Music Information Retrieval Meets Music Education

Christian Dittmar[1], Estefanía Cano[2], Jakob Abeßer[3], and Sascha Grollmisch[4]

1 **Semantic Music Technologies Group, Fraunhofer IDMT**
98693 Ilmenau, Germany
dmr@idmt.fraunhofer.de

2 cano@idmt.fraunhofer.de

3 abr@idmt.fraunhofer.de

4 goh@idmt.fraunhofer.de

Abstract

This paper addresses the use of Music Information Retrieval (MIR) techniques in music education and their integration in learning software. A general overview of systems that are either commercially available or in research stage is presented. Furthermore, three well-known MIR methods used in music learning systems and their state-of-the-art are described: music transcription, solo and accompaniment track creation, and generation of performance instructions. As a representative example of a music learning system developed within the MIR community, the Songs2See software is outlined. Finally, challenges and directions for future research are described.

1998 ACM Subject Classification H.5.5 Sound and Music Computing, J.5 Arts and Humanities–Music, H.5.1 Multimedia Information Systems, I.5 Pattern Recognition

Keywords and phrases Music learning, music transcription, source separation, performance feedback

Digital Object Identifier 10.4230/DFU.xxx.yyy.p

1 Introduction

The rapid development of music technology in the past decades has dramatically changed the way people interact with music today. The way people enjoy and relate to music has changed due to the enormous flexibility given by digital music formats, the huge amount of available information, the numerous platforms for searching, sharing, and recommending music, and the powerful tools for mixing and editing audio.

Consequently, the potential of applying such technologies to music education was recognized. An automatic system that could potentially give instructions and feedback in terms of rhythm, pitch, intonation, expression, and other musical aspects could become a very powerful teaching and learning tool. However, in the early years between the 1980s and the early 2000s, automatic methods for pitch detection, music transcription, and sound separation among other methods, were still in very preliminary stages. Consequently, initial systems for music education, even though innovative and creative, had many restrictions and mainly relied on the possibilities offered by recording studios. In the late 1980s, play-along CDs became popular and offered a number of specially recorded tracks where the user could play with the provided accompaniment. Furthermore, instructional videos were recorded, which mainly featured famous musicians that offered some guidelines in terms of performance and practice. Later on, and mainly aiming for entertainment and not explicitly for music

Multimodal Music Processing, ISBN.
Editors: Meinard Müller, Masataka Goto, and Markus Schedl; pp. 1–25
DAGSTUHL FOLLOW-UPS Dagstuhl Publishing
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Germany

education, the video game community approached music with rhythm games that required the user to follow and repeat patterns of fingering gestures on special hardware controllers. The first commercial music rhythm game dates back to 1996 [1]. Even though these systems were not specifically created as educational tools, they were and still are particularly successful in creating interest in music performance and thus play an educational role.

Interactive applications with a more formal approach to music education have been created such as web services and software tools that guide students through different musical topics like music history or musical instruments. These systems mainly find use in music schools and universities as part of their class work and usually present a set of predefined lectures or practices that the students need to complete.

Recent developments in portable devices like smart-phones and tablets resulted in higher processing power, more powerful audio processing features, and more appealing visuals. As a result, the app market has had an immense growth and everyday more music-related applications are available for both the Android and iOS market. The MIR community has had its share in the development of pitch detection, audio recommendation, and audio identification algorithms necessary for such applications.

The usage of music technology in music education is an ongoing process: on the one hand it completely relies on the accomplishments of the scientific community; on the other hand, it is a process that requires a progressive change of mentality in a community where many processes and techniques still remain very traditional. The development of new music education systems faces many challenges: (1) Development of music technologies robust and efficient enough to be delivered to the final user. (2) Bridging the gap between two communities—music education and music technology—that have completely different environments and mentalities. (3) Design of appealing and entertaining systems capable of creating interest while developing real musical skills.

The remainder of this paper is organized as follows: Section 2 describes some relevant systems for music education, Section 3 presents three Music Information Retrieval (MIR) methods applied in music education applications, Section 4 describes Songs2See—a current music education system developed within the MIR community. Finally, Section 5 discusses future challenges faced by the music information retrieval community and Section 6 draws some conclusions.

2 Related Systems for Music Education

This section gives a general overview of music education systems that are either commercially available or representative of the state-of-the-art in the MIR community. In general, music education systems can be broadly classified in three categories: play along CDs and instructional videos, video games, and software for music education.

2.1 Published Music Education Material

Starting in the 1980s, play-along CDs and instructional videos became popular as an alternative way to practice an instrument. Play-along CDs consist of specially recorded versions of popular musical pieces, where the user plays along to the recorded accompaniment. The main advantage of these systems is that users can practice with their own musical instrument: any progress is directly achieved by real instrumental practice. Furthermore, these systems allow users to get familiar with accompaniment parts—for example, piano or orchestra accompaniments—and as such, they became a popular tool for concert and contest preparation. On the other hand, the amount of available content is limited by the particularly high production costs of such systems: in many cases, large ensembles and long recording sessions are needed for the production of one track. In this sense, play-along CDs are mainly available for very popular songs and for some representative concerts of the instrumental repertoire. Music Minus One[1], for example, offers a large catalog of play-along CDs for different instruments, ensembles, and genres. The Hal Leonard Corporation[2] has published a series of jazz play-alongs for different instruments, with compilations of music of different artists, jazz standards, and thematic editions. Another very popular series of play-alongs is the jazz series published by Jamey Aebersold[3] with a catalog of over a 100 items featuring different artists, playing techniques, and jazz standards.

Instructional videos came out as an educational tool where renowned musicians addressed particular topics—playing techniques, improvisation, warm-up exercises—and offered hints and instructions to help users to improve their skills and to achieve a certain goal. For theses cases, the popularity of a certain musician, as opposed to the popularity of a musical piece, was used as a marketing tool. The idea that you could play like famous musicians do, was very appealing. With time, the catalog of instructional videos grew both in size and diversity, featuring not only famous musicians, but also different playing techniques, learning methods, and the very famous self-teaching videos. The popular VHS tapes from the 1980s and 1990s where slowly replaced by digital formats like the VCD and DVD. Alfred Music Publishing[4], Berklee Press[5], Icons of Rock[6] and Homespun[7] all offer a series of instructional videos for different instruments and styles.

The main weakness of both play-along CDs and instructional videos is that there is no direct feedback for the user in terms of performance evaluation. Users have to completely rely on their own perception and assessment, which in case of beginners, can be a real challenge. However, these types of learning material have played a very important role as they offer an alternative way to practice at home, helping to keep the motivation for learning, and offering the flexibility of practicing on your own time, pace, and schedule.

2.2 Music Video Games

The 1990s was the decade where the development of music rhythm games[8] had a solid start, leading to the great popularity of music games in the next decade. The 1996 release and popularity gain of the game PaRappa the Rapper for Sony PlayStation 1 was an important propeller of the music game development[9]. Later examples of popular releases in the music video game community are Guitar Hero[10], Rock Band[11], and the karaoke game SingStar[12]. Guitar Hero has been

[1] Music Minus One: `http://www.musicminusone.com`

[2] Hal Leonard Corporation: `http://www.halleonard.com/promo/promo.do?promotion=590001&subsiteid=1`

[3] Jamey Aebersold: `http://www.jazzbooks.com/jazz/category/AEBPLA`

[4] Alfred Music Publishing: `http://www.alfred.com/Browse/Formats/DVD.aspx`

[5] Berklee Pree: `http://www.berkleepress.com/catalog/product-type-browse?product_type_id=10`

[6] Icons of Rock: `http://www.livetojam.com/ltjstore/index.php5?app=ccp0&ns=splash`

[7] Homespun: `http://www.homespuntapes.com/home.html`

[8] Type of music video games that challenge a player's sense of rhythm. They usually require the user to press a sequence of buttons shown on the screen.

[9] PaRappa the Rapper: `http://www.gamestop.com/psp/games/parappa-the-rapper/65476`

[10] Guitar Hero: `http://www.guitarhero.com`

[11] Rock Band: `http://www.rockband.com`

[12] Singstar: `http://www.singstar.com`

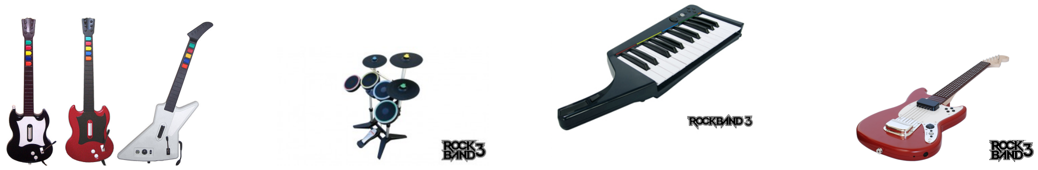
released for different video game consoles like Microsoft Xbox 360, Sony PS3, and also for Windows PCs. The series started as a pure rhythm guitar music game in which the user had to press the correct button at the right time, requiring rapid reflexes and good hand-to-eye coordination [25]. The controller resembles a real guitar but instead of strings and frets, it has five buttons and a strum bar.

