IDMT-SMT-GUITAR

Dataset Description

Authors: Arndt Eppler, Christian Kehling, Andreas Männchen

Semantic Music Technologies Group, Fraunhofer IDMT

Ilmenau, Germany

Contact: jakob.abesser@idmt.fraunhofer.de

January 20, 2016

1 Introduction

1.1 History

- Version 1.0 - February 19, 2014
  - original publication containing datasets 1 and 2
- Version 2.0 - January 20, 2016
  - added datasets 3 and 4
  - fixed minor errors

1.2 Intention

This dataset is intended as a ground truth for the development of diverse analytic algorithms and tools for different music information retrieval tasks. It consists of four subsets that address different application tasks and hence focus on different parameters of musical performances.

1.2.1 Dataset 1

Intended application tasks:

- instrument description
- onset detection
- multi pitch estimation
- string estimation

Note-wise annotated parameters:

- onset
- offset
- pitch
- string/fret
- plucking style
- expression style

1.2.2 Dataset 2

Intended application tasks:

- instrument description
- onset/offset detection
- multi pitch estimation
- partial tracking
- string estimation
- plucking style estimation
- expression style estimation

Note-wise annotated parameters:

- onset
- offset
- pitch
- string/fret
- plucking style
- expression style

1.2.3 Dataset 3

Intended application tasks:

- instrument description
- onset/offset detection
- multi pitch estimation
- partial tracking
- string estimation

Note-wise annotated parameters:

- onset
- offset
- pitch
- string/fret
- plucking style
- expression style

1.2.4 Dataset 4

Intended application tasks:

- genre classification
- tempo detection
- onset detection
- chord recognition
- rhythmic pattern detection
- discrimination of monophony/polyphony

Annotated parameters:

- genre
- tempo
- onsets
- chords
- rhythmic pattern length
- texture (monophonic/polyphonic)

1.3 Content

The recorded material comprises signals from electric and acoustic guitars. Four independent recording sessions by different musicians resulted in the aforementioned four distinct subsets of data. For more information on the various recording setups see section 2.1.
An overview of the structure of the whole database is given in figure 1.1.

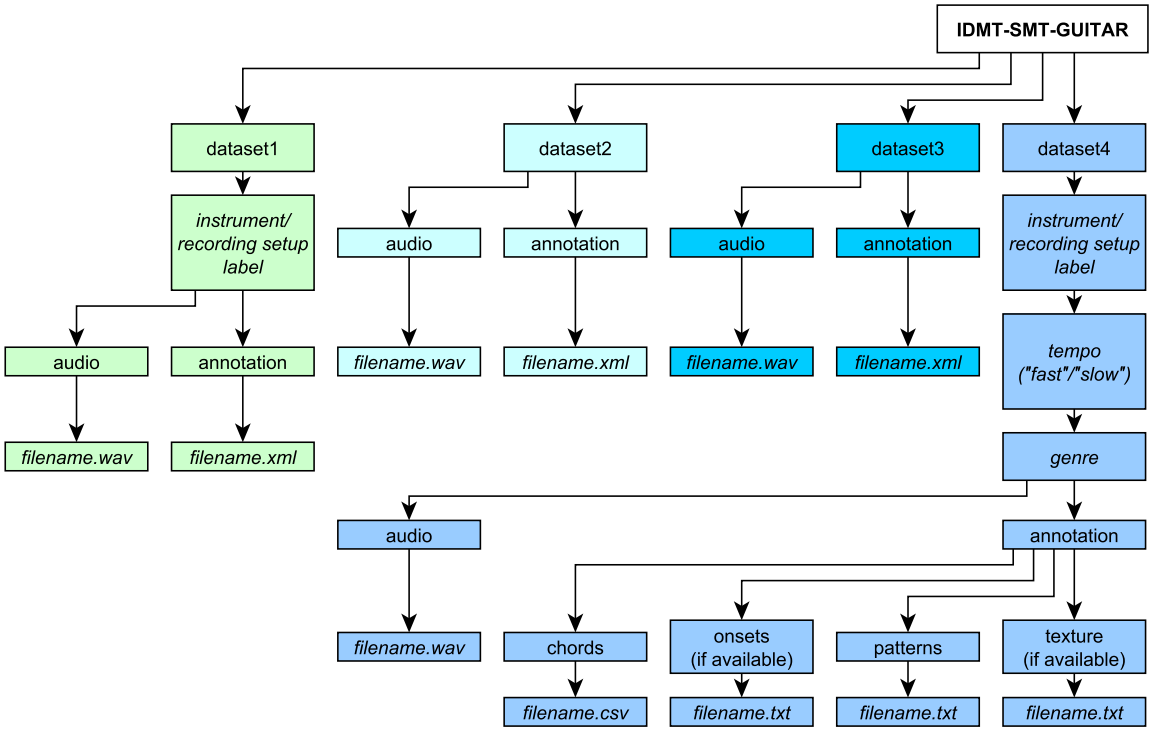

Figure 1.1: Hierarchical visualization of the file structure of the IDMT-SMT-GUITAR dataset.

