

Identificación de BPM (Beats Per Minute) - Presentación de propuesta

Bargas, Cassiet, Grinovero

Un "beat" ("tiempo" en español), significa "un pulso que se repite de forma regular y constituye la base de un patrón musical". La velocidad a la que se reproducen los patrones se llama tempo. El tempo se mide en beats por minuto o BPM. Así que, si decimos que una pieza musical está "a 120 BPM", queremos decir que hay 120 beats (pulsos) por minuto.

Se pretende desarrollar un sistema que permita identificar de manera automática el tempo de una canción, proporcionando una grabación de audio de la pieza musical y determinando la medida en BPM correspondiente a la canción.

Primer paso

1. Aplicar la Transformada de Fourier al archivo de audio:

Para estimar el tempo de la canción, el algoritmo debe usar ciertas características que destaquen de la canción, como cambios de nota, golpes de percusión, etc. porque los beats suelen ocurrir en esos instantes de tiempo.

Para eso se aplica la transformada de fourier.

Obtenemos una curva de energía espectral a lo largo del tiempo

$$E(t_i)$$

2. Calcular la diferencia de energía en el tiempo

Para identificar los cambios rápidos de energía en el tiempo, calculamos

$$E'(t_i) = E(t_i) - E(t_{i-1})$$

Asegurar que sea valor absoluto. (parametro clave tamaño de ventana apropiado)

3. Detección de picos

Identificar los puntos locales donde $E'(t_i)$ alcance un máximo.

Podemos usar un umbral para solo considerar cuando la energía supere ese umbral.

Segundo paso - Estimación del Tempo (BPM) - aca hay varias alternativas:

Alternativa 1

La primera alternativa se basa en el desarrollo de [2], a continuación se describe de manera simplificada el proceso:

1. Los tiempos de los picos detectados indican los posibles beats. Se asume que los **transitorios tienden a ocurrir en los tiempos de beats y sus subdivisiones.**

Los **transitorios** son eventos breves y de alta energía en una señal de audio que suelen indicar cambios rápidos y significativos, como el ataque de un instrumento percusivo (por ejemplo, el golpe de un tambor) o el inicio de una nota musical. En el contexto del análisis de audio y la extracción de BPM, los transitorios son cruciales porque representan los momentos de mayor cambio en la energía del sonido, que a menudo coinciden con los beats o pulsos de la música.

2. Función de verosimilitud: Función de densidad de probabilidad (PDF) para los tiempos de los transitorios, considerando diferentes posibles tempos

$$p_t(t) = \sum_{i=0}^3 p_i G([t]_T - b_i)$$

Donde $G(x)$ es una distribución gaussiana centrada en 0, y los b_i son los tiempos de

los sub-beats (ej, cuarto-beats).

Para seleccionar $G(x)$ Usa una distribución gaussiana con una varianza ajustada según el tiempo.

3. **Máxima Verosimilitud:**
$$L_t(T) = \sum_{i=0}^{N-1} \log p_t(t_i)$$

Se busca el tiempo T que maximiza la verosimilitud $L_t(T)$ en un rango razonable (e.g., 70-140 BPM).

Alternativa 2:

1. **Cálculo de Distancias Temporales:** Una vez que se han identificado los picos, se calcula la distancia temporal entre cada par de picos adyacentes. Esto se hace restando el tiempo de muestra del pico anterior del tiempo de muestra del pico siguiente.
2. **Cálculo de la Media de las Distancias Temporales:** Luego, se calcula la media de todas estas distancias temporales. Esto proporciona una estimación del periodo promedio entre los picos de energía.
3. **Conversión a BPM:** Finalmente, para obtener una medida más intuitiva, la media de las distancias temporales se convierte a BPM. Esto se logra dividiendo el número de muestras por la media de las distancias temporales y multiplicando por la frecuencia de muestreo del audio. Esto nos da el número de beats por segundo, que luego se convierte a BPM multiplicando por 60.

Alternativa 3

Esta alternativa se basa en el desarrollo de [3], a continuación se describe de manera simplificada el proceso:

1. **Selección de Candidatos de Tiempo - Metodo de minimos cuadrados:**
Se utiliza un enfoque de mínimos cuadrados para identificar los mejores candidatos para el tiempo (la velocidad de la música, medida en beats por minuto) y las ubicaciones de los tiempos fuertes (los momentos destacados del ritmo).
Se asume un tiempo R (en beats por segundo) y una ubicación de tiempo fuerte t , y se define una señal de flujo de energía esperada, que representa el conocimiento previo de cómo debería lucir la señal de flujo de energía real.
Se calcula la distancia euclidiana entre la señal de flujo de energía esperada y la real para evaluar qué tan bien coinciden, considerando un factor de escala desconocido K para ajustar la magnitud de la señal.
Se minimiza la norma L_2 para encontrar el valor óptimo de K , que es esencialmente una medida de cuán bien coinciden las señales de flujo de energía esperada y real.
2. **Correlación Cruzada:**
Para calcular esta distancia, se utiliza la correlación cruzada, que compara las dos señales punto por punto y calcula su similitud.
El objetivo es encontrar valores de tiempo y ubicaciones de tiempos fuertes que maximicen la correlación cruzada, lo que indica una buena coincidencia entre la señal esperada y la real.
Se discretiza el tiempo en ciertos valores y se calcula la correlación cruzada para cada combinación posible de tiempo y ubicación de tiempos fuertes, lo que nos da un conjunto de candidatos.

