

Algoritmo para el cálculo del ritmo en una señal de audio digital

Luis Felipe Romero Morales, José Martín Flores Albino

Universidad Autónoma del Estado de México,
Centro Universitario UAEM Valle de México, Estado de México, México

lromero628@alumno.uaemex.mx, jmfloresa@uaemex.mx

Resumen. En este trabajo se presenta un algoritmo para la extracción del ritmo de una señal audio digital. El ritmo está formado por los cambios de intensidad en la señal de audio. El ritmo marca cambios significativos en las señales acústicas. A través de un procedimiento de filtrado para la extracción de la envolvente de la señal de audio, luego a través de una señal cuadrada se marcan los puntos de inicio y su duración de los cambios de ritmo en la señal de audio. El algoritmo se prueba con una señal de un metrónomo de 60pps para validar esta propuesta.

Palabras clave: Extracción del ritmo, señal envolvente, filtro digital.

Algorithm for Rhythm Evaluation in a Digital Audio Signal

Abstract. In this paper presents an algorithm for extraction of audio rhythm of a digital signal. The rhythm is formed by changes of intensity in the audio data signal. The rhythm marks significant changes in the acoustic signals. Through a procedure of filtering for extraction the envelope of the audio signal, then through a square signal marked the start point and duration of changes on the envelope. The algorithm is tested with a signal of a metronome of 60bps to validate this proposal.

Keywords: Rhythm extraction, signal envelope, digital filtering.

1. Introducción

Este artículo presenta un algoritmo para el cálculo del ritmo de una señal digital de audio. El ritmo se expresa en la envolvente de la señal de audio y marca los cambios significativos, sin importar el tono o frecuencia de la misma. El ritmo se experimenta al producirse cambios en el sonido, que ocurren de forma secuencial y organizada durante el paso del tiempo [1]. El algoritmo presentado produce como resultado datos que indican el tiempo y la duración de los cambios asociados al ritmo de la señal.

La detección del ritmo tiene diversas aplicaciones; por ejemplo: En el ámbito médico existe una relación entre el ritmo del corazón y la detección de algunas enfermedades; en los ritmos cerebrales de señales asociadas a la actividad del cerebro, indican su funcionamiento y estado [2]; en la música, el ritmo puede usarse para sincronizar efectos como son de iluminación y movimiento que harían más vívida la experiencia musical. Entonces, se entiende que el ritmo puede ser asociado a diferentes áreas de interés.

Hay trabajos que presentan aplicaciones y técnicas para la extracción del ritmo. En [3], los autores proponen un orden jerárquico de géneros musicales por medio de una clasificación automática de audio; son tres los principios que guían su investigación: la textura del timbre, contenido rítmico y el tono. En [4], se propone un método para representar las periodicidades creadas por un ritmo musical, usan las crestas del sonido para calcular un ritmo de acompañamiento y hacen uso de la transformada de wavelet inversa. En [5], se utiliza la transformada de wavelet y de Fourier para buscar la estructura rítmica y melódica de la música, utilizan un método de escalogramas para percusiones como técnica para analizar ritmos.

El orden de este artículo es el siguiente. En la sección 2, se presenta el algoritmo propuesto y los procedimientos en cada fase. Se inicia con la adquisición de la señal de audio y culmina con la generación de un archivo donde se enlista el tiempo en el que ocurre un cambio significativo en el ritmo y su duración. En la sección 3 se presenta la etapa de pruebas por medio de un ejemplo y los resultados de aplicación del algoritmo. Se utiliza una señal de un metrónomo de 60 pulsos/min. Al aplicar el algoritmo se generan datos que indican los eventos asociados al ritmo de la señal. En la última sección se dan las conclusiones y trabajo futuro de esta investigación.

2. Algoritmo para la detección del ritmo

Para el desarrollo del algoritmo se utiliza el ambiente MatlabTM por ser una herramienta potente de cálculo para actividades de investigación. En la fig. 1, se presenta un diagrama de flujo de los procesos que se ejecutan en el algoritmo.

Proceso 1: El objetivo es cargar los datos de la señal de audio digital en formato mp3. La Frecuencia de Muestreo usada es de $fs=44\ 100$ muestras/segundo [6]. La amplitud está normalizada en el rango de -1 a 1 para los canales izquierdo y derecho. Así entonces un segundo de audio tendrá 88 200 valores de punto flotante en formato de 16 bits.