(a) Guitar Hero **(b)** Singstar **(c)** Rock Band 3 **(d)** Rocksmith

Figure 1 Music Rhythm Games

Rock Band 3 has been released for Microsoft Xbox 360, Nintendo Wii, Sony PS3, and Nintendo DS. It supports up to three singers with a three-part harmony recognition feature. It was released with a set of 83 songs and has full compatibility with all Rock Band peripherals as well as most Guitar Hero instruments.

One important characteristic of the above mentioned rhythm games is that, while being entertaining and successful in creating interest in music performance, they often fail to develop musical skills that can be directly transfered to real musical instruments as game controllers cannot really capture complexities and intricacies of musical instruments.



(a) Guitar Hero guitars **(b)** Rock Band 3 drum set **(c)** Rock Band 3 keyboard **(d)** Rock Band 3 guitar

Figure 2 Music game controllers

SingStar was released for Sony PlayStation 2 & 3. Like conventional karaoke games, it offers the possibility to sing along to the included songs with the lyrics shown synchronously. Additionally, the singing input is automatically rated by the game, which requires the original vocal tracks to be transcribed beforehand by the producers of the game.

The first commercial release in the video game community that allows users to play with real guitars is Rocksmith[13], released in the United States in September 2011 for Microsoft Xbox 360, Windows, and Sony PS3. The system allows users to connect their guitar output via USB interface. The user's performance is rated based on analysis of the audio signal. Like other music games, Rocksmith delivers a set of songs specifically edited for the game. As an additional feature, Rocksmith offers a series of mini-games with scales, finger dexterity,

[13] Rocksmith: `http://rocksmith.ubi.com/rocksmith/en-US/home/`

or chord exercises for developing playing skills. This represents a major leap in the music game community as no special hardware controllers are needed and, instead of following button sequences, the user actually plays a real guitar with the fret and string information provided in the game. Furthermore, the inclusion of additional exercise-games paves the way from mere gaming and entertainment to music education.

A common limitation of the music games mentioned above is that content is entirely limited to a set of songs delivered with each game. Even if a great effort is made to deliver popular songs appealing to a wide audience, personal tastes can never be completely covered. Furthermore, content is in general limited to pop and rock releases, ignoring the large amount of other musical genres and styles.

2.3 Music Education Software

In terms of commercial systems for music education, Music Delta[14], Smart Music[15] and GarageBand[16] present interactive alternatives to music learning. Music Delta is a web based system developed by Grieg Music Education comprising music curricula, content articles, and interactive tools. There are two versions available: (1) Music Delta Master, an online textbook which offers different performance stages and a special tool for composing and remixing. (2) Music Delta planet, specially designed for elementary school children where topics as music history and composers are presented in an entertaining way.

SmartMusic is a Windows and Mac software developed by MakeMusic especially for bands, orchestras, and vocals. Users can play their instruments to the computer microphone and receive immediate feedback from the software. One of the biggest strengths of the system is that teachers can assign tasks for the students to practice at home. The student's progress can be tracked and personal rating system can be generated. Currently, there are around 2000 musical pieces available for the software.

GarageBand is a software released by Apple for Mac and iPad. Among many other features, it provides the possibility to learn how to play piano and guitar with specially designed content, performance feedback, and appealing user interfaces. Users can play directly to the computer microphone or through USB connection.

With a slightly different approach, Songle[17] offers a web service for active music listening. Users can select a song from the list or register to include an audio file via URL. The system uses MIR techniques to analyze the audio file and then displays information regarding melody line, chords, structural segments, and beat grid. As errors are expected in the automatic analysis, users can edit, correct, and include missing information [23]. The main idea behind this system is to allow users to have a deeper understanding of music and enrich the listening experience.

2.4 Music-Related Mobile Apps

As mentioned in Section 1, the development of apps for smartphones and tablets has grown very rapidly in the last years. Many music-related applications are already on the market, some of them dealing with music learning and playing. Rock Prodigy[18] is a guitar playing

[14] Music Delta: `http://www.musicdelta.com`
[15] Smart Music: `http://www.smartmusic.com`
[16] Garage Band: `http://www.apple.com/ilife/garageband/`
[17] Songle: `http://songle.jp/`
[18] Rock Prodigy: `http://www.rockprodigy.com/`

app developed for the iPad, iPhone, and iPod Touch. Users can play their guitar directly to microphone and, based on a lesson plan and a rating system, receive performance feedback from the application. The lesson plan offers chords, rhythm, scales, technique, and theory exercises. There are also popular songs available for purchase that can be played within the app. Tonara[19] is an interactive sheet music iPad app where users can download and view music directly on their iPad. The app records input directly from the microphone and automatically detects the user's position in the score as the user plays. The system can be used with any musical instrument but currently only violin, piano, flute, and cello scores are available for purchase. Wild Chords[20] is a music game developed by Ovelin and designed to help beginners familiarize with guitar chords. It is available as an iPad app and uses appealing visuals and references to animals to help users identify the chords. The game records audio input directly from the microphone and no hardware controllers are needed.

2.5 Research Projects

In the past years a few research projects have dealt with the development of E-learning systems for music education. The IMUTUS [21] (Interactive Music Tuition System), the VEMUS [22](Virtual European Music School), and the i-Maestro [23] (Interactive Multimedia Environment for Technology Enhanced Music Education and Creative Collaborative Composition and Performance), were all European based projects partially funded by the European Commission that addressed music education from an interactive point of view. IMUTUS focused on the recorder with the goal developing a practice environment where students could perform and get immediate feedback from their renditions. VEMUS was proposed as a follow up project of IMUTUS and addressed the inclusion of further musical instruments and the development of tools for self-practicing, music teaching, and remote learning. i-Maestro focused on the violin family and besides offering enhanced and collaborative practice tools, the project also addresses gesture and posture analysis based on audio visual systems and sensors attached to the performer's body.

Music Plus One [10] is a system for musical accompaniment developed in the attempt to make computer accompaniments more aesthetic and perceptually pleasing. It was developed in the School of Informatics & Computing in Indiana University. The idea behind the system is that a computer-driven orchestra listens, learns, and follows the soloist's expression and timing. The system is composed of three main blocks: (1) Listen: based on a Hidden Markov Model, this stage performs real-time score matching by identifying note onsets. (2) Play: generates an audio output by phase vocoding a pre-existing audio recording (3) Predict: predicts future timing by using a Kalman filter-like model.

Antescofo [11] is both a score-following system and a language for musical composition developed by the Music Representation Research Group at IRCAM. It allows automatic recognition of the player's position and tempo in a musical score. Antescofo can be used in interactive accompaniment scenarios and as a practicing tool. It can also be used as a composition tool by synchronizing and programming electronic events and computer generated sounds with instrumental performances. It also serves as a research tool for tempo and performance analysis.

[19] Tonara: http://tonara.com/

[20] Wild Chords: http://www.wildchords.com/

[21] IMUTUS: http://www.exodus.gr/imutus/index.htm

[22] VEMUS: http://www.tehne.ro/projects/vemus_virtual_music_school.html

[23] i-maestro: http://www.i-maestro.org/

Songs2See [9] is a software developed within a project that started in 2010 at the Fraunhofer Institute for Digital Media Technology (IDMT)[24]. The main goal was to apply state-of-the-art MIR techniques in the development of a music education and learning tool. It takes advantage of pitch extraction, music transcription, and sound separation technologies to allow the user to play with real musical instruments—guitar, bass, piano, saxophone, trumpet, flute and voice. The system returns immediate performance feedback based on a rating system, and gives the user flexibility in terms of the content that can be used with the application—the user can load audio files to create content for the game. In Section 4, a thorough description of Songs2See [25] is presented.

A slightly different approach is taken in the project KoMus[26], that started in 2011 at the IDMT. This project deals with measurement of competencies in music. For this matter, a systematic methodology and a proprietary software solution to assign and control music tasks is developed. The outcomes of this project are targeted to German secondary school students. An important aspect is the (semi-)automatic scoring procedure for evaluating different singing and rhythm tasks. By employing MIR methods, the attempt is made to model ratings given by human experts through regression methods.

3 MIR Methods for Music Learning Applications

3.1 Music Transcription

Music transcription refers to the process of automatically extracting parameters such as pitch, onset, duration, and loudness of the notes played by any instrument within a recorded piece of music. Furthermore, rhythmic and tonal content provided by the beat grid and the musical key, are also of importance when transferring note sequences into common music notation. We refer to these parameters as *score parameters* since they generally do not make assumptions on the particular instrument that is notated.

