

Seguimiento en tiempo real de interpretaciones musicales

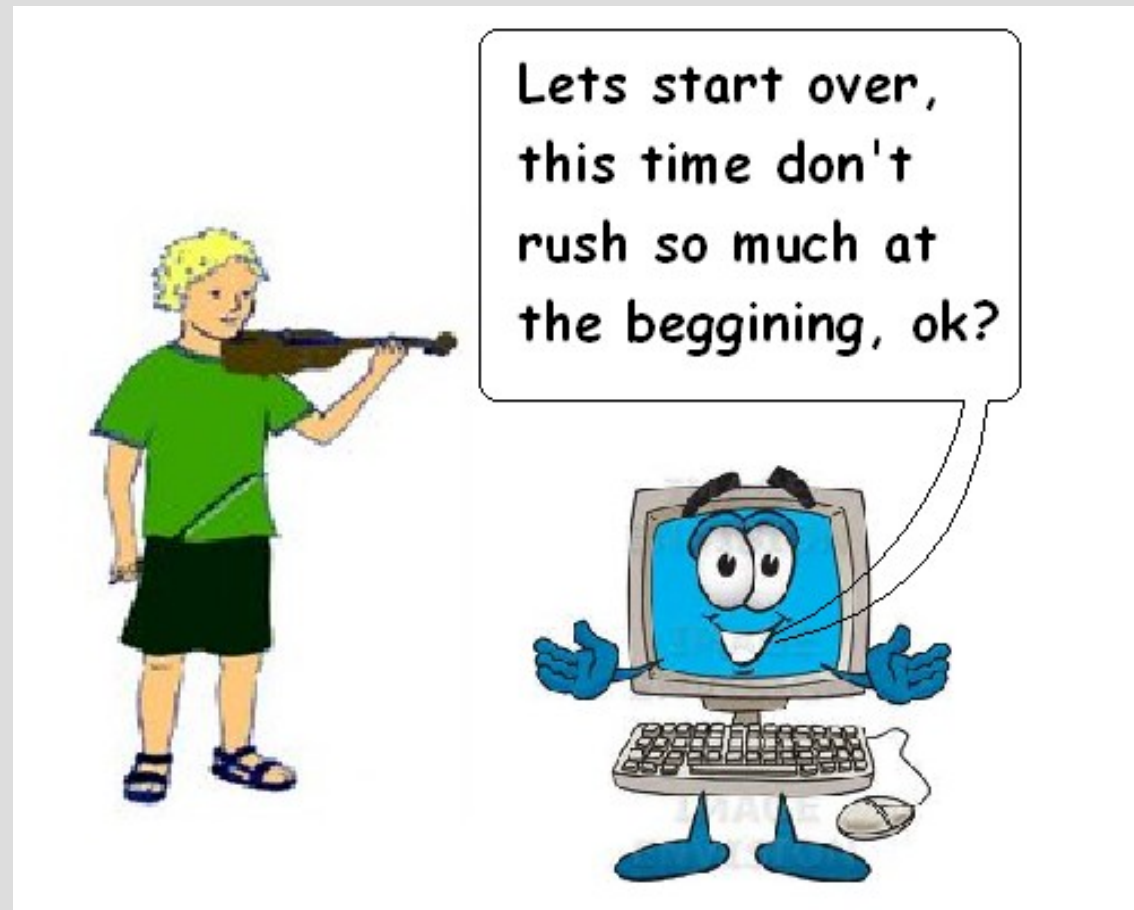
José Antonio Camarena Ibarrola
Facultad de Ingeniería Eléctrica
camarena@umich.mx

IA2010
UMSNH

Motivación

- Aplicaciones interesantes como:
 - El Maestro virtual de música
 - Acompañamiento automático de solistas
 - Efectos especiales automáticos en eventos en vivo

El maestro virtual de música



Acompañamiento automático de solistas (La orquesta de computadoras)



