# A-MAPS: AUGMENTED MAPS DATASET WITH RHYTHM AND KEY ANNOTATIONS

**Adrien Ycart and Emmanouil Benetos**
Centre for Digital Music, Queen Mary University of London

## ABSTRACT

The MAPS dataset is the most used benchmark dataset for automatic music transcription (AMT). We propose here an updated version of the ground truth, containing precise beat, time signature, and key signature annotations.

## 1. INTRODUCTION

Automatic music transcription is a canonical task in music information retrieval [1]. Roughly, music transcription is the task of extracting from an audio recording a music score in staff notation representing what was played. Much of the literature has focused on the simpler, intermediate goal of extracting a symbolic representation describing what notes were played and when, usually in the form of a list of note events (MIDI file), leaving aside tasks such as beat tracking, rhythm transcription, pitch spelling, etc. In what follows, we call the former complete music transcription (CMT) and the latter automatic music transcription (AMT).

The MIDI Aligned Piano Sounds (MAPS) dataset [2] has become a widely-used benchmark dataset for AMT. It contains around 18 hours of classical piano music, along with aligned MIDI transcriptions. However, the information it contains is limited to the pitch, onset and offset times in seconds of the notes in the recording (velocity is also included, but rarely used).

Recently, some systems have incorporated additional musical information into AMT system, such as meter [6]. With the progress of AMT, CMT has also gained attention [3]. To make evaluation and comparison with state-of-the-art AMT systems possible, we propose some augmented annotations for MAPS, that include in particular rhythm (meter, note values), key, as well as other general information, all the while preserving the pitch, onset and offset as they were in the original MAPS files. We make these annotations available for further use [1], in the form of MIDI files.

A. Ycart is supported by a QMUL EECS Research Studentship. E. Benetos is supported by a UK RAEng Research Fellowship (RF/128).

[1] `http://c4dm.eecs.qmul.ac.uk/ycart/a-maps.html`



## 2. THE MAPS DATASET

The MAPS dataset contains MIDI files of polyphonic piano music, along with aligned audio renditions, generated using synthetic pianos and Disklavier acoustic pianos. It also contains recordings of isolated notes and chords, that we do not consider here.

The MAPS MIDI files used to render the audio were taken from the Piano-Midi.de [2] (PM) database. The PM database was made by manually editing the velocities and the tempo curves of quantised MIDI files in order to give them a natural interpretation. These MIDI files originally contained much more information (tempo and note values, but also key, textual indications, sustain pedal...) which, unfortunately, wasn't kept in MAPS. We aim to retrieve this additional information and include it in the MAPS ground-truth MIDI files.

## 3. METHOD

Directly using the PM files as ground truth is not possible, since this dataset is continuously updated by its creator, meaning many files have been slightly modified since the creation of MAPS. We thus resort to a symbolic MIDI-to-MIDI alignment method [4] to align pairs of files. From these alignments, we get a correspondence table that tells us to which PM quantised point each MAPS note is aligned. More precisely, we get a table $T$ such that $T[i] = [t, q]$ where $t$ is a time in seconds and $q$ is a quantised time in quarter notes. We remove the duplicates, so that each $q$ is unique (when a single $q$ value corresponds to many $t$ values, we keep the first $t$ value). From this correspondance table, we get a tempo curve following closely the MAPS files. Let $bpm(t)$ be the tempo value is quarter notes per minute at time t. For each $T[i] = [t_1, q_1]$, $T[i+1] = [t_2, q_2]$ we have the formula:

$$bpm(t) = \frac{q_2 - q_1}{t_2 - t_1} \times 60 \text{ for } t \in [t_1, t_2[$$

In some cases, the ordering is not preserved, *i.e.* $q_1 < q_2$ but $t_1 \geq t_2$. In such cases, we delete one of the two problematic lines, keeping the one for which the quantised time was closest to a 16th note subdivision. When the two quantised times are both on a 16th note subdivision, we keep the one that results in the smallest tempo deviation.

