

Percussion-related semantic descriptors of music audio files

Perfecto Herrera-Boyer
Végarð Sandvold
Fabien Gouyon

The music content of digital music files can be described not only by focusing on melody, harmony or "timbre", as it has been the usual case. Specific instrumentation, specially percussion, provides ways to characterize songs according to different semantic dimensions. We propose four high-level descriptors (Percussion Index, Percussion Profile, Kick-Snare Crossings, and Percussivity) and present our preliminary results with them for music retrieval.

Description versus transcription

- **Transcription:** absolute and score-centered; has to be "exact" compared to a ground truth
- **Description:** ecological and user-centred; compared to a score-based ground truth it is fault-tolerant

Percussion index

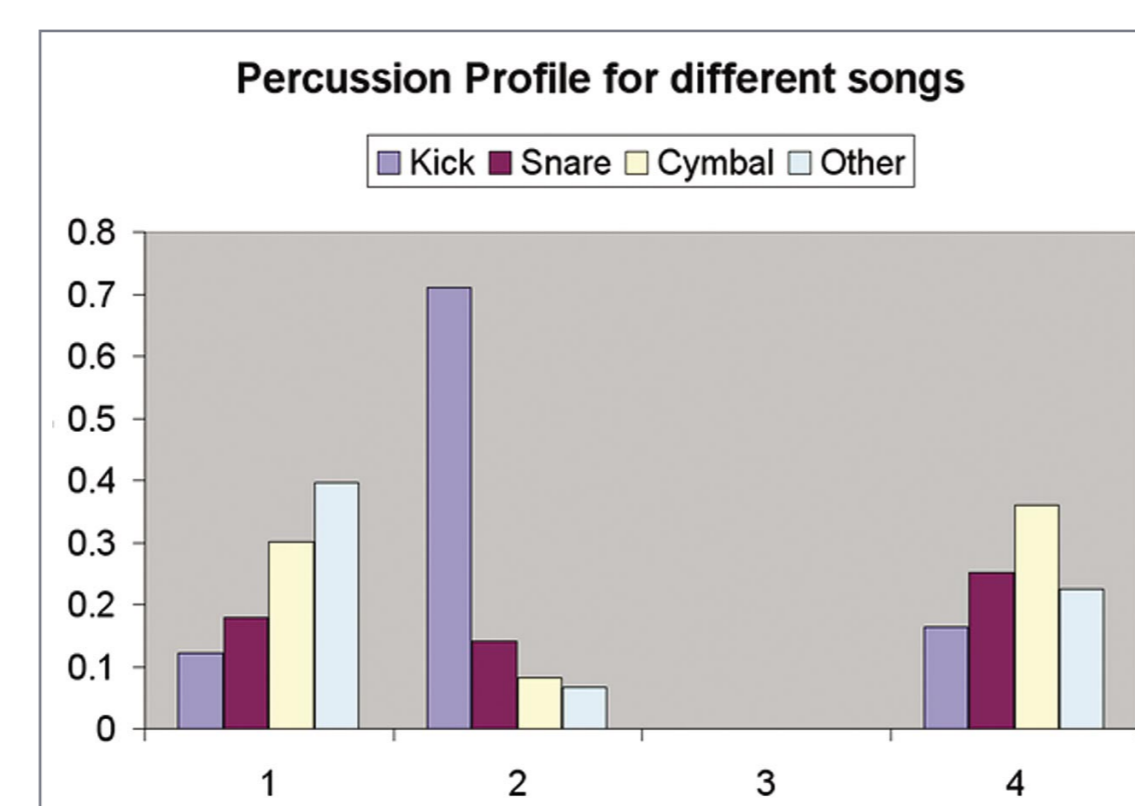
- The PI is computed as the ratio between the number of percussive onsets to the total number of onsets.
- Allows a user to query for music titles containing from "no percussion at all" to "a lot of percussion".

