

Musical Examination to Bridge Audio Data and Sheet Music

Xunyu Pan, Timothy J. Cross, Liangliang Xiao, and Xiali Hei

Department of Computer Science and Information Technologies
Frostburg State University
Frostburg, Maryland 21532, USA

ABSTRACT

The digitalization of audio is commonly implemented for the purpose of convenient storage and transmission of music and songs in today's digital age. Analyzing digital audio for an insightful look at a specific musical characteristic, however, can be quite challenging for various types of applications. Many existing musical analysis techniques can examine a particular piece of audio data. For example, the frequency of digital sound can be easily read and identified at a specific section in an audio file. Based on this information, we could determine the musical note being played at that instant, but what if you want to see a list of all the notes played in a song? While most existing methods help to provide information about a single piece of the audio data at a time, few of them can analyze the available audio file on a larger scale. The research conducted in this work considers how to further utilize the examination of audio data by storing more information from the original audio file. In practice, we develop a novel musical analysis system Musicians Aid to process musical representation and examination of audio data. Musicians Aid solves the previous problem by storing and analyzing the audio information as it reads it rather than tossing it aside. The system can provide professional musicians with an insightful look at the music they created and advance their understanding of their work. Amateur musicians could also benefit from using it solely for the purpose of obtaining feedback about a song they were attempting to play. By comparing our system's interpretation of traditional sheet music with their own playing, a musician could ensure what they played was correct. More specifically, the system could show them exactly where they went wrong and how to adjust their mistakes. In addition, the application could be extended over the Internet to allow users to play music with one another and then review the audio data they produced. This would be particularly useful for teaching music lessons on the web. The developed system is evaluated with songs played with guitar, keyboard, violin, and other popular musical instruments (primarily electronic or stringed instruments). The Musicians Aid system is successful at both representing and analyzing audio data and it is also powerful in assisting individuals interested in learning and understanding music.

Keywords: Audio Analysis, Musical Characteristic, Sheet Music, Music Information Retrieval

1. INTRODUCTION

The development of modern multimedia technologies makes the digitization of audio and video recordings a convenient solution for storage and transmission of large collections of multimedia data. This advancement helps create a wide range of multimedia applications, from reliable audio/video streaming¹ to digital audio/video forensics,^{2,3} that might not have been possible just a decade ago. Meanwhile, the sheet music largely produced in printed form still plays an important role as the guide for both professional and amateur musicians to study and perform a piece of music. Audio recording and sheet music are widely used as two of the most important forms to represent, store, transmit, and experience music. While sheet music describes music in an abstract manner using musical symbols, audio recording of various digital formats directly produces music which can be enjoyed by people. Focusing on different perspectives, these two music representations provide distinct but related ways for music analysis and understanding.

Some practical musical applications require computing the extent of similarity between audio recording and its corresponding sheet music. For example, a musician could ensure what his musical performance by comparing the traditional sheet music with their own playing. This type of comparisons helps to hint the musician exactly

Send correspondence to Xunyu Pan: xpan@frostburg.edu

where he went wrong and how to adjust his mistakes. Due to the distinct manners of interpretation of music, the semantic gap between the audio data and the sheet music should be bridged base on the musical analysis of these two representations.

To address this issue, we present a novel real-time musical analysis system to process musical examination and implement *audio-sheet conversion*. The proposed Musicians Aid system realizes these functionalities by first extracting audio features from an input audio stream. These audio features are further compared to the features collected from the corresponding sheet music aiming to report the extent of similarity. Using the same technique, the system is able to further convert an audio recording to the corresponding sheet music. In an opposite manner, the Musicians Aid can also create computer-generated music from the input of textual representation of musical notes. Based on these functionalities, the proposed system provides professional musicians with an insightful look at the music they created and advances their understanding of their work. Amateur musicians could also benefit from using it solely for the purpose of obtaining feedback about a song they were attempting to play. The system can be extended to assist amateur musicians to conveniently learn music over the web. The Musicians Aid is evaluated with real songs played with guitar and digital audio recordings generated by computer. The quantitative experimental results show that the proposed system is effective in reporting the extent of similarity between an audio recording and the corresponding sheet music. Qualitative experiments also demonstrate the system can reproduce high quality digital sheet music with the input of audio recordings.

