

# Procesamiento, Aprendizaje y Percepción de Música por Computador

Prof. José Manuel Iñesta

Computer Music Laboratory

Grupo de Reconocimiento de Formas e Inteligencia Artificial

Dept. Lenguajes y Sistemas Informáticos

Universidad de Alicante

Esta presentación muestra el trabajo de muchos estudiantes de doctorado y colaboradores:

Pedro J. Ponce de León, David Rizo, Antonio Pertusa, Carlos Pérez-Sancho, David Espí, Plácido Román Illescas, Anssi Klapuri, Pedro Cruz-Alcázar, Thomas Lidý



Universidad de Alicante  
Universitat d'Alacant



# Procesamiento, Aprendizaje y Percepción de Música por Computador

## Contenidos

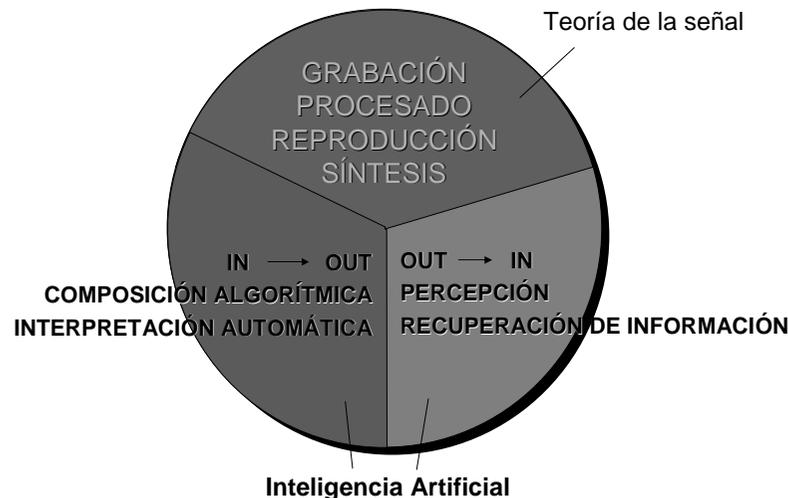
- Introducción
  - P.A.&P. en partituras digitales: señales musicales simbólicas
  - P.A.&P. en audio digital: señales musicales sonoras
  - P.A.&P. mediante el uso de metadatos
  - Comentarios finales

Prof. José Manuel Iñesta  
Grupo de Reconocimiento de Formas e Inteligencia Artificial  
<http://grfia.dlsi.ua.es>  
Depto. Lenguajes y Sistemas Informáticos  
Universidad de Alicante

## Música por computador

¿Qué es?

Amplio dominio con áreas de trabajo muy diferentes:



## Recuperación de información musical

¿Qué es? (en sólo una frase)

La búsqueda de conceptos de alto nivel en señales musicales.

¿Para qué?

Tiene muchas de aplicaciones en diferentes campos:

- Organización y búsqueda en bases de datos multimedia
- Bibliotecas digitales
- Sistemas de procesamiento basados en contenidos
- Gestión de derechos de autor
- Estudios musicológicos
- Herramientas educativas
- Etc., etc., etc.

## Gestión de derechos de autor

1978: Ronald Selle demanda a Maurice Gibb et al.

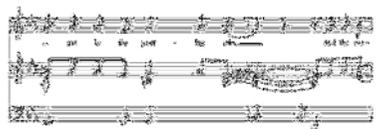
Maurice Gibb  
"How Deep is Your Love" 



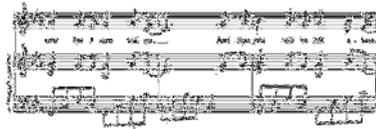
Ronald Selle 

En: "Saturday Night Fever" (1977)

How Deep is Your Love



"Let it End" (1976)



## Gestión de derechos de autor

El juez absolvió a Maurice Gibb. Admitió el parecido pero no quedó demostrado que Gibb hubiera tenido acceso al tema de Selle.

Asunto complejo en el que dar medidas objetivas sólo sirve para emitir informes periciales.

Nadie le ha echado en cara nada a Nino Rota por el tema de El Padrino, ni siquiera la familia de Franz Schubert

El Padrino (Nino Rota, 1972)



Danza Húngara #5 (Franz Schubert, ~1820)



## Recuperación de información musical

¿Qué es? (en sólo una frase)

La búsqueda de conceptos de alto nivel en señales musicales.

¿Para qué?

Tiene muchas de aplicaciones en diferentes campos:

- Organización y búsqueda en bases de datos multimedia
- Bibliotecas digitales
- Sistemas de procesamiento basados en contenidos
- Gestión de derechos de autor
- Estudios musicológicos
- Herramientas educativas
- Etc., etc., etc.

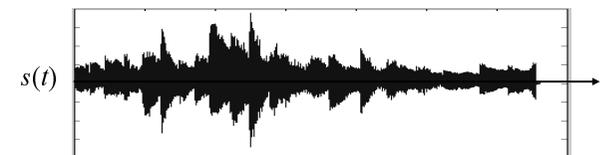
¿Con qué datos trabaja?

- Audio digital
- Secuencias simbólicas (partituras digitales)
- Meta-datos

## Categorías de datos

### • AUDIO DIGITAL

Las fuentes son señales de audio que supuestamente contienen música (formatos de ficheros WAV, AIFF, MP3, OGG, etc.)



¿Qué información usar?

(MPEG-7 es un esfuerzo para la estandarización de descripciones)

- Descriptores temporales
- Descriptores espectrales
- Descriptores psicoacústicos:
  - ✓ altura, sonoridad, timbre, duración, textura, etc.

# Categorías de datos

## • SÍMBOLOS MUSICALES

Las fuentes contienen representaciones simbólicas y estructuradas de música.

Pueden ser partituras digitales (secuencias) en formatos privados (editores de partituras y secuenciadores) o en formatos públicos como MIDI o MusicXML, entre otros.



### ¿Qué información usar?

Información melódica, armónica y rítmica.  
También (menos) sonoridad y timbre.

# Categorías de datos

## • METADATOS

Información que acompaña a la música y que no es la propia música: datos técnicos, administrativos, descriptivos, estructurales, etc.

```
yesterday.MID
Path: /home/bbdd/Repository/Popular/Pop/Beatles/yesterday.MID
Tech: MIDI Type 1; 5 Tracks; 16819 Bytes
Source: © 1964 The Beatles; Sequenced by J. Smith, 1999
Global: 90 BPM; 4/4; F major;
Trk1: Lead vocals ;1; 69; mono; 0.000; 1.000; 1; 1; 0; 1;
Trk2: Acoustic guitar ;3; 42; poly; 0.002; 0.025; 2; 1; 0; 0;
Trk3: Strings 1 ;5; 58; mono; 0.010; 0.825; 2; 18; 0; 0;
Trk4: Strings 2 ;7; 176; poly; 0.500; 0.500; 2; 4; 0; 0;
Trk5: Bass ;9; 74; mono; 0.000; 0.000; 2; 21; 33; 0;
```

### ¿Qué información usar?

- administrativos (autor, título, datos de catalogación,...),
- descriptivos (género, instrumentación, tonalidad, tempo,...),
- técnicos (tipo de fichero, URL, formato,...),
- de uso (derechos de autor, licencias,...),
- etc.

# Procesamiento, Aprendizaje y Percepción de Música por Computador

## Contenidos

- ✓ **Introducción**
- **P.A.&P. en partituras digitales: señales musicales simbólicas**
- **P.A.&P. en audio digital: señales musicales sonoras**
- **P.A.&P. mediante el uso de meta-datos**
- **Comentarios finales**

# Aplicaciones en música simbólica

## Catálogo de problemas abiertos

- Segmentación (motivos, frases y partes estructurales)
- Reconocimiento de:
  - Melodías - versiones
  - Tonalidades
  - Canciones (recuperación por petición)
  - Géneros (estilos musicales)
  - Modos (alegre, triste, melancólica, agresiva, etc.)
  - Autores
  - etc.
- Detección y seguimiento de métrica y tempo
- Y muchos más...

# Representación y descripción simbólica

## Representación

### Formatos de almacenamiento

#### Standard MIDI Files:

- ↑ Enorme volumen de datos existentes
- ↓ Severas limitaciones para la representación de la música (para control)

#### MusicXML:

- ↑ Más orientado a la representación
- ↓ Archivos muy grandes y no está tan extendido (es más "técnico")

#### OTROS:

- Formatos propietarios: Finale, Encore, Sibelius, Band in a Box, ...
- Formatos abiertos: ABC, Humdrum, Essen, Lilypond, etc.