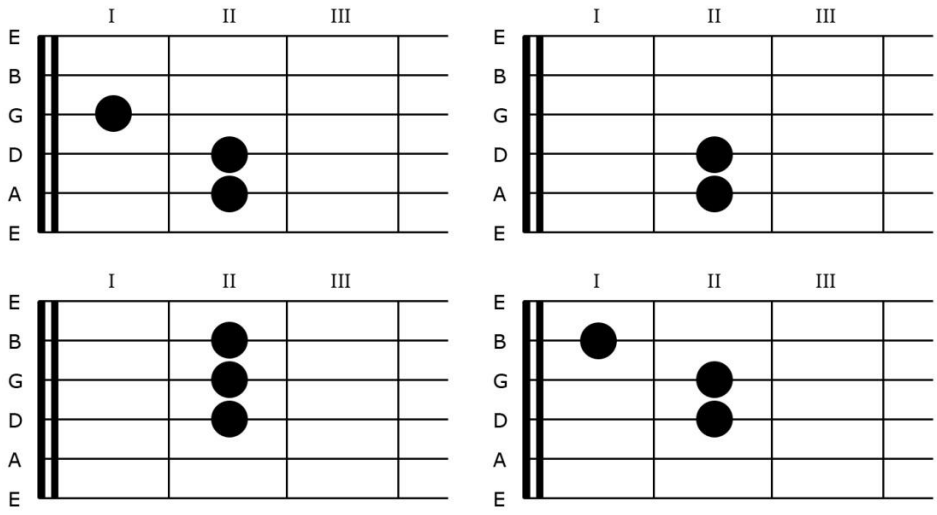

Figure 1.2: Finger positions of the E major chord (top left), the E minor chord (top right), the A major chord (bottom left) and the A minor chord (bottom right).

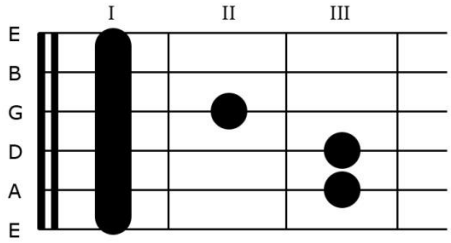

Figure 1.3: Finger Position of the first shift position of the E major type with barré over the first fret.

1.3.1 Dataset 1

The first dataset consists of single note and chord recordings. Each recording contains one beat or pluck. Single notes are played on each string from the 0^{th} fret (empty string) to the 12^{th} fret. In addition, two typical chord fingerings for the A- and the E-chord (see figure 1.2) have been recorded. They are consecutively shifted up by using a barré up to the 10^{th} fret (see figure 1.3). Both minor and major version of each chord are included in the dataset. The parameters listed in section 1.2.1 have been annotated using using XML files.

1.3.2 Dataset 2

The second dataset consists of twelve licks of monophonic and polyphonic parts. Each lick is varied in playing techniques and used guitar. The exact division is shown in figure 1.4. In addition, several initialization files are recorded, which contain prototype versions for each playing technique as well as a plain version of every note on the fretboard for each guitar and each string from fret zero (empty string) to fret 20.

Each parameter introduced in section 1.2.2 is annotated using an XML-based annotation format. Both note onset and note offset values are annotated as absolute time instants in seconds. Pitch information is stored using the according MIDI numbers. The played strings are represented by consecutive numbers starting at the string with the lowest pitch number in empty state (i.e. 1 E2 [MIDI: 40], 2 A2 [MIDI: 45], ..., 6 - E4 [MIDI: 64]).
Three plucking styles are discriminated: picked (PK), finger-Style (FS) and muted (MU). The expression styles are divided in 6 classes: normal (NO), bending (BE), slide (SL), vibrato (VI), natural harmonics (HA) and dead notes (DN). Detailed information about the annotation of each state is given in section 2.2.1.