Efectos especiales automáticos

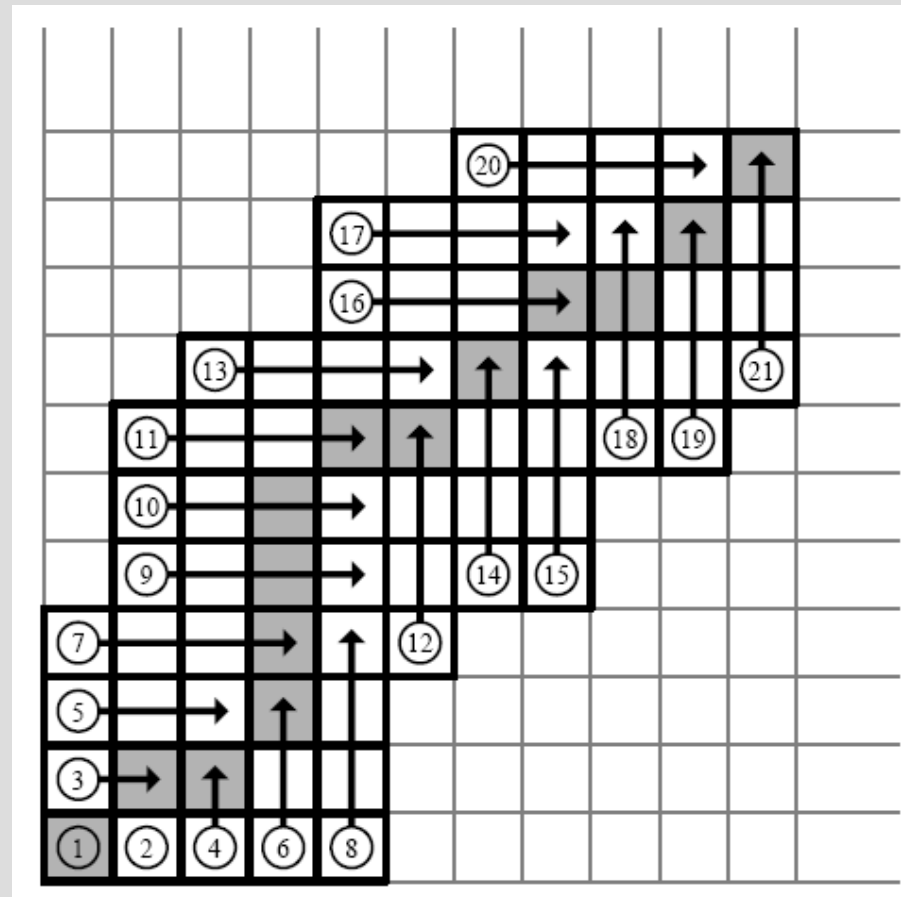


It's time for a little action !!!