- Lista de bases de datos simbólicas:
  - [php.indiana.edu/~donbyrd/MusicTestCollections.HTML](http://php.indiana.edu/~donbyrd/MusicTestCollections.HTML)

# Representación y descripción simbólica

## Representación

### Estructuras de datos

#### CADENAS

- Palabras musicales
- Cadenas de Markov (n-gramas)
- ↑ Algoritmos muy rápidos para construcción y análisis
- ↑ Herramientas disponibles del procesamiento del lenguaje natural
- ↓ Resultados muy sensibles a los códigos seleccionados
- ↓ Contextos (n) de tamaño muy limitado

#### ÁRBOLES

- Árboles melódicos
- Árboles polifónicos (con y sin *multisets*)
- ↑ La duration (ritmo) queda implícitamente codificado en la estructura ⇒ resultados menos sensibles a la codificación
- ↓ Su construcción y análisis son más lentos

# Representación y descripción simbólica

## Representación

### CADENAS

#### Representaciones acopladas

$$\Sigma = \sum_p \times \sum_d$$

#### Alturas con duraciones:

(F<sub>3</sub>,n) (C<sub>4</sub>,c) (D<sub>4</sub>,c) (D<sub>4</sub>,n) (G<sub>3</sub>,n) (S,n) (G<sub>3</sub>,c) (A<sub>4</sub>,c) (B<sub>4</sub>,b)

#### Intervalos con duraciones relativas:

(-,) (+7,1/2) (+2,1) (0,2) (-7,1) (S,1) (0,1/2) (+2,1) (+2,4)

#### Representaciones desacopladas

$$\Sigma = \sum_p \cup \sum_d$$

#### Alturas y duraciones:

F<sub>3</sub> n C<sub>4</sub> c D<sub>4</sub> c D<sub>4</sub> n G<sub>3</sub> n S n G<sub>3</sub> c A<sub>4</sub> c B<sub>4</sub> b

#### Intervalos y duraciones relativas:

- - +7 1/2 +2 1 0 2 -7 1 S 1 0 1/2 +2 1 +2 4



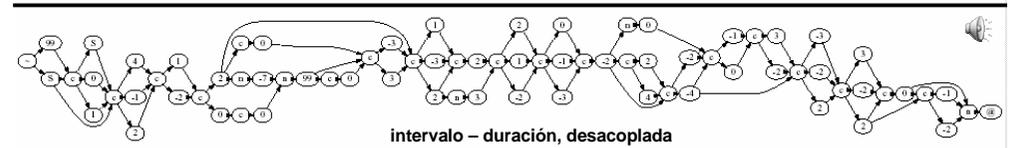
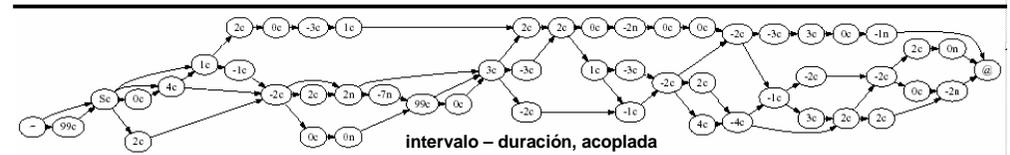
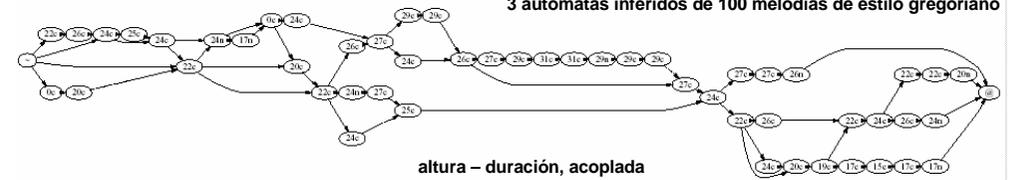
# Representación y descripción simbólica

## Representación

### CADENAS

- Automatas estocásticos finitos deterministas inferidos de cadenas

3 automatas inferidos de 100 melodias de estilo gregoriano



# Aplicaciones en música simbólica

## Métricas de similitud melódica

Se trata de calcular un valor que evalúe el parecido o las diferencias entre dos melodías.

### DISTANCIAS DE EDICIÓN:

- Es el coste mínimo de transformar una secuencia en la otra, dados un conjunto de operaciones posibles y una función que indica el coste de cada una de ellas.
- Puede aplicarse a secuencias de notas o de cualquier tipo de representación.

### DISTANCIAS GEOMÉTRICAS:

- Se calcula la suma de las diferencias locales entre propiedades de ambas secuencias:
  - Alturas, intervalos, duraciones, etc.

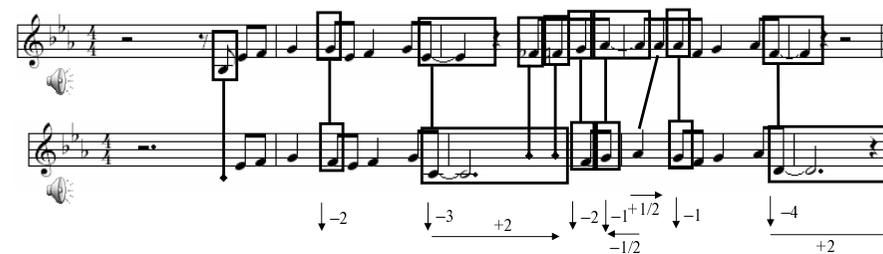
# Aplicaciones en música simbólica

## Métricas de similitud melódica

### Distancia de edición melódica

Mongeau & Sankoff (1990) propusieron una serie de operaciones:

- **Inserción:** de un elemento
- **Borrado:** de un elemento
- **Reemplazo:** de un elemento por uno de la otra secuencia
- **Fragmentación:** Reemplazo de un elemento por una serie de elementos adyacentes de la otra secuencia
- **Consolidación:** Reemplazo de una serie de elementos adyacentes por un único elemento de la otra secuencia



(Mongeau & Sankoff. "Comparison of Musical Sequences". Computers and the Humanities 24, 161-175, 1990)

# Aplicaciones en música simbólica

## Métricas de similitud melódica

### Distancias geométricas

Suma de las diferencias locales entre propiedades de ambas secuencias.

Representación de pianola

Alturas absolutas



(mismas melodías de ejemplo)

# Representación y descripción simbólica

## Representación

### • n-palabras

Construcción de "palabras musicales" mediante la concatenación de magnitudes referidas a  $n$  notas.

- Una posibilidad es usar  $n - 1$  intervalos y duraciones acoplados en una palabra de  $2n - 2$  símbolos.

$n = 3$



$(+7, \underline{2}, +2, \underline{1}) (+2, \underline{1}, 0, \underline{1/2}) (0, \underline{1/2}, -7, \underline{1/2}) (-7, \underline{1/2}, 0, \underline{4}) (0, \underline{4}, +2, \underline{1}) (+2, \underline{1}, +2, \underline{1/4})$

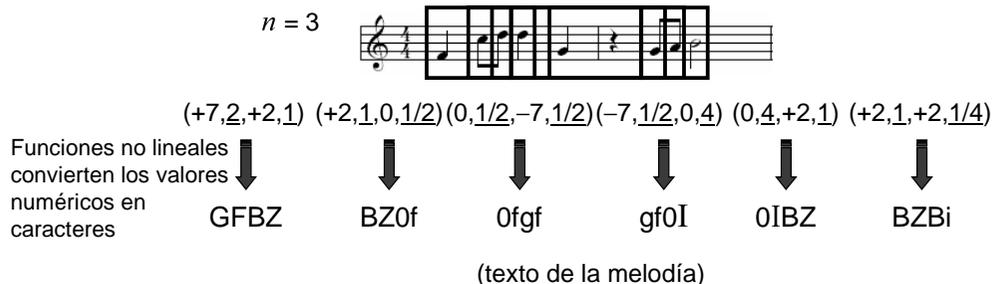
# Representación y descripción simbólica

## Representación

### n-palabras

Construcción de "palabras musicales" mediante la concatenación de magnitudes referidas a  $n$  notas.

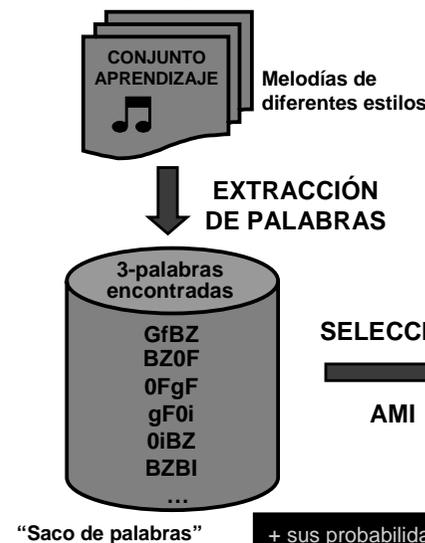
- Una posibilidad es usar  $n - 1$  intervalos y duraciones acoplados en una palabra de  $2n - 2$  símbolos.



# Aplicaciones en música simbólica

## Identificación de géneros musicales

### Aprendizaje



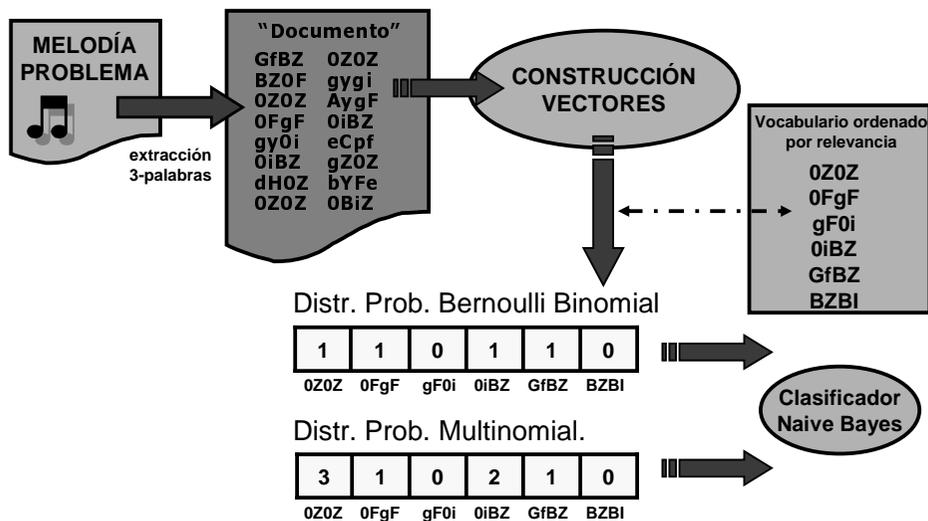
AMI: Averaged Mutual Information

3-palabras	frec <sub>A</sub>	frec <sub>B</sub>	AMI
AZBZ	187	265	0.061
bZbZ	52	2539	0.163
BZAZ	211	266	0.082
bFbF	2	3	0.008
0Z0Z	1978	2016	0.044
BFbf	13	0	0.025
...	...	...	...
	Estilo A	Estilo B	