Proceso 2: La señal obtenida se encuentra dividida en los canales izquierdo y derecho, por ser una señal en “estéreo”. El cambio de ritmo es muy similar en ambos canales, por lo que se procesa una señal monoaural obtenida al promediar los dos canales. Lo anterior además reducirá el número de datos a la mitad.

Proceso 3: La información del ritmo se encuentra en la envolvente de la parte positiva y negativa de la señal monoaural. La envolvente es la representación el ritmo debido a que marca los cambios en la señal sin importar el tono o frecuencia. De manera que se decide eliminar la parte negativa de la señal debido a que, para propósitos prácticos, es la misma que la envolvente de los valores positivos. Para lograrlo hay dos formas: la primera: tomar el valor absoluto; y la segunda: elevar al cuadrado y después

obtener la raíz cuadrada. Se elige el primer método por ser menos demandante de recursos de cómputo.

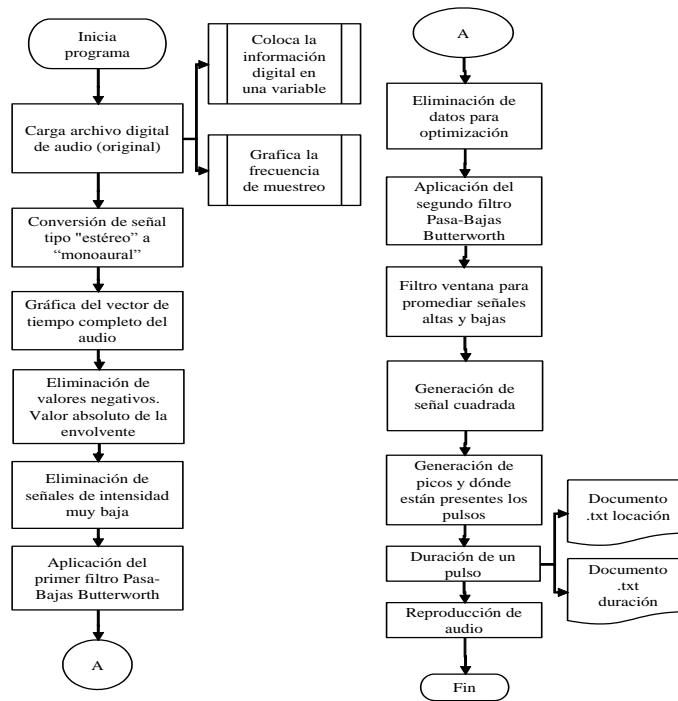


Fig. 1. Diagrama de flujo general del algoritmo para el cálculo del ritmo de una señal digital.

Proceso 4: Dentro de los archivos de audio, en ocasiones se presenta ruido de intensidad muy baja, esto puede afectar el funcionamiento del algoritmo, al detectarlos como parte del audio. Para solucionarlo se utiliza un umbral de nivel bajo para acotar lo que es ruido en la señal.

Proceso 5: La envolvente representa cambios de frecuencia muy baja. La frecuencia de sonido baja es de 20Hz y de 20kHz la frecuencia alta, por lo que se aplica un filtro de elimine los cambios de las señales de audio, y con esto obtener la envolvente. Se aplica un filtro digital tipo Butterworth pasa-bajos con frecuencia de corte de 10 Hz para este propósito.

La función de transferencia un filtro pasa-bajos es:

$$|H(\omega)|^2 = \frac{1}{1+(\omega/\omega_c)^{2N}}, \quad (1)$$

donde ω es la frecuencia angular de la señal, $\omega_c = 2\pi f_c$, f_c es la frecuencia de corte y N orden del filtro. Para llevarlo a su forma digital se utiliza la transformación bilineal.

$$S = \frac{1}{T_s} \frac{z-1}{z+1}, \quad (2)$$

donde T_s es el periodo de muestreo, z es la variable de la transformada z (z^{-1} es un retardo).

Proceso 6: La señal de la envolvente tiene una frecuencia de muestreo de 44 100 muestras por segundo. Se han eliminado las frecuencias por arriba de los 20Hz. Recordando el teorema de Nyquist-Shannon, que indica que para recuperar una señal debe ser muestreada al menos a una tasa de muestreo del doble del ancho de banda de la señal. Se concluye que la señal de la envolvente está en exceso muestreada. El ancho de banda de la señal de envolvente es de a lo más 10 Hz, así que la señal de envolvente se debería submuestrear a 1/2205 veces. Para hacerlo se toma un valor por cada 2205 de la envolvente del audio. Esto además reduce la cantidad de información en la señal.

