**TELECOM
ParisTech**

MAPS - A piano database for multipitch estimation and automatic transcription of music

MAPS - Base de données de sons de piano pour l'estimation de fréquences fondamentales multiples et la transcription automatique de la musique

Valentin Emiya
Nancy Bertin
Bertrand David
Roland Badeau

2010D017

Juillet 2010

MAPS

-

A piano database for multipitch estimation and automatic transcription of music[*]

Valentin Emiya[†], Nancy Bertin, Bertrand David[‡], Roland Badeau

http://www.tsi.telecom-paristech.fr/aao/en/category/database/

July 2010

MAPS - A piano database for multipitch estimation and automatic transcription of music

Abstract

MAPS – standing for MIDI Aligned Piano Sounds – is a database of MIDI-annotated piano recordings. MAPS has been designed in order to be released in the music information retrieval research community, especially for the development and the evaluation of algorithms for single-pitch or multipitch estimation and automatic transcription of music. It is composed by isolated notes, random-pitch chords, usual musical chords and pieces of music. The database provides a large amount of sounds obtained in various recording conditions.

Keywords: Audio, database, piano, pitch, multipitch, transcription, music, MAPS

MAPS - Base de données de sons de piano pour l'estimation de fréquences fondamentales multiples et la transcription automatique de la musique

Résumé:

MAPS (*MIDI Aligned Piano Sounds*) est une base de données de sons de pianos enregistrés et annotés sous format MIDI. MAPS a été conçue pour la recherche d'information musicale et a vocation à être utilisée dans la communauté de chercheurs associée. Elle est tout particulièrement appropriée pour le développement et l'évaluation d'algorithmes d'estimation de fréquences fondamentales simples ou multiples et de transcription automatique de la musique. Elle comporte des enregistrements de notes isolées, d'accords aléatoires, d'accords usuels et de morceaux du répertoire de piano, proposés dans différentes conditions d'enregistrement.

Mots clés: Audio, base de données, piano, fréquence fondamentale, transcription, musique, MAPS.

Contents

1   Introduction

In the field of multipitch estimation (MPE) and automatic transcription of music (ATM), annotated sound databases are needed both to develop and to evaluate the algorithms. Public databases are useful for individual works while private databases are used for contests like MIREX [1]. In the former case addressed here, a number of issues are commonly faced: a little amount of sounds is available due to production, copyright or distribution reasons; the ground truth is often generated *a posteriori*, with some inaccurate or erroneous values of pitch or onset and offset times.

Thus, few databases are currently available (*e.g.* [2, 3, 4]). They are usually made up of isolated tones from various musical instruments and/or musical recordings. Then, when necessary, isolated tones may be added by the user to generate chords to be analyzed. These databases provide a large quantity of sounds and were generally obtained after considerable efforts, but may still suffer from some of the drawbacks previously mentioned. In particular, the annotation process is time-consuming when dealing with numerous events. Several strategies may be adopted: manual annotation of the recordings [5], semi-automatic annotation [6, 7] or entertaining systems [8]. In this work, we use a reverse process in which the ground truth is first created as standard MIDI files and then generated in an automatic way, somehow similar to [7], resulting in a fully-automatic and reliable annotation.

In this documentation, we describe the contents and the generation of the new database called MAPS (standing for MIDI Aligned Piano Sounds). The main features provided in MAPS are described in Section 2. The contents of the database are then detailed in Section 3. In Section 4, the recording devices and processes are explained. Instructions on how to get and cite MAPS are finally given in Sections 5 and 6.

2   Main features of MAPS

MAPS provides recordings with CD quality (16-bit, 44-kHz sampled stereo audio) and the related aligned MIDI files as ground truth[1]. The overall size of the database is about 40GB, *i.e.* about 65 hours of audio recordings. The database is available under a Creative Commons license.

A large amount of sounds and a reliable ground truth are provided thanks to some automatic generation processes, consisting in the audio synthesis from MIDI files. The use of a Disklavier (MIDIfied piano) and of high quality synthesis software based on libraries of samples permitted a satisfying trade-off between the quality of the sounds and the time consumption needed to produce such a quantity of annotated sounds.

