

FACULTAD DE INGENIERÍA DE LA
UNIVERSIDAD DE LA REPÚBLICA

PROCESAMIENTO DIGITAL DE SEÑALES DE
AUDIO

CURSO 2012

Reconocimiento de Tempo

Autores:

Gonzalo GUTIÉRREZ

Matías TAILANIÁN



7 de octubre de 2012

Segunda etapa

De estudios psicoacústicos sobre la percepción del *beat* se deduce que hay ciertas manipulaciones y simplificaciones que se pueden realizar sobre las señales sin alterar la percepción del *beat*.

Consideremos el siguiente experimento: se divide en bandas de frecuencia y se detecta la envolvente de cada banda. Por otro lado se genera un ruido gaussiano y se lo multiplica por la envolvente anterior en cada banda de frecuencia. Se suman las señales de las distintas bandas y se obtiene así un ruido de amplitud modulada. Lo interesante de este experimento es que todavía es posible identificar las características métricas de la señal original. Otros experimentos como tomar solamente la salida de uno de los filtros del banco o combinar linealmente las envolventes de las subbandas, indican que no se pueden realizar dichas simplificaciones, ya que alteran el contenido métrico de la señal.

Técnicas clásicos

Para la detección del *beat* comúnmente se utilizan técnicas basadas en una detección estadística, simplemente utilizando la energía de la señal o analizando la energía en una banda de frecuencias seleccionada, o técnicas de filtrado y detección del ritmo mediante bancos de filtros y procesamiento de frecuencias seleccionadas.

Una de las técnicas más utilizadas, por casi todos los algoritmos de detección y tracking del *beat*, es la división en subbandas de frecuencias. A modo de ejemplo se muestra un banco de 6 filtros que puede ser utilizado: el primero es un pasabajos de frecuencia de corte en 200Hz, el siguiente un pasabanda entre 200Hz y 400Hz, otro entre 400Hz y 800Hz, otro de 800Hz a 1600Hz, de 1600Hz a 3200Hz, y por último un pasaaltos de frecuencia 3200Hz. Una vez pasada la señal por el banco de filtros se realiza la detección de envolvente en cada subbanda.

Luego de extraída la envolvente y procesada con alguna técnica en cada canal, una técnica clásica para determinar el *tempo* de la señal es utilizar un banco de filtros resonadores. Considerando un filtro de retardo T y ganancia α , si la señal de entrada tiene amplitud A y periodo K , la señal de salida va a presentar valores altos

si $T=K$ (o eventualmente algún divisor). De lo contrario se obtiene una salida chica. Al final del algoritmo se obtiene una especie de puntaje para cada filtro resonador. El filtro que obtenga más puntaje será el que tenga un retardo más parecido al período de la señal.

El ritmo como concepto musical es intuitivamente fácil de entender pero de alguna manera difícil de definir. La experiencia del ritmo involucra movimiento, regularidad, agrupamiento, acentuación y diferenciación (Handel, 1989, p. 384). No hay un “ground truth” que se pueda encontrar en simples medidas de la señal. El único *ground truth* es el acuerdo que se realiza entre las percepciones de distintas personas que escuchan la pieza musical. Pensando en técnicas de tracking de *beat* automáticas, más allá de todas las técnicas utilizadas, es necesario dar en algún momento, un salto semántico que nos permita obtener un resultado similar a nuestro *ground truth*: la percepción humana. Para lograr este objetivo se pueden realizar algunas consideraciones heurísticas como acotar el rango de tiempo que una persona tiene a marcar al escuchar una pieza. Una regla simple es por ejemplo poner un umbral donde todas las piezas que tengan un tempo detectado superior, se utilice la mitad del tempo detectado.

Musical Beat Tracking via Kalman Filtering and Noisy Measurements Selection

En la etapa de preprocesado se realiza la detección del comienzo (*Onset Detection*) y estimación del período (*Periodicity Estimation*).