Estado del Arte

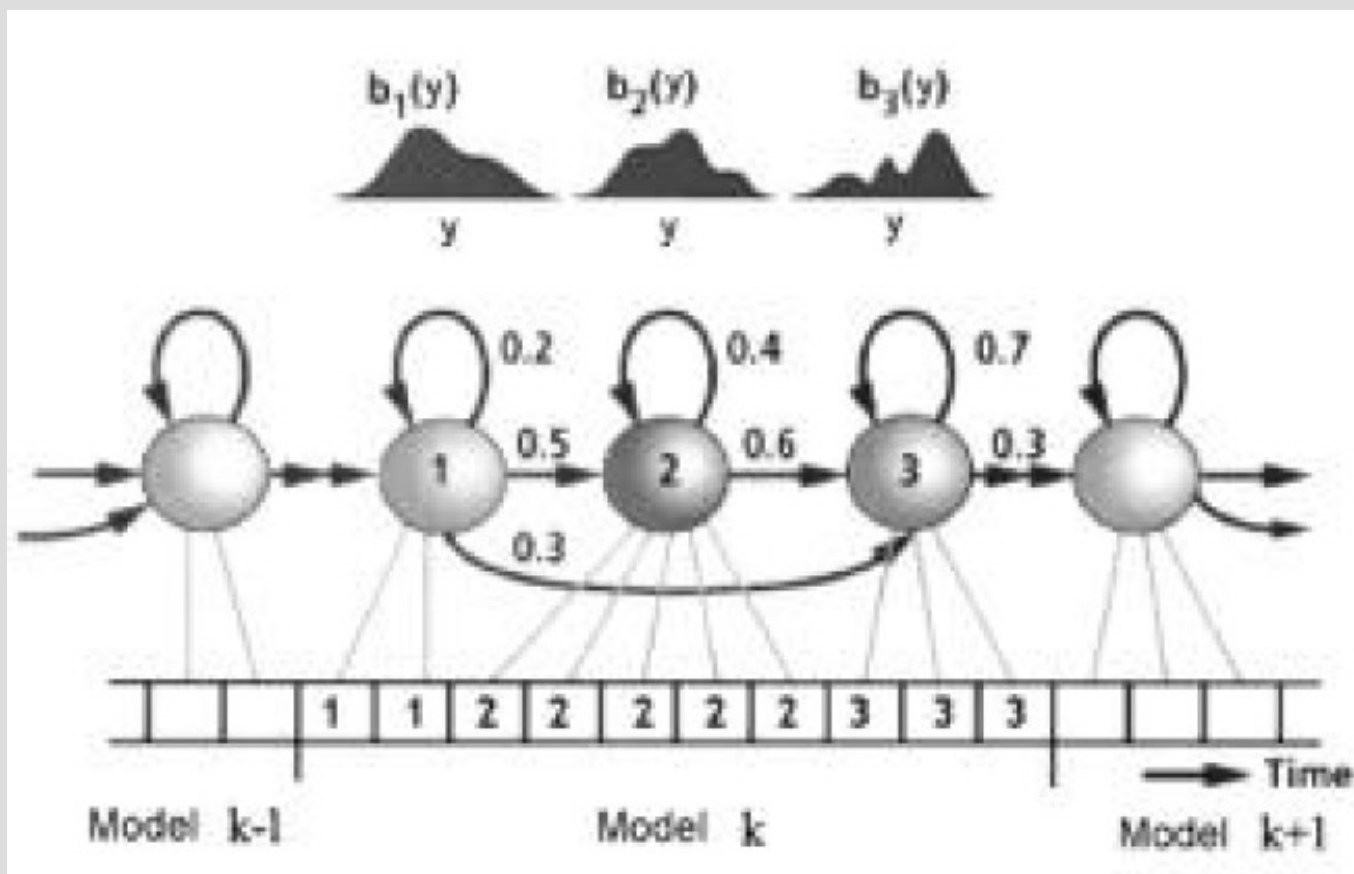
“On-line time warping” [1]



