Automatic Music Transcription with User Interaction

Christian Dittmar, Jakob Abeßer

Fraunhofer Institute for Digital Media Technology, Ilmenau, Germany, {dmr, abesjb}@idmt.fraunhofer.de

Introduction

This publication describes a software toolbox bundling different algorithms for automatic extraction of symbolic note information from digitized music excerpts. In the field of Music Information Retrieval (MIR), this process is often referred to as automatic transcription. Seemingly a research object of purely scientific interest, the automatic transcription of music potentially provides various fruitful application scenarios, such as novel music search paradigms and enhanced possibilities for creative communities evolving around music production, karaoke games and so on. However, automatic transcription is a very challenging task, both from the signal-processing as well as from the musicology point of view. There still exist no definitive solutions for problems like mimicking the human perception or decision making between ambiguous note candidates. Therefore, this toolbox uses specialized algorithms for the detection and classification of drum notes, bass notes, main melody notes and chord structures from real-world music in conjunction with possibilities for intervention of a human user.

Problem Definition

A symbolic transcription representation is an organized sequence of consecutive notes and rests. A note is characterized by its pitch (note name), starting time (onset) and ending time (offset). This notation has to be automatically extracted from digitized excerpts of real world music, where usually a complex mixture of harmonic sustained (e.g. melody) instruments as well as the percussive un-pitched (e.g. drum) instruments is present. Automatic transcription in its original meaning also implies the need for the extraction of the musical context, e.g. the tempo, the bar measure, repetitions etc. However, the MIR community has adopted the term transcription for the process of detecting and classifying notes. For the special case of drum instrument classification this term will also be used because drum patterns are usually notated on the different staves of a score sheet. An overview of related work on automatic music transcription can be found in [1]. Since an in-depth description of the single transcription stages is not feasible in this paper, only the most important facts will be given and further reading will be referred to [1].

Transcription Toolbox

The transcription toolbox has a graphical user interface (GUI) encapsulating four different transcription technologies as well as appropriate methods to display and manipulate the results (i.e. notes in a piano-roll view). It also features a very basic synthesizer for rendering the detected notes and drum instruments in sync with the original music excerpt. The GUI enables the user to load an audio file (WAV or MP3), to select an excerpt for analysis, to run all four transcription methods and to view and listen to the results. Additionally, the user is able to intervene and adjust the transcription results if they are unsatisfactory and save all extracted note information to a MIDI file. Per default all audio data is internally converted to 44.1 kHz sampling rate and 16 bit per sample on loading. In addition, a pre-analysis is conducted on loading. This analysis comprises the deduction of a beat grid hypothesis as well as an estimation of the most likely root key. During this program step, several global parameters like tempo, bar measure and key are detected for later application. Furthermore, a beat grid is identified as the temporal base for a metronome accompaniment to the audio file.

Drum Transcription

The drum transcription algorithm [2] is optimized towards popular music where only a limited set of percussive un-pitched instruments is presumed to be present. The algorithm described in this paper is able to identify up to 17 distinct drum and percussion instruments in real-world music and to generate a MIDI notation of their onsets. The analysis process mainly consists of an algebraic decomposition of so-called onset-spectra in conjunction with a subsequent classification of drum spectrograms.

Bass Transcription

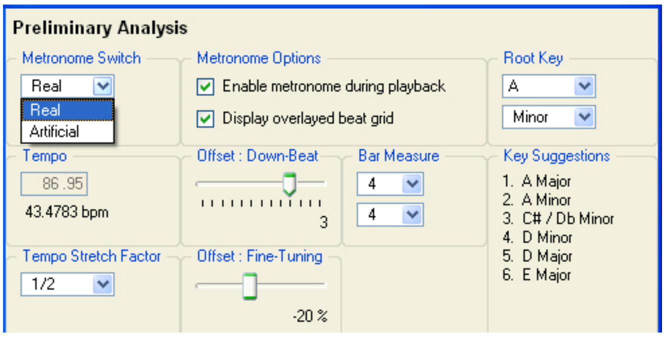
The bass transcription is the fastest extraction procedure of the toolbox, because it only concentrates on the low-frequency musical events and thus allows for a rather strong sub-sampling of the audio signal, discarding irrelevant high-frequency content. The process [3] consists of an onset candidate detection stage followed by pitch modeling. The pitch candidates are extracted from focused areas, where the influence of low frequency drum sounds is minimal. Based on a note quantization, the preliminary onset times are either validated or discarded.

Melody Transcription

The melody transcription algorithm [4] is specialized to identify the notes of preferably monophonic instrument phrases in polyphonic and multi-timbral music excerpts. An efficient implementation of a Multi-Resolution Fast Fourier Transform (MRFFT) is used as time-frequency transform. Subsequently, the most prominent sinusoidal components are identified and subjected to pitch estimation. The pitch candidates together with timbral descriptors are merged to note objects in the end.