# Aplicaciones en música simbólica

## Identificación de géneros musicales

### Reconocimiento



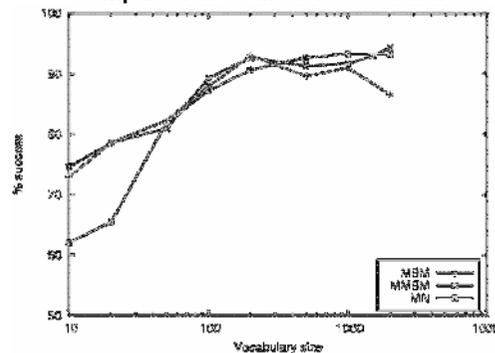
# Aplicaciones en música simbólica

## Identificación de géneros musicales

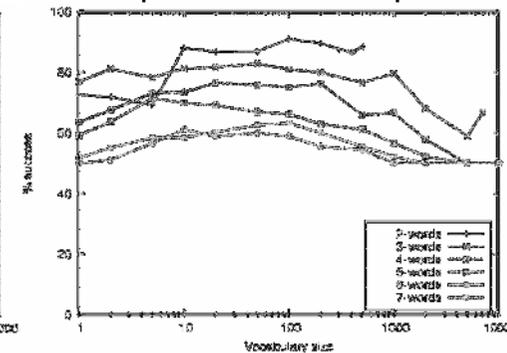
### Resultados del reconocimiento

- Tamaños de n-palabras,  $n \in [2, 7]$
- Tamaños de vocabularios:  $|V| \in [10, >10000]$
- 10% validación cruzada: resultados promediados para 10 sub-experimentos

Comparación de modelos estadísticos



Comparación de tamaños n-palabras



# Representación y descripción simbólica

## Representación

### • n-gramas (Cadenas de Markov)

- Cada parámetro es la probabilidad de que aparezca el símbolo  $s_n$  tras la aparición de los símbolos anteriores  $s_{i,n-1}$ .
- Los parámetros se estiman a partir de las frecuencias de aparición observadas en un conjunto de aprendizaje.

- Secuencia de longitud  $N$ :  $s_{i,N}$ 
  - ↳ Último símbolo :  $s_N$
  - ↳ Los  $N-1$  anteriores :  $s_{i,N-1}$

⇒ Probabilidad de observar  $s_N$  después de ver los  $s_{i,N-1}$  es:

$$P(s_N | s_{i,N-1}) = \frac{n(s_{i,N-1}, s_N)}{\sum_{s \in \Sigma} n(s_{i,N-1}, s)}$$

⇒ Probabilidad de observar una cadena de  $N$  símbolos es:

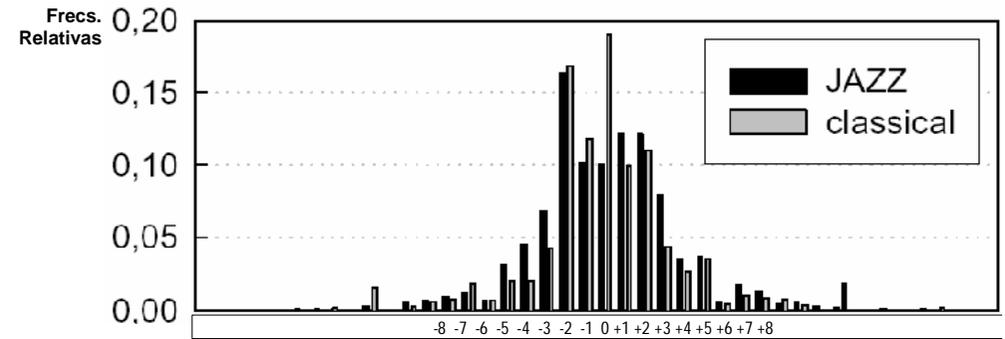
$$P(s_{i,N}) = \prod_{i=1}^N P(s_i | s_{i-N+1} \dots s_{i-1})$$

- Problema del suavizado si un símbolo no aparece nunca!

# Representación y descripción simbólica

## Representación

### • n-gramas (Cadenas de Markov)



Estas diferencias son  $n = 1$   
la base del clasificador

$$P_c(0) = 0.191 \quad P_c(+12) = 0.01769$$

$$P_f(0) = 0.103 \quad P_f(+12) = 0.00175$$

$$P(\text{unísono}) \quad P_c(+11) = 0.00064$$

$$P_f(+11) = 0.00182$$

+12 = +octava → Clásica!  
+11 = +séptima → Jazz!

# Representación y descripción simbólica

## Identificación de géneros musicales

### • n-gramas (Cadenas de Markov)

- Usar las secuencias de acordes en vez de las melodías



### ▪ Ejemplo:

| Am7 D7 | G | C#m7 F#7 | B |

El uso de acordes relativos a la tonalidad permite invarianza a la transposición

| IIm7 V7 | I || IIm7 V7 | I |

pero el precio es la necesidad de conocer la tonalidad

- Existe este metadato y es de fiar
- Detección automática

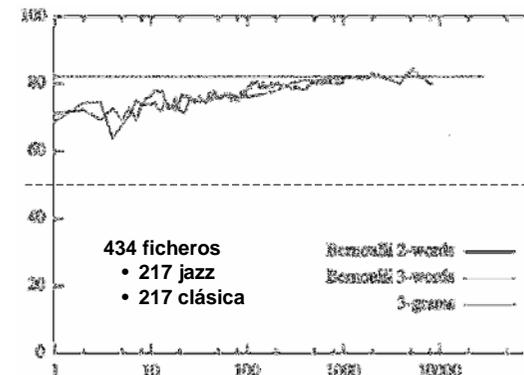
# Representación y descripción simbólica

## Identificación de géneros musicales

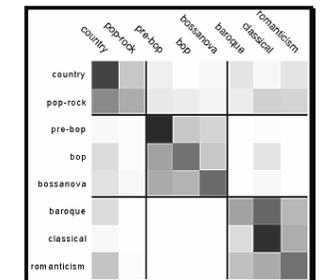
### • n-gramas (Cadenas de Markov)

- Comparación con Naïve Bayes y n-palabras

- Se construyen modelos de lenguaje para cada clase según el training set.
- Para una entrada problema se construyen todos los 3-gramas
- Se clasifica según la clase que maximiza la probabilidad de ese conjunto de 3-gramas



	Jazz	Clásica	Total
2-gramas	68,5	92,6	80,5
3-gramas	71,7	92,7	82,2
4-gramas	69,3	95,1	82,2



# Aplicaciones en música simbólica

## Identificación de géneros musicales

### • Mediante descriptores estadísticos

#### Objetivo:

Evaluar la capacidad de llevar a cabo esta tarea usando estadística descriptiva de notas, alturas, duraciones y silencios, evaluando características estadísticas de tipo melódico, armónico y rítmico.

#### Herramientas:

- Descriptores estadísticos de las secuencias de notas
- Selección de características
- Diferentes clasificadores:
  - Bayesiano
  - k-vecinos más cercanos (k-NN)
  - Self-organizing Maps (SOM)
  - Support vector machines (SVM)
  - Multi-layer Perceptrons (MLP)

# Aplicaciones en música simbólica

## Identificación de géneros musicales

### • Descriptores estadísticos

#### ▪ Contadores:

- Notas y silencios (cortos y largos –  $> 1/8 t$  (de fusa) –)

#### ▪ Melódicos:

##### ▪ Vertical:

- Alturas: rango, media relativa y desviación
- Intervalos: rango, media relativa y desviación

##### ▪ Horizontal:

- Duraciones de notas: rango, media relativa y desviación
- Tiempo entre inicios: rango, media relativa y desviación
- Duraciones de silencios: rango, media relativa y desviación

#### ▪ Armónicos:

- Notas no diatónicas: cuenta, grado media y desviación de grados

#### ▪ Rítmicos:

- Cuenta de notas sincopadas

#### ▪ Normalidad de las distribuciones de:

- Alturas, duraciones de notas y silencios, inter-inicios, intervalos y grados no diatónicos

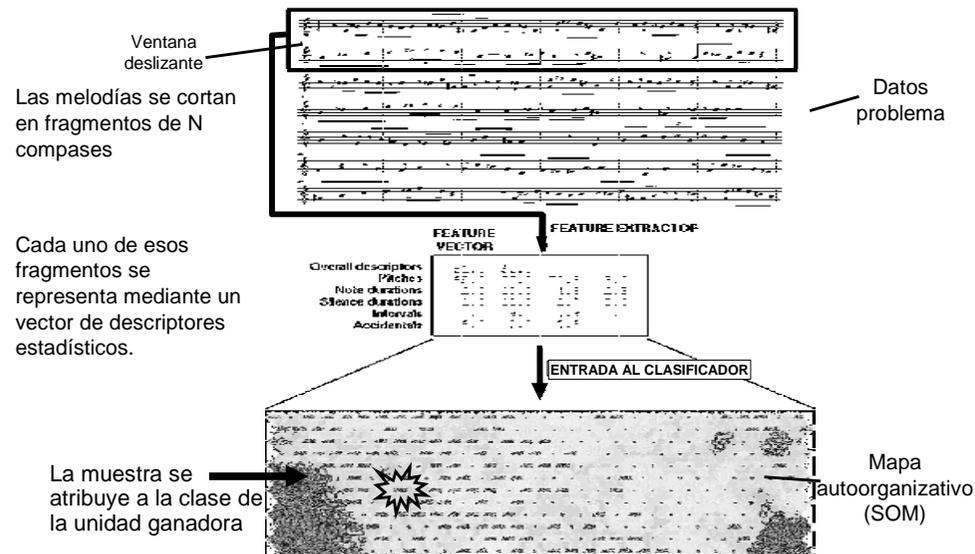
28  
descriptores

(Ponce de León & Iñesta, "Feature-driven recognition of music styles". LNCS 2652: 773-781, 2003)

# Aplicaciones en música simbólica

## Identificación de géneros musicales

### • Esquema general



# Aplicaciones en música simbólica

## Identificación de géneros musicales

### • Trabajamos con líneas melódicas

#### Problema:

Dado un fichero MIDI ¿Qué pista contiene la melodía?



#### Pero ¿Qué es la melodía?

- Lo que se puede cantar o tararear
- Secuencia de alturas con sentido musical coherente
- Lo que se te queda de una canción

# Aplicaciones en música simbólica

## Identificación de la pista de melodía

- **Descriptores estadísticos de pista**

La necesidad de evaluar las propiedades de cada pista en relación al fichero en el que se encuentra lleva a la necesidad de calcular descriptores relativos.

