GTZAN-RHYTHM: EXTENDING THE GTZAN TEST-SET WITH BEAT, DOWNBEAT AND SWING ANNOTATIONS

Ugo Marchand, Quentin Fresnel, Geoffroy Peeters
STMS IRCAM-CNRS-UPMC
marchand,fresnel,peeters@ircam.fr

ABSTRACT

We present here the GTZAN-Rhythm test-set which is an extension of the GTZAN test-set with rhythm annotations. While there is a controversy related to the use of the GTZAN test-set for music genre classification [13], its audio content is however representative of real commercial music of various music genre and its audio is also already widely shared. Also, considering the various genres included in it, it provides a good balance between various types of rhythm including tracks with and without swing. We manually annotated the 1000 audio excerpts into beat, downbeat and swing annotations. We describe here this new test-set, as well as the validation of the annotations and the various uses one can make out of it (swing analysis, beat and downbeat tracking).

1. INTRODUCTION

The GTZAN-Rhythm test-set [1] is an extension of the GTZAN test-set with rhythm annotations (beat, downbeat and swing). While there exist several test-set related to rhythm annotations (see Table below), none of them provides annotation at the swing level (systematic displacement of the eight note).

We therefore decided to create a new test-set with the full metrical information including the swing factor for our research on swing estimation [9]. Rather than starting from scratch, we have decided to extend an existing corpus, the GTZAN test-set, because its audio represents real commercial music and because its audio is already widely shared.

2. GTZAN-RHYTHM TEST-SET

In the following we describe this new test-set using the recommendation made by [11] for "the description of annotated MIR corpora". The letters and numbers in brackets (such as "(B31)") refer to the description of [11].

Corpora Name	Accessibility	Tempo	Meter	Beat	Downbeat	Swing
MIREX-2006	no	(T)	-	B	-	-
ISMIR-2004 Songs	yes	T	-	-	-	-
ISMIR-2004 Loop	yes	T	-	-	-	-
ISMIR-2004 Ballroom [7]	yes	(T)	(M)	B	D	-
Klapuri [6]	on-demand	(T)	-	B	D	-
Hainsworth	on-demand	(T)	-	B	-	-
Mazurka (MIREX)	no	(T)	-	B	-	-
Isophonic-Beatles [10]	yes	(T)	(M)	B	D	-
RWC [2]	yes	(T)	(M)	B	D	-
SMC / MIREX-2012 [4]	yes	(T)	-	B	-	-
Last-FM [8]	yes	T	-	-	-	-
HJDB [3]	no	-	-	-	D	-
GTZAN-Rhythm [9]	yes	T	M	B	D	S

2.1 (A) Raw corpus (audio):

The GTZAN-Rhythm test-set is based on the audio of a widely used test-set for music genre recognition research: the GTZAN test-set [14]. This test-set is made of 1000 audio excerpts of 30-seconds duration in mono at 22.050 Hz. Those are equally distributed into 10 music genres: blues, classical, country, disco, hip-hop, jazz, metal, pop, reggae and rock.

2.2 (B) Annotations

(B1) Annotation origin: Each track of the test-set has been manually annotated (B15).

(B21) Concepts definition: The concepts being annotated are the downbeat, beat/tactus (quarter-note) and eight-note/tatum positions [2].

(B22) Annotation rules: The annotation was performed on the whole duration of each excerpt in a semi-automatic way. Beat and downbeat positions have been first automatically estimated using ircambeat software [12]. Then the resulting beat and downbeat positions have been manually corrected. The subdivision (by 2 or 3) of the beat positions have been used to generate the tatum positions which have been displaced by hand to fit locally the audio content.

(B31) Annotators: The annotations have been made by two different annotators, both are practicing musicians and researchers. One annotator did all the tracks numbered from 0 to 49, the other annotator did the tracks from 50 to 99. Each annotator then verified the work of each other, and reached a consensus when they were not in agreement.

---

[2] We only annotated the eight-note position when swing exist or in case of a ternary track.

(B32) Corpus Validation/Reliability: We tested the inter-annotator agreement between the two annotators. For this a part of the test-set (tracks from 95 to 99 from each genre) was annotated by both annotators.

For *beat positions*, we tested the inter-annotator agreement using the 3 followng metrics: F-measure, AMLt and Information Gain. We used the beat-tracking-evaluation-toolbox presented in [1] to compute these metrics. The results are presented in Table 1. The good F-measure (91%), AMLt (95%) and information gain (4 bits) show that both annotators are most of the times in agreement for the positions of the beats.

	F-measure	AMLt	Inf. Gain
Agreement	90.7%	94.9%	3.99 bits

Table 1. *Agreement between the annotators on beat-tracking, evaluated on a 5% subset of our base.*

For *downbeat positions*, we tested the agreement for the subset of the most difficult tracks, the 153 tracks with swing. Knowing that the beat positons are shared by the two annotators, the two annotators agreed for the ones being the downbeat for 150/153 tracks. After discussion, they always reached a consensus.