Chord Transcription

The extraction of the harmonic structure [5] comprises the detection of as many chords as possible in a piece. That includes the characterization of chords with a key and type as well as a chronological sequence with onset and duration of the chords. The current implementation is able to identify the five most common triads (major, minor, diminished, augmented, suspended) as well as 10 more exotic quads (that augment the triads with sevenths and ninths). A so-called Warped FFT serves as front end time-frequency transform. Chord candidates are found in between consecutive borders of a beat grid. An elaborate evaluation scheme chooses the most probable chord sequence based on criteria such as fitness to the root-key and reliability of the chords.



Preliminary Analysis

Metronome Switch: Real (Real / Artificial)

Metronome Options:
☑ Enable metronome during playback
☑ Display overlayed beat grid

Root Key: A, Minor

Tempo: 86 .95
43.4783 bpm

Offset : Down-Beat: 3

Bar Measure: 4 / 4

Key Suggestions:
1. A Major
2. A Minor
3. C# / Db Minor
4. D Minor
5. D Major
6. E Major

Tempo Stretch Factor: 1/2

Offset : Fine-Tuning: -20 %

Abbildung 1: **Detail of the Expert Mode GUI**

User Interaction

In the majority of cases, the users of the toolbox have a certain pre-knowledge concerning the musical structure and content of the audio file that shall be transcribed. Consequentially, the software offers some ways to correct the transcription results obtained by the automatic algorithms. These functionalities related to user interaction are bundled inside an additional GUI-element, the so-called **Expert Mode** window (shown in figure 1).

Preliminary Analysis

The section called **Preliminary Analysis** refers to the above mentioned pre-processing step during audio file loading. The **Real** metronome is based on the detected beat grid that generally has non-equidistant beat onset times due to temporal fluctuations within real-world music. In contrast, the **Artificial** metronome features equidistant time points. Both can be manipulated and fine-tuned with regard to tempo, global offset shift as well as beat and bar structure. During the preliminary analysis of the audio file, the six most probable keys are estimated and are displayed as key suggestions. The user can select the root key to influence the optional note correction that is described later on.

Transcription Results

There are four additional tabs directly related to the four different transcription algorithms. In the tab called **Drum Transcription**, all detected drum instruments are listed. If a wrong instrument is detected, the user can manually correct the instrument for a certain drum track from a given instrument list. Another useful feature for all four transcription types is denoted **Snap Notes To Rhythmical Grid**. It performs a temporal quantization of the detected notes based on the metronome beat grid and is again offered as an option. Besides, up to four different quantization grid values can be chosen, each one between 4 and 64. For bass and melody transcription, the quantization is conducted under the constraint of keeping the notes monophonic. A note-wise pitch correction is given by the option **Only Notes in Root Key**. It is available for the bass, melody and harmony transcription. Depending on the selected root key, all detected note pitches that are outside the root key are corrected towards the nearest adjacent pitch inside the key. The direction of the note correction is achieved by investigating the pitch of the previous note. This is done to maintain a melodically fluency within the transcribed melodies. This option applies a note-wise pitch correction towards adjacent notes that are contained in the root key. For the harmony transcription the corrected tones of a chord are furthermore aspired to still have a third distance towards each other if possible. This is mainly because regular structured chords with three or four voices are assumed and therefore a third distance between adjacent chord tones. By applying to the option **Harmony Bass**, only the lowest chord notes are visible and audible transposed two octaves below in the main window. If the analyzed audio file does not contain a bass track, this option can offer a useful alternative.

Conclusions

In this publication, a software toolbox comprising automatic transcription methods for the important music aspects drums, bass, melody and chords has been described. Existing concepts for user interaction with the analysis results have been presented. The additional user interaction allows the achievement of comprehensive and satisfying transcription results even in difficult cases. Future work will include the opportunity to manually change the onset, offset and pitch of individual notes by a sequencer-like operability. Furthermore the user will be able to adjust parameters of the algorithms and reapply them to achieve better transcription results.

Literatur

[1] Dittmar, C., Dressler, K., Rosenbauer K.: A Toolbox for Automatic Transcription of Polyphonic Music. Proceedings of Audio Mostly, 2nd Conference on Interaction with Sound (2007), 58-65

[2] Dittmar, C.: Drum Detection from polyphonic Audio via detailed Analysis of the Time Frequency Domain. Proceeding of the 7th Int. Conf. Music Information Retrieval (2005)

[3] Korn, T.: Untersuchung verschiedener Verfahren zur Transkription von Basslinien aus dem Audiosignal. Diploma Thesis Technical University Ilmenau (2005)

[4] Dressler, K.: An Auditory Streaming Approach on Melody Extraction. Proceeding of the 7th Int. Conf. Music Information Retrieval (2005)

[5] Rosenbauer, K.: Entwicklung eines Verfahrens zur Analyse der Harmoniestruktur von Musikstücken unter Einbeziehung von Rhythmus- und Taktmerkmalen. Diploma Thesis Technical University Ilmenau (2006)