Bartok: Music for Strings, Percussion and Celesta IV Allegro molto	0.22
Mancini: Pink Panther Theme	0.46
Pet Shop Boys: It's a Sin	0.56
The Beatles: Help!	0.66
Portishead: Strangers	0.91

Evidence from listening experiments grants a 4-level quantization: nopercussion, few percussion, moderate, and lots-of-percussion.

Percussion profile

- The PP is a refinement of the PI.
- It has been devised in order to yield an index that gives relative information about specific classes of percussion instruments.
- The PP makes possible to describe songs as having "lot of cymbals" or "being short on snares".



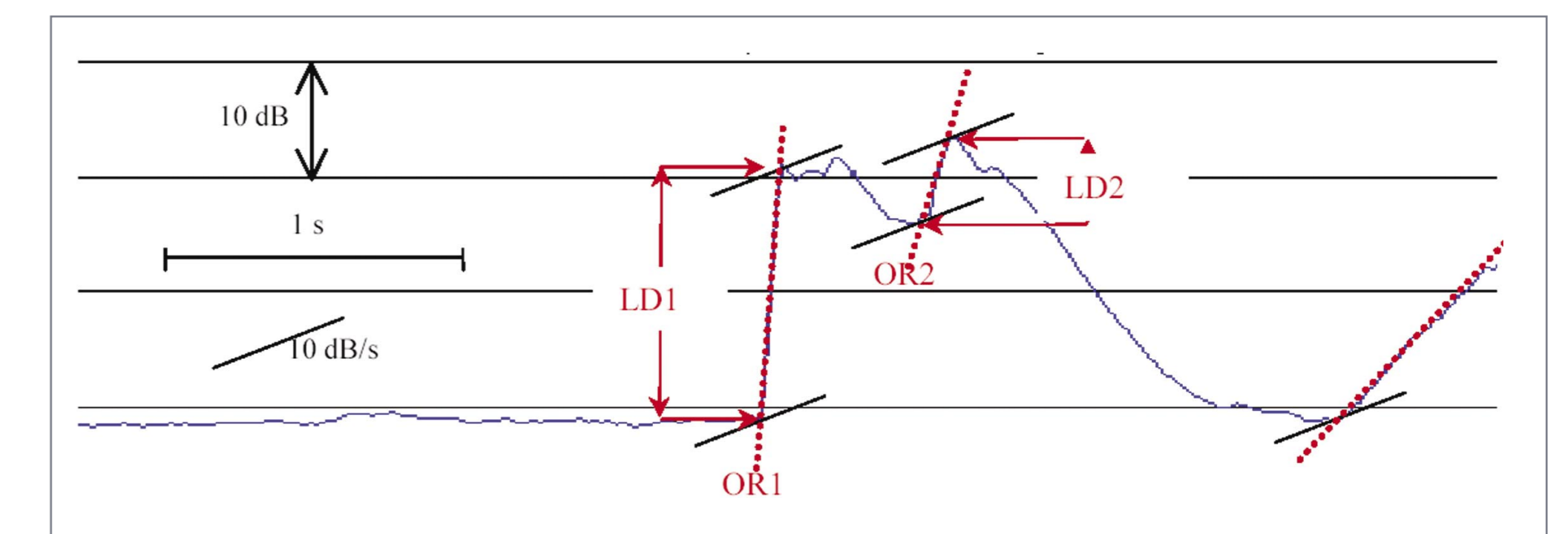
- 1: Beguine the beguine
- 2: Brazil
- 3: J.C. Bach's Ach da ich Wassers gnug hotte
- 4: Riders on the storm

Kick-snare crossings

- KSC is the average number of changes from kick to snare and the other way round. It can be taken as a rough estimation of the complexity of the rhythm pattern.