Adapting the tempo curve does not yield perfect results: some notes were modified in the PM files after the creation

[2] `http://piano-midi.de/`

| $\delta_{mean}$ | | | $\delta_{max}$ | | |
|---|---|---|---|---|---|
| min | med | max | min | med | max |
| 0.0 | 0.028 | 3.18 | 0 | 1 | 384 |

**Table 1**. Evaluation of the beat annotations: minimum, median and maximum values in MIDI ticks of $\delta_{mean}$ and $\delta_{max}$ across the dataset

of MAPS, or displaced during the re-alignment process in the making of MAPS. To get the exact notes from MAPS, we match the notes in the MAPS files with those in the tempo-aligned PM files using the `mir_eval` toolbox [5]. We then remove all the PM notes that are unmatched, and add the MAPS notes that are unmatched. We add the missing notes to the staff holding the PM note to which they were aligned by the alignment algorithm [4]. When adding a MAPS note aligned to no PM note, we add it to the right hand staff if its MIDI pitch is above 60, to the left hand staff otherwise (this concerns 2.7% of the notes). Eventually, all the correct notes are retrieved, and their onset and offset values are within 5ms of the original.

Because of this second step, some notes in the aligned files are not exactly quantised to beat subdivisions, although they were in the original PM files. To know how much, we compare the quantised versions of the original PM files and aligned files. We match the notes of the two files with [4], and for each pair of notes, we compute the deviation between their onsets in MIDI ticks (1 tick is a $480^{\text{th}}$ of a quarter note). Then, for each file, we compute the mean ($\delta_{mean}$) and maximum ($\delta_{max}$) deviations. We report the minimum, median, and maximum values of $\delta_{mean}$ and $\delta_{max}$ across the dataset in Table 1. It has to be noted that the large maximum values recorded are most likely due to notes that were modified in later versions of the PM MIDI files.

## 4. CONTENTS

All the annotations that were in the original PM files were retrieved. They include:

- Tempo curve
- Time signature
- Durations of notes in fraction of a quarter note (some of them are approximate)
- Key signature (always written as the major relative)
- Sustain pedal activation
- Separate left and right hand staff
- Text annotations from the score (tempo indications, coda...).

Yet, they do not include all the information needed for CMT, mostly due to limitations of the MIDI format. In particular, they do not include pitch spelling (C♯ is the same as D♭), rhythm spelling (a quarter note tied to an eighth note is the same as a dotted quarter note), ornaments (trills are written as a succession of notes, not as a separate symbol), dynamics information (only the velocity of individual notes is available), or short key modulations (only changes in key signature are written).

Some data from the MAPS dataset was lost throughout the process. Overall, we lost around 13 minutes of data (1.2% of total duration). The details of the lost data can be found on the website [3] .

## 5. APPLICATIONS

These annotations can be useful in the development of musically-informed AMT systems (for instance, key- or meter-dependent systems). They can be used directly as input to the systems to test their performance in the ideal case, or to assess the performance of sub-modules of the system.

They can also be used for other tasks than AMT, such as beat tracking, and meter or key detection from audio. Although the note values given in the annotations are not always exact, they can provide some useful estimates for rhythm quantisation. The textual annotations could also be used for structure analysis.

Finally, even though some elements are missing to reconstruct a full score, we hope that all the information contained in these annotations can help working towards CMT, by providing a big amount of annotated audio data and allowing comparison with previous AMT systems.

## 6. REFERENCES


[1] E. Benetos, S. Dixon, D. Giannoulis, H. Kirchhoff, and A. Klapuri. Automatic music transcription: challenges and future directions. *Journal of Intelligent Information Systems*, 41(3):407–434, 2013.

[2] V. Emiya, R. Badeau, and B. David. Multipitch estimation of piano sounds using a new probabilistic spectral smoothness principle. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(6):1643–1654, 2010.

[3] E. Nakamura, E. Benetos, K. Yoshii, and S. Dixon. Towards complete polyphonic music transcription: Integrating multi-pitch detection and rhythm quantization. In *ICASSP*, 2018.

[4] E. Nakamura, K. Yoshii, and H. Katayose. Performance error detection and post-processing for fast and accurate symbolic music alignment. In *ISMIR*, 2017.

[5] C. Raffel, B. McFee, E. J. Humphrey, J. Salamon, O. Nieto, D. Liang, and D. P. W. Ellis. mir_eval: A transparent implementation of common MIR metrics. In *ISMIR*, 2014.

[6] A. Ycart and E. Benetos. Polyphonic music sequence transduction with meter-constrained LSTM networks. In *ICASSP*, 2018.


[3] `http://c4dm.eecs.qmul.ac.uk/ycart/a-maps.html`