In order to favor generalization to many audio scenes, several grand pianos and upright pianos have been played in various recording conditions, including various rooms and close/ambient takes. Table 1 details each of the nine configurations in terms of instrument, recording conditions and code reference. It also specifies the origin of the recording, which may be high quality synthesis software based on sample libraries or a Disklavier. For each of these configurations, similar but not equal contents have been produced and can be stored in one 4.7GB DVD.

The contents of MAPS is divided in four sets, which are detailed in section 3:

- the ISOL set: isolated notes and monophonic excerpts;
- the RAND set: chords with random pitch notes;
- the UCHO set: usual chords from Western music;
- the MUS set: pieces of piano music.

3   Detailed contents

3.1   ISOL: isolated notes and monophonic excerpts

The *ISOL* set specifically provides monophonic excerpts. It thus aims at testing single-pitch estimation algorithms or at training multipitch algorithms when isolated tones are required.

---

[1]In order to make the use of MAPS easy in various contexts, the ground truth is also available as text files, including onset times, offset times and pitches.

Code	Instrument model	Recording conditions	Real instrument or software
StbgTGd2	Hybrid	Software default	The Grand 2 (Steinberg)
AkPnBsdf	Boesendorfer 290 Imperial	church	Akoustik Piano (Native Instruments)
AkPnBcht	Bechstein D 280	concert hall	Akoustik Piano (Native Instruments)
AkPnCGdD	Concert Grand D	studio	Akoustik Piano (Native Instruments)
AkPnStgb	Steingraeber 130 (upright)	jazz club	Akoustik Piano (Native Instruments)
SptkBGAm	Steinway D	"Ambient"	The Black Grand (Sampletekk)
SptkBGCl	Steinway D	"Close"	The Black Grand (Sampletekk)
ENSTDkAm	Yamaha Disklavier Mark III (upright)	"Ambient"	Real piano (Disklavier)
ENSTDkCl	Yamaha Disklavier Mark III (upright)	"Close"	Real piano (Disklavier)

Table 1: MAPS: instruments and recording conditions.

Each sound file is characterized by a playing style ps, by a loudness $i0$, by the use/no use of the sustain pedal s and by the pitch m. The related file is named

$$MAPS\_ISOL\_ps\_i0\_Ss\_Mm\_instrName.wav$$

The playing style ps can be:

- NO: 2-second long notes played normally;

- LG: long notes (the duration varies from 3 seconds for the highest-pitch notes to 20 seconds for the lowest-pitch notes);

- ST: staccato;

- RE: repeated note, faster and faster, from about 1.4 to 13.5 notes per second;

- CHd: chromatic ascending and descending scales, with various note duration indexed by d;

- TRi: trills, faster and faster, up to a half tone ($i = 1$) or to one tone ($i = 2$), from about 2.8 to 32 notes per second.

The loudness $i0$ can be: P (piano), M (mezzo-forte), F (forte). The sustain pedal is pressed in half of the cases, as specified by the binary variable s ($s = 1$ when the pedal is pressed). When it is used (50% of the cases), the pedal is pressed 300ms before the beginning of the sequence and released 300ms after the end[2]. The field $instrName$ is a code defined in Table 1.

Except for chromatic scales, the pitch is coded as a MIDI code $m \in [\![21; 108]\!]$, each note of the piano scale $[\![21; 108]\!]$ being recorded.

3.2 RAND: random chords

The $RAND$ set provides chords composed of randomly-chosen notes. It was designed in order to evaluate the algorithms in an objective way, without any a $priori$ musical knowledge, which is commonly performed in the papers on multipitch estimation. The generation process is:

---

[2]Although the pedal is not commonly pressed before playing a note in a musical context, this way of playing is chosen here in order to separate the sound effects due to the pedal and to the note.

Algorithm 1 RAND-set MIDI-file generation process
---
for each polyphony level x **do**
   for each pitch range $m1$-$m2$ **do**
      for a number of outcomes indexed by n **do**
         randomly choose x notes in the pitch range $m1$-$m2$
         randomly and individually choose their loudness in the range $i1$-$i2$
         randomly choose the chord duration and the use/no use of the sustain pedal
         generate the resulting MIDI file
      end for
   end for
end for

Each chord is stored in a file named

$$MAPS\_RAND\_Px\_Mm1\text{-}m2\_Ii1\text{-}i2\_Ss\_nn\_instrName.wav,$$

where

- the polyphony level x varies from 2 to 7;