The automatic transcription of a music piece is a very demanding task since it embraces many different analysis steps such as instrument recognition, multiple fundamental frequency analysis, and rhythmic analysis [32]. Depending on the type of musical instrument to be transcribed, music transcription algorithms are often associated with melody transcription, polyphonic transcription, drum transcription, or bass transcription [13]. The challenges in automatically transcribing music pieces are diverse. First, music recordings usually consist of multiple sound sources overlapping constructively or destructively in time and frequency. Mutual dependencies between the different sources exist due to rhythm and tonality. Furthermore, the number of sound sources is in general unknown and not easily extracted and consequently, often needs to be given as a parameter to the algorithms. Second, all sound sources have very diverse spectral and temporal characteristics that strongly depend on the instrument type and the applied playing techniques (see also Section 3.1.1). Finally, different instruments can be associated with different functional groups such as the main melody or the bass line.

Transcribing the main melody is the most popular task due to the various applications in music education or karaoke systems. If multiple melodies are played simultaneously, different perceptual criteria need to be considered by the transcription algorithm in order to identify the main melody. A selection of existing transcription algorithms are thoroughly described

[24] Fraunhofer IDMT: `http://www.idmt.fraunhofer.de/en.html`
[25] Songs2See: `http://www.songs2see.com`
[26] KoMus: `http://www.idmt.fraunhofer.de/de/projekte/laufende_projekte/komus.html`

in [22], [5], and [4]. In general, state-of-the art automatic music transcription algorithms consist of the following parts:

- **Time-frequency representation:** In order to separately analyze frequency components of different instruments, a suitable time-frequency representation needs to be computed. Commonly used techniques are the Short-time Fourier Transform (STFT), the Multi-Resolution FFT (MR-FFT) [14], the Constant-Q Transform [6], or the resonator time frequency image (RTFI) [56].
- **Spectral decomposition:** Often based on harmonic templates, techniques such as Non-Negative Matrix Factorization (NMF) [35] or Probabilistic Latent Component Analysis (PLCA) [48] are used to decompose the time-frequency representation into the contributions of different (harmonic) instruments. Spectral decomposition yields one or multiple fundamental frequency estimates for each time frame.
- **Onset detection & note tracking:** Probabilistic models such as Hidden-Markov Models (HMM) [42] are applied to model the temporal progression of notes and to estimate their onset and offset time. Based on the frame-wise fundamental frequency estimates, the pitch can be extracted for each note event.

The Music Information Retrieval Evaluation eXchange (MIREX[27]) contest offers several transcription-related tasks such as "Audio Melody Extraction", "Multiple Fundamental Frequency Estimation & Tracking", and "Audio Onset Detection". In this annual contest, various algorithms based on signal processing techniques are evaluated and compared.

In the context of music education, automatic music transcription is an indispensable tool for the automatic generation of music exercises from arbitrary recordings. Music transcription applications allow to detect playing errors in real-time. Thus, the musical performance can be evaluated immediately. By using these applications, music students are not restricted to take lessens in the environment of a music school anymore, Instead, they can use automatic learning tools always and everywhere, which increases their motivation and enhances their musical experience.

3.1.1 Extraction of Instrument-Specific Parameters

In contrast to the *score parameters* discussed in the previous section, there are other parameters that describe performance aspects on a specific instrument. These *instrument-specific* parameters provide cues about the applied playing techniques such as finger-style guitar play or slap-style bass guitar play. They can also describe different techniques such as vibrato or string bending used by the musician as expressive gestures during the performance. These techniques alter the fundamental frequency of a note in a characteristic way and can be parametrized with different levels of granularity, depending on the context of application [2]. Some studies focus solely on estimating *instrument-specific* parameters from audio recordings whereas other studies use these parameters to improve music synthesis algorithms. In this section, four selected studies are briefly discussed that focus on the clarinet, the classical guitar, and the bass guitar.

Sterling et al. [49] presented a physical modelling synthesis algorithm that incorporates two typical performance gestures for clarinet synthesis—tonguing and the pitch bend. Pitch bending allows the musician to alter the fundamental frequency within a range of about a

[27] http://www.music-ir.org/mirex/wiki/2011:Main_Page

semitone. Tonguing relates to the note articulation and controls the onset and offsets while maintaining a constant blowing pressure.

In [39], Özaslan and Arcus focused on two expression styles performed on the classical guitar—the legato technique, which corresponds to the hammer-on and pull-off techniques, and the glissando technique, which corresponds to the slide technique. Both techniques result in an ascending or descending pitch after the note was initially plucked. The authors use a high frequency content (HFC) measure to detect the pluck onset with its percussive characteristics. Between note onsets, the YIN fundamental frequency estimation algorithm [12] was used to characterize the pitch progression. The note release part was segmented into five segments to analyze whether a continuous or an abrupt change of the fundamental frequency appears.

Laurson et al. [34] presented a method for synthesizing the use of rasgueado technique on the classical guitar, which is a rhythmically complex strumming technique primarily used in flamenco music. The technique is characterized by fast consecutive note plucks with the finger nails of all five fingers of the plucking hand and an upwards and downwards movement of the plucking hand. The authors extract signal characteristics—timing and amplitude of individual note attacks—from real-world recordings and use it to re-synthesize the recording.

Abeßer et al. [3] presented a feature-based approach for automatically estimating the plucking style and expression style from isolated bass guitar notes. The authors used a taxonomy of ten playing techniques and described several audio features that capture specific characteristics of the different techniques. The parameter estimation is interpreted as a classification problem.

3.1.2 Spatial Transcription Using Different Modalities

In addition to the score parameters and instrument-specific parameters discussed in the previous sections, two more aspects need to be considered in order to get a better semantic description of a performed instrument track. Firstly, when music is performed on a string instrument such as the bass or the guitar, a given set of notes can usually be played on different positions on the instrument neck. Due to the tuning of the strings and the available number of frets, the assignment between note pitch and *fretboard position*—fret number and string number—is ambiguous. Second, different fingers of the playing hand can be used in order to play a note within a fixed fretboard position. Consequently, various *fingerings* are possible to play the same set of notes. This fact holds true also for other instruments such as the piano and the trumpet. The process of finding the optimal fingering is discussed in Section 3.3.2.

In order to chose suitable fretboard positions and fingerings, musicians usually try to minimize the amount of physical strain that is necessary to play an instrument. For string instruments, this strain is caused by finger stretching and hand movement across the instrument neck. We refer to *spatial transcription* as the process of estimating the applied fretboard positions. In order to automatically estimate the fretboard positions and the fingering from a given musical recording, the sole focus on the audio analysis is often not sufficient. This is mainly due to the fact that the change of fingering barely alters the sonic properties of the recorded signal. In the following sections, we discuss different approaches that include methods from computer vision as a multi-modal extension of audio analysis.

3.1.2.1 Audio-visual Approaches

Already in 2003, Smaragdis and Casey [47] proposed the use of audio-visual data for the extraction of independent objects from multi-modal scenes. They utilized early fusion by means of concatenating audio spectra and corresponding image frames and using Independent Component Analysis (ICA). Although merely a proposal, their paper included an early practical example: onset detection of piano notes by analyzing both a recorded video of the played keyboard and the corresponding audio signal.

Hybryk and Kim [27] proposed a combined audio and video analysis approach to estimate the fretboard position of chords that were played on an acoustic guitar. First, the authors aim at identifying all played chords in terms of their "chord style", i.e., their root note and mode such as E minor. The Specmurt algorithm [46] is used to analyze the spectral distribution of multiple notes with their corresponding harmonics. The outcome is a set of fundamental frequencies that can be associated to different note pitches. The amplitude weights of the harmonic components are optimized for different pitch values. Based on the computed "chord style" (e.g., E minor), the "chord voicing" is estimated by tracking the spatial position of the gripping hand. The chord voicing describes the way the chords are played on the fretboard.

Paleari et. al. [40] presented a method for multimodal music transcription of acoustic guitars. The performing musicians were recorded using two microphones and a digital video camera. In addition to audio analysis, the visual modality was used to track the hand of the guitar player over time and to estimate the fretboard position. The fretboard position is initially detected by analyzing the video signal and then spatially tracked over time. Furthermore, the system detects the position of the playing hand and fuses the information from audio and video analysis to estimate the fretboard position.

3.1.2.2 Visual Approaches

Other approaches solely use computer vision techniques for spatial transcription. Burns and Wanderley [7] presented an algorithm for real-time finger-tracking. They use cameras attached to the guitar in order to get video recordings of the playing hand on the instrument neck. Kerdvibulvech and Saito [30] use a stereo camera setup to record a guitar player. Their system for finger tracking requires the musician to use *colored fingertips*. The main disadvantage of these approaches is that both the attached cameras as well as the colored fingertips are unnatural for the guitar player. Therefore,this may influence the user's expressive gestures and playing style.

3.1.2.3 Enhanced Instruments

A different approach is followed when using *enhanced music instruments* that comprise additional sensors and controllers to directly measure the desired parameters instead of estimating them from an audio or video signal. The major disadvantage of enhanced instruments is that despite of their high accuracy in estimating performance and spatial parameters, they are obtrusive to the musicians and may affect their performance on the instrument [27].

