ACOUSTICS AND SIGNAL PROCESSING IN THE DEVELOPMENT OF MUSIC EDUCATION SOFTWARE

Estefanía Cano, Christian Dittmar, and Sascha Grollmisch

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

ABSTRACT

Interactive software applications have proved to be powerful means for music education as they provide the user with real time performance assessment, unlimited orientation and practice time, variable set of user selected features, entertaining ways of displaying content and variable levels of difficulty in a single application. The algorithms behind these tools require in general solid rhythmic, harmonic and melodic processing approaches as well as flexibility in terms of parameter selection and processing time. In the past years, many efforts have been made by the Music Information Retrieval (MIR) community to improve algorithm performance. Even though good results have been obtained in systems for harmonic-percussive separation, singing voice extraction, accompaniment track creation, pitch tracking, rhythm and beat extraction, audio segmentation and chord transcription among others, there is still much room for improvement and exploration. In particular, better understanding of instrument acoustics and models and their implications in signal processing is needed. Many studies have attempted to include instrument dependent spectral information in their algorithms. However, most of them are trained systems that use a set of collected data but that include very limited or no real acoustical models and parameters. In contrast, some few systems have attempted to use instrument specific acoustical models but suffer in general from lack of robustness due to the great variability found in terms of playing styles, setups and performers that necessarily create deviations from the model. It is then important and relevant to explore possibilities of building algorithms that combine the flexibility of trained systems with current signal processing techniques and real structured acoustical information from the treated instruments. Furthermore real testing data, recordings, measurements and design parameters used by acousticians and instrument makers can also be important in terms of generality and flexibility in any built system.

1. INTRODUCTION

In the past years, several efforts have been made to build interactive tools, web-based applications, e-learning solutions and games for music learning. Music games such as RockBand [1] and GuitarHero [2] offer an entertaining approach to play music through specially designed game controllers. The user is presented with different tasks with varying levels of difficulty and a set of songs is available in each game [3]. Similar approaches are offered by singing games as Singstar [4]and Mikestar[5] where the user scores points depending on the accuracy of their performance. Leaning to a more structured music education process, initiatives like Music Delta [6], i-Maestro[7] and SmartMusic [8]offer a wide set of possibilities to enhance and complement education methods. Production tools that allow you to design and personalize music exercises and lessons, cooperative work support that allows groups of musicians or students to collaborate in a single project, servers especially designed to distribute material in schools and music academies, self learning applications that provide audio and gesture feedback to the students and possibilities to learn about the different musical instruments, genres or periods in music history are among the possibilities offered by these initiatives. In general, these applications try to make the process of learning music a more efficient and engaging one.

Our approach in the development of Music Education software tries to take advantage of the flexibility and accessibility of web-based applications, with the entertaining and engaging possibilities of Music Games, while allowing the user to develop real musical skills by using their own musical instrument, getting feedback of real musical elements as rhythm, pitch, tempo and dynamics and by offering wider access to musical content as no restriction is set in terms of the music pieces that can be used. The ability to extract semantic content from any audio recording or music file that the user wants to practice requires specialized signal processing techniques often linked to the Music information Retrieval research field.

Music Information Restrieval was defined in [9] as follows:

> Music Information Retrieval (MIR) is a multidisciplinary research endeavor that strives to develop innovative content-based searching schemes, novel interfaces, and evolving networked delivery mechanisms in an effort to make the world's vast store of music accessible to all.

It is a research effort that includes the fields of computer science, information retrieval, musicology, music theory, audio, signal processing and cognitive science among others. Common applications of MIR research are music indexing and classification, automatic music transcription, audio remixing, restoration and coding, music perception, cognition and emotion, music archives and libraries, music queries, acoustic fingerprinting, metadata and semantic web, human-computer interaction and interfaces [10].

Songs2See [1] is our current project for development of Music Education software which in consequence has working teams in music information retrieval, musicology, education and interface development. The remainder of this paper will only refer to the music information retrieval aspects concerning the Songs2See project.

The paper is organized as follows: Section 2 presents an overview of the project Songs2See, its main research fields and challenges. Section 3 presents some conclusions and future approaches. Sections 4 and 5 present acknowledgements and references.

[1] http://www.idmt.fraunhofer.de/de/projekte_themen/songs2see.html

2. SONGS2SEE

Supported by Thuringian Ministry of Economy, Employment and Technology, the project Songs2See addresses the development of a web-based application that assists music students in their individual practice time. The student has the option to load any audio track or recording for processing and practicing. Commonly used digital audio formats as .mp3, .wav are supported. The possibility of loading music scores in MusicXML [11] format and tracks as midi files is currently being studied. MusicXML is supported by scorewritting softwares as Finale[12] and Sibelius[13] and by music sequencer softwares as Cubase[14]. Using their own musical instruments, the student plays the selected music piece to the computer's microphone. This removes the need for specially designed game controllers and hardware and allows the user to develop real musical skills. Using signal processing techniques and music information retrieval algorithms, the application provides immediate assessment and feedback in terms of dynamics, rhythm and pitch. In order to personalize and make practice sessions more efficient and entertaining the application allows the user to select options as tempo, key, playback loops, difficulty level, instrument and volume. Furthermore, the application provides different symbolic representations of the processing results, i.e., traditional score notation, piano roll view and dynamic animations of fingerings for a certain instrument. Figure 1 shows an image of the prototype application.