- the pitch range $m1$-$m2$ can be $21 - 108$ or $36 - 95$; the former range is the full, $7^1/_4$-octave piano range while the latter spreads over the centered 5 octaves and is commonly used to evaluate multipitch algorithms;

- the loudness is chosen, independently for each note, in two possible ranges: $60 - 68$ (*mezzo-forte*, which may represent a typical chord situation with similar note intensities) or $32 - 96$ (from *piano* to *forte*, which may reflect the polyphonic contents when several tracks/melodic lines are played, resulting in heterogeneous loudnesses);

- s denotes the use/no use of the sustain pedal, as in the ISOL set (see section 3.1);

- n denotes the outcome index;

For a given configuration of the parameters, 50 outcomes are actually generated. For instance, the database provides 50 random 3-notes chords for which pitches are chosen between 36 (C_2) and 95 (B_6), with a *mezzo-forte* loudness, around half of the chords being played using the sustain pedal.

3.3 UCHO: usual chords

The UCHO set provides usual chords from Western music such as jazz or classical music. Thus, these chords are useful to assess the performances with an *a priori* knowledge and are made with notes that are harmonically related.

The 2-note "chords" are all the intervals from 1 to 12 semitones, plus the 13$^{\text{th}}$ (fifth at the upper octave) and the 16$^{\text{th}}$ (two octaves), as detailed in Table 2. In polyphony 3, the database provides major, minor, diminished and augmented triads. The seven usual 7$^{\text{th}}$ chords are available in polyphony 4, while the ten usual 9$^{\text{th}}$ chords are recorded in polyphony 5. In polyphony 3, 4 and 5, all inversions are provided as detailed in Tables 3, 4 and 5 respectively. In a given chord, each note is coded according to the distance in semitones from the root of the chord. For instance, a major triad is coded by $0 - 4 - 7$.

A chord with p notes is stored in a file named

$$MAPS\_UCHO\_Cc_1\text{-}\ldots\text{-}c_p\_Ii1\text{-}i2\_Ss\_nn\_instrName.wav,$$

where

- c_1-$\ldots$-c_p denotes the contents of the chord: for $1 \leq k \leq p$, c_k is an integer related to distance in semitones from the root of the chord and note k.

- $i1$-$i2$ is the pitch range, as in the RAND set (see section 3.2);

- s denotes the use/no use of the sustain pedal, as in the ISOL and RAND sets (see section 3.1);

- n is the outcome index, the root of the chord being randomly and uniformly chosen among the possible notes (*e.g.* between 21 and 101 for the major triad);

- additionally, the chord duration is set to one second.

For a given configuration of the parameters, 10 outcomes with different roots are actually generated and are indexed by $n \in [\![1; 10]\!]$. For chords with 4 notes and more, only 5 outcomes are generated.

Interval		Interval	
minor 2nd	0-1	minor 6th	0-8
major 2nd	0-2	major 6th	0-9
minor 3rd	0-3	major 7th	0-11
major 3rd	0-4	perfect 8^{ve}	0-12
perfect 4th	0-5	perfect 13th	0-19
diminished 5th	0-6	two octaves	0-24
perfect 5th	0-7		

Table 2: Intervals.

Triads	Root position	Inversion 1	Inversion 2
major	0-4-7	0-3-8	0-5-9
minor	0-3-7	0-4-9	0-5-8
diminished	0-3-6	0-3-9	0-6-9
augmented	0-4-8	0-4-8	0-4-8

Table 3: Three-note chords: triads and related codes.

7th chords	Root position	Inversion 1	Inversion 2	Inversion 3
major 7th	0-4-7-11	0-3-7-8	0-4-5-9	0-1-5-8
minor 7th	0-3-7-10	0-4-7-9	0-3-5-8	0-2-5-9
dominant 7th	0-4-7-10	0-3-6-8	0-3-5-9	0-2-6-9
half diminished 7th	0-3-6-10	0-3-7-9	0-4-6-9	0-2-5-8
diminished 7th	0-3-6-9	0-3-6-9	0-3-6-9	0-3-6-9
minor major 7th	0-3-7-11	0-4-8-9	0-4-5-8	0-1-4-8
augmented major 7th	0-4-8-11	0-4-7-8	0-3-4-8	0-1-5-9

Table 4: Four-note chords: tetrads and related codes.

