

## Corpora SELECCION DE PISTAS MIDI

Este documento describe el formato de estos corpora.

Cada corpus está compuesto por una colección de ficheros MIDI en formato 1 ←  
**COMPROBAR**

Cada corpus corresponde a un género musical diferente

Distinguimos entre corpora de entrenamiento y corpus de validación

Corpora de entrenamiento:

nombres paper ISMIR 2006: CL200, JZ200, KR200

nombre fichero/directorio: clas200(.tgz), jazz200(.tgz), kar200(.tgz)

Corpora de validación:

nombres paper ISMIR 2006: CLA, JAZ, KAR

nombre fichero/directorio: clasica(.tgz), jazz(.tgz), kar(.tgz)

Cada corpus **contendrá** un fichero en formato CSV donde se especifica cuales son los diferentes tipos de pistas que el fichero contiene.

Distinguimos entre los siguientes tipos de pistas:

melody	Pista(s) que contienen la melodía, independientemente de qué instrumento se trate.
bass	Pista(s) que contienen la línea de bajo
piano_rh, piano_lh	Pista(s) que corresponden a la mano derecha (piano_rh) o izquierda (piano_lh) del piano. Reservado para ficheros MIDI donde la distinción entre mano izquierda y mano derecha del piano es clara. La pista de mano derecha será piano_rh sí no está etiquetada como 'melody'.
mixdown	Pistas que contienen partes de diferentes instrumentos.

TABLA 1 : Tipos de pista MIDI

El formato del archivo CSV es el siguiente:

- Cada línea corresponde a un fichero MIDI del corpus, excepto la primera, que contiene los nombres de los campos.
- Usamos el carácter coma ',' para separar campos
- El primer campo es el nombre del fichero MIDI
- El segundo campo indica el número de pistas que contiene el fichero. ATENCION: este es el número de pistas en total, contando tanto las que contienen eventos de nota como las que no (llamadas 'pistas vacías').

- Los siguientes campos contienen listas de números de pista, comenzando en 1, que corresponden a cada tipo, en el orden en el que aparecen en la TABLA 1. Los números de la lista se separan por espacios en blanco.

- Ejemplo de fichero CSV:

```
Nombre_fichero,N_Pistas,melody,bass,piano_rh,piano_lh,mixdown
fichero_ejemplo,16,1 3,5,,,
otro_ejemplo,2,,,1,2,
otro_ejemplo_mas,5,2,,,,,3 4
```

'fichero\_ejemplo' contiene dos pistas de melodía y una de bajo  
 'otro\_ejemplo' contiene una pista de mano derecha de piano y otra de mano izquierda de piano.  
 'otro\_ejemplo\_mas' contiene una pista de melodía y dos pistas mixdown

**OJO: deberíamos plantearnos el usar comillas dobles para los nombres de fichero:**  
**PIERRE 20120212:** OpenOffice pone comillas dobles a cualquier campo que contenga algo más que un número (p. ej. "1 3 5"). Como lo estoy usando para mezclar mis datos de melodía con los de Octavio, lo que estoy generando lleva comillas en esos casos.

20120210: PIERRE dixit

Así pues, tomamos como referencia los ficheros CL200, JZ200 y KR200 que Octavio envió en su último mensaje. Estos sólo tienen etiquetada internamente las pistas de melodía. Desde este momento, bautizo a este conjunto de corpus como melody\_tagged\_2012 y lo dejaré en Verdi (carpeta Training\_sets). Hay dos cosas que hacer a partir de aquí:

1. Creo un documento Google donde podamos ir poniendo las decisiones que tomamos (por centralizarlo en un sitio y no andar buscando mensajes en el correo): PIERRE
2. Subir melody\_tagged\_2012 a verdi. PIERRE
3. Crear el fichero CSV con el formato y la información que hemos discutido:

OCTAVIO.

Voy a ver cual es la mejor manera de hacer esto automáticamente. Para la melodía, tendré que extraer esa info de los ficheros MIDI. Octavio parece que tiene su información en una hoja excel o así. Prepararé un script basado en smf2txt para lo de la melodía y se lo paso a Octavio. Luego Octavio tendrá que importar la información de la que el dispone. UPDATE 20120212: Lo hemos hecho al revés: yo he importado la info sobre melodías a los CSV de Octavio.

**20120212:** Los tres corpora clas200, jazz200, kar200 junto a sus archivos CSV están ahora en verdi ([verdi.dlsi.ua.es:/disco/bbdd/Training\\_sets/pierre/melody\\_tagged\\_2012](http://verdi.dlsi.ua.es:/disco/bbdd/Training_sets/pierre/melody_tagged_2012))