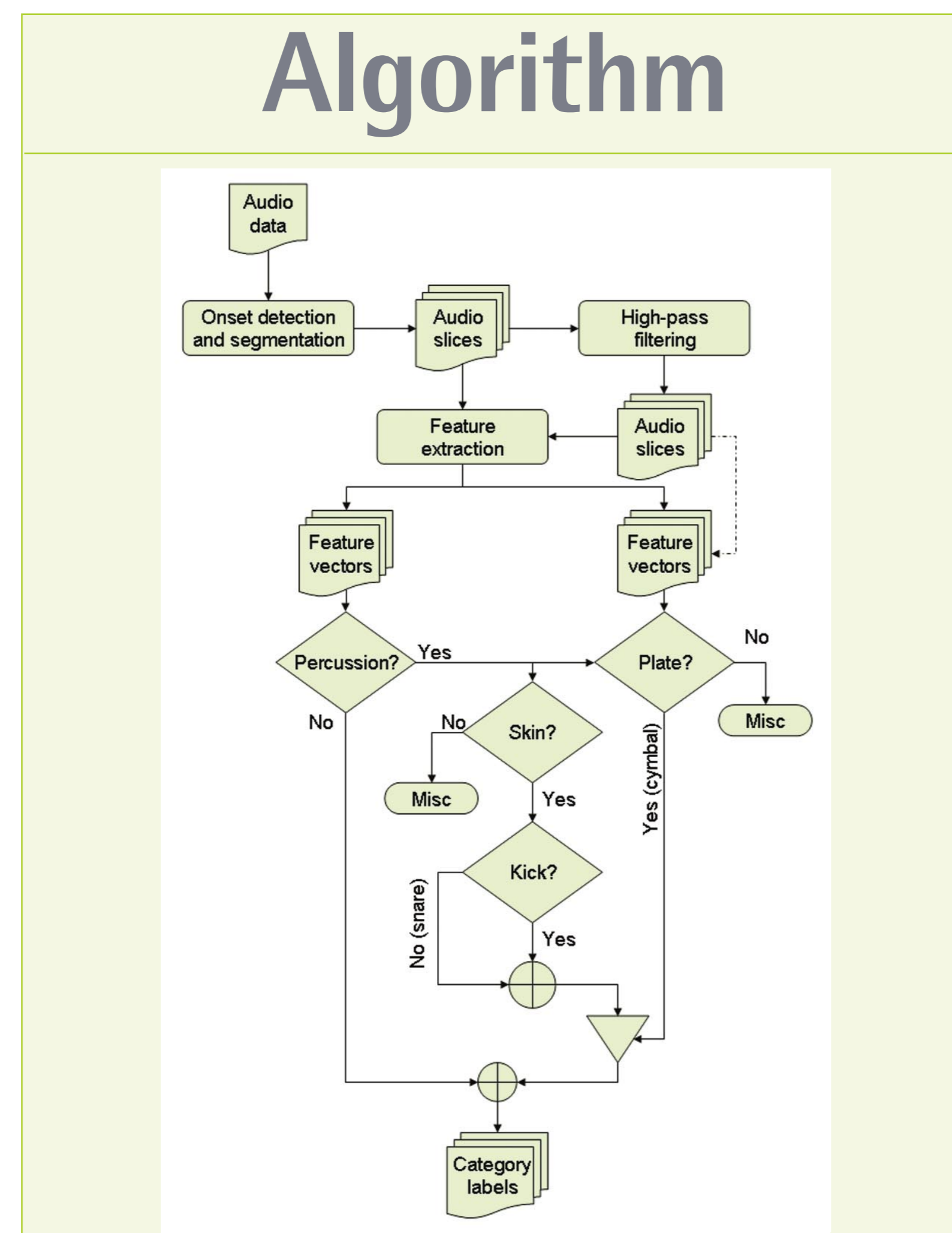
Percussivity

- Perceptual sensation of onset prominence, depending on:
 - The onset slope
 - The onset level
 - The spectrum of the sound at the onset

