ENST-Drums: an extensive audio-visual database for drum signals processing

Olivier Gillet, Gael Richard

ENST-Drums: an extensive audio-visual database for drum signals processing

Olivier Gillet and Gaël Richard
GET / ENST, CNRS LTCI, 37 rue Dareau, 75014 Paris, France
[olivier.gillet, gael.richard]@enst.fr

Abstract

One of the main bottlenecks in the progress of the Music Information Retrieval (MIR) research field is the limited access to common, large and annotated audio databases that could serve for technology development and/or evaluation. The aim of this paper is to present in detail the ENST-Drums database, emphasizing on both the content and the recording process. This audiovisual database of drum performances by three professional drummers was recorded on 8 audio channels and 2 video channels. The drum sequences are fully annotated and will be, for a large part, freely distributed for research purposes. The large variety in its content should serve research in various domains of audio signal processing involving drums, ranging from single drum event classification to complex multimodal drum track transcription and extraction from polyphonic music.

Keywords: Research database, Automatic drum transcription, Drum event detection in polyphonic music, Source separation, Multimodal music transcription.

1. Introduction

The field of Music Information Retrieval (MIR) is receiving an ever growing interest from the research community, leading to numerous new approaches and algorithms to solve specific indexing and retrieval problems. However, one of the main bottlenecks in this field is the limited access to common, large and annotated audio databases that could serve for both technology development and evaluation. McGill University Master Samples (MUMS)[1], IRCAM Studio Online collection (SOL) [2], and the University of Iowa Musical Instrument Samples [3] are three examples of such databases. Although they are limited to isolated notes, they are widely used by the community, especially for musical instrument recognition tasks. More recently, a large and remarkable database, the RWC Music Database [4], was built and distributed by the Real World Computing Partnership of Japan. As for percussive instruments and drum processing in particular, no large database is publicly available, although several interesting private databases have been built internally by several teams and used in a recent evaluation campaign. For example, the database used for the MAMI drum transcription project [5] has been used during the latest MIREX campaign.

To cope with the limitations of the previous databases for drum signal processing, a large audiovisual drum database was recorded and fully annotated, in order to cover as many applications as possible in the general framework of automatic drum signal analysis. For this purpose, three professional drummers were recorded on eight audio tracks and simultaneously filmed by two cameras (front and right-side views) which shall allow studies on multimodal music transcription and automatic scene and gesture analysis. This approach overcame two common hurdles in the building of music databases: copyrights - the recorded material is original - and annotation - as the availability of individual tracks and video feedback greatly eases the annotation process. For parts of this database, the drummers played on background music to produce material suitable for studies on drum event detection in polyphonic music or single or multiple sensor audio source separation. A significant part of this database will be publicly released for research purposes while a part of it will remain in our premises and could serve for future evaluation campaigns.

The content of the database is described in section 2. Section 3 details the recording and annotation process. The distribution terms and modalities are given in section 4. Finally, some conclusions and perspectives are given in section 5.

2. Database content

The ENST-Drums database is a large and varied research database for automatic drum transcription and processing. For this database, three professional drummers specialized in different music genres were recorded. The total duration of audio material recorded per drummer is around 75 minutes. Each drummer played his own drum kit, and for each sequence, used either sticks, rods, brushes or mallets to increase the diversity of drum sounds. The drum kits themselves are varied, ranging from a small, portable, kit with two toms and 2 cymbals, suitable for jazz and latin music ; to a larger rock drum set with 4 toms and 5 cymbals.

2.1. Detailed content played by each drummer

For each drummer, five different kinds of sequences were recorded. We underline that for all of these items, the drummers never had to follow a score or imitate a reference pattern, but rather had to freely interpret the set of constraints given to them. While it made annotation more difficult and cross-checking impossible, this decision ensured that the musicians always played naturally, producing all kinds of combinations and situations likely to be encountered in real drum playing.

2.1.1. *Individual strokes or ”hits”*

The drummers were asked to play sequences of several strokes separated by a few seconds of silence on each element of the drum kit, for each kind of stick available (plain sticks, rods, mallets and brushes).

2.1.2. *Phrases*

About sixty short drum sequences in various popular styles, without accompaniment, were played by each drummer. Each drummer was given a list of styles: bossa, disco, afro, reggae, jazz, swing, salsa, cha-cha, oriental, rock, blues, metal, hard rock, waltz, funk, country, and was asked to pick his favorites. Due to the different music backgrounds and preferences of the three drummers, only nine of these styles are common to all of them.