El *onset* de una señal proporciona el cambio de intensidad de la pieza musical a lo largo del tiempo. Básicamente se distinguen 2 tipos de cambios que producen la aparición de un *onset*:

- pulsos causados por la percusión, que se reflejan como pulsos ruidosos de elevada energía
- cambios de pitch y armonías debido a la llegada de una nueva nota

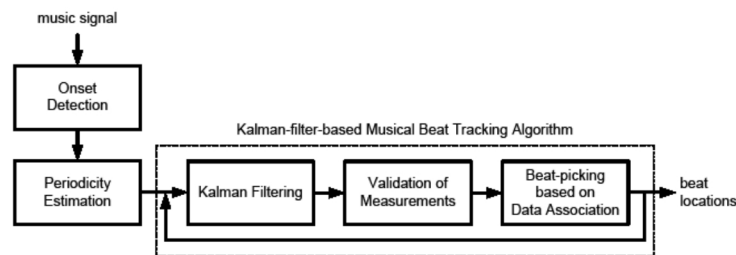


Figura 1: Diagrama de bloques

Se utiliza el método de la distancia cepstral para identificar los inicios musicales¹ y la función de autocorrelación para determinar el período.

¹L. Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition”, Proceedings of the IEEE, vol. 77, No. 2, pp. 257-286, 1989.

Para el filtrado de Kalman se modela al sistema de Beat Tracking como un sistema dinámico de ecuaciones lineales donde el Tempo musical puede ser modelado como una variable de estado de un sistema dinámico estocástico. Aquí, se utiliza el modelo dinámico propuesto por Cemgil² para formular el problema del Beat Tracking en un marco probabilístico.

Selección de los Beats

El rendimiento del Beat Tracking basado sólo en el filtrado de Kalman no alcanza una muy buena performance. Para mejorar su rendimiento, necesitamos una forma más inteligente para localizar beats reales a partir de la señal de audio. Se aplicará entonces una técnica de asociación probabilística de datos (**PDA**), o la mejorada, **Enhanced PDA (EPDA)**.

Improved Perceptual Tempo Detection of Music

Un esquema general muy ilustrativo del algoritmo propuesto se puede ver en la figura 2.

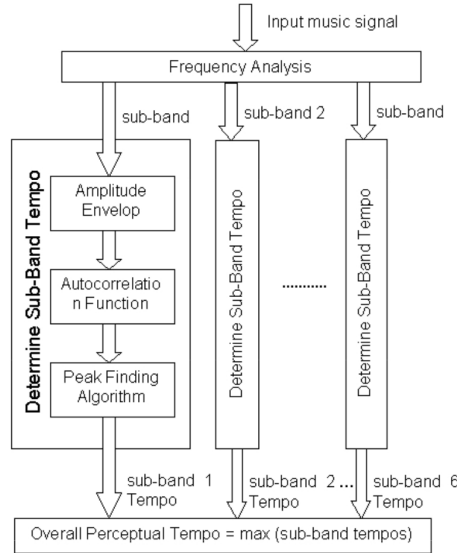


Figura 2: Esquema

1. **Frequency Analysis:** Se divide la señal a analizar en seis bandas diferentes (0-200Hz, 200-400Hz, 400-800Hz, 800-1600Hz, 1600-3200Hz y más de 3200Hz)
2. **Autocorrelation function:** La periodicidad de las señales de cada sub-banda se detecta utilizando la función de autocorrelación (ACF).
3. **Peak Finding Algorithm:** Se aplica el algoritmo para encontrar el *tempo* de cada sub-banda. El algoritmo estima el *tempo* mediante la búsqueda del mayor pico con tiempo de retardo mayor a cero. Sin embargo, este pico por lo general no suele reflejar el *tempo perceptual*. El algoritmo de búsqueda de pico

² A. T. Cemgil, B. Kappen, P. Desain, and H. Honing, "On tempo tracking: tempogram representation and Kalman filtering," J. New Music Res., vol. 28, pp. 259-273, 2001.

propuesto debe determinar el *tempo* en cada sub-banda mediante la búsqueda del pico con el retardo de tiempo más corto mayor a cero, por encima del 20 % de la energía total. El umbral de 20 % se determinó empíricamente. tiempo perceptual de la señal musical está determinada por la búsqueda del mayor *tempo* entre los *tempos* de cada sub-banda.

IPTE: descripción del algoritmo PTE mejorado

1. Se divide a la pieza en segmentos de longitudes fijas (por ejemplo de 10 segundos)
2. Se aplica el algoritmo de PTE para cada segmento para obtener el tiempo de ese segmento.
3. Se encuentra el tempo ocurrido con mayor frecuencia entre los tempos del segmento, según la determinación de la similitud en peso en cada segmento.
4. Determinar la probabilidad del peso para cada segmento de tempo para que sea más probable que el tempo del segmento anterior sea seleccionado como tempo perceptual global.
5. Determinar el tempo general perceptual de la pieza musical a partir de la información obtenida en los pasos 2, 3 y 4.