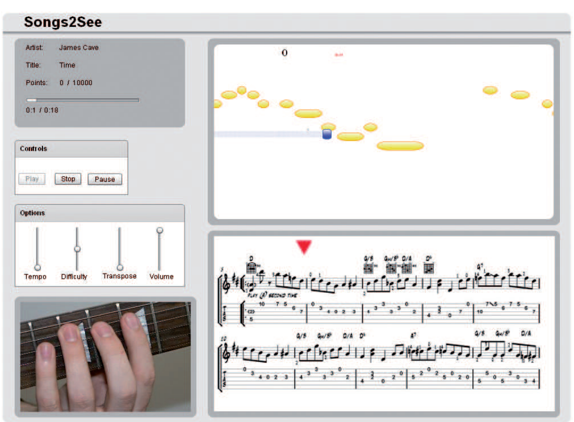

Figure 1: Screenshot of the Songs2See prototype application.

2.1. Music Information Retrieval & Songs2See

In terms of the technical aspects of the application, two main fields of research, both of them developed within the scope of the Music Information Retrieval (MIR) community, are essential: Sound Source Separation and Automatic Music Transcription.

2.1.1. Sound Source Separation

Sound Source Separation deals with the extraction of individual musical signals from an observed mixture by computational means, e.g., obtaining separate audio tracks for the violin and piano from a recording of Beethoven's Kreutzer Sonata. In the scope the Songs2See project, the capability of separating or simplifying audio tracks is desirable as it allows the inclusion of speciel features into the application:

- Isolation of individual instruments, e.g., voice, percussion or soloist, in the track. This can be used for analysis and better understanding of the different voices in a musical piece.
- Removal or attenuation of a certain instrument in the track to create accompaniment or play along versions of real recordings, e.g., remove the soloist from a recording of a certain concerto and obtain a track of the remaining orchestra accompaniment to play along.
- Automatic transcription of individual instruments or musical lines when no musical score is available

Currently, three different sound separation scenarios have been studied and developed for the project: Harmonic-Percussive decomposition, voice extraction and soloist removal for accompaniment track creation.

Harmonic Percussive Decomposition: Being able to separate percussive and harmonic instruments not only allows the creation of percussion tracks for transcription, rhythmic analysis and play along options, but also makes the analysis of musical elements in the harmonic instruments a much simpler task. In general, the characteristics both in the frequency and in the time domain, of percussive and harmonic instruments are too different in nature to be processed with the same types of algorithms. Furthermore, pitch tracking, see Section 2.1.2, time stretching and chord transcription algorithms notoriously decrease performance in the presence of percussive material in the audio signal.

Voice Extraction: Detecting and isolating the singing voice allows the creation of voice tracks for analysis and performance. Given a certain voice track, the algorithm can compare the user's recoded performance with the extracted voice and assess elements as rhythm and intonation. Voice extraction has also been thoroughly used in singing games and karaoke applications.

Soloist Removal: Deals with the detection and removal of soloist instruments from real recordings. Both piano and orchestra accompaniments have been studied and prototype algorithms have already been designed[15]. The main benefit of this feature is that it would give the flexibility to the performer to play along his favorite recording of a certain piece or concerto. The need to transcribe full orchestra scores and the use of midi files to play along would be removed.

2.1.2. Automatic Music Transcription

Automatic Music Transcription deals with the extraction of symbolic representations, e.g., music scores, from real-world music recording. It is the direct connection between the application and the user as both user renditions and application results have to be translated to symbolic representations. Only after renditions and recordings have been transcribed can they be displayed to the user. Automatic music transcription is very closely related to pitch tracking and multi-pitch detection algorithms.

Pitch Tracking: Describes the process of determining the frequency that a human listener would agree to be the same pitch as the input signal[16]. The outcome of such an algorithm is most often a sequence of frequencies played over a time interval. Multipitch tracking algorithms attempt to detect fundamental frequencies played simultaneously by different sound sources. In general, pitch tracking algorithms need to deal with complexities inherent in spectral content and time evolution of musical tones. Attack transients, time-frequency resolution of processing algorithms, overlapping of different sources and acoustical characteristics of the recordings are some of the main difficulties faced by pitch detection algorithms.