Proceso 7: Con una menor cantidad de datos se aplica una vez más un filtro pasabajas con el propósito de suavizar la envolvente.

Proceso 8: El ritmo está expresado en la forma de la envolvente. Un efecto no deseado de aplicar el filtro pasa-bajos es que se acumula un valor residual que va elevando el nivel promedio de la señal de la envolvente. Para eliminar este aumento de nivel se aplica un filtro del tipo media móvil (*moving average*) para el cálculo del valor medio y posteriormente eliminarlo a través de una simple resta.

Proceso 9: En esta etapa se ha obtenido la señal del ritmo. Pero con el propósito de cuantificar los cambios en la señal del ritmo. Se marca con una señal cuadrada en el momento donde hay un cambio en la envolvente y su duración. Para ello se calculan los máximos y mínimos en la envolvente. Usando los puntos de máximo para indicar el tiempo en el que inicia un cambio en el ritmo y los puntos mínimos marcan cuando termina el cambio detectado, es decir su duración. Cuando la señal cuadrada pasa a nivel alto indica el inicio de cambio en la envolvente. La duración de los cambios serán el tiempo que tarda en alcanzarse los mínimos de la señal de la envolvente. Así se obtendrán los cuándo y cuánto duran los cambios en el ritmo.

Proceso 10: Para medir los tiempos se cuenta el número de muestras de la señal cuadrada y tiempo el muestreo de la señal. Estos resultados se registran en dos archivos de texto plano (txt). El primero contiene las marcas de tiempo que corresponde al inicio de un cambio en el ritmo y el segundo la duración del cambio en la señal del audio digital inicial.

3. Pruebas y resultados

Se presentan los resultados al aplicar el algoritmo anterior para una señal de audio. La señal seleccionada es la señal de un metrónomo a 60 bpm (bits por minuto), el archivo es de formato mp3 y una frecuencia de muestreo de 44.1k muestras/segundo, la magnitud de los valores tiene un rango entre -1 y 1, con una codificación digital de 16 bits de resolución. La razón por se toma está señal es que resulta claro el resultado esperado y permitiendo evaluar si el algoritmo entrega resultados coherentes.

En la fig. 2, se presenta la señal sin procesar. En la figura 2.a es el detalle de un solo pulso del metrónomo, se observa su complejidad. En la figura 2.b se presentan 10 pulsos del metrónomo.

En la fig. 3, se presenta la señal después de convertirla en monoaural y tomar su valor absoluto para dejar solo la parte positiva. Esto prepara la señal para el cálculo del ritmo.

En la fig. 4, se muestra el resultado de filtrar las frecuencias altas (tono de audio) de la señal anterior. El filtro usado es de tipo IIR (Respuesta al Impulso Infinita) digital pasa-bajos tipo Butterworth. Para el primer filtro se utilizaron los siguientes parámetros: frecuencia de corte: $f_c=10$ Hz y $N=4$ es el orden del filtro. El resultado es una señal que representa la envolvente de la señal. La señal resultante tiene una cantidad de 44.1K datos por segundo que impactaría en el tiempo de procesamiento, por eso se realiza un submuestreo de la señal, al tomar 1 de cada 2205 datos, reduciendo la cantidad de datos a procesar.

Observar que la envolvente muestra los pulsos de la señal y notar que se va incrementando su nivel al paso del tiempo, esto produce un “efecto de escalera”, que se debe a la acumulación del valor promedio de la señal. Para eliminarlo el efecto escalera se calcula el valor medio de la señal de la envolvente a través de un filtro de valor medio, ver fig. 5.

Para eliminar efecto escalera, se cancela el valor medio por simple resta de la envolvente. La señal resultante muestra los pulsos del metrónomo continuo a través del contorno, en la fig. 6, se observan los pulsos del sonido del metrónomo con claridad.