For each style, six phrases are played, at different tempi (slow, medium, fast) and at two complexity levels: straight without ornaments, and complex with fill-ins and ornaments. The tempi are not absolute and do not correspond to a given beat per minute (BPM) value, but are rather relative to each genre - e.g., a slow disco phrase would be played at 110 BPM, while a slow Jazz would be played at 70 BPM. Similarly, each drummer interpreted the notion of ”complexity” differently, taking into account his preferences and the targeted style.

2.1.3. *Soli*

Each drummer played a minimum of five soli in the styles of his choice. The instructions given were the following: a typical solo should last about 30s, should use all the drum instruments of the kit and contain some very complex sequences (in terms of number of drum instruments involved, in terms of rhythmic content or/and in terms of tempo).

2.1.4. *Accompaniment*

Seventeen (17) sequences are played by each drummer on top of a pre-recorded accompaniment extracted from ”minus one” CDs [6, 7]. Such CDs are used for the teaching of drumming, and allow students to practice on top of a music accompaniment from which the drum track has been removed. The ”minus one” excerpts are about one minute long, cover various styles (blues, twist, metal, funk, celtic...) and are mostly played by acoustic instruments with a few synthetic keyboards. Additionally, twenty-four (24) shorter sequences were also recorded, in which the drummers played on top of pre-recorded synthetic accompaniments generated from MIDI files (the MIDI drum sounds being muted). A summary of the content available for each drummer is given in table 1.

2.2. Video recordings

For each sequence, two video files are available, corresponding to the front (angle 1) and right side (angle 2) views. Examples are shown in figure 1.

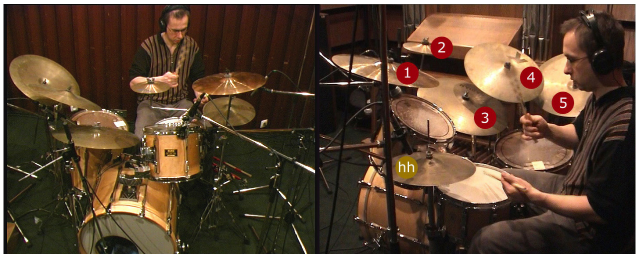

Figure 1. Examples of images recorded by camera 1 (top view) and camera 2 (right side view). The numbering used for cymbal events is overlaid on image 2.

2.3. Audio recordings

For each drum sequence played, a number of audio tracks are recorded or generated which allow the tackling of various drum signal processing applications. This leads to ten (or eleven) audio files per sequence. First, 8 monophonic files corresponding to the 8 microphones: bass drum, snare drum, hi-hat, mid tom, low-mid (if available), low tom track, left overhead, right overhead.Then, 3 stereophonic files: a dry stereo mix of the aforementioned tracks, a ”wet” stereo mix of the aforementioned tracks (see section 3.4 for the list of processings applied); and finally, a stereo file contains the accompaniment (either ”minus one” music background or synthetic MIDI audio files) without drums.

2.4. Annotation

The annotation for each sequence is available as a text file containing a list of $(time, event)$ pairs. Events are identified by the labels listed in table 2. For events associated to cymbals, the number of the cymbal (cymbals are numbered from left to right, from the drummer’s point of view, see figure 1) is also added. For example, **rc3** indicates a ride cymbal hit, the 3rd cymbal for this particular drummer.

3. Building the ENST-Drums database

3.1. Audio recording

8 microphones were used to record the performances: A Beyerdynamic M-88 for the bass drum, a Shure SM57 for the snare drum, a Schoeps CMC body with a cardioid capsule for the hi-hat, two Shure SM58 for the mid and low-mid toms, a Sennheiser 441 for the low tom and two Audio-Technica AT4040 for the overheads. The microphones were amplified by 4 Behringer Ultragain Pro Mic2200 dual preamplifiers. The signals were recorded on a Tascam MX2424