Music game controllers as depicted in Figure 2 were introduced as parts of music games such as Guitar Hero or Rockband. These controllers imitate real instruments in shape and functions and are usually purchased in combination with the corresponding game. However, the controllers often simplify the original musical instruments. The Guitar Hero controller, for instance, reduces the number of available frets on the instrument track from 22 to 4 and furthermore encodes each fret with a different color. The player does not pluck strings but instead presses colored buttons. These simplifications reduce the complexity of performance instructions within the music games and guarantee faster learning success for beginners. The main disadvantage of such controllers is that even though they have a similar instrument shape, their way of playing differs strongly from real instruments. Thus, learning to use the controllers does not necessarily help when learning to play a real instrument.

Hexaphonic pickups allow guitar players and bass guitar players to use their instruments as MIDI input devices, a feature otherwise only available to keyboard players using MIDI keyboards. Since each instrument string is captured individually without any spectral overlap or additional sound sources, a fast and robust pitch detection with very low latency and very high accuracy can be realized using the individual pickup signals as input. This transcription process is usually implemented in an additional hardware device. This way, hexaphonic pickup signals can be converted into MIDI signals nearly in real-time. These MIDI signals allow the musician to intuitively play and control sequencer software, samplers, or virtual instruments in real time.

3.2 Solo & Accompaniment Track Creation

The idea that users can take any recording of their choice—no matter how, where, and when it was created—and obtain solo and accompaniment tracks to play along with, represents a very appealing and flexible approach for practicing music. Whether it be playing along to the Berlin Philharmonic or to the Count Basie Orchestra, all would be possible. Besides being a powerful practicing aid, solo and accompaniment tracks can also be used for performance and musicological studies. We consider sound source separation as a common ground for solo and accompaniment track creation algorithms and further describe it in the next section.

3.2.1 Source Separation

In the context of solo and accompaniment track creation, some separation systems have specifically focused on singing voice extraction from polyphonic audio—the solo instrument is always assumed to be the singing voice. In [36], a system based on classification of vocal/non-vocal sections of the audio file, followed by a pitch detection stage and grouping of the time frequency tiles was proposed. In [45], voice extraction is achieved by main melody transcription and sinusoidal modeling. A system based on pitch detection and non-negative matrix factorization (NMF) is proposed in [52]. Others have focused on the separation of harmonic from percussive components of an audio track [38], [21]. Similarly, a system is proposed in [8] to specifically address the extraction of saxophone parts in classical saxophone recordings and in [18], a score-guided system for left and right hand separation in piano recordings is proposed. More general algorithms have also been proposed for main melody separation regardless of the instrument used: Durrieu [16] proposes a source/filter approach with a two-stage parameter estimation and Wiener filtering based separation. In [33], Lagrange proposes a main melody extraction system based on a graph partitioning strategy—normalized cuts, sinusoidal modeling and computational auditory scene analysis (CASA).

In [9], a system for solo and accompaniment track creation is presented. This algorithm is included in the Songs2See application (see Section 4). The system is composed of five building blocks shown in the diagram in Figure 3. It was designed with the goal of taking audio files from commercial recordings and by means of a pitch detection algorithm and a sound separation scheme, identify the predominant melody in the track, extract the main melody and deliver two independent tracks for the accompaniment and solo instrument. The different processing blocks are briefly explained:

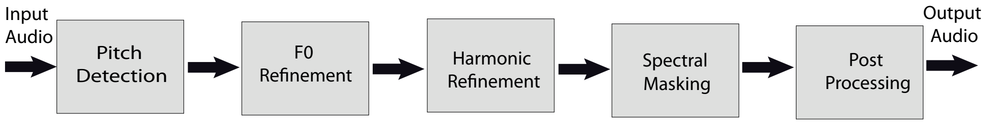
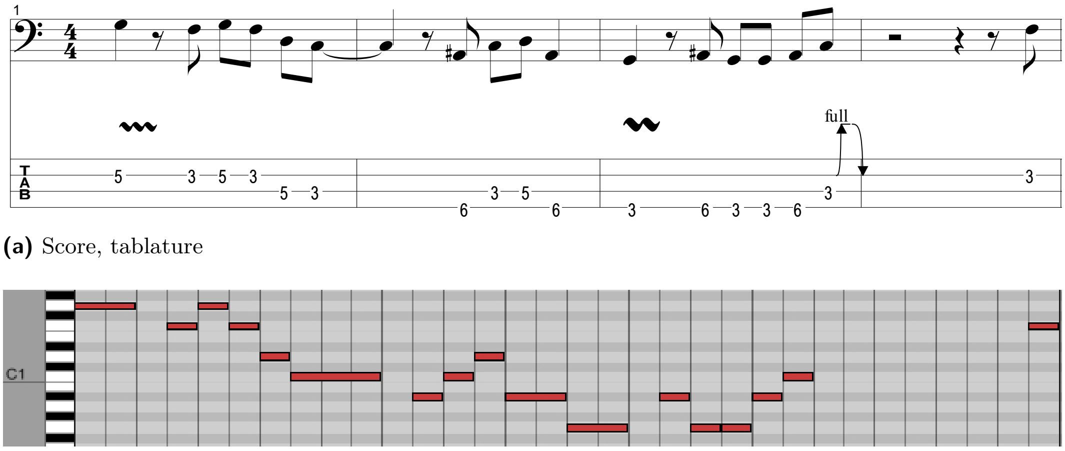


Figure 3 Block diagram of the solo and accompaniment track creation algorithm

- **Pitch Detection:** The pitch detection algorithm proposed in [15] is used, which uses a multi-resolution FFT as a front end. After an initial peak detection stage, pitch candidates are obtained based on a pair-wise evaluation of spectral peaks. Tones are formed based on information of past analysis frames, gathered to analyze long term spectral envelopes, magnitude and pitch information. The main voice is obtained from the pitch candidates using a salience measure.
- **F0 Refinement:** To further improve F0 estimations, a refinement stage is proposed where the magnitude spectrogram is interpolated in a narrow band around each initial F0 value and its constituent harmonics. To obtain a more realistic estimate of the harmonic series, an inharmonicity measure is introduced where harmonic components are not expected to be exact integer multiples of the fundamental frequency but may slightly deviate from the theoretic values.
- **Harmonic Refinement:** The underlying principle at this stage is that the higher the harmonic number of a partial, the more its location will deviate from the calculated harmonic location, i.e., multiple integer of the fundamental frequency. Three main aspects are considered here: (1) Each harmonic component is allowed to have an independent deviation from the calculated harmonic location. (2) Each partial is allowed to deviate from its harmonic location a maximum of one quarter tone. (3) Acoustic differences of string and wind instruments are considered.
- **Spectral Filtering:** After the complete harmonic series has been estimated, an initial binary mask is created where each time-frequency tile is defined either as part of the solo instrument or the accompaniment. To compensate for spectral leakage in the time frequency transform, a tolerance band is defined where not only the specific time frequency tile found in the harmonic refinement stage is filtered out but also the tiles within a band centered at the estimated location.
- **Post Processing:** A final refinement stage is implemented where the different tones are independently processed to remove attacks and possible artifacts caused by inaccurate classification of the time-frequency tiles. Two cases are considered: (1) Due to the particular characteristics of the pitch detection algorithm, a few processing frames are necessary before a valid F0 value is detected. This sometimes causes the pitch detection algorithm to miss the attack frames that belong to each tone. To compensate for the inherent delay in the pitch detection algorithm, a region of 70 ms before the start of each tone is searched for harmonic components that correlate with the harmonic structure of each tone. The binary mask is modified accordingly to include the attack frames found for each tone. (2) Percussion hits are often mistakenly detected as being part of a tone. To reduce the perceptual impact of these inaccuracies, a final analysis of the tone is performed where the harmonic series is analyzed as a whole and transients occurring



(a) Score, tablature

(b) Piano roll

Figure 4 Score, tablature, and piano-roll representation of a bass-line

in several harmonic components simultaneously are detected. For these time-frequency tiles, the spectral mask is weighted and is no longer binary. Finally, the complex valued spectrogram is masked and independent solo and accompaniment tracks are re-synthesized by means of an inverse short term Fourier transform.

3.3 Performance Instructions

3.3.1 Music Representations

In this section, three different symbolic music representations are compared. First, we briefly review the *score* and the *tablature* representations since they are the two most popular written representations of a music pieces. Afterwards, we discuss the *piano-roll* representation, which is often used in music production software, music education applications, as well as in music games. In the three representations each note is described as a temporal event that is characterized by a set of distinct parameters such as note pitch or duration. As an example, a bass line is illustrated as score, tablature, and piano-roll representation in Figure 4.

The score notation is the oldest and most popular written representation of music. It offers a unified and well-established vocabulary for musicians to notate music pieces for different instruments. Furthermore, the score notation provides a compact visualization of the rhythmic, harmonic, and structural properties of a music piece.