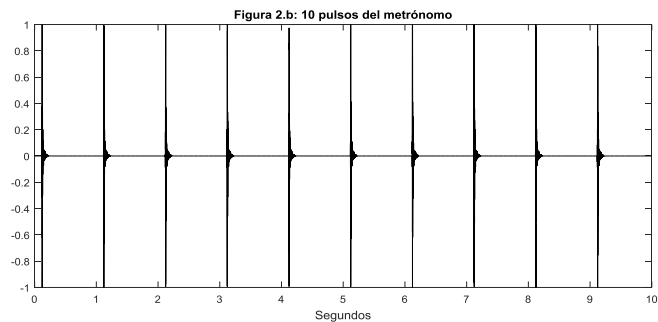
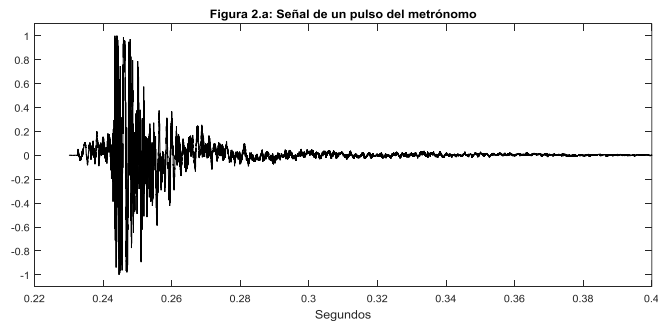


Fig. 2. Señal de audio de sonido de un metrónomo a 60bps. Señal sin procesar: 2.a) es la señal de un solo pulso. 2.b) Trazo de diez pulsos.

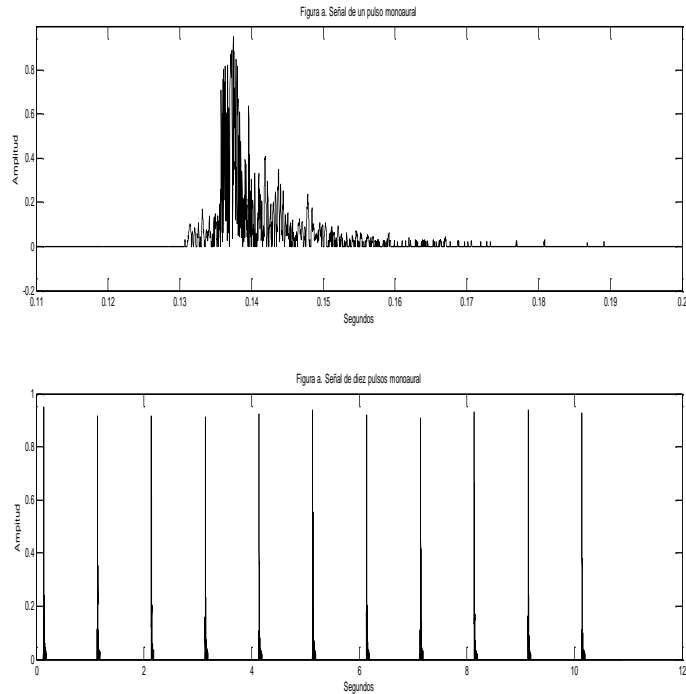


Fig. 3. Señal de audio de sonido de un metrónomo a 60bps. Valor absoluto de la señal, recortada su parte negativa: 3.a) Un pulso. 3.b). Trazo de diez pulsos.

El siguiente paso es generar una señal que sea capaz de indicar el principio y el fin de un pulso rítmico dentro de la señal. Esto es posible con la generación de una señal cuadrada. En el resultado que se observa en la **fig. 7.**, donde se forman los pulsos que indican el tiempo donde el pulso se presenta y su duración. Este paso es muy importante porque la generación de la señal cuadrada concede información del momento en que hay un cambio significativo en el audio. Para ello se evalúan los puntos de máximos y mínimos de la señal de la envolvente. En los máximos se indican cuando hay un cambio en la señal de audio y el punto mínimo cuando la señal cuadrada regresa al nivel de referencia. La duración de un cambio en la señal de la envolvente es equivalente al tiempo que la señal cuadrada está en nivel alto.

Resultados: La señal de audio para la prueba del algoritmo es de un metrónomo a 60 bpm, con duración de 10 minutos y 11 segundos, dando un total de 606.6 pulsos. El algoritmo arroja una cantidad total de 606 pulsos con una duración de 0.4 segundos. Los archivos generados contienen la información digital del tiempo de cuándo hay un cambio de la señal y su duración. En la tabla 1 se muestran los resultados. Se observa que el primer pulso inicia a los 0.1 segundos, desde ese tiempo en adelante, cada segundo se produce otro pulso y la duración de los pulsos es de 0.4 segundos. Estos resultados están de acuerdo a lo esperado.