2. RELATED WORK

The analysis of distinguished characteristics of a piece of music largely relies on the audio features extracted from the audio recording and its corresponding sheet music. Consequently, the studies of the relationship between these two musical representations share the similar processing procedure with most music information retrieval (MIR) technologies. Recently, MIR has become an crucial research area aiming to effectively and efficiently retrieve and manage the vast amount of digital music created every day in the world.⁴

Content-based musical analysis is one of the major MIR techniques where the critical information about music is automatically collected and processed for the purpose of finding the music of interest to users.⁵ This technique involves the extraction of low level audio features including timbre features for instrument recognition and temporal features for capturing the variation of timbre over time.⁴ The collection of musical information starts with the splitting of input audio signal into individual frames. Using standard signal processing techniques like Fourier transform (FFT) or discrete wavelet transform (DWT), timbre features such as Mel-frequency cepstrum coefficient (MFCC) can be directly obtained from each local frame of input audio recordings.⁶⁻⁸ The temporal feature is typically acquired with the integration of timbre features extracted from a series of frames. Low level features are used for various MIR applications. However, some MIR tasks require the extraction of high level features such as rhythm, pitch, melody, and harmony, which are typically the intrinsic musical properties perceived by human. One of the most challenging MIR problems is *similarity retrieval* where the goal of the task is to seek similar song in a music collection by a given query song. Early music retrieval systems predominantly rely on detecting the similarity of low level features such as timbre.^{7,9,10} More recent techniques require the similarity comparisons for the higher level features such as melody and harmony.¹¹⁻¹⁵

Advanced from the traditional musical similarity retrieval technologies, recent studies on the relationship between audio recording and sheet music become an attractive research topic in the MIR field. Audio recording and sheet music emphasize two different but related perspectives of a piece of music. Most research problems of this topic involve with how to link these two forms of musical interpretation for the purpose of solving some specific tasks. Instead of using traditional symbolic score data (e.g. MIDI), optical music recognition (OMR) technique is used to realize sheet music-audio synchronization aiming to link regions in images of scanned scores to musically corresponding sections in an audio recording of the same piece.¹⁶⁻¹⁸ Another audio feature based work describes a method to link musically relevant sections in a score of a piece of music to the corresponding time intervals of an audio recording.^{19,20} However, the studies so far have been rather limited in terms of showing the correctness of a given stored/real-time audio recording when comparing to the corresponding sheet music, which is typically very useful information in assisting individuals interested in learning and understanding music.

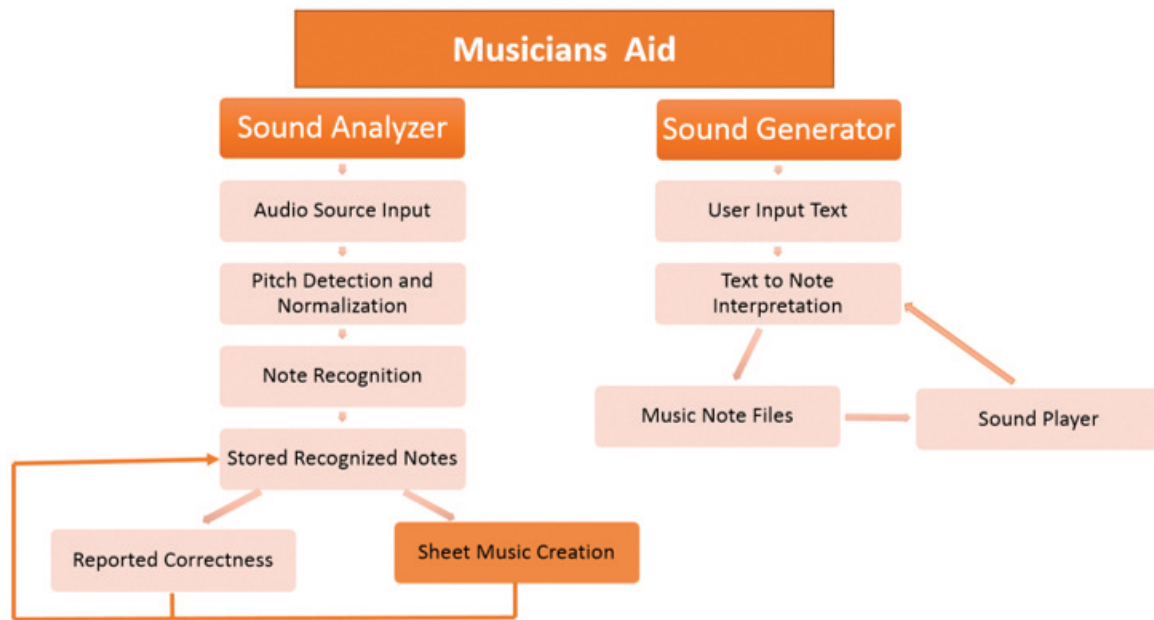


Figure 1. The Musicians Aid system has two major functional components: (a). The *sound analyzer* responsible for the analysis of the audio stream from the audio input; and (b). The *sound generator* responsible for the creation of music based on the textual input of user.