The tablature representation, on the other hand, is specialized on the geometry of fretted string instruments such as the guitar or the bass guitar. Each note is visualized according to its fretboard position, i.e., the applied string number and fret number. Due to the instrument construction and tuning of most string instruments, notes with the same pitch can be played in different fretboard positions. The choice of the fretboard position usually depends on the stylistic preferences of the musician and the style of music. Tablatures often include additional performance instructions that indicate the playing techniques for each note. These techniques range from frequency modulation techniques such as vibrato or bending to plucking techniques such as muted play or slap style for the bass guitar. The main advantage of the tablature representation is that it resolves the ambiguity between note pitch and fretboard position. This benefit comes along with several problems: (1) Tablatures are hard to read for musicians who play other instruments such as the piano, trumpet, or saxophone.

(2) Tablatures often do not contain any information about the rhythmic properties of notes. Different note lengths are sometimes encoded in different distances between the characters but this method is often ambiguous. (3) Tablatures, which are nowadays easily accessible from the Internet, are often incomplete or erroneous because they were written by semi-professional musicians. Without the help of a teacher, music students often cannot identify the erroneous parts. Instead, the students might adopt inherent mistakes without even being aware of it.

Finally, the piano-roll representation is often used in music sequencer programs where each note is represented as a rectangle. The rectangle's length encodes the note duration and its horizontal and vertical coordinates encode the note onset and pitch, respectively.

All three representations discussed in this section fail to provide information on micro-tonality or micro-timing of a music piece. These aspects can usually be neglected in basic music education where the main focus of study is more on learning to play melodies than to imitate a particular performance style. Once the student reaches a certain skill level, the precise imitation of a given performance and artistic shaping becomes more important. Here, even slight differences in terms of micro-tonality (intonation) or micro-timing can be of high importance.

3.3.2 Automatic Generation of Fingering Instructions

Fingering instructions indicate which finger or set of fingers of the playing hand are to be used in order to play a particular note on an instrument. Both score notation and tablatures can provide this information by means of a number printed on top of each note. This number encodes the index of the finger that is to be used. However, most often, this information is not given. Since multiple fingerings are usually possible, finding the most convenient fingering is a challenging task. In general, fingering instructions can be generated manually or automatically. Trained music experts can derive optimal fingerings manually based on their experience. Even though this approach leads to proper performance instructions, it is time consuming and inapplicable for a fast analysis of a large number of songs. Furthermore, this manual process clearly stands in contrast to the idea of a complete automatic music transcription system. Therefore, automatic algorithms for fingering generation were developed based on the same criteria that musical experts apply.

In terms of applicable fingerings, musical instruments can be categorized into three different types. Instruments of the first type, such as the piano, have an 1-to-N relationship between note pitch values and possible fingerings. Each note pitch is generated by one distinct key but each key can be pressed by different fingers. Instruments of the second type, such as the saxophone or the trumpet, have an N-to-1 relationship between pitch and fingering. These instruments can produce each pitch with possibly different fingerings but each fingering has a unique finger assignment. For instance, each key on the saxophone is associated to one finger but the same note pitch can be played by using different key combinations. These combinations require a different amount of physical strain both for gripping and blowing the instrument. Similar to the fretboard positions for string instruments discussed in Section 3.1.2, the choice of the fingering depends on the stylistic preferences, performance level, and the musical context within a music piece. The most complex case is the third type of instruments with an N-to-N relationship such as the guitar or the bass guitar. On these instruments, each note can be played at different positions on the instrument neck as well as using different fingers.

Algorithms for automatic fingering generation have to be tailored towards the specific properties of musical instruments, including geometry, tuning, and pitch range. Furthermore, these algorithms have to be scalable to music pieces of different lengths. Usually, a cost value is assigned to each possible fingering in order to automatically derive the optimal fingering.

The process of manually selecting the optimal fingering is influenced by different factors that are part of an underlying cognitive process of the musician [41]. In this study, the authors focused on piano fingerings but discussed several factors that can be applied to other instruments:

- **Biomechanical criteria:** These criteria relate mostly to physical strain, i.e., the necessary effort to play a note on the instrument.
- **Cognitive criteria:** These criteria are often related to the given musical context such as the rhythmic structure of a piece. As an example, strong accents are usually played with stronger fingers.
- **Conventional criteria:** One of these criteria, for example, indicates that musicians prefer using fingering patterns that they already learned over new patterns.
- **Additional criteria:** These criteria comprise musical style, timbre, and intonation and describe how these factors are affected by different fingerings.

Furthermore, the skill level of the musician strongly influences the choice of fingerings. Algorithms for automatic fingering generation need to include these criteria for generating usable results.

Most methods found in the literature focus on the guitar and the piano likely due to their high popularity. For polyphonic guitar fingerings, the presented methods usually distinguish between two different types of costs as in [29], [43], and [44]. Horizontal costs describe the difficulty of changing between different fretboard positions and vertical costs describe the complexity of a chord shape at a fixed position. Kasimi et al. [29] use the Viterbi Algorithm to determine the optimal fingering for a given note sequence. In contrast, Tuohy and Potter [50] follow a two-step approach. First, they use a genetic algorithm to generate a tablature representation. Then, they apply a neural network that was trained based on expert knowledge to derive the optimal fingering. Radisavljevic et al. [44] introduce the "path difference learning" approach, which allows to adapt the cost factors to the individual needs and abilities of the musician.

Presented methods for piano fingerings focus usually on short monophonic melody phrases. Hart et al. [26] assigned cost values to different transitions between white and black keys on the piano. In [41], Parncutt et al. empirically found that pianist in average, read eight notes ahead when playing monophonic melodies. The authors assigned cost values based on the size of intervals and used a set of twelve rules to derive an optimal fingering. Yonebayashi et al. [54] use a Hidden Markov Model to model different fingering positions and apply the Viterbi Algorithm to derive the final fingering.

Similar approaches to obtain optimal fingerings were discussed for the flute in [20] and for the trumpet in [28]. A detailed comparison of the presented methods can be found in [53].

4 Songs2See

Songs2See[28] is an application software for music learning, practice, and gaming which integrates various state-of-the-art MIR algorithms. In developing this system, the following requirements were taken into consideration:

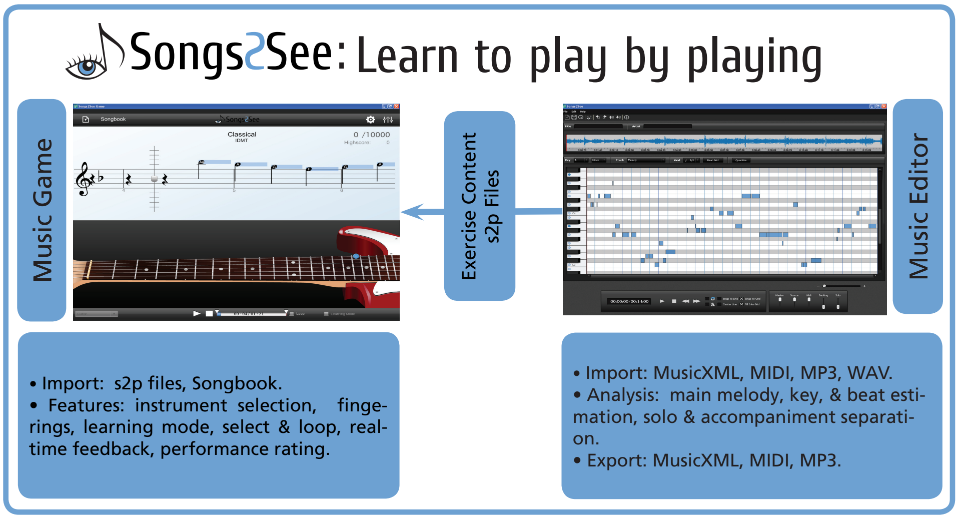
1. The system should allow the use of real musical instruments or the voice without requiring special game controllers.

[28] http://www.songs2see.com

2. Users should be able to create their own musical exercise-content by loading audio files and using the analysis features available in the software.
3. The system should provide the entertainment and engagement of music video games while offering appropriate methods to develop musical skills.

4.1 System Architecture

Songs2See consists of two main applications: the Songs2See Editor, used to create content for the game, and the Songs2See Game, used at practice time. Figure 5 shows a block diagram of the system.



• Import: s2p files, Songbook.
• Features: instrument selection, fingerings, learning mode, select & loop, real-time feedback, performance rating.

• Import: MusicXML, MIDI, MP3, WAV.
• Analysis: main melody, key, & beat estimation, solo & accompaniment separation.
• Export: MusicXML, MIDI, MP3.