Category	Descriptors
Descripción de pista	Track information
	Relative duration
	Number of notes
	Occupation rate
Pitch	Polyphony rate
	Highest
	Lowest
	Mean
Descripción de contenido	Standard deviation
	Pitch intervals
	Number of different intv.
	Largest
	Smallest
Note durations	Mean
	Mode
	Standard deviation
	Longest
Syncopation	Shortest
	Mean
	Standard deviation
	Number of Syncopated notes

$$\frac{v_i - m}{M - m}$$

$M$  = máximo valor para las pistas del fichero

$m$  = mínimo valor para las pistas del fichero

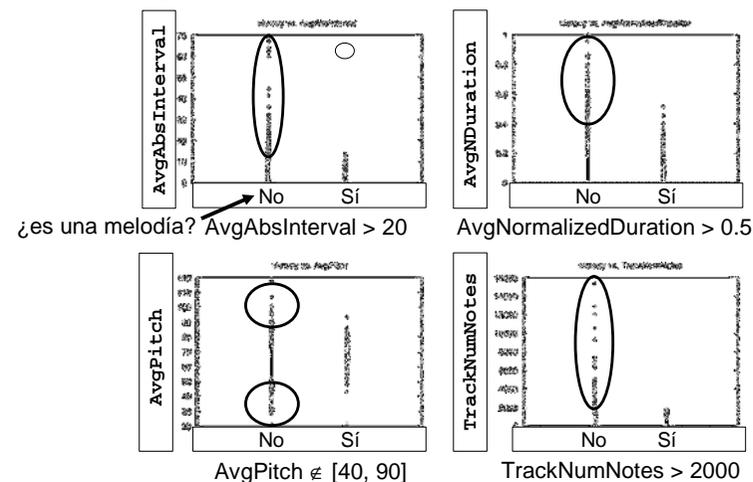
(D. Rizo et al. "A Pattern Recognition Approach for Melody Track Selection in MIDI Files". ISMIR 2006)

# Aplicaciones en música simbólica

## Identificación de la pista de melodía

- **Descriptores estadísticos de pista**

Algunos descriptores muestran su utilidad:

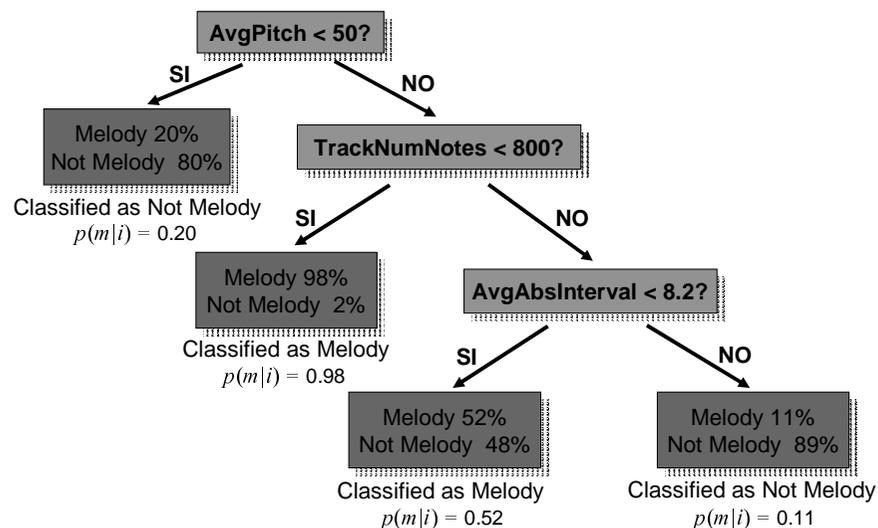


La combinación adecuada de estos indicios permitirá atribuir a la pista  $i$  una probabilidad  $p(m|i)$  de ser una melodía

# Aplicaciones en música simbólica

## Identificación de la pista de melodía

- **Clasificación mediante árboles de decisión**

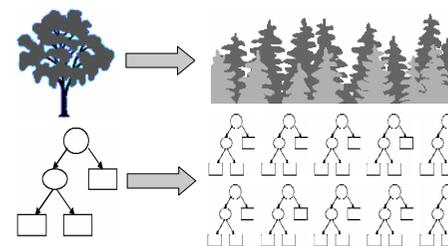


# Aplicaciones en música simbólica

## Identificación de la pista de melodía

- **Clasificación mediante *random forests***

- Colección de árboles de decisión
- Cada nodo usa una selección aleatoria de descriptores



• Para clasificar una pista la decisión de cada árbol es un voto  
Se elige la clase más votada

Capacidad para distinguir entre pistas melódicas y no melódicas

Estilo	% OK
Clásica	99.2
Jazz	96.0
Pop	94.8

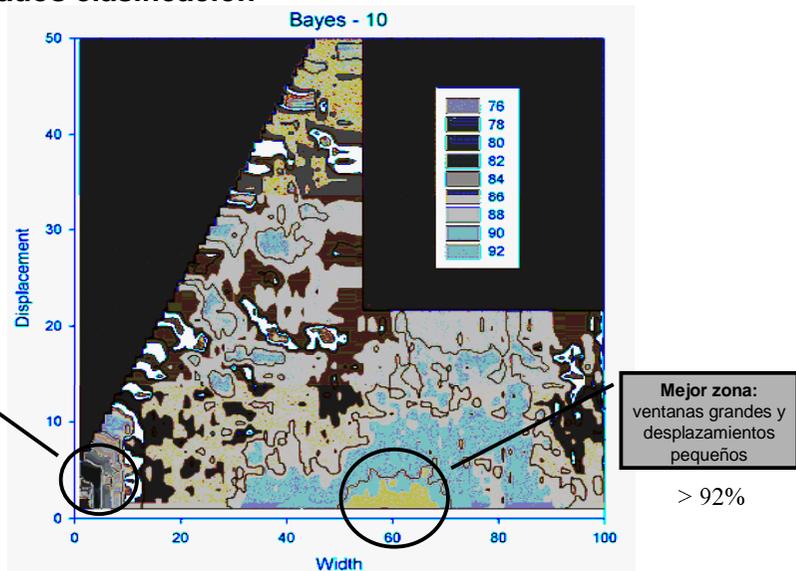
Estudio de generalización y dependencia con el estilo

Training	Test	% OK
Todos	Clas	76.3
Todos	Jazz	95.6
Todos	Pop	79.9

# Aplicaciones en música simbólica

## Identificación de géneros musicales

- Resultados clasificación



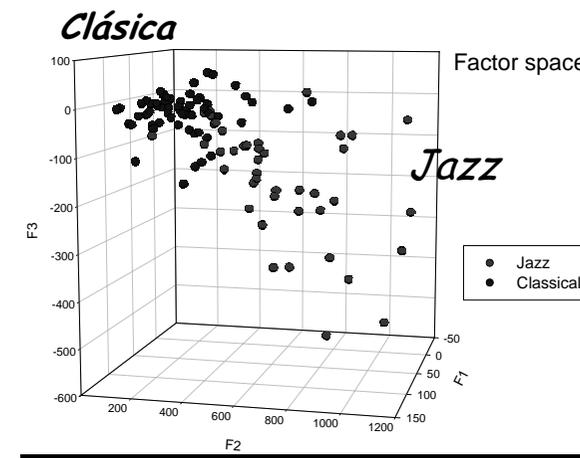
# Aplicaciones en música simbólica

## Identificación de géneros musicales

- Visualización del espacio de características

Algunas herramientas permiten descubrir cómo se organizan los puntos representativos de las melodías estudiadas en el espacio de características.

### Análisis de componentes principales



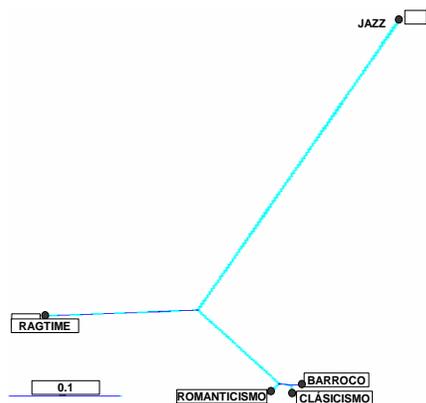
# Aplicaciones en música simbólica

## Identificación de géneros musicales

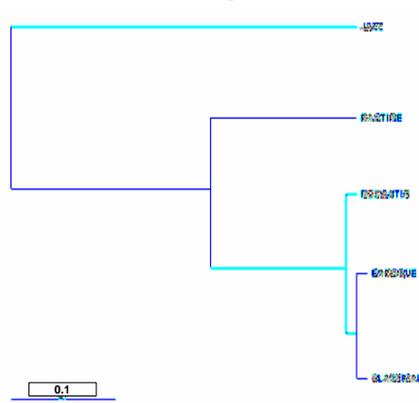
- Visualización del espacio de características

Algunas herramientas permiten descubrir cómo se organizan los puntos representativos de las melodías estudiadas en el espacio de características.