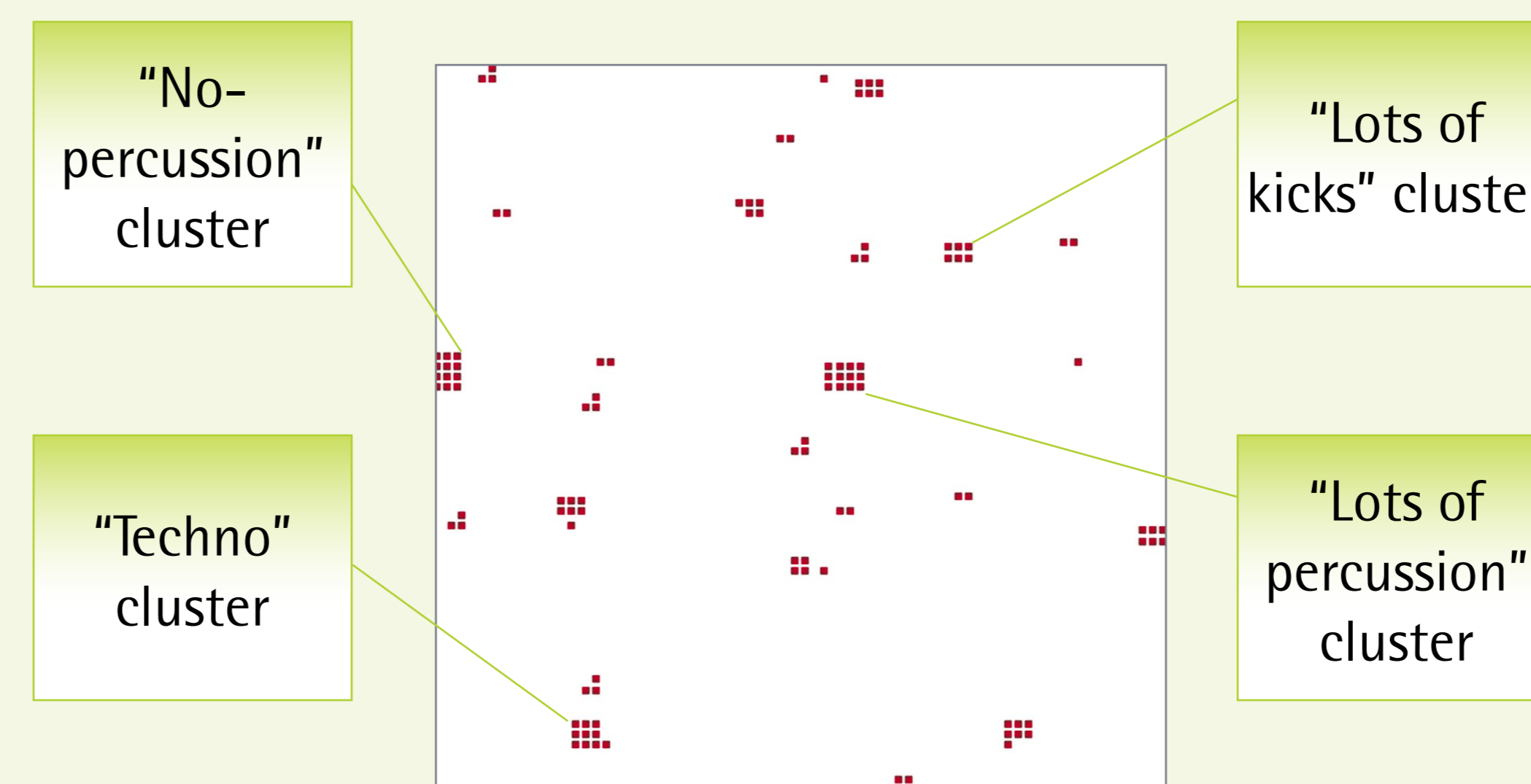


$$P = \alpha \cdot \log(\text{onset rate}/[\text{dB/s}]) + \beta \cdot \log(\text{level difference}/[\text{dB}])$$

Algorithm



Song retrieval



SOM of a 110 songs collection using percussion-related descriptors