Figure 5 Songs2See block diagram

The Songs2See Game is a platform-independent application based on Adobe Flash. The Songs2See Editor is a stand-alone application currently available for Windows PCs. The only additional hardware needed for Songs2See are speakers or headphones and a computer microphone to capture the performances. The standard work-flow in Songs2See is as follows: (1) Choose or create the content to be played: either select a track from the delivered songbook or load an audio file to the Songs2See Editor and use the analysis tools to create content for the game. The Editor sessions can be exported for the Game as .s2p files. (2) Load file in the Songs2See Game. (3) Select the desired instrument from the drop-down menu. (4) Run the Game and start playing. Besides the options already outlined, both the Songs2See Editor and Game offer several processing and performance options that will be further explained in the next sections.

4.2 Songs2See Game

The Songs2See Game is an application where users can practice a selected musical piece on their own musical instrument. There are several features in the Songs2See Game:

- **Score-like display of the main melody:** The Songs2See Game View, shown in Figure 6, combines elements both from standard piano roll views and score notation. The main goal was to include as many musical elements as possible without requiring the user to be able to read music beforehand. The length of the note is displayed both by using music notation symbols—sixteenth notes, eight notes, quarter notes, half notes, whole notes, triplets, and their corresponding rests—and by displaying colored bars of different lengths behind each symbol. Note pitch is displayed both by placing the note objects in the correct position on the staff, and by writing the note names inside the note heads. The clef and key signature are also displayed on this view.

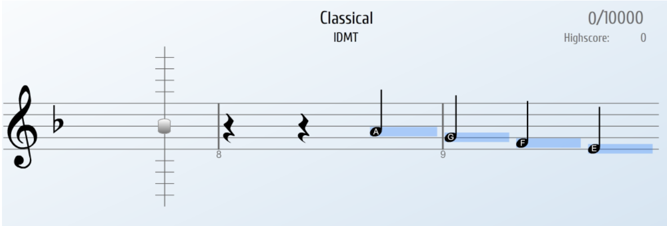
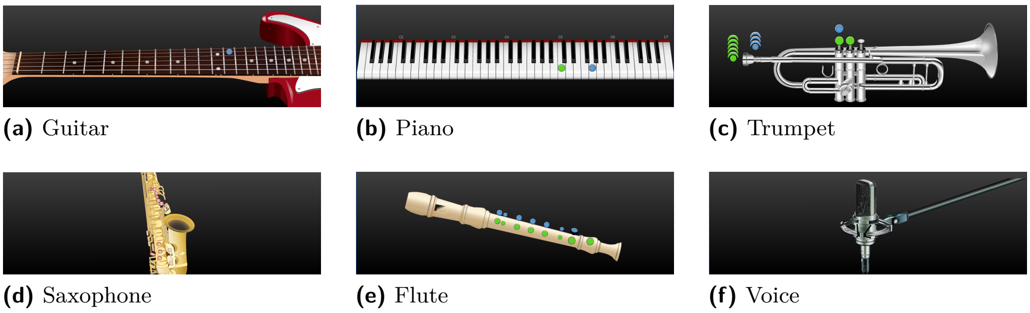


Figure 6 Game View

- **Different instrument support:** The Songs2See Game currently supports bass, guitar, piano, saxophone, trumpet, flute, and singing voice. The user can select any of these instruments from the drop-down menu and an image of the instrument will be shown in the Instrument View. Figure 7 shows the different options for the instrument selection.


(a) Guitar **(b)** Piano **(c)** Trumpet

(d) Saxophone **(e)** Flute **(f)** Voice

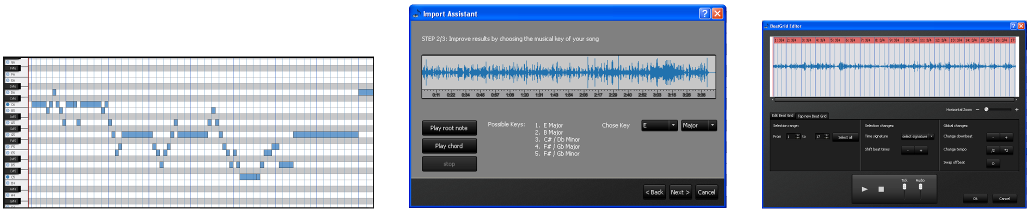
Figure 7 Instruments supported in Songs2See

- **Instrument-specific fingering generation and display:** Every time the user loads a musical piece into the Game, the system automatically generates a fingering animation that describes the most common fingering sequence for the loaded melody. The fingering generation algorithm is instrument-specific and combines several criteria discussed in Sect. 3.3.2 for fingering selection. For instance, in the case of the trumpet, the blowing requirements are also displayed (see Figure 7c). For all instruments, the fingering for the current note in the note sequence is displayed in blue and the next note is displayed in green. This allows the user to prepare in advance for the next note in the melody sequence. In the event that the user wants to play a song on a musical instrument whose register does not allow to play all the notes—some notes might be too high or too low to be played in that particular instrument—a red sign will be displayed over the instrument to warn the user about register problems (see Figure 7d).
- **Real-time performance feedback:** Based on a real-time pitch detection algorithm [24], the user's performance is rated based on the pitch information extracted from the original audio track. When the user hits the note, the note will be painted in green and the user will score points with a maximum of 10000 points per song. When the user plays a wrong note, the error will be displayed and a reference for correction will be given.

- **Selection and looping option:** The user can select particularly difficult segments within the song and practice them separately. There is also a looping option where the selected segment will be played repeatedly.
- **Learning mode:** This option is meant to help users familiarize with the fingerings and performance requirements of a new song. When the Learning Mode is selected, the Game will be halted on each note until the user plays it correctly. This will give the user enough time to check finger positions, pitch, and all other performance details needed in the piece.
- **Songbook and loading options:** The user has two possibilities in terms of loading content into the Game: (1) The Songs2See Game is delivered with a set of songs that can be accessed through the Songbook icon. (2) Users can create their own content using the Songs2See Editor and exporting the session as an s2p file—the proprietary Songs2See format.
- **Other options:** Through the Options Menu, users can access several performance and set-up options: adjust delay between audio and visual to perfectly synchronize the Game with the performance, select the microphone input and adjust the gain, choose language, show or hide note names, enable or disable left hand mode for left-handed guitar players. In the Audio Menu, users can adjust the playback level as well as the balance between the accompaniment and solo tracks.

4.3 Songs2See Editor

The Songs2See Editor is a software component that allows users to create exercise content for the game. Additionally, it offers many general features that can potentially be used for many other applications outside Songs2See . The following options are available:



STEP 2/3: Improve results by choosing the musical key of your song

Possible Keys:
1. E Major
2. B Major
3. C# / Db Minor
4. F# / Gb Major
5. F# / Gb Minor

Chose Key: E, Major

Selection range: From 1 to 17

Time signature: select signature

(a) Piano Roll View **(b)** Key Detection Assistant **(c)** Beat Grid Editor

Figure 8 Songs2See Editor: Piano Roll View, Key Assistant, and Beat Grid Editor.

- **Import options:** One of the most powerful features of the Songs2See Editor is that it allows the user to create material for the Game starting from different types and formats of audio material: WAV, MP3, MIDI, and MusicXML are supported. These four import possibilities make it possible to combine the use of the Songs2See Editor with other powerful processing tools as score-writing and sequencer software.
- **Main melody extraction:** Every audio file imported into the Songs2See Editor is automatically analyzed to extract the main melody. The employed algorithm [15] detects the most salient voice in the track regardless of the instrument played or the type of music. Results are displayed in the form of note objects in a Piano Roll View (see Figure 8a). As errors are expected, the user is allowed to create new notes, delete or merge existing notes, and adjust the length or the pitch of existing note objects. Audible feedback is

provided every time the user creates a new note or adjusts the pitch of an existing one. These options facilitate the process of correcting the melody extraction results.

- **Key detection:** The audio material imported into the Editor is also analyzed to extract key information. By using the Import Wizard (see Figure 8b), the user is presented with the five most probable keys and has the option to play the corresponding chords or root notes to help decision making. The key of the song can be changed at any time. The piano keys in the Piano Roll View that correspond to the notes in the selected key will be marked to guide the user through the editing process. The selected key is also important in terms of notation as this will be the key used when using the MusicXML export options.
- **Tempo and beat-grid extraction:** Tempo and beat extraction as presented in [13] are also performed when an audio file is loaded into the Editor. The automatic beat extraction currently only supports 4/4 time signatures. Considering possible extraction errors and the large amount of musical pieces written in other time signatures, the Songs2See Editor offers a processing tool called the Beat Grid Editor (see Figure 8c), where every beat can be independently modified and shifted. The downbeats can be modified and quick options for doubling or halving the tempo are available. A tapping option is also available, where the user can tap the beats of the song or sections of it.
- **Solo and accompaniment track creation:** The sound separation algorithm described in Section 3.2.1 is used to create independent solo and accompaniment tracks. The results from the main melody extraction are directly used in the separation stage. Therefore, the quality of the separation results is directly dependent on the quality of the extracted melody. The algorithm has been optimized to allow users to correct information in the melody extraction and receive immediate separation results based on the included changes.
- **Export options:** Every Songs2See Editor session can be exported for the Game as an s2p file—the proprietary format of Songs2See . All the necessary information for playback and performance is contained within this file. Furthermore, the intermediate results in the processing chain can also be exported to be used in other applications. For example, the solo and accompaniment tracks created can be exported as MP3 and then be used as learning and practicing material outside the Songs2See environment. The results from the main melody extraction can be exported as MusicXML or MIDI files and be used with score-writing or sequencer software.