1.3.3 Dataset 3

The third dataset is made up of five short musical pieces that are either completely monophonic or contain at least some polyphonic note events. All five pieces have been recorded with the same instrument, applying no special expression styles.
Annotations for this dataset use the same format as datasets 1 and 2.

1.3.4 Dataset 4

Dataset 4 consists of 64 short musical pieces grouped into the 8 genres shown in figure 1.5. Each piece has been performed at two different tempi (*slow* and *fast*) with various guitars and recording setups (for details, see section 2.1). Each recording starts with a full bar of silence followed by a control signal. The control signal consists of two bars of dead notes (strings muted with left hand) played on each beat at the tempo of the current piece. The first beat in each bar of the control signal is heavily emphasized.
The audio files are accompanied by a number of different annotations regarding the parameters introduced in section 1.2.4. A detailed description of the annotations for dataset 4 can be found in section 2.2.2. Please note that onset and texture annotations for some audio files are missing.

2 Database

2.1 Recording

- USB audio interfaces used for the recordings

Subset	**Hardware**
Dataset 1	M-Audio Fast Track Ultra Pro
Dataset 2	Tascam US-1641
Dataset 3	M-Audio Fast Track Pro
Dataset 4	Steinberg UR22, M-Audio Fast Track Pro

The audio interfaces were connected to standard PCs running recording software.

	PS **ES**	**PS** **ES**	**PS** **ES**	**PS** **ES** **(Lage)**	**PS** **ES** **(Lage)**	**PS** **ES** **(Lage)**	**PS** **ES**
Lick 1 *monophon*	fingered *no*	picked *no*	palm muted *no*	fingered *no* (1 Lage)	fingered *v, s, dn*	palm muted *v, s, dn*	picked *v, s, dn*
Lick 2 *monophon*				picked *no* (1 Lage)	fingered *v, s, h*	palm muted *v, s, h*	picked *v, s, h*
Lick 3 *monophon*				palm muted *no* (1 Lage)	fingered *s, b, dn*	palm muted *s, b, dn*	picked *s, b, dn*
Lick 4 *monophon*				fingered *no* (2 Lagen)	fingered *s, b, h*	palm muted *s, b, h*	picked *s, b, h*
Lick 5 *monophon*				picked *no* (2 Lagen)	fingered *b, v, dn*	palm muted *b, v, dn*	picked *b, v, dn*
Lick 6 *monophon*				palm muted *no* (2 Lagen)	fingered *b, v, h*	palm muted *b, v, h*	picked *b, v, h*
Lick 7 *polyphon*				fingered *no* (1 Lage)	picked *no* (1 Lage)	palm muted *no* (1 Lage)	
Lick 8 *polyphon*							
Lick 9 *polyphon*				fingered *no* (1 Lage)	picked *no* (1 Lage)	palm muted *no* (1 Lage)	
Lick 10 *polyphon*							
Lick 11 *monophon*				fingered *no, b, v,* *h, s, dn*	palm muted *no, b, v,* *h, s, dn*	picked *no, b, v,* *h, s, dn*	
Lick 12 *polyphon*							

Figure 1.4: PS .. plucking style, ES .. expression style, no .. normal, v .. vibrato, s .. slide, b .. bending, h .. harmonics, dn .. dead notes, x. Lage .. playing the same pitches on the x higher possible fret position.

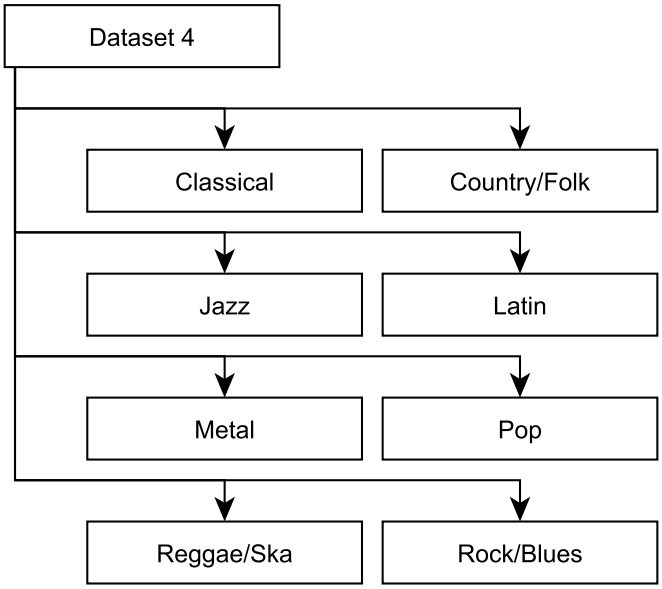

Figure 1.5: Genres found in dataset 4.