[1] Dixon, S. “Live tracking of musical performances using on-line time warping”. 8th International Conference on Digital Audio Effects (DAFx 2005), Austrian Research Institute for Artificial Intelligence, Vienna (September 2005)

Estado del arte (2)

Modelos Ocultos de Markov[2-3]



[2] Cano, P., Loscos, A., Bonada, J.: Score-performance matching using hmms. InICMC 1999, Audiovisual Institute, Pompeu Fabra University, Spain (1999)

[3] Orio, N., D'échelle, F.: Score following using spectral analysis and hidden Markov models. In: Proceedings of the ICMC, pp. 151–154 (2001)

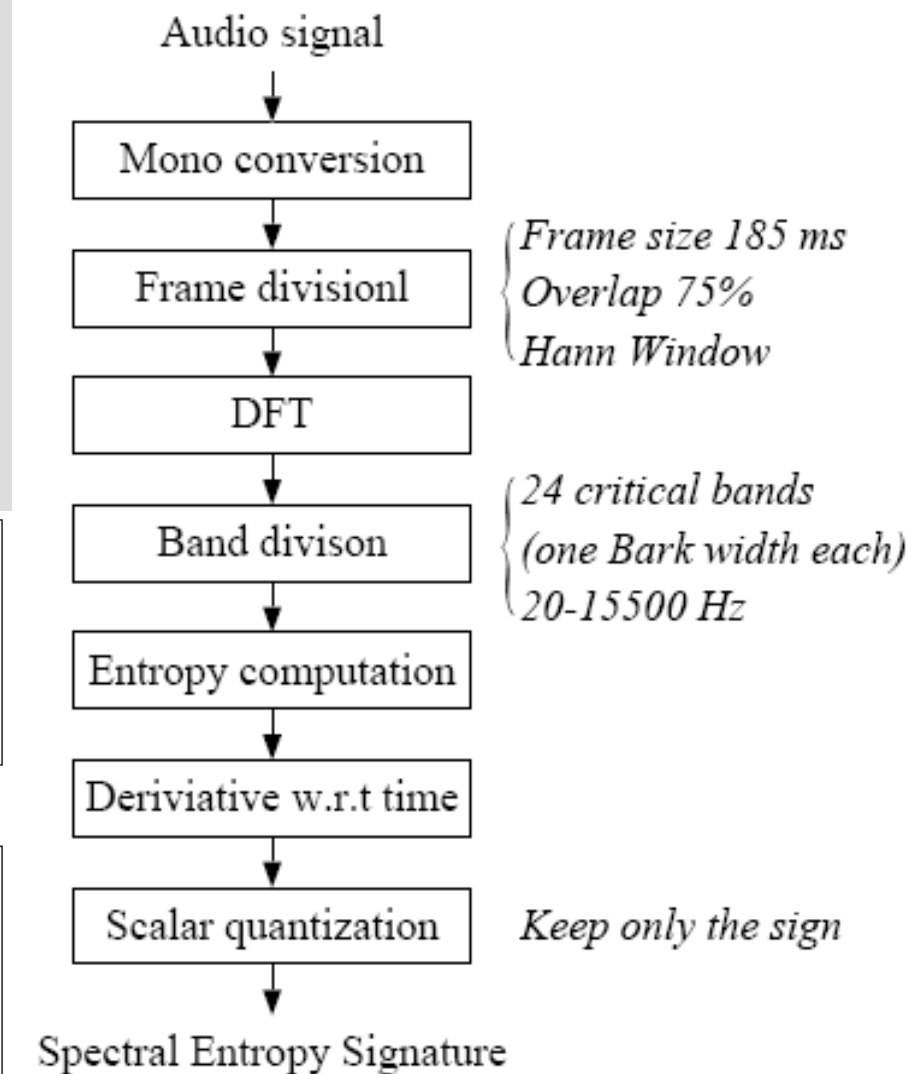
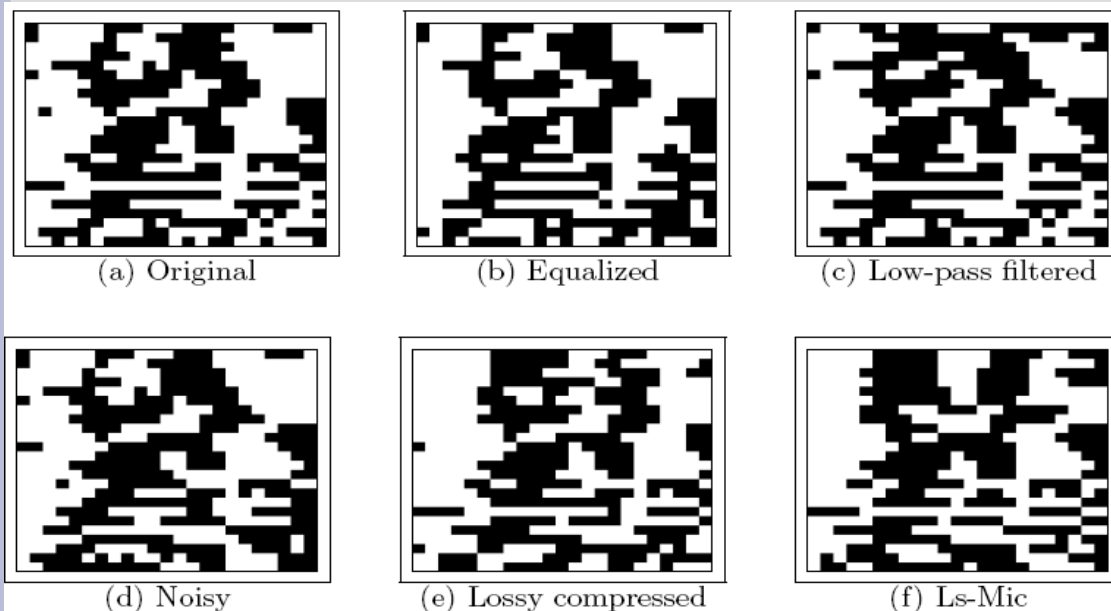
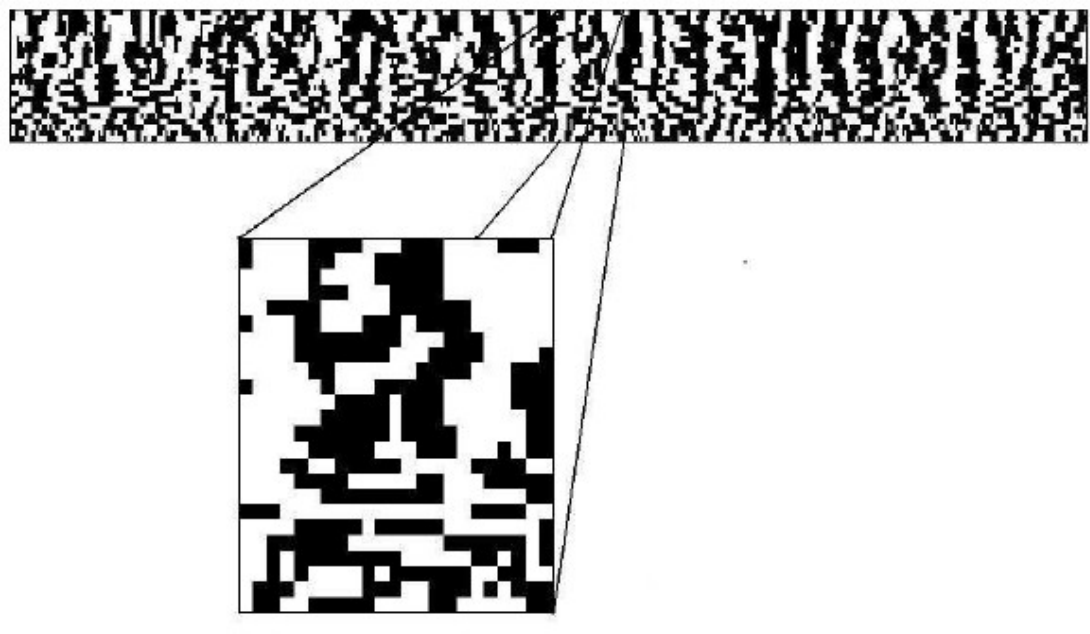
Desventajas de los enfoques del estado del arte

