicación de los puntos que asocian a los diferentes triángulos que fueron generados por el programa.

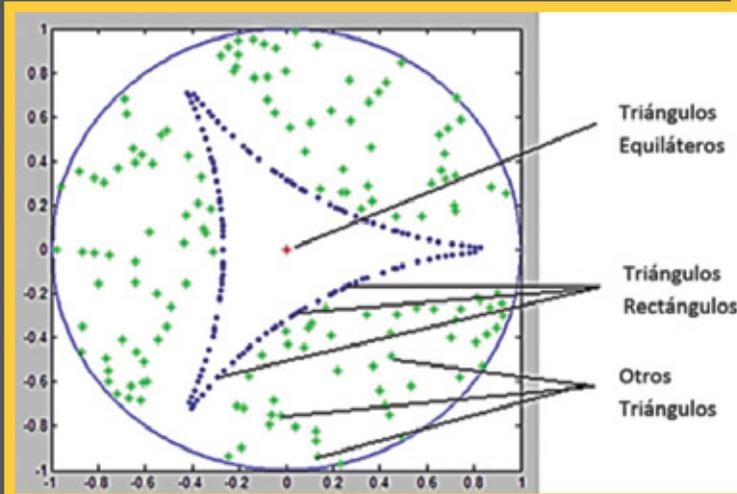


Figura 1. Mapeo de diferentes tipos de triángulos sobre el círculo unitario en el plano complejo.

A manera de ejemplo, considere la serie de tiempo que es mostrada en la Figura 2 y que fue obtenida a partir de una señal de audio. Usando el criterio mínimo-máximo-mínimo puede generarse una secuencia de triángulos sobre esta serie, tal como se muestra en la parte inferior de la Figura 2. Dado que es posible proyectar la secuencia de triángulos dentro del círculo unitario ubicado en el plano complejo, el resultado sería un círculo con diferentes puntos en su interior (puntos rojos) el cual estaría representando por la HA de la señal de audio (ver lado izquierdo de la Figura 3).

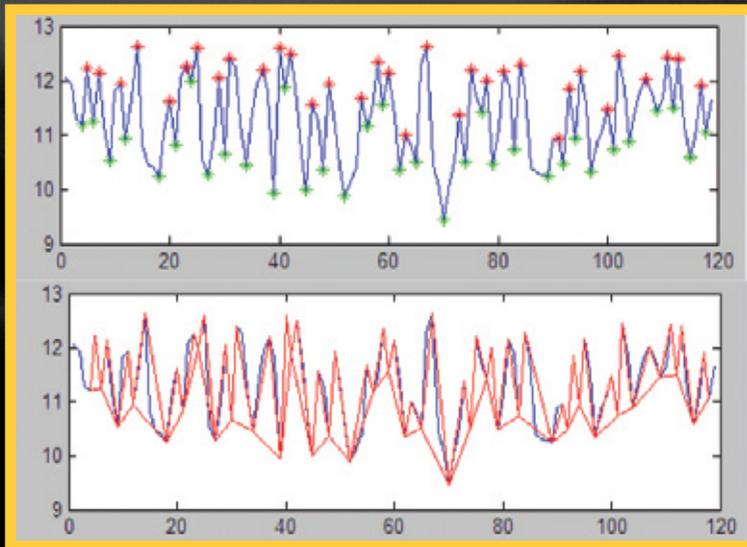


Figura 2. Criterio de máximos y mínimos (puntos rojos y verdes, respectivamente) para generar una secuencia de triángulos sobre una serie de tiempo.

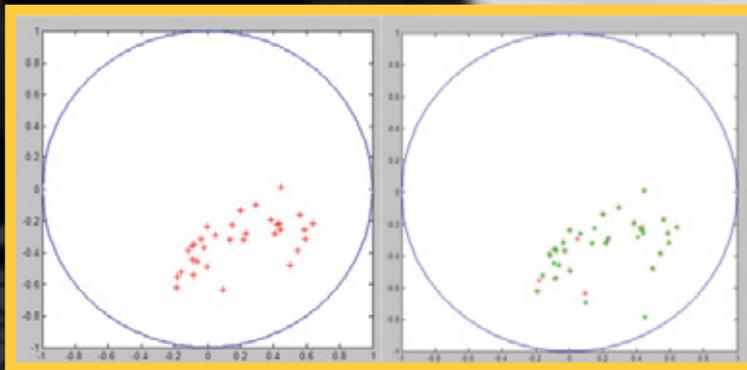


Figura 3. Representación de una HA basada en el mapeo de triángulos sobre un círculo unitario en el plano complejo (lado derecho). Los puntos verdes sobre la gráfica corresponden a la HA generada a partir de una versión degradada de la señal original de audio (lado izquierdo).

Es posible demostrar que la versión degradada de una señal de audio puede generar una HA muy similar a la HA de su versión original utilizando este procedimiento. En la parte derecha de la Figura 3, mediante puntos verdes se muestra la HA generada a partir de una versión degradada de la señal de audio original. Como era de esperarse, existe una correspondencia en cuanto a la ubicación de los puntos rojos y verdes. Esto se logra sólo si los triángulos son congruentes o semejantes entre sí en las secuencias obtenidas a partir de las series de tiempo de ambas piezas de audio. Como último paso de esta técnica de reconocimiento de audio se puede usar una métrica para determinar si la versión degradada de una pieza de música puede ser identificada como el mismo contenido perceptual con respecto a su versión original.

Referencias

- Cano, P., Battle, E., Kalker, T., Haitsma, J., (2002). *A Review of Algorithms for Audio Fingerprinting*, *Proceedings on IEEE Workshop Multimedia Signal Processing*, 169-173.
- Allamanche, E., Jurgen, J., Hellmuth, O., Froba, B., Cremer, M., (2001). *AudiID: Towards Content-Based Identification of Audio Material*, *110th AES-Convention*, Amsterdam, Convention Paper 5380.
- Camarena, A., Chavez E., (2006). *A Robust Entropy-Based Audio Fingerprint*, *IEEE International Conference on Multimedia and Expo*.