- Guitars used for the recordings

Subset	**Instrument(s)**
Dataset 1	Fender Stratocaster, Ibanez RG2820
Dataset 2	Gibson Les Paul, Fender Stratocaster, Aristides 010
Dataset 3	Ibanez RG2820
Dataset 4	Career SG, Ibanez RG2820, Crafter GAE-8/N acoustic guitar

Each guitar was tuned to standard tuning (E2, A2, D3, G3, B3, E4). For the recordings the guitars were mostly plugged directly into the Line-In channel of the audio interface, the only exception being one part of dataset 4 that features the acoustic guitar recorded with a Røde NT1-A condenser microphone.
The datasets are provided in 1 channel RIFF WAVE format with a sample rate of 44100 Hz. The first, third, and fourth set have been recorded with a bit depth of 16 Bit and the second set with a bit depth of 24 Bit.

- Guitarists in each dataset

Subset	**Musician(s)**
Dataset 1	Eppler, Männchen
Dataset 2	Kehling
Dataset 3	Eppler
Dataset 4	Eppler, Männchen

2.2 Annotations

2.2.1 Datasets 1, 2, and 3

The general structure of the XML annotations for datasets 1, 2, and 3 is defined as shown in the following example:

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<instrumentRecording>
  <globalParameter>
    <audioFileName>violin_1.wav</audioFileName>
    <recordingDate>13.12.2005</recordingDate>
    <recordingMusician>John</recordingMusician>
    <instrument>VN</instrument>
    <!-- further "global" parameter here -->
  </globalParameter>
  <transcription>
    <event>
      <pitch>34</pitch>
      <onsetSec>3.44</onsetSec>
      <offsetSec>3.90</offsetSec>
      <excitationStyle>BO</excitationStyle>
      <expressionStyle>VI</expressionStyle>
      <!-- further note parameters here ...>
    </event>
    <event>
      <pitch>38</pitch>
      <onsetSec>4.1</onsetSec>
      <offsetSec>4.55</offsetSec>
      <excitationStyle>PK</excitationStyle>
      <expressionStyle>NO</expressionStyle>
      <!-- further note parameters here ...>
    </event>
    <!-- etc... (further notes) -->
  </transcription>
</instrumentRecording>
```

- List of global parameters:

Parameter	Description	Mandatory
audioFileName	file name of corresponding audio file	yes
instrument	instrument name, see table below	yes

instrumentModel	e.g. "Fender Stratocaster"	no
pickUpSetting	pick-up combination used	no
instrumentTuning	tuning of all open strings given as MIDI pitch values, e.g. 28,33,38,43 (4 string bass guitar)	no
audioFX	audio effects used	no
recordingDate	date of recording	no
recordingArtist	person playing the instrument	no
instrumentBodyMaterial	instrument body material	no
instrumentStringMaterial	instrument string material	no
composer	composer of recorded melody / music piece	no
recordingSource	source of recording	no

- List of note event parameters:

Parameter	**Description**	**Mandatory**
pitch	MIDI pitch value	yes
onsetSec	absolute note onset in seconds	yes
offsetSec	absolute note offset in seconds	yes
fretNumber	fret number where a note was played, starts with 0 for notes played on the open string	no, only relevant for guitar, bass guitar
stringNumber	string number where a note was played, starts with 1 for the lowest string	no, only relevant for guitar, bass guitar
excitationStyle	style which is used to excite the note (commonly referred to as plucking style for string instruments), see table below	no
expressionStyle	expression style which is used after note was excited, see table below	no
loudness	dynamic level expressed in "classical notation" (p, f, mf, ...)	no
modulationFrequencyRange	modulation frequency range in cent (e.g. quarter-tone bending $\rightarrow$ 50 (cent))	no
modulationFrequency	(average) modulation frequency in Hz	no
(instrument)	in case notes from multiple instruments are annotated in one WAV file $\rightarrow$ instrument must be set as note event parameter, if it is not set $\rightarrow$ value of global parameter (instrument) is used	no

The instrument tag entry is unchanged over the whole database:

