Acoustic Cues to Beat Induction: A Machine Learning Perspective

Fabien Gouyon, Gerhard Widmer

11th April 2005

According to Honing, "there seems to be a general consensus on the notion of discrete elements (e.g. notes, sound events or objects) as the primitives of music ... but a detailed discussion and argument for this assumption is missing from the literature." [Contemporary Music Review (9), 1993]

While early computational models of beat induction often entail the processing of discrete events as parsed scores or MIDI events [Longuet-Higgins and Lee, Perception (11), 1982; Desain and Honing, CMJ (13:3), 1989], recent systems tend to deal directly with acoustic signals. Part of them intend to derive similar note-like representations [Dixon, JNMR (30:1), 2001] while others refer to a data granularity of a lower level of abstraction and a different timescale: acoustic features computed on consecutive short signal frames (typically 10 ms-long).

Dealing with discrete note representations, different musical cues to beat induction have been studied: note time, duration, pitch, harmony [Snyder and Krumhansl, Music Perception (18:4), 2001; Dixon and Cambouropoulos, ECAI, 2000]. On the other hand, few different lower level features have been considered, mainly energy variations in several frequency bands [Scheirer, JASA (103:1), 1998; Klapuri et al., IEEE-TSAP (in press), 2005].

In this study, we address the question of which acoustic features are the most representative of musical beats. We consider 274 different acoustic features and evaluate systematically the worth of individual features as well as feature subsets in the task of providing reliable cues on the presence of beats in musical signals. Evaluation of features is based on a machine learning methodology entailing a large corpus of beat-annotated musical audio pieces covering 10 musical genres (1360 instances).