3.4   MUS: pieces of music

The MUS set provides pieces of music generated from standard MIDI files available on the Internet[3] under a Creative Commons license. These high quality files have been carefully hand-written in order to obtain a kind of musical interpretation as a MIDI file. The note location, duration and loudness have thus been adjusted by hand by the creator of the MIDI database. About 238 pieces of classical and traditional music were actually available when MAPS was created.

For each set of recording conditions (*i.e.* each line in Table 1), 30 pieces of music are randomly chosen and recorded. The database thus provides a number of different musical pieces, some of them being available several times in various recording conditions.

Each file is named using a description of the musical piece as

$$MAPS\_MUS\_description\_instrName.wav$$

---
[3]B. Krueger, "Classical Piano MIDI files", http://www.piano-midi.de.

9$^{\text{th}}$ chords	Root position	Inversion 1	Inversion 2	Inversion 3	Inversion 4
dominant 7$^{\text{th}}$ and major 9$^{\text{th}}$	0-4-7-10-14	0-3-6-8-10	0-3-5-7-9	0-2-4-6-9	0-2-5-8-10
dominant 7$^{\text{th}}$ and minor 9$^{\text{th}}$	0-4-7-10-13	0-3-6-8-9	0-3-5-6-9	0-2-3-6-9	0-3-6-9-11
minor 7$^{\text{th}}$ and major 9$^{\text{th}}$	0-3-7-10-14	0-4-7-9-11	0-3-5-7-8	0-2-4-5-9	0-1-5-8-10
minor 7$^{\text{th}}$ and minor 9$^{\text{th}}$	0-3-7-10-13	0-4-7-9-10	0-3-5-6-8	0-2-3-5-9	0-2-6-9-11
half diminished 7$^{\text{th}}$ and minor 9$^{\text{th}}$	0-3-6-10-13	0-3-7-9-10	0-4-6-7-9	0-2-3-5-8	0-2-5-9-11
major 7$^{\text{th}}$ and major 9$^{\text{th}}$	0-4-7-11-14	0-3-7-8-10	0-4-5-7-9	0-1-3-5-8	0-2-5-9-10
major 7$^{\text{th}}$ and augmented 9$^{\text{th}}$	0-4-7-11-15	0-3-7-8-11	0-4-5-8-9	0-1-4-5-8	0-1-4-8-9
diminished 7$^{\text{th}}$ and minor 9$^{\text{th}}$	0-3-6-9-13	0-3-6-9-10	0-3-6-7-9	0-3-4-6-9	0-2-5-8-11
minor major 7$^{\text{th}}$ and major 9$^{\text{th}}$	0-3-7-11-14	0-4-8-9-11	0-4-5-7-8	0-1-3-4-8	0-1-5-9-10
augmented 7$^{\text{th}}$ and major 9$^{\text{th}}$	0-4-8-11-14	0-4-7-8-10	0-3-4-6-8	0-1-3-5-9	0-2-6-9-10

Table 5: Five-note chords and related codes.

4 Recording devices

Two procedures were used to record the database: a software-based sound generation and the recording of a Disklavier piano (see Table 1). In both cases, the MIDI files had been created beforehand and were automatically "performed" by one of the devices.

The software-based generation was performed using three steps:

1. concatenating the numerous MIDI files into a low number of large files;

2. generating the audio using a sequencer (Steinberg's Cubase SX);

3. segmenting the large audio files into individual files related to the original MIDI files[4].

(a) Block diagram.