- Ambos enfoques (HMM y “on-line time warping”) tienen la propiedad de Markov
- Estos métodos asumen que el seguimiento hasta el punto actual ha sido correcto
- Solo utilizan información local en cada paso del proceso de seguimiento
- Como no se contempla la posibilidad de cometer un error, la recuperación no está contemplada
- Los errores son acumulativos y estos sistemas pueden perder la pista por completo
- Problemas de diseño tanto en HMM como en “on-line time warping”

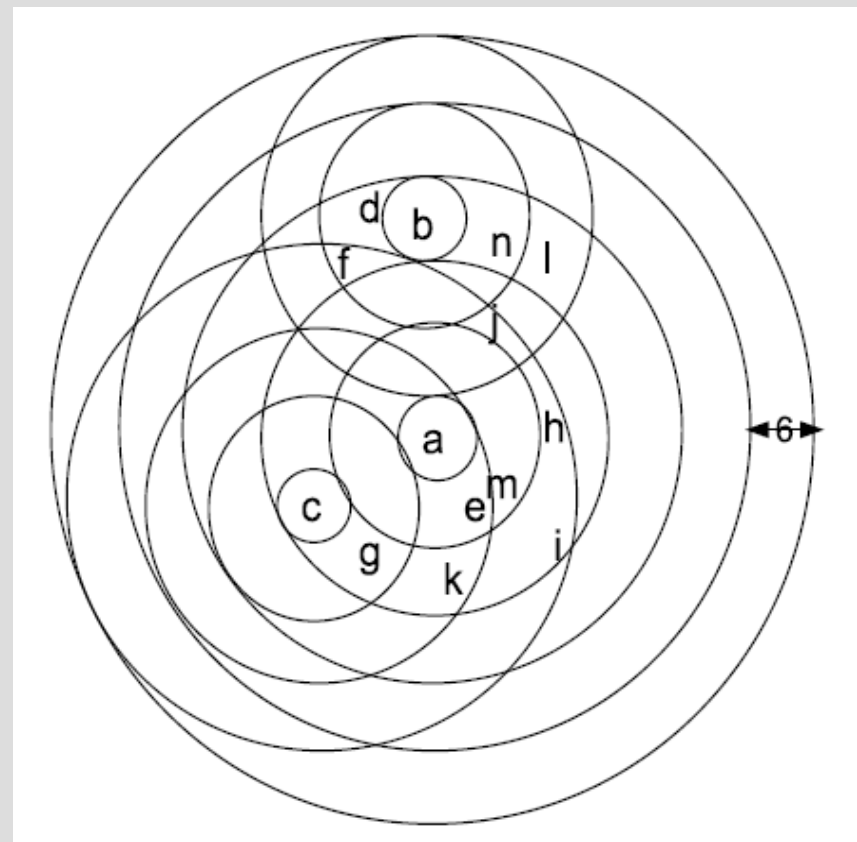
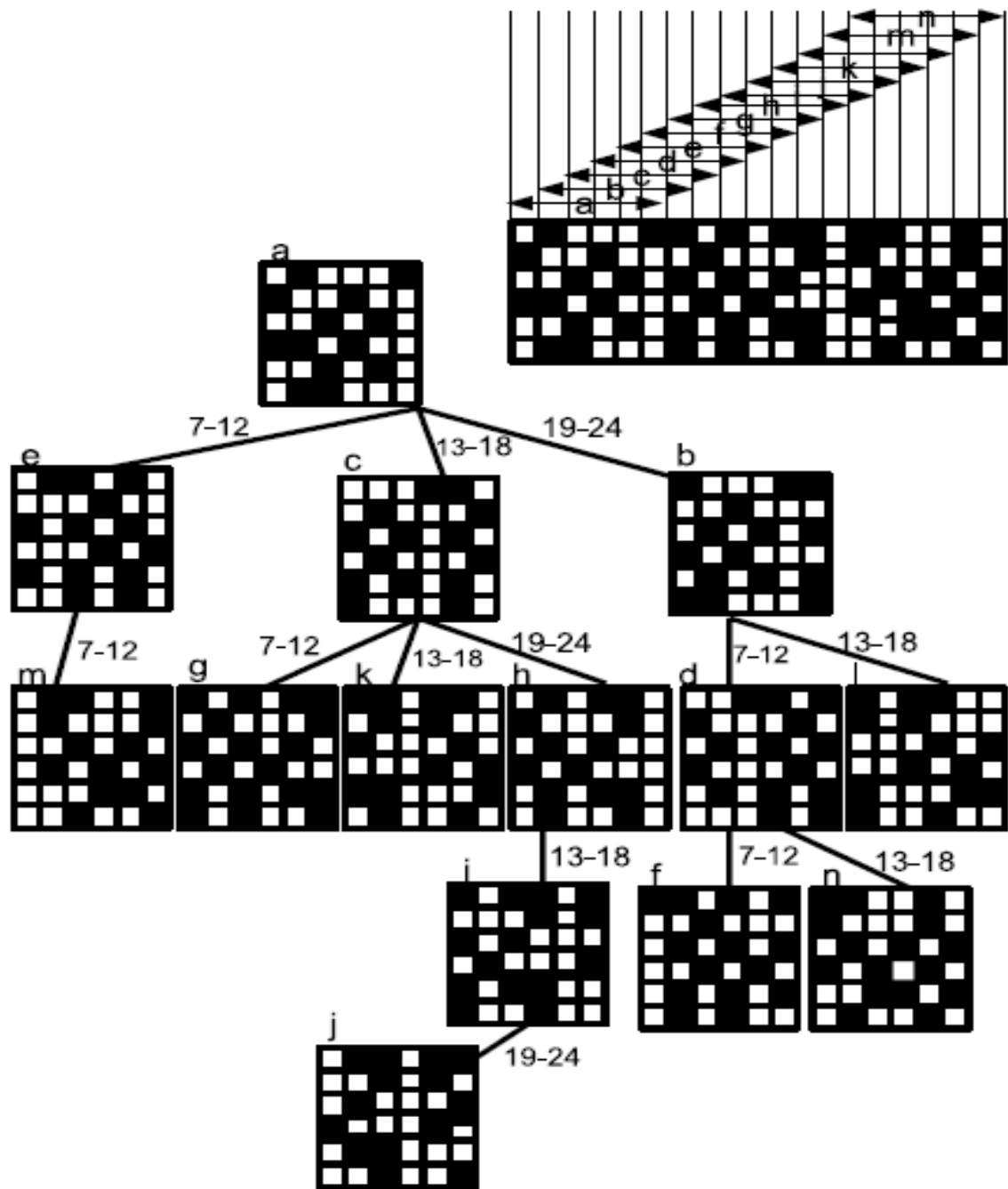
La propuesta