Beat Tracking based on Multiple-agent Architecture

La arquitectura de múltiples agentes se basa en que estos agentes llevan diferentes estrategias en paralelo e interactúan y compiten entre sí para seguir a los *beats*.

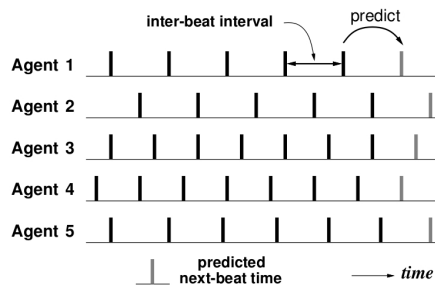


Figura 3: Agentes

Los agentes deben cumplir que:

1. El agente interactúa con otros agentes para llevar a cabo una determinada tarea
2. El agente evalúa su propio comportamiento
3. El agente se adapta a la entrada mediante el ajuste de su propio comportamiento

Cada agente mantiene una hipótesis de posición de beat, que consiste en una predicción del tiempo de *beat* siguiente.

Todos los agentes se agrupan en pares que tienen diferentes estrategias para el Beat Tracking. Cada agente en el par examina el mismo inter-beat interval usando los mismos resultados de análisis de frecuencia. Para predecir los tiempos de compás próximos cooperativamente, un agente interactúa con el otro agente en el mismo par a través de una predicción de

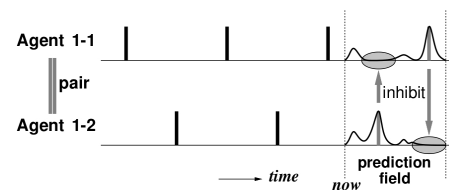


Figura 4: Par de agentes

campo. La predicción de campo es una curva de esperanza que representa cuando el ritmo próximo se espera que ocurra.

Descripción del sistema

Las dos etapas principales del proceso son el **Análisis de Frecuencia**, en la que varias señales utilizadas por los agentes se detectan y la **Predicción de Beat**, en el que múltiples hipótesis de las posiciones de beat son examinados por varios agentes.

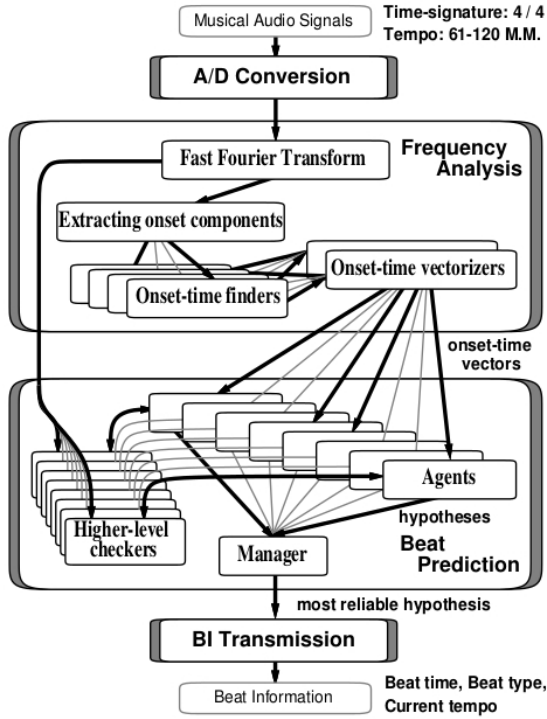


Figura 5: Diagrama de bloques

hipótesis y luego determina la salida final sobre la base de la más fiable.

En la etapa de Análisis de Frecuencia, el sistema usa múltiples buscadores de onsets, detectándolos en varios rangos de frecuencia diferentes. Estos resultados se transforman en una representación vectorial (llamado onset-time vectors).

En la etapa de predicción del *beat*, el sistema gestiona múltiples agentes que, de acuerdo con diferentes estrategias, hacen hipótesis en paralelo basados en estos vectores tiempo de *onsets*. Cada agente, primero calcula el *inter-beat interval* y predice el siguiente tiempo de *beat*, de donde luego se deduce el tipo de compás mediante la comunicación con un corrector de nivel superior, y evalúa la confiabilidad de su propia hipótesis. El gestor reúne todas las