Algoritmo para el cálculo del ritmo en una señal de audio digital

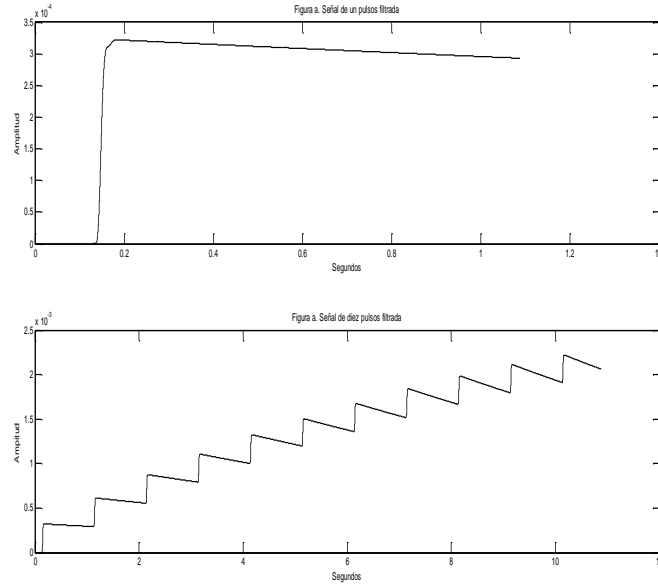


Fig. 4. Aplicación del primero y segundo filtro. 4.a Trazo de un pulso. 4.b Trazo de diez pulsos (*efecto escalera*).

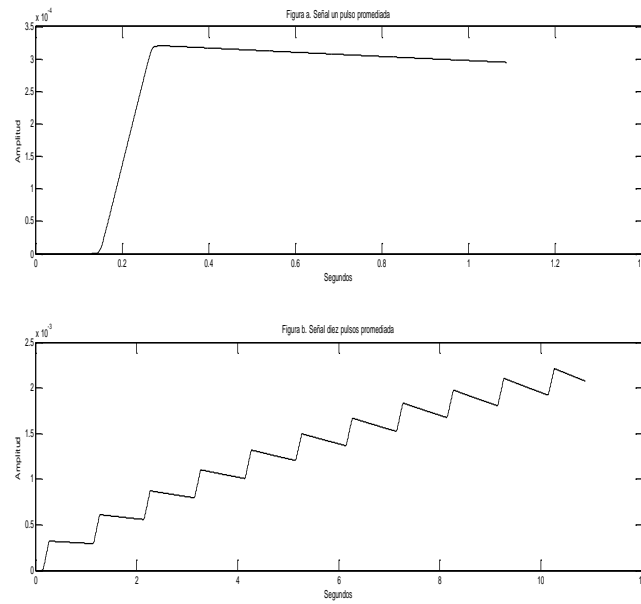


Fig. 5. Valor promedio de la señal: Se obtiene al calcular el valor medio de la señal. 5.a Trazo de un pulso. 5.b Trazo de diez pulsos (*efecto escalera*).

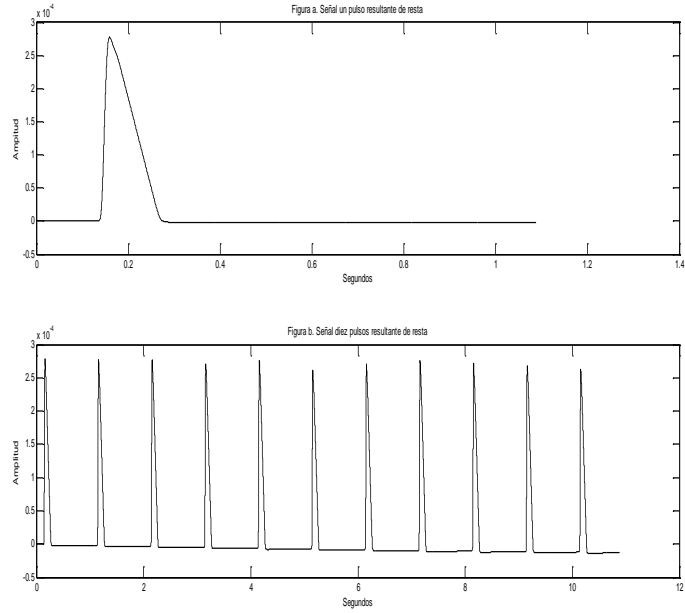


Fig. 6. Señales de ritmo del metrónomo: Resultante de restar el valor promedio de la señal y la filtrada. 6.a Trazo de un pulso. 6.b Trazo de diez pulsos.