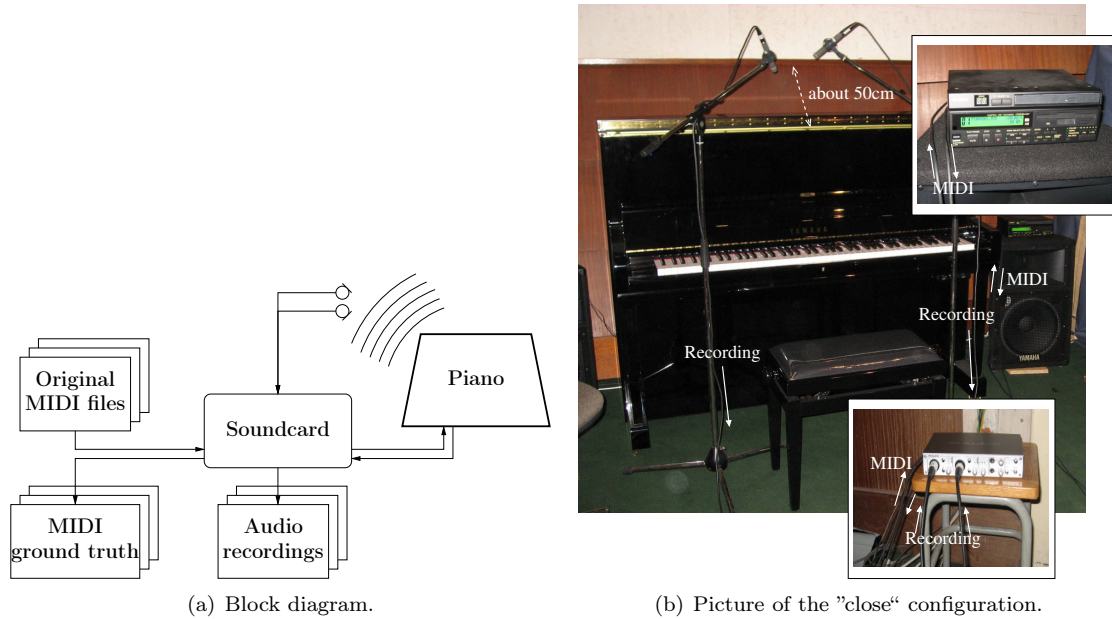(b) Picture of the "close" configuration.

Figure 1: Disklavier recording device: MIDI files are sent from the sound card to the MIDI input of the Disklavier. The generated audio and MIDI signal are recorded using the same sound card.

The Disklavier recording device is illustrated in Figure 1. The room is a studio with a rectangular shape and dimensions equal to about 4×5 meters. It has been designed to perform recordings and its walls are covered with wood and absorbent panels. The distance between the piano and the microphones is about 50cm in the "close" position and about $3-4$m in the "ambient" position. Unlike in the previous

---

[4]This three-step process was performed since the sequencer could not be managed by scripts and thus implied a human action for each MIDI file.

software-based process, the individual MIDI files are here sent one by one from the computer sound card (M-Audio FireWire 410) to the Disklavier via a MIDI link using home-made software. The audio is recorded using two omnidirectional Schoeps microphones and the audio input ports of the same sound card. Since the performance of the Disklavier is improved when a 500ms delay is automatically inserted by the instrument, a MIDI link from the Disklavier to the sound card is set up, which provides the audio-synchronized MIDI files.

5 How to get MAPS?

MAPS is under a Creative Commons License[5] and is freely available. MAPS can be downloaded from:

<div align="center">http://www.tsi.telecom-paristech.fr/aao/en/category/database/</div>

6 How to cite MAPS?

Any use of MAPS should be reported by citing one of the following references:

- 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, (to be published);

- V. Emiya, *Transcription automatique de la musique de piano,* Thèse de doctorat, Telecom Paris-Tech, 2008 (in French).

References

[1] International Music Information Retrieval Systems Evaluation Laboratory, "Multiple fundamental frequency estimation & tracking," in *Music Information Retrieval Evaluation eXchange (MIREX)*, Philadelphia, PA, USA, Sept. 2008.

[2] F. Opolko and J. Wapnick, "Mcgill university master samples," 1987.

[3] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka, "RWC music database: Music genre database and musical instrument sound database," in *Proc. of ISMIR*, Baltimore, MD, USA, Oct. 2003.

[4] The University of Iowa Musical Instrument Samples, "http://theremin.music.uiowa.edu/."

[5] M. Goto, "AIST annotation for the RWC Music Database," in *Proc. of ISMIR*, Victoria, Canada, Oct. 2006.

[6] O. Gillet and G. Richard, "ENST-Drums: an extensive audio-visual database for drum signals processing," in *Proc. of ISMIR*, Victoria, Canada, Oct. 2006.

[7] C. Yeh, N. Bogaards, and A. Roebel, "Synthesized polyphonic music database with verifiable ground truth for multiple f0 estimation," in *Proc. of ISMIR*, Vienna, Austria, Sept. 2007.

[8] D. Turnbull, R. Liu, L. Barrington, and G. Lanckriet, "A game-based approach for collecting semantic annotations of music," in *Proc. of ISMIR*, Vienna, Austria, Sept. 2007.

---

[5]http://creativecommons.org/licenses/by-nc-sa/2.0/