Table 1. Number of sequences and events (strokes) recorded per drummer

	Drummer 1		Drummer 2		Drummer 3	
Item	Sequences	Events	Sequences	Events	Sequences	Events
Hits	29	139	31	180	48	283
Phrases	66	5339	74	9305	68	10467
Soli	7	1420	5	1613	5	1983
Accompaniment (Minus one CD)	17	8856	17	8788	17	9382
Accompaniment (MIDI file)	24	8224	24	6274	24	7357
Total	143	23978	151	26160	162	29472

Table 2. Labels used in the annotation

Label	Description	Label	Description
bd	Bass drum	lmt	Low-mid tom
sweep	Brush sweep	mt	Mid tom
sticks	Sticks hit together	mtr	Mid tom, hit on the rim
sd	Snare drum	lt	Low tom
rs	Rim shot	ltr	Low tom, hit on the rim
cs	Cross stick	lft	Lowest tom
chh	Hi-hat (closed)	rc	Ride cymbal
ohh	Hi-hat (open)	ch	Chinese ride cymbal
cb	Cowbell	cr	Crash cymbal
c	Other cymbals	spl	Splash cymbal

digital multitracker, with a resolution of 16 bits and a sampling rate of 44100 Hz. The click and background tracks were played to the drummers through headphones during the recording of the accompaniment sequences.

3.2. Video recording

Two cameras were used for the video recording (see figure 1 for examples of images). The front view (angle 1) was recorded with a Canon XL1 professional DV camera. The camera was fixed on a tripod mounted on a table, for a total elevation of 2.10m. The right side view (angle 2) was recorded by a Sony DCR-TRV30E DV camcorder, mounted on a tripod. Both cameras recorded at a spatial resolution of 720x576, at 25 frames per second, on mini-DV tapes. Though the recording conditions for this database were well controlled, it is important to mention that no visual clues such as coloured gloves, sticks or backgrounds were used.

3.3. Editing and synchronization

About 3 hours of raw audio material was recorded for each drummer. A first stage in the editing process consisted in editing the audio tracks to remove bad takes and long gaps between sequences. This resulted in 9 edited master audio tracks (8 mono tracks corresponding to the 8 microphones, 1 stereo track corresponding to the accompaniment) per drummer.

Then, two master video tracks, one per camera, in DV format, were built by trimming and aligning the video sequences to match the master audio tracks. We did not observe time base drifting, frame loss, or desynchronization between the audio and video tracks recorded by distinct devices. Consequently, no time-stretching had to be performed. The actual alignment was manually performed by matching sharp and short peaks in the master audio tracks signals, and in the audio signals recorded by the cameras' built-in microphones.

3.4. Mixing

Additionally, two stereo audio mixes were made from the master audio tracks. The ”dry” mix consisted in simply panning and adjusting the level of each instrument, without any further processing. On the ”wet” mix, each instrument was processed by an appropriate equalization and compression. A slight reverberation was added to the result, along with a dynamic processing (Waves L3 Ultramaximizer).

3.5. Annotation

3.5.1. The semi-automatic annotation process

The availability of individual audio tracks eased the annotation process, since each class of drum sound is predominant on the corresponding recording channel. Especially, the bass drum, snare drum, and toms tracks, on which the other instruments of the kit are the most attenuated, could be easily annotated by a same semi-automatic process consisting in detecting all note onsets with the onset detection algorithm presented in [8], building from this onset list a marker file for an audio editor (Wavelab), and finally manually fixing the detection mistakes in the audio editor.

The hi-hat track was annotated using a similar process, but required many more manual corrections, as the snare drum was also present in this track. Moreover, the annotation of this track required the discrimination between closed and open hi-hat strokes. The cymbals were similarly annotated from the pair of overheads. In all cases, a video file adapted to the annotated instrument (angle 1 for cymbals and toms, angle 2 for hi-hat and snare drum) was opened simulatenously, and was extremely helpful in disambiguating strokes.

3.5.2. Special cases

The availability of a video feedback and the mismatch between the audio and video signals we sometimes experienced raised some questions during the annotation process, about which events should be annotated, and which events should not. We encountered:

- Missed strokes, for example when a drummer stretches out his arm to hit a cymbal, but the head of the drum stick misses the cymbal by a few centimeters. These events were not annotated.

- Moves used purely for time keeping which do not cause any sound, or cause extremely quiet artefacts. For example, one of the drummers tapped the base of the hi-hat pedal on odd beats - which resulted in a slight metallic click very distinct from a *closed hi-hat* sound. These events were not annotated.