3. METHOD

In this section, we introduce a new musical representation and analysis system to bridge the semantic gap between audio recording and the corresponding sheet music. The proposed application system Musicians Aid can read either stored or real-time audio recordings. The input audio stream is then processed to detect each individual audio pitch which can be further identified as a musical note. These detected musical notes are stored in the system serving for two purposes: (a). The corresponding sheet music is used as the reference for comparison with the detected musical notes. Based on the comparison results, the system reports the overall correctness of the input audio recording. (b). The observed musical notes are used for the generation of sheet music upon the request from the user. Hence the sheet music can be automatically reproduced from even an unknown music recording. In addition to these two core features, the Musicians Aid provides a module for generating digital music directly from the textual input of the user. This feature helps musicians quickly review the musical composition just written. As shown in Figure 1, the overall logical structure of the Musicians Aid system can be divided into two major functional components:

1. **Sound Analyzer:** The music analyzing module of the Musicians Aid system is responsible for the processing of input audio recording and the analysis of the processed audio stream for musical note recognition and sheet music reproduction. Each of the recognized notes is matched individually with the corresponding sheet music to compute the overall quality of the input music.
2. **Sound Generator:** The music generation module is responsible for the creation of music and the later playback of the created music based on the textual representation of musical notes from user input. The computer-generated music is obtained with the integration of a series of individual sound files retrieved from the standard library of musical notes.

When Musicians Aid is set in the *sound analyzer* state, the system first read a real-time music stream from the input port or a stored audio recording from a specific music collection. The audio stream is then processed to estimate the pitch period using the classic auto-correlation pitch detection algorithm.²¹ The dominating

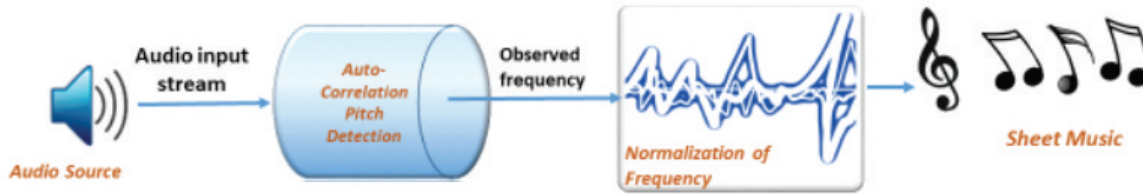


Figure 2. Major phases involved with the generation of a sheet music from the input of an audio recording.

frequency of the audio clip is identified by computing the inverse of the estimated pitch period. All frequency values are then normalized to a specific range based on Table 1.²² Following the normalization procedure, each musical note corresponding to a specific frequency subrange as shown in Table 1 is then stored as an identified musical tone. The stored musical notes are finally utilized to compute the overall correctness of the input music in terms of its similarity with the standard sheet music.

Another powerful feature of the proposed system is that a sheet music representation of the observed musical notes can be generated and displayed at any time upon the request from the user. As shown in Figure 2, after the pitch detection and the frequency computation, each note of the input audio stream is drawn onto a sheet music. More specifically, the placement of each note on the sheet music is identified based on the matchup between the audio frequency subranges and the corresponding musical notes as shown in Table 1. Although not present in this example, the system can also identify sharp notes by placing a symbol # right after the note whenever it is raised. Compatible with wide variety of audio input, Musicians Aid currently performs most efficiently when directly receiving input from a stringed instrument or an audio file in digital format.

When Musicians Aid is set in the *sound generator* state, the system awaits a textual input from the user, where the input could be any text representation of musical notes. Each individual input note is mapped to a corresponding music file stored in the standard library of musical notes. The sound player of the system then accesses the library to playback the textual input as music. For example, a textual version of the popular song *Mary Had a Little Lamb* is shown below:

Example text input: B A G A B B B A A A B D D B A G A A B B B B A A B A G.

After the above textual input is entered, the sound player turns it into a real song. Obviously, the quality of the computer-generated songs relies on the musical note clips stored in the library. This music generation functionality helps musicians quickly review the newly written musical composition.

Musical Notes	Audio Frequency
F	174.61 HZ
E	164.81 HZ
D#	155.56 HZ
D	146.83 HZ
C#	138.59 HZ
C	130.81 HZ
B	123.47 HZ
A#	116.54 HZ
A	110.00 HZ
G#	103.83 HZ
G	98.00 HZ
F#	92.50 HZ

Table 1. The relationship between chromatic scale notes and their corresponding audio frequency values.

