An unsupervised system for the synthesis of variations from audio percussion patterns

Marco Marchini and Hendrik Purwins

Universitat Pompeu Fabra
Music Technology Group - Department of Information and Communications Technologies. Roc Boronat, 138, 08018 Barcelona, Spain.
{marco.marchini3@gmail.com,hendrik.purwins@upf.edu}

Abstract. A system is introduced that learns the structure of an audio recording of a rhythmical percussion fragment in an unsupervised manner and synthesizes musical variations from it. The procedure consists of 1) segmentation, 2) symbolization (feature extraction, clustering, sequence structure analysis, temporal alignment), and 3) synthesis. The symbolization step yields a sequence of event classes. Simultaneously, representations are maintained that cluster the events into few or many classes. Moreover, a tempo estimation procedure is used to preserve the metrical structure in the generated sequence. Employing variable length Markov chains, the final synthesis is performed recombining the audio material derived from the sample itself. In particular, the level of refinement of the clustering procedure is selected, choosing a representation that displays maximal regularity. Examples synthesized from percussion patterns such as the amen break and beat boxing are available on the web. For a broad variety of musical styles the musical characteristics of the original are preserved. At the same time, considerable variability is introduced.

Key words: music analysis, machine listening

As a front-end of the system an onset detector yields a event-wise representation of the signal (cf. Figure 1). On one side, the events are stored as an indexed sequence of audio fragments which will be then used at the end for the resynthesis. On the other side, they are compared one with respect to the other to get a reduced score representation of the percussion pattern. Then a structure discovery procedure is used to find a relevant metrical structure (cf. Fig. 1)

The symbolization consists of a transcription of the audio recording to a fuzzy symbolic representation. Each onset is analyzed by a feature extraction method. According to different clustering thresholds that are selected with a regularity maximization procedure, the sequence of onsets is represented at multiple levels. In each of those, the two sounds can be represented by the same symbol if they are part of the same cluster formed by a clustering procedure (single linkage) with the threshold selected for the level.

This approach comes from an attempt to solve the polyphony problem and results to work fine for percussion sounds.

The symbolic sequence is analyzed statistically employing Variable Length Markov Chains. In [2], a general method for inferencing long sequences is described. For a faster computation, we use a simplified implementation as described in [4]. We employ the construction of a suffix tree for each level based on the sequence of that level. Each node of the tree represents a specific context that has appeared in the past. It carries a list of continuation indices corresponding to block indices matching the *context*.

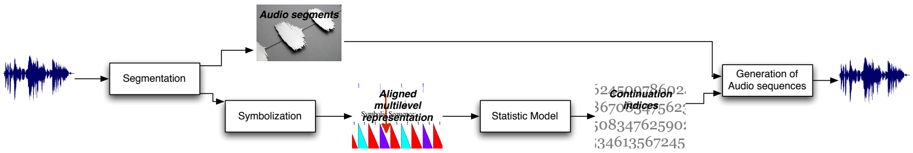

Fig. 1. General architecture of the system.

Some examples are available on the web site [1]. Four examples were taken from the ENST database (see [3]), one from FreeSound.org and two examples were recorded with the percussionist. From ENST we have selected medium/high complexity examples that we numbered according to our collection list. They are example 15, 21, 28 and 31 that correspond respectively to the following files of the ENST database.

From FreeSound we have selected a very popular loop, the “Amen Break”, because of it’s common use and manipulations during improvisation sets.

The system effectively generates sequences which respect the structure and the tempo of the original sound for medium/high complexity rhythmic patterns.

A descriptive evaluation of a professional percussionist confirmed, from one side, that the metrical structure is correctly managed and, from the other, that the statistical representation generates meaningful sequences. For example, he noticed explicitly that the *drum fills* (short musical passages which helps to sustain the listener’s attention during a break between the phrases) were managed correctly by the system.

References

1. www.youtube.com/user/audiocontinuation.
2. Peter Buhlmann and Abraham J. Wyner. Variable length markov chains. *Annals of Statistics*, 27:480–513, 1999.
3. Olivier Gillet and Gaël Richard. Enst-drums: an extensive audio-visual database for drum signals processing. In *ISMIR*, pages 156–159, 2006.
4. Francois Pachet. The continuator: Musical interaction with style. In ICMA, editor, *Proceedings of ICMC*, pages 211–218. ICMA, September 2002. best paper award.