2.2. Research Challenges

Some of the biggest challenges faced in the development of this application are:

- In any sound separation task, the proper and accurate detection of the source to be isolated is critical. The spectral and temporal content of a source needs to be determined before any attempt of separation can be made. Overlapping spectral components, mixing conditions of the recording, complexity of the musical line played by the source, frequency and time resolution of the algorithms are a few of the intricacies of the process. Performance of any given separation algorithm is strongly dependant on the quality and accuracy of the source detection.
- All musical instruments have different acoustical characteristics. Building a general, robust and flexible algorithm that would perform under all conditions is a complicated task. Harmonic and percussive instruments are completely different in nature[17]. Dealing with percussive instruments requires fine time resolution and dealing with noise-like signals. Harmonic instruments have slower time evolutions but require finer frequency resolutions to be properly characterized. Furthermore, characteristics of musical instruments are very variable throughout their registers and are further modified by dynamics and style of the performance[18]. Attacks transients in harmonic instruments are difficult to characterize and are often misclassified as belonging to percussive sources. The singing voice also poses a complex mix between harmonic and percussive material.
- Overlapping of spectral information is a given in all separation applications. Musical signals, especially in Western music where a strong harmonic relation between the sources can be expected, will share frequency bands in the spectrum. Separation in the strict sense of the overlapped components is a problem for which a general solution has not yet been proposed. Several estimation techniques have been developed but there is still a lot of room for developments and research. Better understanding of the characteristics, behavior and evolution in time and frequency of musical signals is needed before any real solution can be proposed.
- When it comes to dealing with commercial recordings, a great deal of variability can be found in terms of mixing conditions, acoustic characteristics of the recording, instrumental set-ups and digital audio effects used in the recording. All of these elements influence performance of tracking and separation algorithms and need to be taken into consideration when building such an application.
- Being a web-based application, there are processing time constraints that need to be considered. In this regard, separation and tracking algorithms have to be kept simple and efficient, processing times and delays have to be minimized and memory usage has to be controlled.
- For the design and development stage of all algorithms, appropriate audio signals are needed for testing and evaluation. Even though many music databases are available now[19][20], separation algorithms require both original signals and mixtures to be properly evaluated. Multi-track recordings are often used for these tasks, but availability of such signal collections is scarce. Furthermore, the diversity of characteristics of any given instruments will only be revealed by having very thorough signal collections recorded in different dynamics, set-ups, registers, instruments and performers. Building such datasets is a time consuming and expensive task.
- Better understanding in terms of the acoustic characteristics of musical instruments and their impact on musical signals is needed in order to incorporate such parameters in classification and separation algorithms[21]. Being each instrument such a complicated system, good characterization in terms of frequency and time information and their variability in terms of registers, instruments, performers and setups is needed.

3. CONCLUSIONS

It is our belief that further progress and development can be achieved if joint efforts between the different communities are made. Expertise of the MIR community in terms of signal processing and information retrieval, acoustical knowledge and data from the acoustics field, musical knowledge and feedback from musicians and music schools can be a compelling mix to develop software as the one here described. Our final objective is in benefit of music, trying to develop better and more efficient tools that will lead to better musical experiences for the final user.

4. 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.

5. REFERENCES

[1] RockBand [online]. Available:http://www.rockband.com/

[2] GuitarHero [online]. Available: http://www.guitarhero.com/

[3] S. Grollmisch, C. Dittmar, and G. Gatzsche, "Concept , Implementation and Evaluation of an improvisation based music video game," Proceedings of IEEE Consumer Electronics Society's Games Innovation Conference (IEEE GIC), 2009.

[4] SingStar [online]. Available: http://www.singstargame.com/de-de/

[5] MikeStar [online]. Available: http://www.mikestar.com/

[6] Music Delta. Internet Based Music Education. [online]. Available: http://www.musicdelta.com/main.php

[7] i-maestro. [online]. Available:http://www.i-maestro.org/

[8] SmartMusic. Music learning software for band, orchestra, and voice. [online]. Available: http://www.smartmusic.com/

[9] J.S. Downie, "The Scientific Evaluation of Music Information Retrieval Systems : Foundations and Future," Computer Music Journal, vol. 28, 2004, pp. 12-23.

[10] N. Orio, "Music Retrieval : A Tutorial and Review," Foundations and Trends in Information Retrieval, vol. 1, 2006, pp. 1-90.

[11] MusicXML Definition. [online]. Available: http://www.musicxml.org/xml.html

[12] Finale Music. Music Notation Software. [online]. Available:http://www.finalemusic.com

[13] Sibelus Music. [online]. Available: http://www.sibelius.com

[14] Cubase.[online]. Available: http://www.steinberg.net/en/products/cubase/start.html

[15] E. Cano and C. Cheng, "Melody Line Detection and Source Separation in Classical Saxophone Recordings," 12th international Conference on Digital Audio Effects (DAFx-09), Como, Italy: 2009, pp. 1-6.

[16] C. Roads, The Computer Music Tutorial. Cambridge: The MIT Press, 1996.

[17] 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," EUSIPCO, Lausanne, Switzerland: 2008, pp. 1-4.

[18] J.J. Burred, "From Sparse Models to Timbre Learning : New Methods for Musical Source Separation," 2009, p. 201.

[19] University of Iowa Musical Instrument Samples Database.[online]. Available: http://theremin.music.uiowa.edu/

[20] RWC Music Database: Musical Instrument Sound.[online]. Available: http://staff.aist.go.jp/m.goto/RWC-MDB/rwc-mdb-i.html

[21] M. Bay and J.W. Beauchamp, "Harmonic Source Separation Using Prestored Spectra," 6th International Conference on Independent Component Analyisis and Blind Signal Separation (ICA06), Charleston, USA: 2006, pp. 561 - 568.