```
<instrument>EGUI</instrument> // electric guitar
```

- List of excitation styles:

Abbreviation	Style	Explanation
FS	finger-style	
MU	muted	
PK	picked	using a plastic pick

- List of expression styles:

Abbreviation	Style	Explanation
BE	bending	
DN	dead-notes	
FL	flutter	(flute technique)
HA	harmonics	flageolet tones
NO	no expression style	note sounding
SL	(pitch) slide	
ST	staccato	
TR	tremolo	
VI	vibrato	

2.2.2 Dataset 4

A number of different formats is used for the annotations of different musical parameters in dataset 4.
First of all, the genre and global tempo of a piece can be found in the filename, as seen in the following example:

```
latin_1_200BPM.wav
```

Here, the genre is Latin and the tempo is 200 beats per minute. The number after `latin_` is just a running index.

Chords have been annotated for each beat in a musical piece. The chord annotations are stored in CSV files that have the following structure:

```
beat_position,bar_number.beat_number:chord_name/bass_note
.
.
.
beat_position,bar_number.beat_number:chord_name/bass_note
repetition_start_bar.repetition_start_beat
```

```
repetition_end_bar.repetition_end_beat
time_signature_numerator,time_signature_denominator
```

Here, beat_position is the temporal position of the beat within the audio file in seconds, bar_number indicates the current bar number in the piece, and beat_number corresponds to the current beat number within the bar. Whenever there is a chord change (and at the beginning of the piece), beat_number is followed by a colon and chord_name, which indicates the chord symbol of the new chord. Optionally, a base note is added after chord_name. Each chord annotation file ends with three lines that contain the following information: repetition_start_bar and repetition_start_beat mark the beginning bar and beat of the first occurrence of the repeated chord sequence in the musical piece. repetition_end_bar and repetition_end_beat mark the final bar and beat of that same chord sequence. Finally, time_signature_numerator and time_signature_denominator indicate the time signature of the musical piece.

Onset information is stored in a number of different formats: Sonic Visualiser project file (.svl), Sonic Visualiser event layer (.sv), CSV and TXT. In both CSV and TXT files, the first value in each line corresponds to the onset position in seconds.

The rhythmic pattern annotations are stored as TXT files and simply contain the dominant pattern length in bars (1, 2 or 4) of the corresponding audio file.

Texture information has been annotated for each onset in a musical piece and is stored in TXT files. Each line in the annotation files follows this format: The first value indicates whether the onset has a monophonic (1) or polyphonic (2) texture, and the second value is the onset position in seconds.

Additional Information

The data sets have been created for research purposes as part of the following five student theses written and supervised at the Technical University of Ilmenau, Germany in cooperation with the Fraunhofer Institute of Digital Media Technology:

Title: Entwicklung eines Verfahrens zur Audiorestauration basierend auf Re-Synthese von Gitarrenaufnahmen
Author: Arndt Eppler
Year: 2013

Title: Entwicklung eines echtzeitfähigen Verfahrens zur automatischen Saitenerkennung in monophonen und polyphonen Gitarrenaufnahmen
Author: Andreas Männchen
Year: 2013

Title: Entwicklung eines parametrischen Instrumentencoders basierend auf Analyse und Re-Synthese von Gitarrenaufnahmen
Author: Christian Kehling
Year: 2013

Title: Entwicklung und Implementierung eines Verfahrens zur automatischen und echtzeitfähigen Erkennung von wiederholten rhythmischen Patterns sowie der rhythmischen Stilistik von Gitarrensignalen
Author: Arndt Eppler
Year: 2015

Title: Entwicklung und Implementierung eines Verfahrens zur automatischen und echtzeitfähigen Erkennung von Akkorden sowie wiederholten Harmoniefolgen in Gitarrensignalen
Author: Andreas Männchen
Year: 2015

Parts of these theses have been published in:

Title: REAL-TIME GUITAR STRING DETECTION FOR MUSIC EDUCATION SOFTWARE
Authors: Christian Dittmar, Andreas Männchen and Jakob Abeßer
Conference: 14th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)
Year: 2013

Title: AUTOMATIC TABLATURE TRANSCRIPTION OF ELECTRIC GUITAR RECORDINGS BY ESTIMATION OF SCORE- AND INSTRUMENT-RELATED PARAMETERS
Authors: Christian Kehling, Jakob Abeßer, Christian Dittmar, Gerald Schuller

Conference: Proc. of the 17th Int. Conference on Digital Audio Effects (DAFx-14), Erlangen, Germany, September 1-5, 2014
Year: 2014