- Quiet strokes played periodically for time keeping (for example, played for each quarter note). These events were not annotated.

- Attenuated ”Ghost notes” played off-beat and used to create a feeling of ”groove”, especially in styles such as Funk or Shuffle-Blues. These events were annotated. This latter class of events, which is usually ignored by studies on drum transcription, can be filtered out by computing, for each stroke, its energy, and by removing from the transcription all the strokes whose energy falls below a given threshold, or by clustering the strokes in different classes according to their energy and their position within the metric structure.

3.5.3. Verification

The annotation process (which mostly consisted in correcting the output of the onset detection algorithm) was performed by one individual(the first author of this paper). In order to correct mistakes and to homogenize the handling of the special cases described above, the result of this first annotation step was verified once again by the same annotator. Finally, all the verified annotations, for each instrument, were merged in a single master annotation file per performance, whose format is described in 2.4.

3.6. Segmentation

The final step consisted in segmenting the master files (be it annotations, audio or video tracks) into individual files, in order to isolate each sequence into one individual file. For this purpose, a list of markers defining the beginning and end of each sequence was created from the master tracks. A chain of Python and Sylia (VirtualDub's own scripting language) scripts processed this list and created individual files for each segment.

4. Distribution

A large part of the ENST-Drums database will be freely distributed for research purposes. For this purpose, we have received the acceptance for such a distribution (i.e. limited to research purposes) from the three professional drummers and from PDG Music Publishing, who has edited the ”minus one” background music used. The procedure for the distribution is not yet finalized but it should consist in a two step mechanism similar to the one used for the distribution of the RWC Music Database [4]. Firstly, prior to database download, a letter of engagement will need to be signed in which the database usage restriction will be specified. The database web site on which updated information will be posted and from which the database will be downloadable is `http://www.enst.fr/˜grichard/ENST-drums/`. At the time of publication, the web site will be fully operational. The remaining part of the database will remain private to serve in particular future evaluation campaigns.

5. Conclusion

In this paper, we provided a detailed description of the ENST-Drums database. This audiovisual database of drum performances is fully annotated and will be, for a large part, freely distributed for research purposes. The large variety of its content should serve research in different domains of audio signal processing involving drums, ranging from single drum event classification to complex multimodal drum track transcription and extraction from polyphonic music. Future work will be dedicated to the setting up of the distribution procedure and to the definition of evaluation protocols using this database to facilitate future cross comparisons between research studies on this database.

6. Acknowledgements

The authors wish to acknowledge the support of the French ministry of research (ACI-MusicDiscover project) and of the European Commission under the FP6-027026-K-SPACE contract.

References

[1] F. Opolko J. Wapnick. McGill University Master Samples. http://www.music.mcgill.ca/resources/mums/html, 1987-1989.

[2] G. Ballet, R. Borghesi, P. Hoffmann, and F. Levy. Studio online 3.0: An internet killer application for remote access to ircam sounds and processing tools. In *Proc. of Journes d'Informatique Musicale (JIM'99)*, 1999.

[3] L. Fritts. University of Iowa Musical Instrument Samples. http://theremin.music.uiowa.edu/.

[4] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka. RWC Music Database: Popular, Classical, and Jazz Music Databases. In *Proc. 3rd International Conference on Music Information Retrieval (ISMIR 2002)*, pages 287–288, October 2002.

[5] K. Tanghe, M. Lesaffre, S. Degroeve, M. Leman, B. De Baets, and J.-P. Martens. Collecting Ground Truth Annotations for Drum Detection in Polyphonic Music. In *Proc. 6th Int. Conf. on Music Information Retrieval (ISMIR 2005)*, pages 50–57, September 2005.

[6] E. Thiévon. *Batterie mode d'emploi - Playbacks*. PDG Music Publishing, 2004.

[7] E. Thiévon and P. Argentier. *Drums Training Session - Métier et variété*. PDG Music Publishing, 1999.

[8] M. Alonso, G. Richard, and B. David. Extracting Note Onsets from Musical Recordings. In *Proc. IEEE Int. Conf. Multimedia and Expo*, 2005.

[9] O. Gillet and G. Richard. Automatic Transcription of Drum Loops. In *Proc. 2004 International Conference on Acoustics, Speech, and Signal Processing (ICASSP'04)*, May 2004.