Radial tree



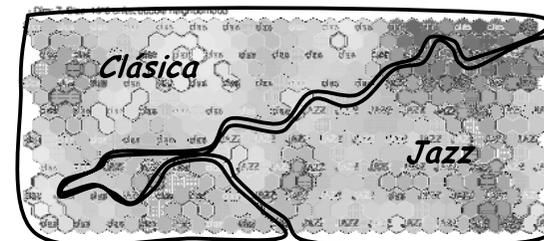
Dendrograma



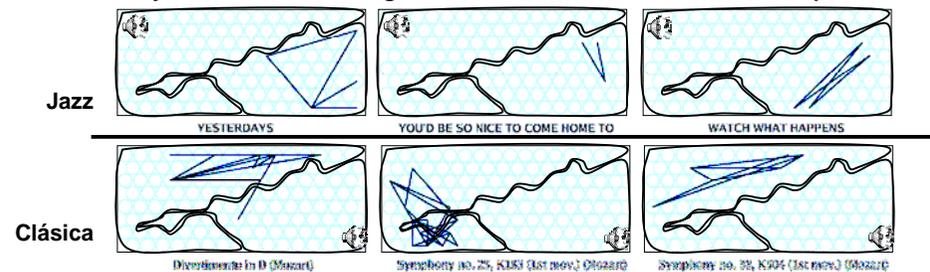
# Aplicaciones en música simbólica

## Identificación de géneros musicales

- Visualización con SOM



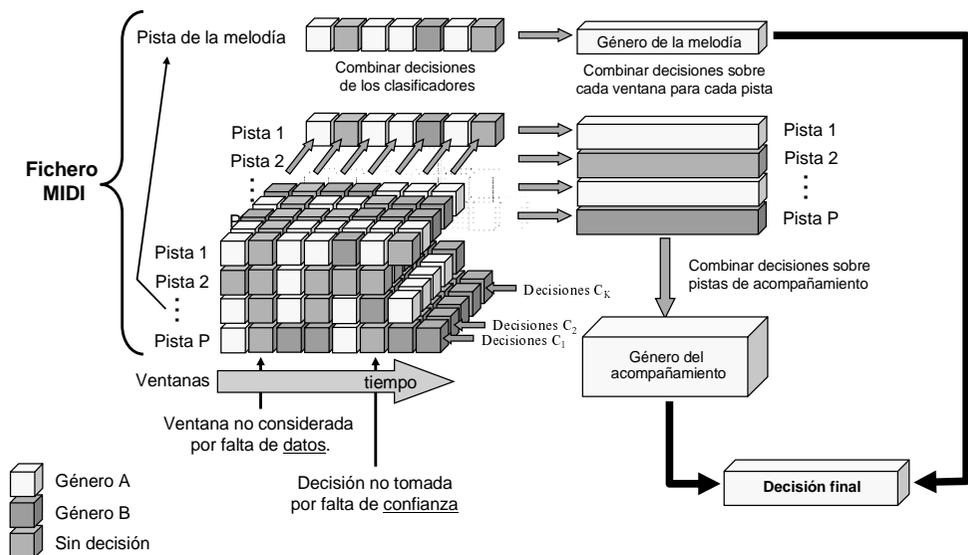
Trayectoria de la unidad ganadora usando una ventana de 8 compases



# Aplicaciones en música simbólica

## Identificación de géneros musicales

### Combinación de decisiones parciales

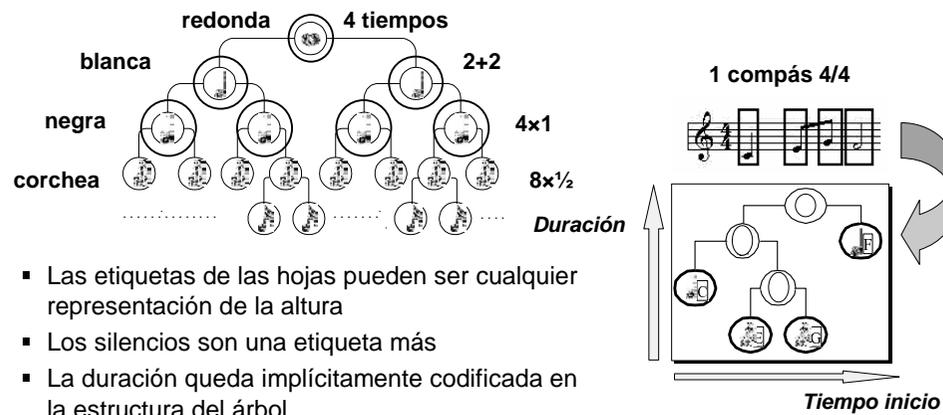


# Representación y descripción simbólica

## Representación

### ÁRBOLES

- Basados en la naturaleza proporcional de la notación musical
- Cada nivel del árbol es una subdivisión del superior



- Las etiquetas de las hojas pueden ser cualquier representación de la altura
- Los silencios son una etiqueta más
- La duración queda implícitamente codificada en la estructura del árbol

(D. Rizo et al. "Tree-Structured Representation of Musical Information". LNCS, 2652: 838-846, 2003)

# Representación y descripción simbólica

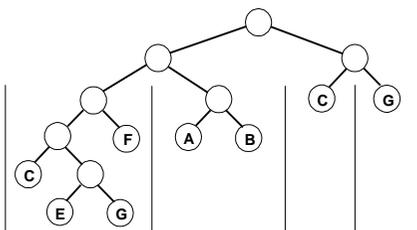
## Representación

### ÁRBOLES

- La melodía completa se compone de compases agrupados secuencial o jerárquicamente



Agrupamiento jerárquico:



Los nodos internos necesitan ser etiquetados → Reglas de propagación de etiquetas

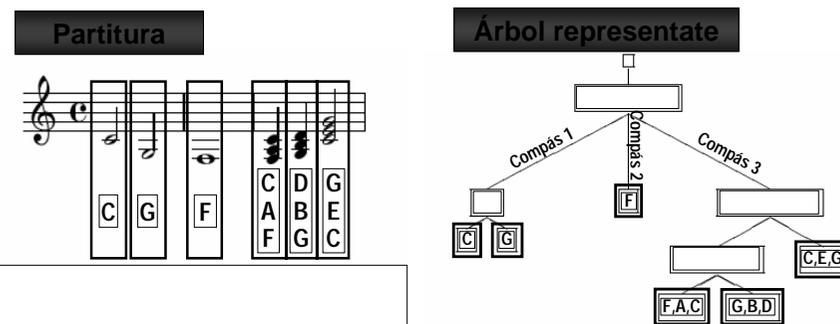
- Notas cortas → árboles grandes
- Elevado coste de manipulación
- Reglas de poda de ramas no relevantes
- Un sistema de análisis melódico se encarga de priorizar en los conflictos

# Aplicaciones en música simbólica

## Representación

### ÁRBOLES POLIFÓNICOS (I)

- Las diferentes pistas se colapsan en una única que contiene todas las notas (sin repeticiones).
- Cada hoja del árbol se etiqueta con las notas que suenan en ese momento.



Las hojas se etiquetan con un conjunto de alturas  $P = \{p_1, p_2, \dots, p_{|P|}\}$

(Rizo D. Iñesta JM. Ponce de León PJ. "Tree Model of Symbolic Music for Tonality Guessing" IASTED Int. Conf. AIA 2006)

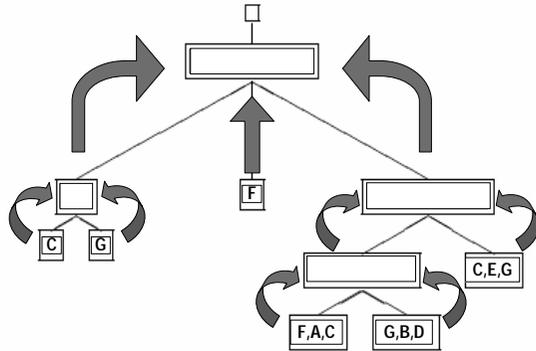
# Aplicaciones en música simbólica

## Representación

### ● ÁRBOLES POLIFÓNICOS (I)

#### Propagación hacia arriba de las etiquetas:

Para etiquetar los nodos internos, dado un subárbol  $\tau$  con hijos  $c_i$ :  $P(\tau) = \bigcup_i P(c_i)$



# Representación simbólica

## Estructuras de datos

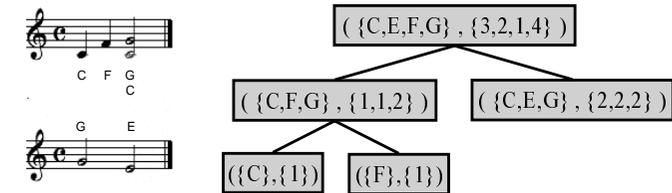
### ● ÁRBOLES POLIFÓNICOS (II)

#### Con multiset

$M = (X, f)$ ;  $X = \{x \mid 0 \leq x \leq 11\}$  (semitonos)

$f: X \rightarrow \mathbb{N} \mid f(x) = 2^{h-l}$  si  $x \in L(\tau)$

$f(x) = 0$  en otro caso



Para los nodos internos, dado un subárbol  $\tau$  con hijos  $c_i$ :  $M(\tau) = \bigcup_i M(c_i)$

donde ahora se usa la unión entre multisets, sumando las multiplicidades de los elementos de los hijos.

# Aplicaciones en música simbólica

## Métricas de similitud melódica

### Distancia de edición melódica entre árboles

Una extensión para árboles del concepto de cadenas. La distancia se halla como el número de operaciones que hay que efectuar para transformar un árbol en otro.