(B4) Annotation tools: The annotation have been performed using the Audiosculpt 3.0 graphical software which allows to represent graphically the downbeats, beats and tatums positions over a spectrogram and allows dedicated editing of those marker types.

2.3 (C) Documenting and storing

(C1) Corpus Identifier: The identifier of the annotated corpus is GTZAN-Rhythm-V02.

(C2) Storage: The GTZAN-Rhythm test-set can be downloaded at `http://anasynth.ircam.fr/home/media/GTZAN-rhythm/`. The test-set contains:

• a set of 1000 musicdescription XML files (one for each audio track) containing the beat, downbeat and tatum positions,

• a single CSV file describing on each row the global information associated to a track: artist, title and a summary of the rhythm information (mean tempo, mean meter, mean swing ratio . . .). The summary of the rhythm information has been obtained using the included python script `generate.py`. This script also exemplifies the reading of the musicdescription xml files,

• a set of 1000 JAMS [3] files containing the beat, downbeat, tatum, artist, title and mean tempo, ...

The artist and the title information have been kindly provided by Bob Sturm. If you use the GTZAN-Rhythm test-set, please cite [9].

---

[3] JSON Annotated Music Specification for Reproducible MIR Research [5]

3. APPLICATIONS

As there is currently no test-set on which the *swing ratio* is annotated, our main contribution is to provide one. This test-set has been already used for our study on swing ratio estimation [9]. We hope it could serve to quantitatively evaluate other related works on swing.

This test-set also provides *beat and downbeat annotations* on a large (1000 tracks) set of commercial audio tracks. It could therefore be used for the evaluation of beat-tracking, downbeat-tracking, tempo or meter estimation algorithms. Even though some audio tracks of the GTZAN-Rhythm test-set are quite simple for beat and downbeat tracking, some other (especially in the jazz, blues and classical section) are quite challenging.

4. REFERENCES

[1] Matthew EP Davies, Norberto Degara, and Mark D Plumbley. Evaluation methods for musical audio beat tracking algorithms. *Queen Mary University of London, Centre for Digital Music, Tech. Rep. C4DM-TR-09-06*, 2009.

[2] Masataka Goto. Aist annotation for the rwc music database. In *ISMIR*, pages 359–360, 2006.

[3] Jason Hockman, Matthew EP Davies, and Ichiro Fujinaga. One in the jungle: Downbeat detection in hardcore, jungle, and drum and bass. In *ISMIR*, pages 169–174, 2012.

[4] Andre Holzapfel, Matthew EP Davies, José R Zapata, João Lobato Oliveira, and Fabien Gouyon. Selective sampling for beat tracking evaluation. *Audio, Speech, and Language Processing, IEEE Transactions on*, 20(9):2539–2548, 2012.

[5] Eric J Humphrey, Justin Salamon, Oriol Nieto, Jon Forsyth, Rachel M Bittner, and Juan P Bello. Jams: a json annotated music specification for reproducible mir research. In *Int. Society for Music Information Retrieval Conf.(ISMIR 2014)*, 2014.

[6] Anssi P Klapuri, Antti J Eronen, and Jaakko T Astola. Analysis of the meter of acoustic musical signals. *Audio, Speech, and Language Processing, IEEE Transactions on*, 14(1):342–355, 2006.

[7] Florian Krebs, Sebastian Böck, and Gerhard Widmer. Rhythmic pattern modeling for beat and downbeat tracking in musical audio. In *ISMIR*, pages 227–232, 2013.

[8] Mark Levy. Improving perceptual tempo estimation with crowd-sourced annotations. *Proc. of ISMIR,(Miami, USA)*, 2011.

[9] Ugo Marchand and Geoffroy Peeters. Swing ratio estimation. In *Proc. of the 18th Int. Conference on Digital Audio Effects (DAFx-15)*, December 2015.

[10] Matthias Mauch, Chris Cannam, Matthew Davies, Simon Dixon, Christopher Harte, Sefki Kolozali, Dan Tidhar, and Mark Sandler. Omras2 metadata project 2009. In *Proc. of 10th International Conference on Music Information Retrieval*, 2009.

[11] Geoffroy Peeters and Karen Fort. Towards a (better) definition of the description of annotated m.i.r. corpora. October 2012.

[12] Geoffroy Peeters and Helene Papadopoulos. Simultaneous beat and downbeat-tracking using a probabilistic framework: theory and large-scale evaluation. *Audio, Speech, and Language Processing, IEEE Transactions on*, 19(6):1754–1769, 2011.

[13] Bob L Sturm. The gtzan dataset: Its contents, faults, and their effects on music genre recognition evaluation. *IEEE Transactions on Audio, Speech and Language Processing*, 2013.

[14] George Tzanetakis and Perry Cook. Musical genre classification of audio signals. *Speech and Audio Processing, IEEE transactions on*, 10(5):293–302, 2002.