3. Normalización y Selección de Candidatos:

Se normalizan las correlaciones cruzadas para obtener una puntuación local máxima de 1 para cada par candidato de tempo y ubicación de tiempos fuertes.

Para reducir el costo computacional, se seleccionan solo los 10 a 15 mejores candidatos de tempo y, para cada uno de estos, se retienen solo las 10 a 15 mejores ubicaciones de tiempos fuertes.

Esto resulta en un conjunto más pequeño pero aún representativo de candidatos de tempo y ubicaciones de tiempos fuertes, que se utilizarán en las etapas posteriores del algoritmo.

4. Cálculo de BPM:

Después de haber seleccionado los mejores candidatos para el tempo y las ubicaciones de los tiempos fuertes en el paso anterior, el algoritmo calcula el tempo en beats por minuto (BPM) utilizando la información de los tiempos fuertes detectados.

El cálculo del BPM se realiza generalmente calculando la diferencia de tiempo entre los tiempos fuertes consecutivos y luego convirtiendo esta diferencia de tiempo en una medida de BPM.

Aquí está el proceso básico:

a. Diferencia de Tiempo entre Tiempos Fuertes Consecutivos:

Se calcula la diferencia de tiempo entre dos tiempos fuertes consecutivos. Esto implica medir el tiempo transcurrido entre dos puntos en los que se detecta un golpe fuerte en la música.

b. Conversión a BPM:

Una vez que se tiene la diferencia de tiempo entre los tiempos fuertes consecutivos, se convierte esta diferencia a una medida de BPM.

Para convertir la diferencia de tiempo a BPM, se utiliza la fórmula:

$$BPM = 60 / \text{Diferencia de tiempo}$$

Esto convierte la diferencia de tiempo en una medida de beats por minuto, que es la velocidad a la que se reproduce la música.

c. Promedio de Múltiples Mediciones:

Para obtener un valor de BPM más preciso y robusto, generalmente se toma el promedio de las mediciones de BPM obtenidas de varias parejas de tiempos fuertes consecutivos a lo largo de la pista de música.

Propuesta a desarrollar:

Estimamos que la **Alternativa 2** tendrá problemas al momento de estimar los BPM, ya que no tiene en cuenta el peso de los sub-beats al momento de calcular el tempo, por lo que estos contarán para el cálculo de la media y afectarán el resultado real.

Por lo que se desarrollará el Trabajo Práctico Integrador basándonos en la **Alternativa 1** y la **Alternativa 3**, eligiendo durante el desarrollo las aplicaciones que nos parezcan más óptimas y prácticas.

Cálculo de la certeza:

Según lo establecido por [1], la certeza será evaluada como el cociente entre la cantidad de tempos calculados de manera correcta respecto al total de piezas musicales.

Se considerará que la estimación del tempo será correcta si difiere en un 5% con el tempo real siguiendo el principio:

$$0.95\alpha T < T_R < 1.05\alpha T \text{ con } \alpha \in \{1/2, 1, 2\} \quad (1)$$

El beat, al ser un concepto subjetivo, dependiente de la manera en que la gente “siente” la música, es difícil definir la posición correcta del beat de una manera objetiva.

Por ejemplo, es normal que al escuchar una pieza con compás 4/4, las personas identifican los beats cada cuarto de nota o cada media nota. Por lo que es posible que el tempo identificado sea el doble o la mitad del tempo real.

De esta manera, se tomará el valor de α de $\frac{1}{2}$, 1, o 2 según corresponda.

BIBLIOGRAFÍA:

[1] ALONSO, M. Tempo and beat estimation of musical signals. En *Int. Soc. for Music Information Retrieval Conf. (ISMIR 2004)*. 2004. p. 158-163.

<https://www.ee.columbia.edu/~dpwe/ismir2004/CRAFILES/paper191.pdf>

[2] LAROCHE, Jean. Estimating tempo, swing and beat locations in audio recordings. En *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No. 01TH8575)*. IEEE, 2001. p. 135-138.

<https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=516880b7f48e1dba92343e1b14935f4272ee8d4d>

[3] LAROCHE, JEAN. Efficient Tempo and Beat Tracking in Audio Recordings. *J. Audio Eng. Soc.*, 2003, vol. 51, no 4, p. 227.

<https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=e2fdf3918394adbd2eb9fb8e523895e703da2986>