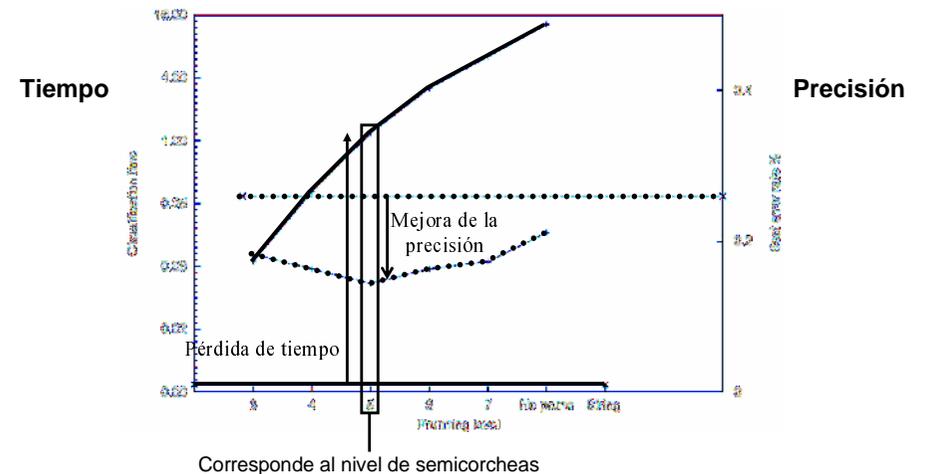


!! Distancia edición árboles  $\in O(|T_1| \times |T_2| \times h(T_1) \times h(T_2))$

# Aplicaciones en música simbólica

## Identificación de melodías mediante árboles

### En función de la profundidad de poda

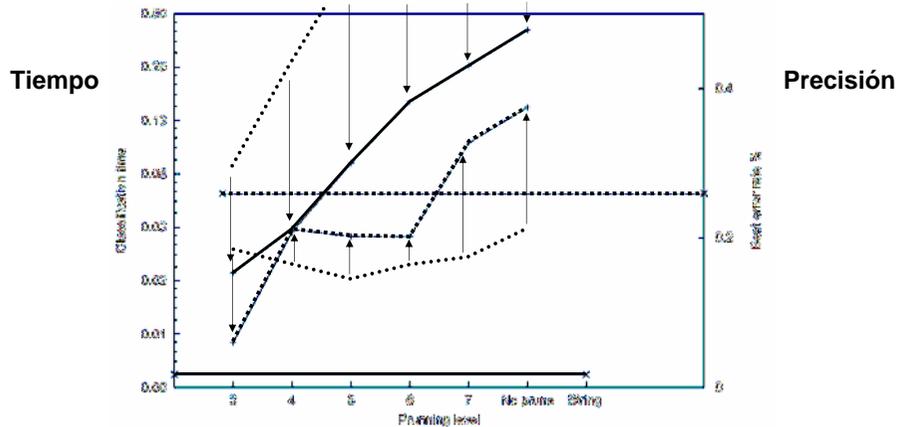


Problema del tiempo  $\rightarrow$  uso de distancias aproximadas

# Aplicaciones en música simbólica

## Identificación de melodías mediante árboles

Tiempo y precisión usando una distancia de edición aproximada (Selkow, 1977)



Búsqueda de compromiso entre precisión y tiempo

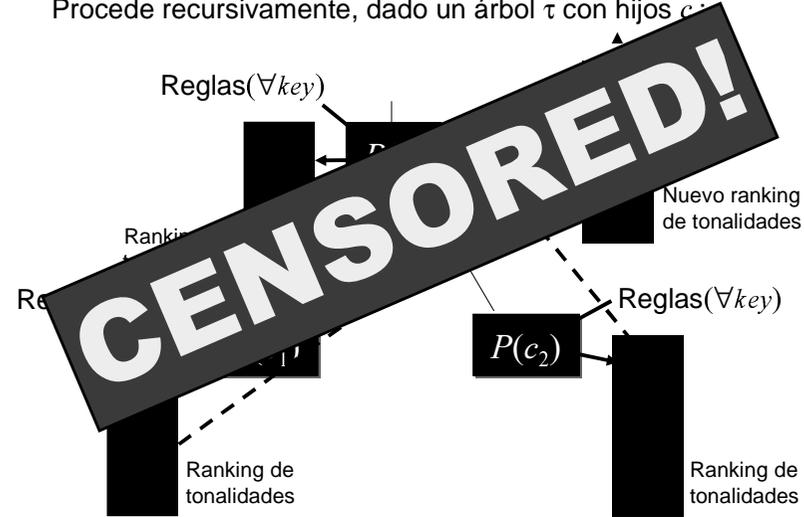
(Selkow, The tree-to-tree editing problem. Information Processing Letters, 6:184-186, 1977)

# Aplicaciones en música simbólica

## Identificación automática de la tonalidad

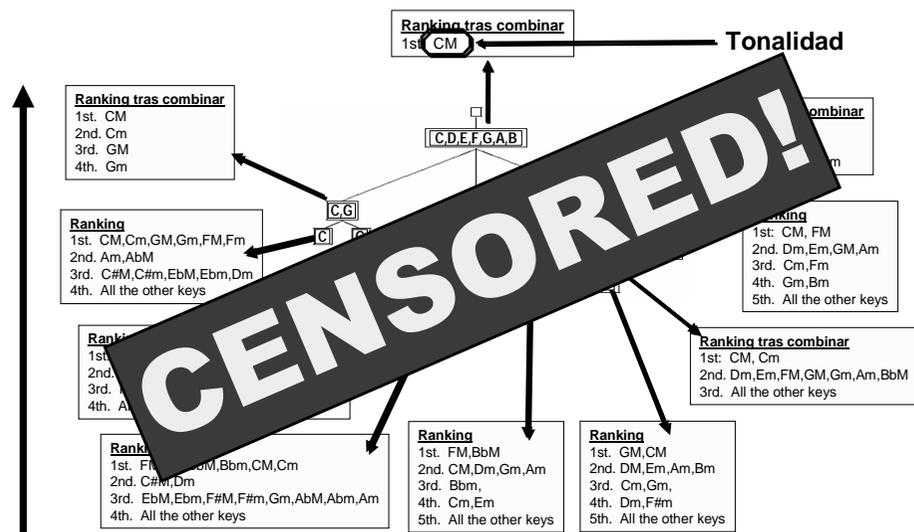
Algoritmo:

Procede recursivamente, dado un árbol  $\tau$  con hijos  $c_1$



# Aplicaciones en música simbólica

## Identificación de tonalidad mediante árboles



Evaluación en post-orden

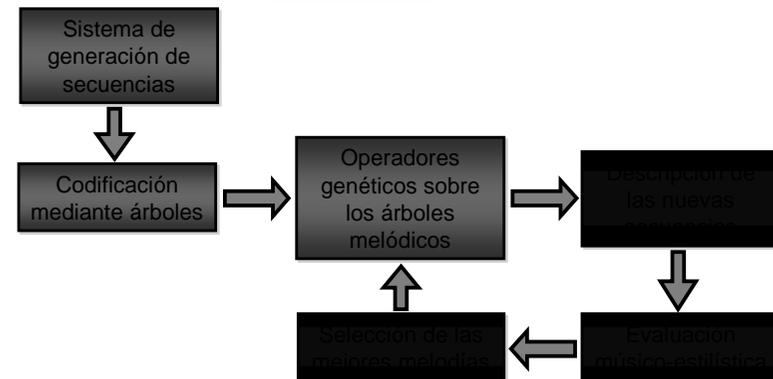
# Aplicaciones en música simbólica

## Composición algorítmica

Combinación de Algoritmos Genéticos y Rec. Pat.

Sistema genético basado en estilo

Compositor  $\leftrightarrow$  interacción  $\leftrightarrow$  Crítico especializado

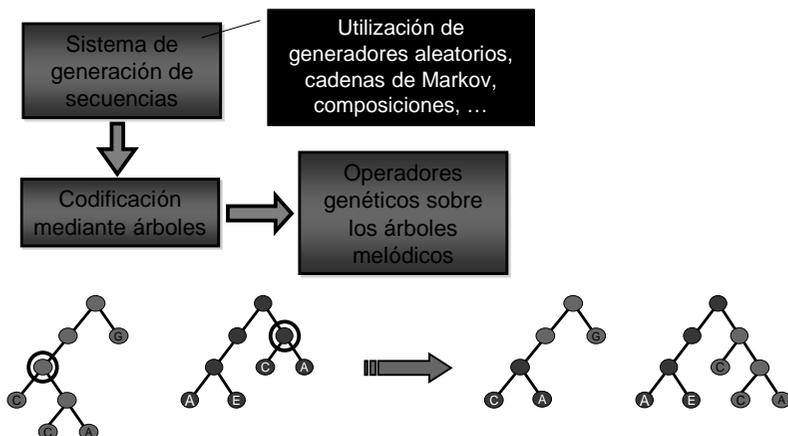


# Aplicaciones en música simbólica

## Composición algorítmica

Combinación de Algoritmos Genéticos y Rec. Pat.

El “compositor”

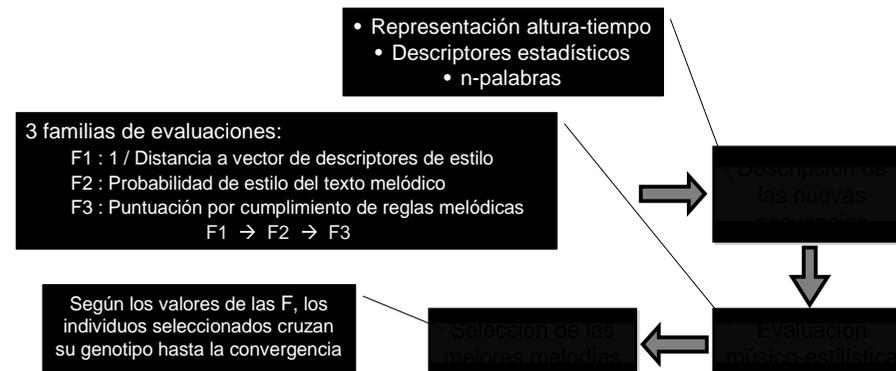


# Aplicaciones en música simbólica

## Composición algorítmica

Combinación de Algoritmos Genéticos y Rec. Pat.

El “crítico especializado”



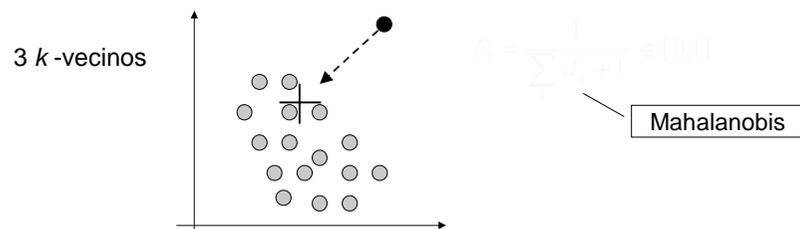
# Aplicaciones en música simbólica

## Composición algorítmica

Combinación de Algoritmos Genéticos y Rec. Pat.