- Trasladar el problema al de búsquedas en espacios métricos
- Considerar la interpretación musical como una secuencia de segmentos de audio de un segundo.
- Buscar cada segmento dentro del audio objetivo a medida que llega
- La búsqueda se realiza entre todos los posibles segmentos del audio objetivo
- Usar un índice de proximidad para realizar la búsqueda rápidamente (Ej BK-tree)
- Buscar los K vecinos mas cercanos y elegir entre ellos basándose en alguna heurística
- Este enfoque permite la recuperación en caso de cometer errores y estos no se acumulan
- El seguimiento se puede realizar aún si la interpretación musical comenzó antes (no requiere sincronización)

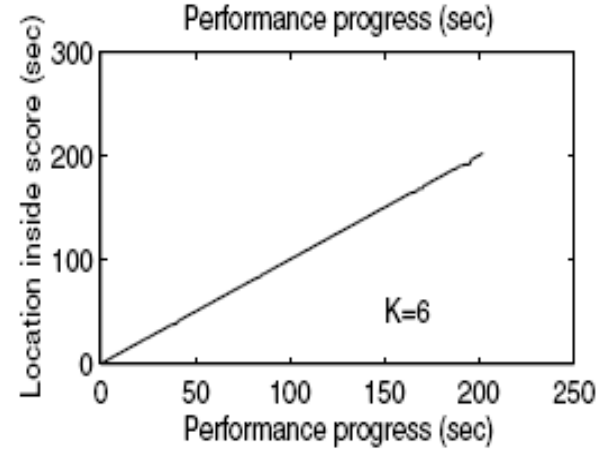
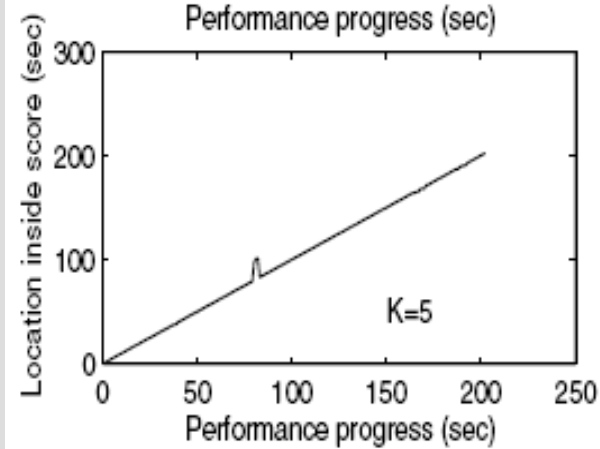
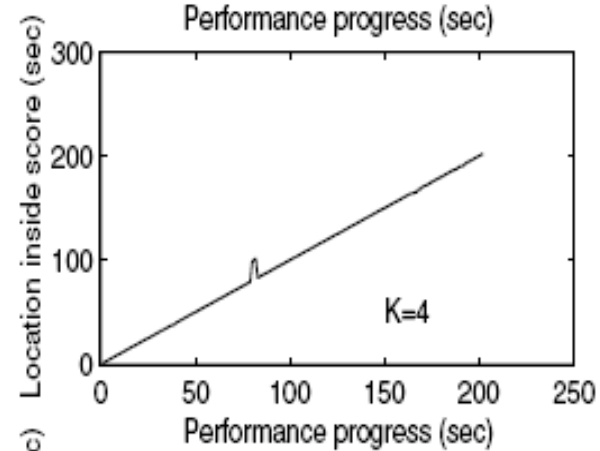
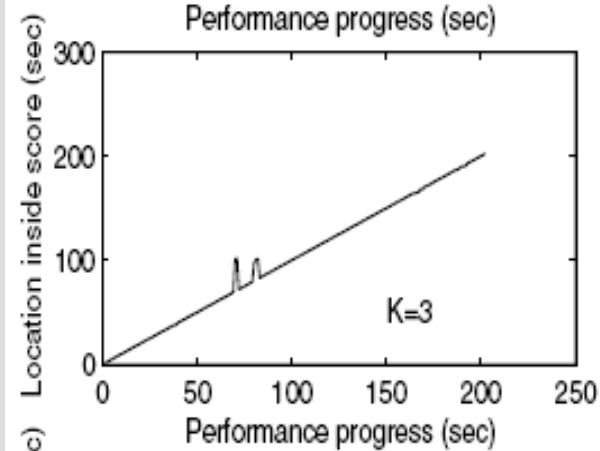
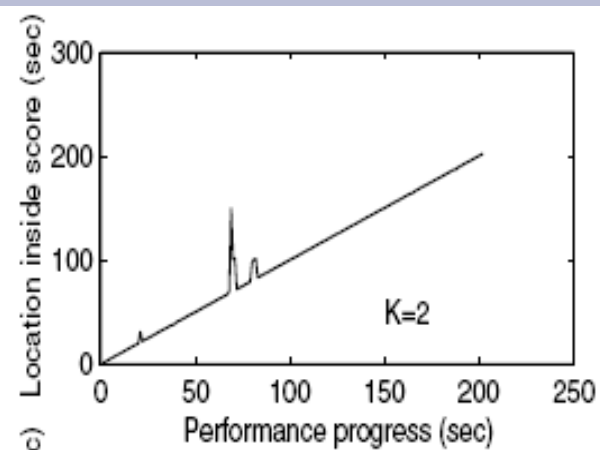
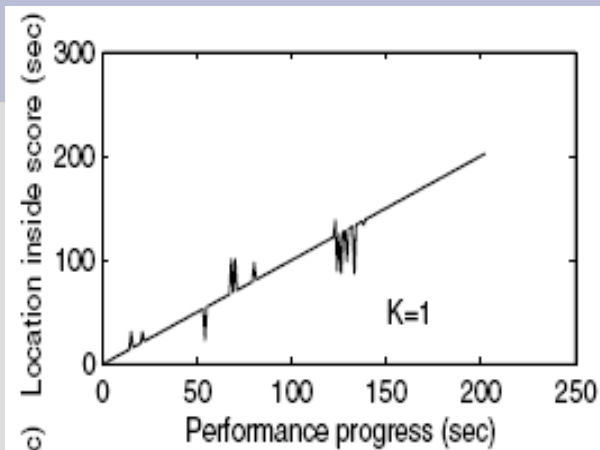
Extracción de la firma de audio basada en entropía espectral multibanda (MBSES)