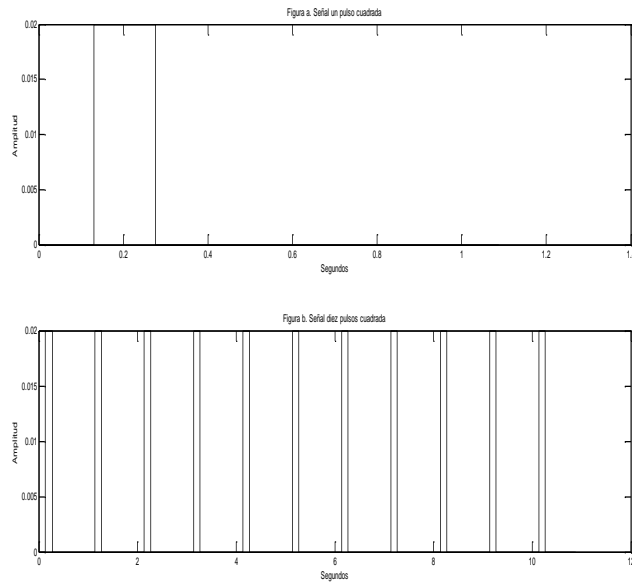


Fig. 7. Señal Cuadrada: Los pulsos cuadrados indican la detección hecha por el algoritmo en la señal. 7.a Trazo de pulso. 7.b. Trazo de diez pulsos.

Tabla 1. Valores en tiempo de la señal de audio procesada (únicamente los primeros 10 pulsos).

Inicio de pulso (x100 milisegundos)	Duración (x50 milisegundos)
1=0.1 segundos	4=0.4 segundos
11=1.1 segundos	4=0.4 segundos
21=2.1 segundos	4=0.4 segundos
31=3.1 segundos	4=0.4 segundos
41=4.1 segundos	4=0.4 segundos
51=5.1 segundos	4=0.4 segundos
61=6.1 segundos	4=0.4 segundos
71=7.1 segundos	4=0.4 segundos
81=8.1 segundos	4=0.4 segundos
91=9.1 segundos	4=0.4 segundos

4. Conclusiones y trabajo futuro

Se concluye que los pulsos contenidos en el audio de prueba y su duración corresponden a la señal usada, pues se están presentando en la posición y con la duración correcta de todos los pulsos presentes. Esta señal sirvió para una validación inicial del algoritmo ya que los pulsos se identificaron de manera precisa; sin embargo, es necesario realizar las pruebas para evaluar la capacidad del algoritmo y los ajustes que deba de tener según la aplicación. Al probarse con una señal de audio digital compleja se observaron resultados que mostraban sincronía con los datos obtenidos. Como trabajo futuro se debe probar exhaustivamente el algoritmo con diferentes señales de audio y realizar un estudio comparativo con otras metodologías existentes para la extracción del ritmo, además de usar el algoritmo en aplicaciones donde esta técnica resulte útil.

Referencias

1. Encyclopædia Britannica. <https://global.britannica.com/art/rhythm-music> (2016)
2. Andrea, B.: Técnicas de procesamiento de EEG para detección de eventos. Argentina: Universidad Nacional de la Plata (2013)
3. George T., Perry C.: Musical genre classification of audio signals. IEEE Transactions on Speech and Audio Processing, Vols. 1 de 210, Issue 5, ISSN: 1063-6676, pp. 293-302 (2002)
4. Leigh, S., Henkjan, H.: Frequency Representation of musical rhythm by continuous wavelets. Journal of Mathematics And Music, vol. 2, Issue 2, pp. 81-97 (2008)
5. Cheng, X., Jarod, H., James, W.: Time-frequency Analysis of Musical Rhythm. Notices of the AMS, vol. 56, n. 3, pp. 256-372 (2009)
6. Jesualdo, F.: Tema 3: Audio, Tecnologías para los Sistemas Multimedia. dis.um.es/~jfernand/0405/tsm/tema3.pdf, 1-65 (2004)
7. Kamenov, A.: Digital Signal Processing for Audio Applications. Create Space Independent Publishing (2013)
8. Joshua, R., McPherson, A.: Audio Effects: Theory, Implementation and Application. CRC Press Taylor & Francis Group (2015)

Luis Felipe Romero Morales, José Martín Flores Albino

9. Aggelos, G., Vassilis, K., George, C.: Towards Multi-Purpose Spectral Rhythm Features: An Application to Dance Style, Meter and Tempo Estimaion. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Vol. 24, Issue 11, pp. 1885–1896 (2016)
10. Shiratori, T., Nakazawa, A., Ikeuchi, K.: Detecting dance motion structure through music analysis. In: *Automatic Face and Gesture Recognition*, DOI: 10.1109/AFGR.2004.130164 (2004)