El “crítico especializado”

Criterio de distancia a estilo (global, una especie de condición necesaria):  
Los descriptor estadísticos melódicos, armónicos y rítmicos



+ Criterio local de probabilidad con n-palabra:  
Problema de la clasificación de 1 sola clase

+ Análisis melódico:  
Sistema de reglas

Clásica: 🎻

Jazz: 🎷

# Procesamiento, Aprendizaje y Percepción de Música por Computador

## Contenidos

- ✓ Introducción
- ✓ P.A.&P. en partituras digitales: señales musicales simbólicas
- P.A.&P. en audio digital: señales musicales sonoras
  - P.A.&P. mediante el uso de metadatos
  - Comentarios finales

## ... en audio digital: señales musicales sonoras

### Datos

#### ▪ Categorización de características (4 criterios):

- **Temporalidad:**
  - Estáticas: tomadas en un instante dado (*frame*, decenas de ms)
  - Dinámicas (p.ej., medias o desviaciones de las estáticas a lo largo del tiempo)
- **Extensión temporal:**
  - Globales: descripción de toda la señal (p.ej., sonoridad)
  - Locales: sólo de una parte (p.ej., tiempo de ataque)
- **Nivel de abstracción:**
  - Según lo intuitivas que sean.
- **Proceso de extracción:**
  - Directamente de la forma de onda (p.ej., cruces por cero)
  - De una transformación de la onda (Fourier, wavelets,...)
  - Relacionadas con algún modelo de la señal (fuente, filtro, auditivo...)

## ... en audio digital: señales musicales sonoras

### Catálogo de problemas abiertos:

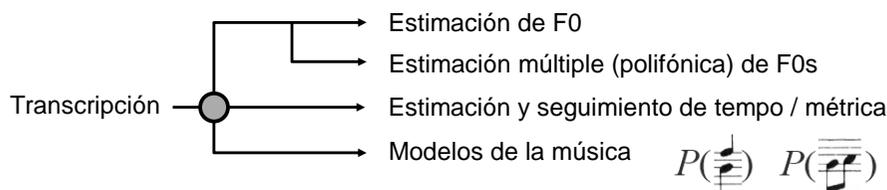
- Identificación de la frecuencia fundamental
- Detección de inicios de notas
- Transcripción automática
- Clasificación de géneros musicales
- Organización de bases de datos musicales
- Segmentación de audio
- Identificación y separación de instrumentos
- Organización de efectos sonoros
- Entre muchos otros...

## Procesamiento y percepción en audio digital

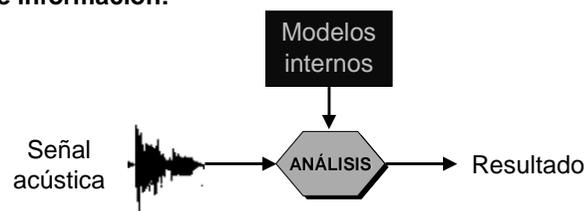
### Transcripción automática

#### Descomposición del problema de transcripción

...en otros menores y más "tratables"



#### ▪ Dos fuentes de información:



## Procesamiento y percepción en audio digital

### Identificación de la frecuencia fundamental (F0)

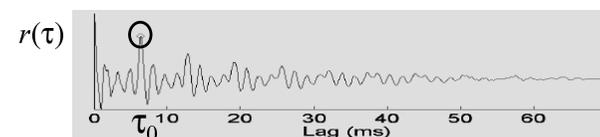
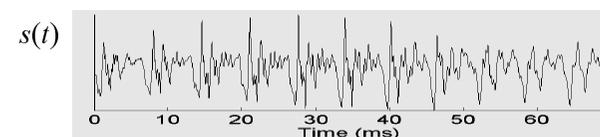
- Ya conocido en otros campos como en el reconocimiento del habla (prosodia)

#### Periodicidad en el tiempo: búsqueda del período fundamental

- Algoritmos basados en la función de autocorrelación

La función de autocorrelación a corto plazo de una señal  $s(t)$ :

$$r(\tau) = \frac{1}{N} \sum_{t=0}^{N-\tau-1} s(t)s(t+\tau)$$



$$1/\tau_0 = f_0$$



# Procesamiento y percepción en audio digital

## Transcripción automática

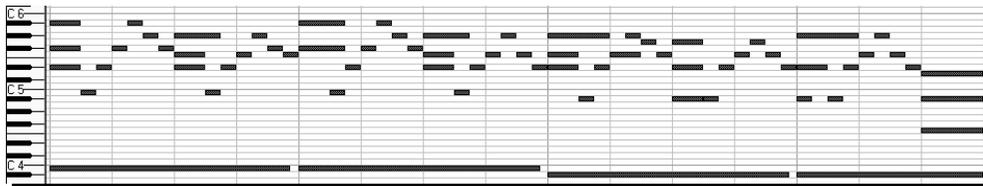
### Ejemplo de transcripción



- Partitura de una melodía ejemplo...



...y su correspondiente pianola.

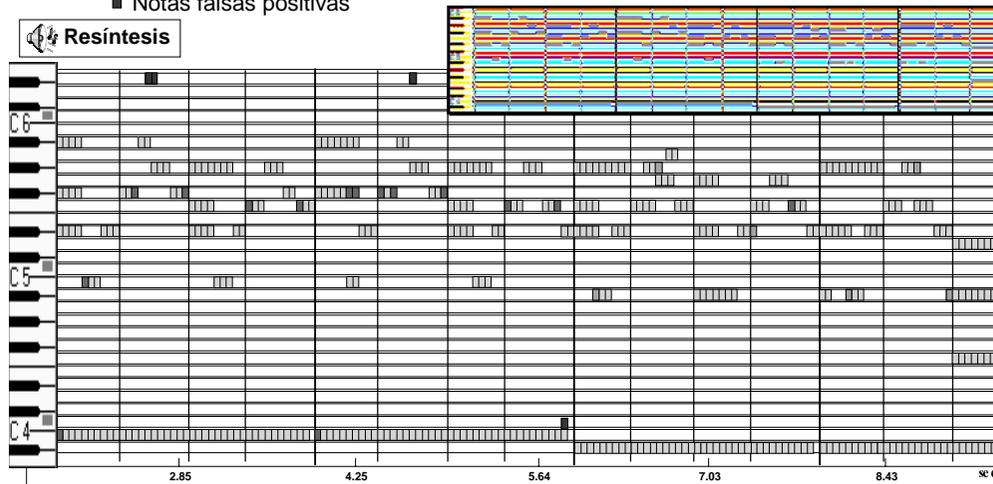


# Procesamiento y percepción en audio digital

## Transcripción automática

### Comparación entre la pianola original y el resultado de la transcripción

- Eventos detectados (output) 94.3% de éxito en el reconocimiento de eventos
- Eventos falsos negativos 100 % de notas detectadas
- Eventos falsos positivos 3 notas falsas positivas
- Notas falsas positivas



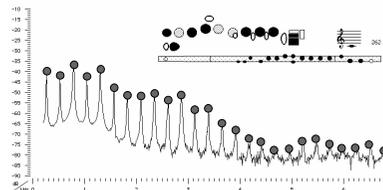
# Procesamiento y percepción en audio digital

## Transcripción automática

### Método basado en modelos armónicos:

#### ■ HIPÓTESIS:

- Armonicidad de la mayor parte de los instrumentos musicales afinados
- “Suavidad” espectral: las amplitudes de los parciales decaen suavemente en función de la frecuencia



#### ■ Esquema general:

- Identificar F0 candidatas (alturas) en cada *frame*
- Generar todas las posibles combinaciones entre candidatos (acordes)
- Seleccionar la combinación que maximiza su sonoridad armónica y su suavidad (“saliencia”).

# Procesamiento y percepción en audio digital

## Transcripción automática

### Método basado en modelos armónicos:

#### ■ Evaluación:

- Usando una base sintética de 4000 acordes y 13000 notas:
  - TP = 9.052 notas, FP = 3.103 y FN = 3.948
  - Precisión = TP / (TP+FP) = 0,75; Recall = TP / (TP+FN) = 0,70
- Usando una base de grabaciones extraídas de CDs y sintetizadas [MIREX 2007]:
  - Precisión = TP / (TP+FP) = 0,83; Recall = TP / (TP+FN) = 0,61

#### Ejemplos:

	Piano clásico	Piano jazz	Piano + bajo	Politimbria	Pop-rock
Original . . . . . :					
Resíntesis de la transcripción . . . :					

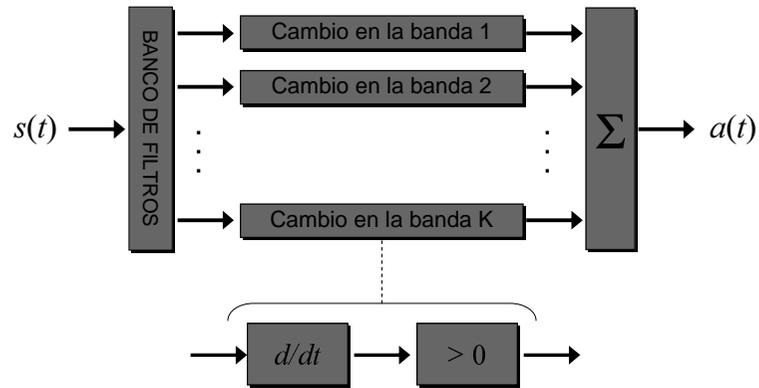
No hay separación de fuentes!