5 Future Challenges

As described in the preceding sections, the state-of-the-art of MIR techniques for music education is already quite advanced. However, there are still numerous challenges present that motivate further research into specific directions. Some of them will be discussed in the following sections.

5.1 Polyphonic Music

Despite years of research, the automatic transcription of polyphonic music is still considered the holy grail in the MIR community. Many algorithms have been proposed that address polyphonic music played on monotimbral instruments, such as piano or guitar. The software

Celemony Melodyne[29] incorporates polyphonic transcription and sound separation of such data and is successfully used in music recording studios. During the last years, a multitude of novel signal processing front-ends for multi-pitch estimation have been proposed. Examples are Specmurt analysis [46] and systems as the ones presented in [31],[55],[5]. However, these methods only reveal pitch candidates and do not explicitly consider interference of un-pitched sounds such as drums. The majority of the proposed methods exhibit only straight forward post-processing based on empirical thresholds. One promising research direction is the usage of data-driven post-processing methods to train probabilistic models—such as HMMs. These methods rely on huge amounts of manually annotated audio material which can then be used to denoise the extracted multi-pitch candidates. These kinds of data can be derived from semi-automatic alignment of existing MIDI files to real-world recordings [19].

5.2 Sound Separation

Even though good results can be achieved in experiments under controlled or restricted conditions, most sound separation algorithms still lack robustness in real-world scenarios. A general solution, capable of identifying independent sound sources regardless of their timbre, salience, or recording conditions, has not been found.

With the recent development of perceptual measures for quality assessment [17], the sound separation community made an important step towards the development of structured and meaningful evaluation schemes. However, further research has to be conducted in terms of perceptual quality of extracted sources in order to better characterize algorithm artifacts and source interferences in terms of their perceptual impact. Separation evaluation campaigns as the Data Analysis Competition 05 have also been conducted since 2005[30]. The first separation campaign [51] specifically addressing Stereo Audio Source Separation (SASSEC) was conducted in 2007[31] and the SISEC (Signal Separation Evaluation Campaign) also included an audio separation task in 2011. The LVA/ICA[32] 2012 (International Conference on Latent Variable Analysis and Signal Separation), included a special session on real-world constraints and opportunities in audio source separation. These efforts represent an important step to push current state-of-the-art both in theoretical and practical aspects.

5.3 Improved Parametrization/Transcription

An improved parametrization of music recordings will be beneficial for various applications ranging from sound synthesis, music transcription, as well as music performance analysis. Each musical instrument has a unique set of playing techniques and gestures that constitute the "expressive vocabulary" of a performing musician. These gestures result in various sound events of different timbral characteristics. A detailed parametrization needs to be based on sophisticated models of sound production for each particular instrument. The applicability of these models can be evaluated by applying sound synthesis algorithm that are based on these models in order to re-synthesize the parametrized recordings. The main evaluation question remains: To what extend does a given model re-synthesize detected notes from a given recording while at the same time capturing the most important timbral characteristics of an instrument?

[29] Celemony Melodyne Editor: `http://www.celemony.com/`
[30] Data Analysis Competition 05: `http://mlsp2005.conwiz.dk/index.php@id=30.html`
[31] SASSEC 07: `http://www.irisa.fr/metiss/SASSEC07/`
[32] LVA/ICA 2012: `http://events.ortra.com/lva/`

5.4 Non-Western Music

MIR methods have mainly addressed the characteristics of Western music genres. This is mainly due to the fact that state-of-the-art technologies were not robust enough to propose generalized solutions and constraints had to be placed in most methods to consistently address a particular problem. However, both due to the advances in the MIR community and to the spread of MIR research to Asian, African, and Latin American countries, a very strong interest in addressing other musics has emerged. New methods have to be developed to properly address the often complicated and intricate rhythmic and harmonic characteristics of these music styles. Collaboration between research facilities and experts from the different communities such as MIR and musicology is crucial to overcome current limitations.

5.5 Improvement of Computational Efficiency

One important obstacle for the inclusion of many MIR algorithms in commercial products is computational cost. Processing and time requirements are, in many cases, still very demanding, and even when they grant performance robustness, they also prevent the algorithms from being included in commercial applications. As in any research process, initial stages are always result-oriented. However, an effort has to be made to streamline algorithms to allow robust performance under standard computational capacities and facilitate real-time applications. As an example, under practical considerations, it is often sufficient to replace constant-Q spectra by conventional spectra obtained by Fast Fourier Transform subsequently resampled to a logarithmic frequency axis [37]. Although problematic from a signal theoretic point of view, this computationally efficient approach is often successfully used in music transcription methods.

5.6 Multi-User Applications

An important development direction for current music video games is the inclusion of multi-user applications that not only bring the entertainment and information contained within the game, but also further competition, engagement, and immersion from the interaction with other users. In music in particular, interaction with others is an expected scenario. Musicians rarely play alone and in most cases, they have to learn to interact and communicate with other musicians in order to produce an artistic ensemble performance. However, for multi-user applications to be feasible, algorithm efficiency has to be improved, real-time conditions have to be met, and latency and algorithmic delays reduced to the minimum. Furthermore, in the case of music learning applications, new feedback, rating and instruction approaches have to developed in order to properly assess the interaction and interplay with regard to intonation and timing.

5.7 Music Technology & Music Education

The inclusion of music technologies in both formal and informal music education is still fairly new. However, new generations grow up and live submerged in a digital era where possibilities are endless. This poses an important challenge to the music education community, as in order to reach the new generations, education methods have to evolve correspondingly. Nonetheless, changing mentalities and opening minds to new approaches is never an easy process and even less in a community as traditional as the music education community. This necessarily implies that music technology and music education have to work together to reach a common goal: develop systems for music education that can be flexible, appealing, and suitable for developing real musical skills.

6 Conclusions

A general overview of the use of MIR technologies in music learning applications has been presented. Both the evolution of the community over time and its current state-of-the-art suggest that music education will be rapidly and dramatically influenced by computer-based music technologies in the next years. Systems get more robust and flexible every day, a multitude of platforms is available, and there is a growing interest for pushing forward research in the field. Nonetheless, the community still faces many challenges in terms of future research directions, many of them pointing out to the imminent need for collaboration between different fields and communities. Music technologies swiftly evolve and consequently, the way people interact with music. In the same manner, music education and learning systems have to evolve and take advantage of the many possibilities provided by new technologies.

7 Acknowledgements

The Thuringian Ministry of Economy, Employment and Technology supported this research by granting funds of the European Fund for Regional Development to the project Songs2See, enabling transnational cooperation between Thuringian companies and their partners from other European regions.

References

1 25 Jahre Musikspiele. *M! Games*, 5:24–25, 2009.

2 J. Abeßer, C. Dittmar, and G. Schuller. Automatic recognition and parametrization of frequency modulation techniques in bass guitar recordings. In *Proceedings of the 42nd Audio Engineering Society (AES) Conference: Semantic Audio*, Ilmenau, Germany, 2011.

3 J. Abeßer, H. Lukashevich, and G. Schuller. Feature-based extraction of plucking and expression styles of the electric bass guitar. In *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Dallas, US*, 2010.

4 F. Argenti, P. Nesi, and G. Pantaleo. Automatic transcription of polyphonic music based on the constant-q bispectral analysis. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(6):1610–1630, 2011.

5 E. Benetos and S. Dixon. Joint multi-pitch detection using harmonic envelope estimation for polyphonic music transcription. *Selected Topics in Signal Processing, IEEE Journal of*, 5(6):1111 –1123, oct. 2011.

6 J. C. Brown. Calculation of a constant Q spectral transform. *The Journal of the Acoustical Society of America*, 89(1):425–434, 1991.

7 A.M. Burns and M.M. Wanderley. Visual methods for the retrieval of guitarist fingering. In *Proceedings of the 2006 International Conference on New Interfaces for Musical Expression (NIME06)*, pages 196–199, Paris, France, 2006.

8 E. Cano and C. Cheng. Melody Line Detection and Source Separation in Classical Saxophone Recordings. In *12th International Conference on Digital Audio Effects (DAFx-09)*, pages 1–6, Como, Italy, 2009.

9 E. Cano, C. Dittmar, and S. Grollmisch. Songs2See: Learn to Play by Playing. In *12th International Society for Music Information Retrieval Conference (ISMIR)*, Miami, USA, 2011.

10 R. Christopher. Music Plus One and Machine Learning. In *27th International Conference on Machine Learning*, Haifa, Israel, 2010.

11 A. Cont. ANTESCOFO: Anticipatory Synchronization and Control of Interactive Parameters in Computer Music. In *International Computer Music Conference (ICMC)*, Belfast, Ireland, 2008.