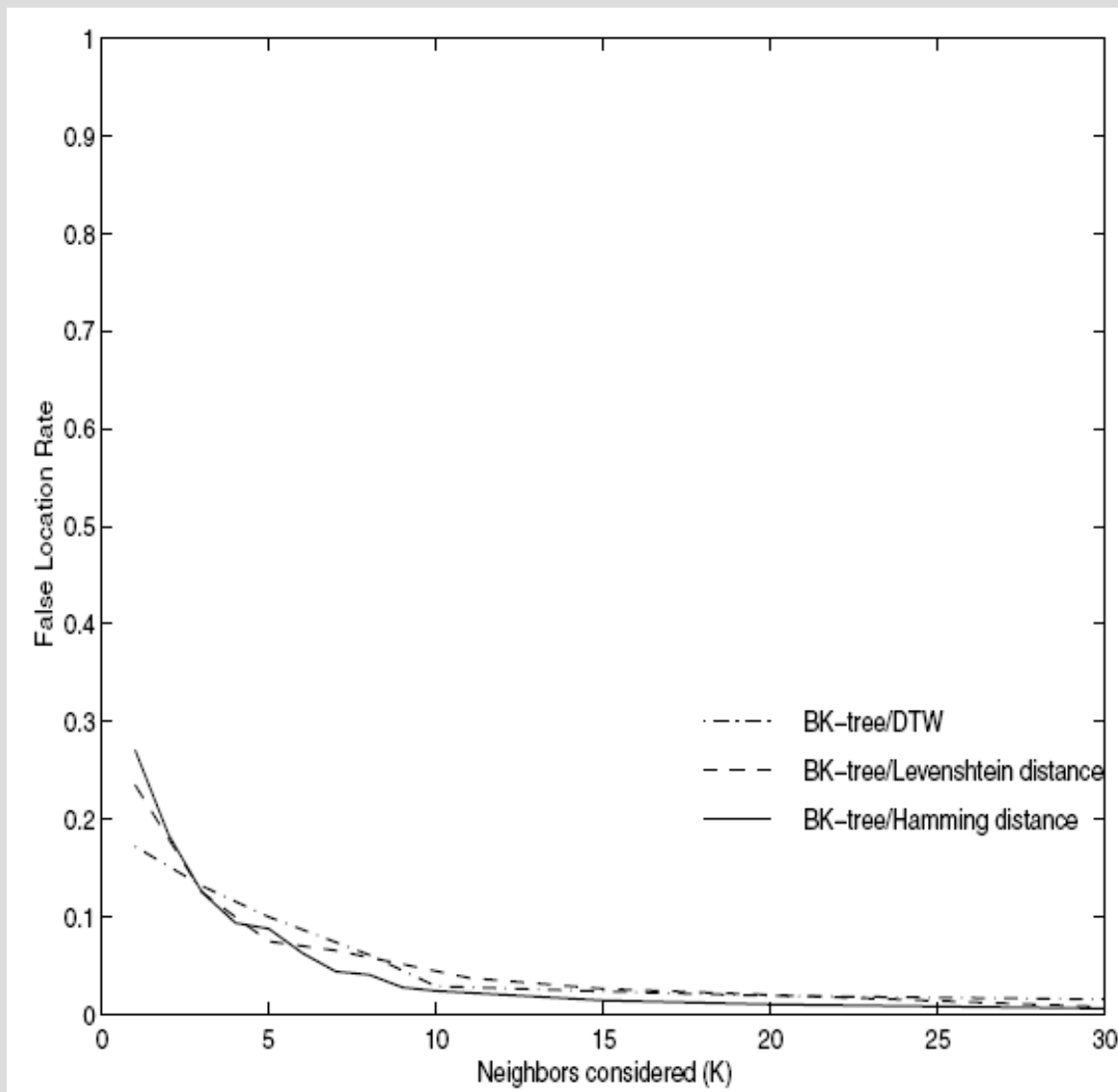
BK-tree



Seguimiento de una interpretación musical



Errores de localizacion vs K



Evitando errores de localización

- Los errores de localización no deben conducir a la ejecución de acciones.
- Los errores aparecen debido al hecho de que ninguno de los K vecinos más cercanos ocurren cerca de la posición actual de seguimiento.
- Cuando ninguno de los K vecinos más cercanos ocurren cerca de la posición actual de seguimiento, esta no se deberá mover de la misma manera que si el último segmento de audio no se hubiera recibido.
- Con esta modificación, los picos no aparecen ni siquiera para $K=1$.

Análisis de tiempo

- Se requieren dos pasos para encontrar la ubicación dentro del audio objetivo de un segmento de audio de un segundo perteneciente a la interpretación musical a la que se le está dando seguimiento
- 130 ms para determinar la firma de audio
- 10 ms adicionales para buscar los K vecinos más cercanos utilizando el índice de proximidad tipo BK-tree
- Dual core pentium laptop con 1GB of RAM
- La construcción del índice del audio objetivo de una canción de 4 minutos aproximadamente implica la indexación de cerca de 5300 huellas de audio de un segundo, esto toma alrededor de 10 segs. Sin embargo, este es un proceso fuera-de-línea

Conclusiones

- Es posible hacer seguimiento de una interpretación musical mediante búsqueda de segmentos de audio en lugar de la técnicas tradicionales de alineamiento.
- A diferencia del uso de HMMs no hay necesidad de entrenamiento (estimación de parámetros)
- A diferencia de la técnica “on-line time warping”, se aprovecha el hecho de que el audio objetivo se conoce a priori
- No Markoviano
- Errores no acumulativos (la recuperación es automática)
- El seguimiento puede iniciar en cualquier momento y no solo al inicio de la interpretación musical

Trabajo futuro

- Probar otros índices de proximidad
- Probar POMDPs (Procesos de decisión de Markov parcialmente observables)
- Aplicar al etiquetado automático de películas para débiles auditivos

Gracias!

- camarena@umich.mx

Etiquetado automático de películas para débiles auditivos