# Procesamiento y percepción en audio digital

## Detección de notas (inicios)

Detección de cambios rápidos y significativos en el espectro de bandas agrupadas por semitonos

### Esquema general

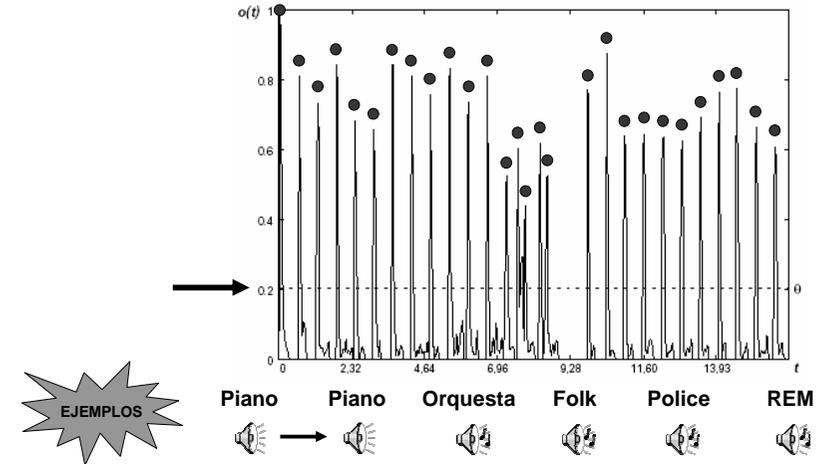


(A. Pertusa et al., "Recognition of note onsets in digital music using semitone bands", CIARP, 2005)

# Procesamiento y percepción en audio digital

## Detección de notas (inicios)

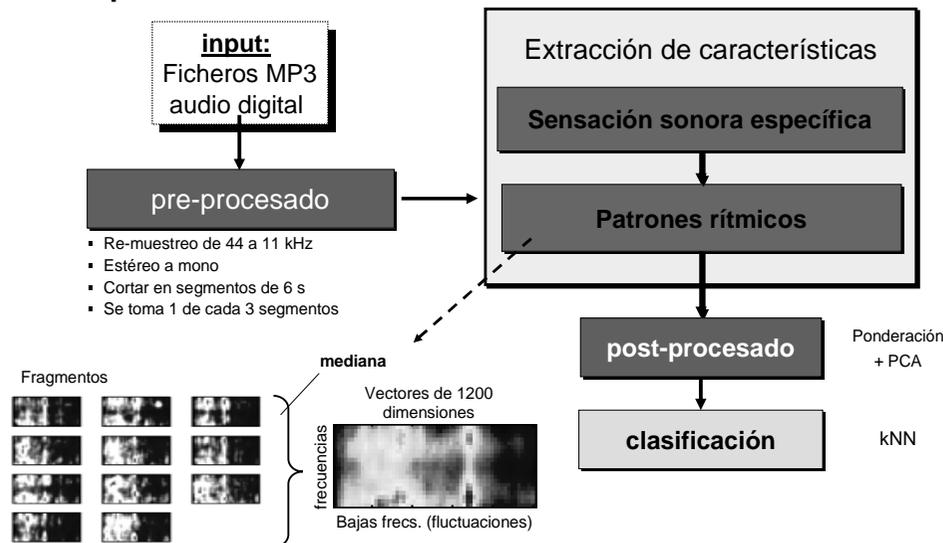
- $a(t)$  se normaliza con la suma para las bandas (en un  $t$  dado)  $\Rightarrow o(t)$
- Se aplica un umbral y los máximos locales son considerados como principios de notas:



# Procesamiento y percepción en audio digital

## Reconocimiento de estilos musicales

### Esquema



# Procesamiento y percepción en audio digital

## Reconocimiento de estilos musicales

### Clasificación

- 1129 piezas de música de baile (~56 horas)
- 10 bailes:

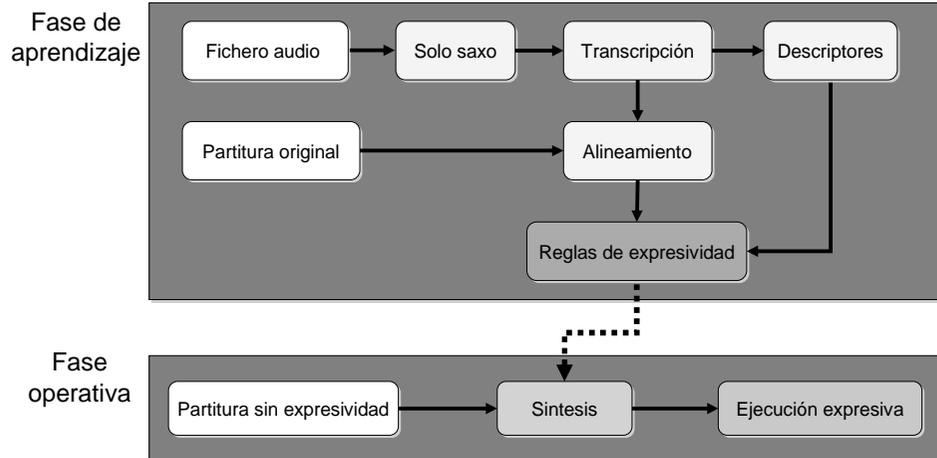
LATIN		BALLROOM	
Samba	SA	Vals lento	SW
Cha-Cha-Cha	CC	Tango	TG
Rumba	RU	Vals vienés	VW
Paso Doble	PD	Foxtrot lento	SF
Jive	JJ	Quickstep	QS

k-NN (k = 5)										
	CC	JJ	SW	PD	QS	RU	SA	SF	TG	VW
CC	110	1	0	8	1	0	0	2	2	0
JJ	0	93	2	0	1	4	0	0	1	2
SW	0	0	144	0	1	1	0	5	0	6
PD	6	0	0	36	0	0	0	2	9	0
QS	0	0	3	0	113	8	0	2	0	0
RU	0	0	10	0	5	80	0	2	0	0
SA	0	0	0	0	4	3	92	0	0	1
SF	0	0	11	0	0	0	0	180	0	4
TG	1	0	1	0	1	0	0	7	89	0
VW	0	1	13	0	2	1	0	9	0	40

## Procesamiento y percepción en audio digital Interpretación humanizada

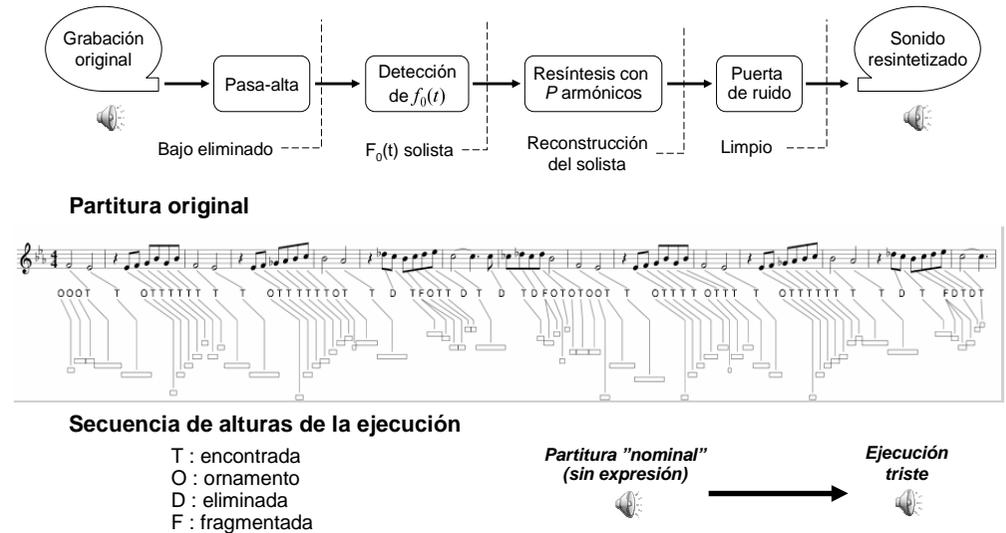
Crear modelos de interpretación estudiando cómo tocan los humanos

### Esquema general del sistema



## Procesamiento y percepción en audio digital Interpretación humanizada

### Extracción del solista



## Procesamiento, Aprendizaje y Percepción de Música por Computador

### Contenidos

- ✓ Introducción
- ✓ P.A.&P. en partituras digitales: señales musicales simbólicas
- ✓ P.A.&P. en audio digital: señales musicales sonoras
- P.A.&P. mediante el uso de meta-datos
- Comentarios finales

## Procesamiento, Aprendizaje y Percepción de Música por Computador

### Contenidos

- ✓ Introducción
- ✓ P.A.&P. en partituras digitales: señales musicales simbólicas
- ✓ P.A.&P. en audio digital: señales musicales sonoras
- ✓ P.A.&P. mediante el uso de meta-datos
- Comentarios finales

# Procesamiento, Aprendizaje y Percepción de Música por Computador

---

## Comentarios finales

- o Los métodos de la Inteligencia Artificial en general y del reconocimiento de patrones / machine learning en particular son útiles en este dominio.
- o *Music Information Retrieval* es un área emergente y existen muchas técnicas de otros campos relacionados (Procesamiento del Lenguaje Natural, Reconocimiento del Habla, Teoría de Lenguajes Formales, Teoría de la señal, etc.) que pueden ser aplicadas o extendidas a ella.
- o Tanto audio, símbolos, como meta-datos presentan retos importantes para el investigador.
- o El reconocimiento y percepción de la música es un tema de investigación interesante y multi-disciplinar, y las industrias y los profesionales de la música puede beneficiarse de sus progresos.

# Procesamiento, Aprendizaje y Percepción de Música por Computador

---

## Comentarios finales

- o **CONGRESOS ESPECÍFICOS:**
  - o Int. Computer Music Conference (ICMC)
  - o Int. Symposium on Music Information Retrieval (ISMIR)
  - o Computer Music Modelling and Retrieval (CMMR)
  - o Int. Conf. Music Perception and Cognition (ICMPC)
- o **REVISTAS SCI:**
- o **ESPECÍFICAS:**
  - o Journal of New Music Research
  - o Computer Music Journal
  - o Journal of Mathematics and Music
- o **RELACIONADAS:**
  - o Journal of the Acoustic Society of America
  - o IEEE Transactions on Systems, Man & Cybernetics
  - o IEEE Transactions on Audio, Speech & Signal Processing
  - o Computers and the Humanities
  - o .....

# Procesamiento, Aprendizaje y Percepción de Música por Computador

**Prof. José Manuel Iñesta**

Computer Music Laboratory

Grupo de Reconocimiento de Formas e Inteligencia Artificial

Dept. Lenguajes y Sistemas Informáticos

Universidad de Alicante

Gracias por vuestra atención



Universidad de Alicante  
Universitat d'Alacant