12 A. de Cheveigné and H. Kawahara. YIN, a fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America*, 111(4):1917–1930, 2002.

13 C. Dittmar, K. Dressler, and K. Rosenbauer. A Toolbox for Automatic Transcription of Polyphonic Music. In *2nd Conference on Interaction with Sound- Audio Mostly*, Ilmenau, 2007.

14 K. Dressler. Sinosoidal extraction using an efficient implementation of a multi-resolution fft. In *Proceedings of the 9th International Conference on Digital Audio Effects (DAFx-06), Montreal, Canada*, pages 247–252, 2006.

15 K. Dressler. An Auditory Streaming Approach For Melody Extraction from Polyphonic Music. In *12th International Society for Music Information Retrieval Conference (ISMIR)*, number Ismir, pages 19–24, Miami, USA, 2011.

16 J. L. Durrieu, G. Richard, and B. David. An Iterative Approach to Monaural Musical Mixture De-Soloing. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, number 1, pages 105–108, 2009.

17 V. Emiya, E. Vincent, N. Harlander, and V. Hohmann. Subjective and Objective Quality Assessment of Audio Source Separation. Technical report, Institut National de Recherche en Informatique et en Automatique, Rennes, 2010.

18 S. Ewert and M. Müller. Score-informed voice separation for piano recordings. In *Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR)*, Miami, USA, 2011.

19 S. Ewert, M. Müller, and R. B. Dannenberg. Towards reliable partial music alignments using multiple synchronization strategies. In *Proceedings of the International Workshop on Adaptive Multimedia Retrieval (AMR)*, Madrid, Spain, 2009.

20 R. Fiebrink. Modeling flute fingering difficulty. Technical report, The Ohio State University, 2004.

21 D. Fitzgerald. Harmonic/Percussive Separation Using Median Filtering. In *13th International Conference on Digital Audio Effects (DAFx -10)*, number 1, page 10, 2010.

22 B. Fuentes, R. Badeau, and G. Richaed. Adaptive Harmonic Time-Frequency Decomposition of Audio Using Shift-Invariance PLCA. In *Proc. of the IEEE Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 401–404, 2011.

23 M. Goto, K. Yoshii, H. Fujihara, M. Mauch, and T. Nakano. Songle: A web service for active music listening improved by user contributions. In *Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR)*, Miami, USA, 2011.

24 S. Grollmisch, C. Dittmar, E. Cano, and K. Dressler. Server based pitch detection for web applications. In *AES 41st International Conference: Audio for Games*, London, UK, 2011.

25 S. Grollmisch, C. Dittmar, and G. Gatzsche. Concept , Implementation and Evaluation of an improvisation based music video game. 2009.

26 M. Hart, R. Bosch, and E. Tsai. Finding optimal piano fingerings. *The UMAP (Undergraduate Mathematics and Its Applications) Journal*, 21:167–177, 2000.

27 A. Hrybyk and Y. Kim. Combined audio and video for guitar chord identification. In *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR), Utrecht, Netherlands*, pages 159–164, 2010.

28 D. Huron and J. Berec. Characterizing idiomatic organization in music: A theory and case study of musical affordances. *Empirical Musicology Review*, 4, 2009.

29 A. A. Kasimi, E. Nichols, and C. Raphael. A simple algorithm for automatic generation of polyphonic piano fingerings. In *Proc. of the 8th International Conference on Music Information Retrieval (ISMIR), Vienna, Austria*, 2007.

30 C. Kerdvibulvech and H. Saito. Vision-based guitarist fingering tracking using a bayesian classifier and particle filters. *Advances in Image and Video Technology*, pages 625–638, 2007.

31 A. Klapuri. A method for visualizing the pitch content of polyphonic music signals. In *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR), Kobe, Japan*, 2009.

32 A. Klapuri and M. Davy, editors. *Signal Processing Methods for Music Transcription.* Springer Science+Business Media, 2006.

33 M. Lagrange, L. G. Martins, J. Murdoch, and G. Tzanetakis. Normalized Cuts for Predominant Melodic Source Separation. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(2):278–290, February 2008.

34 M. Laurson, V. Välimäki, and H. Penttinen. Simulating idiomatic playing styles in a classical guitar synthesizer: Rasgueado as a case study. In *Proc. of the 13th Int. Conference on Digital Audio Effects (DAFx-10), Graz, Austria*, 2010.

35 D. D. Lee and H. S. Seung. Algorithms for Non-negative Matrix Factorization. In *Advances in Neural Information Processing Systems 13, Papers from Neural Information Processing Systems (NIPS), Denver, CO, USA*, pages 556–562. MIT Press, 2000.

36 Y. Li and D. Wang. Separation of Singing Voice From Music Accompaniment for Monaural Recordings. *IEEE Transactions on Audio, Speech and Language Processing*, 15(4):1475–1487, May 2007.

37 M. Müller, D. Ellis, A. Klapuri, and G. Richard. Signal processing for music analysis. *IEEE Journal of Selected Topics in Signal Processing*, 5(6):1088 –1110, oct. 2011.

38 N. Ono, K. Miyamoto, J. L. Roux, H. Kameoka, and S. Sagayama. Separation of a Monaural Audio Signal into Harmonic/Percussive Components by Complementary Diffusion on Spectrogram. In *EUSIPCO*, pages 1–4, Lausanne, Switzerland, 2008.

39 T. H. Özaslan and J. L. Arcos. Legato and glissando identification in classical guitar. In *Proc. of Sound and Music Computing Conference (SMC), Barcelona, Spain*, 2011.

40 M. Paleari, B. Huet, A. Schutz, and D. Slock. A multimodal approach to music transcription. In *Proc. of the 15th IEEE International Conference on Image Processing (ICIP)*, pages 93–96, 2008.

41 R. Parncutt, J. A. Sloboda, E. F. Clarke, M. Raekallio, and P. Desain. An ergonomic model of keyboard fingering for melodic fragments. *Music Perception*, 14:341–382, 1997.

42 L. R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, feb 1989.

43 D. P. Radicioni. *Computational Modeling of Fingering in Music Performance.* PhD thesis, Department of Psychology, University of Torino, Italy, 2005.

44 E. Radisavljevic and P. Driessen. Path difference learning for guitar fingering problem. In *Proc. of the International Computer Music Conference (ICMC)*, 2004.

45 M. Ryynänen, T. Virtanen, J. Paulus, and A. Klapuri. Accompaniment Separation and Karaoke Application Based on Automatic Melody Transcription. In *IEEE International Conference Multimedia Expo*, pages 1417–1420, 2008.

46 S. Saito, H. Kameoka, K. Takahashi, T. Nishimoto, and S. Sagayama. Specmurt Analysis of Polyphonic Music Signals. *Audio, Speech, and Language Processing, IEEE Transactions on*, 16(3):639–650, February 2008.

47 P. Smaragdis and M. Casey. Audio/visual independent components. In *Proc. of the 4th International Symposium on Independent Components Analysis and Blind Signal Separation, Nara, Japan*, 2003.

48 P. Smaragdis, B. Raj, and M. Shashanka. A Probabilistic Latent Variable Model for Acoustic Modeling. In *Proc. of the 20th Annual Conference on Neural Information Processing Systems (NIPS)*, 2006.

49 M. Sterling, X. Dong, and M. Bocko. Pitch bends and tonguing articulation in clarinet physical modeling synthesis. In *Proc. of the IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, 2009.

50 D. R. Tuohy and W. D. Potter. Guitar tablature creation with neural networks and distributed genetic search. In *Proc. of the 19th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems, IEA-AIE06, Annecy, France*, 2006.

51 E. Vincent, H. Sawada, P. Bofill, S. Makino, and J. Rosca. First Stereo Audio Source Separation Evaluation Campaign: Data, Algorithms and Results, 2007.

52 T. Virtanen, A. Mesaros, and M. Ryynänen. Combining Pitch-Based Inference and Non-Negative Spectrogram Factorization in Separating Vocals from Polyphonic Music. In *ISCA Tutorial Res. Workshop Statist. pecept. Audition.*, Brisbane, Australia, 2008.

53 D. Wagner. Implementierung und Evaluation einer interaktiven Fingersatz-Animation in Musiklernsoftware. Master's thesis, Ilmenau University of Technology, 2011.

54 Y. Yonebayashi, H. Kameoka, and S. Sagayama. Automatic decision of piano fingering based on hidden markov models. In *Proc. of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2007.

55 K. Yoshii and M. Goto. Infinite latent harmonic allocation: A nonparametric bayesian approach to multipitch analysis. In *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR), Utrecht, Netherlands*, pages 309–314, 2010.

56 R. Zhou and M. Mattavelli. A new time-frequency representation for music signal analysis: Resonator time-frequency image. In *Proc. of the 9th International Symposium on Signal Processing and Its Applications (ISSPA)*, pages 1–4, feb. 2007.