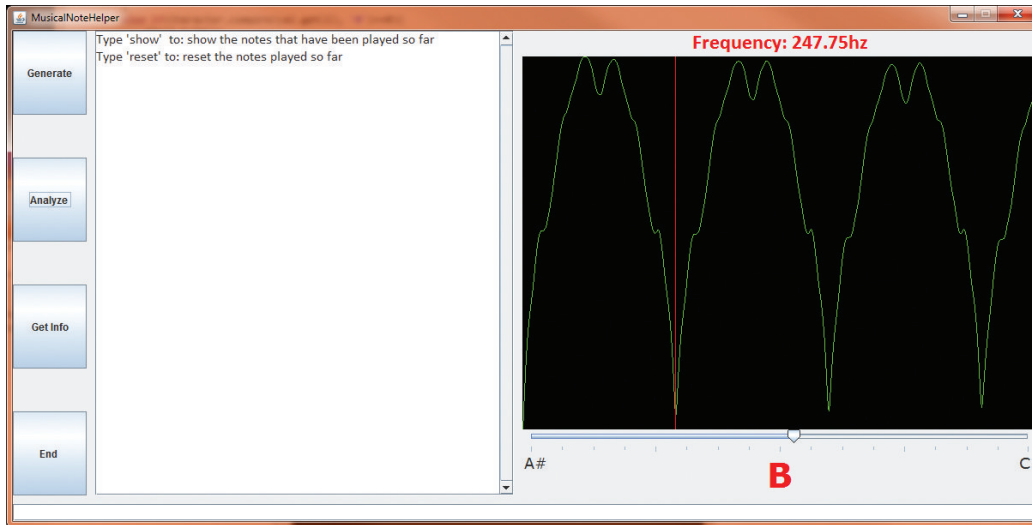


Figure 3. The main window of the proposed Musicians Aid system in *sound analyzer* state.

4. RESULTS

Our Musicians Aid system is implemented with the Java language based on the Eclipse platform. As shown in Figure.3, the GUI for the proposed system is quite straightforward. The buttons located on the left side of the main window help users to switch the system to different states (e.g. sound generator, sound analyzer):

1. **Generate:** The *generate* button allows user to create a song when a textual input of musical notes is entered inside the text field located at the bottom of the main window.
2. **Analyze:** The *analyze* button opens a graphical display to show the plot of the generated auto-correlation function with the identified musical note displayed below. In addition, typing 'show' visualizes the observed notes as sheet music. Typing 'reset' clears all the notes that the program has observed so far.
3. **Get Info:** The *Get Info* button helps identify available information about the current state of the program.
4. **End:** The *End* button clears all temporary data and then closes the program.

One regular laptop computer is used for music analyzing experiments. The machine has an Intel Core i5-2500k 3.3-GHz processor with 4 GB memory and run on Windows 7(x64) Professional platform. During the experiments, two types of music recordings, human-generated songs and computer-generated songs, are collected. The songs of the first type are directly played by an amateur guitar player. These songs are then converted to music recordings which can be saved on the computer. The songs of the second type are digitally generated by computer program. More specifically, we use *Audacity*,²³ a free open source audio editor and recording computer application, to create digital songs. For each of these songs, the musical tones are randomly generated and further integrated as a music file. Both of these two types of songs are later read into the Musicians Aid system for the experiments of performance analysis.

We further quantitatively evaluate the efficacy of the proposed Musicians Aid system. For testing purposes, we collect 100 real songs played by an amateur guitar player and 100 digital recordings randomly generated by the *Audacity* program, resulting in a total 200 test songs. Each song contains 12 musical notes. These songs are all analyzed by the proposed system aiming to report the correctness of input song in terms of the similarity to the corresponding sheet music. We use two quantitative measures to evaluate the performance of our system: *Absolute Correctness* (AC) and *Scaled Correctness* (SC). We define AC value as the fraction of notes in a song that are correctly played and SC value as the overall relative correctness of notes in a song. For example, if a note 'F' is played as 'E', the relative correctness of this note is 11/12. Obviously, *Scaled Correctness* is a more



Mary Had a Little Lamb

(a)



(b)

Figure 4. The comparison of (a) The printed sheet music of the popular song *Mary Had a Little Lamb* and (b) The reproduced hard copy version of the original sheet music using the proposed Musicians Aid system.

accurate way of evaluating the quality of a played song. In addition to note-level testing, we also show song-level performance of our system with the number of songs detected by our system as being created perfectly. Here, a perfectly created song is defined as all notes of the song, either played by a human player or generated by a computer program, are exactly same as the original notes on the corresponding sheet music. The experimental results as shown in Table 2 indicate that the proposed system can accurately detect note errors in each individual song and the total number of songs perfectly created. We also observe that the computer-generated songs are more similar to the original sheet music than human-generated songs. This is expected, as a song played by an amateur guitar player is typically not as perfect as that generated by a computer program.

Finally we qualitatively evaluate the performance of the proposed system. The popular song *Mary Had a Little Lamb* is used for this experiment. The top panel of Figure 4 shows the printed sheet music of the song. The bottom panel of Figure 4 shows the sheet music representation of the given song created by the music generation module of the system. The experimental result demonstrates that the proposed system can perfectly reproduce the hard copy version of the original sheet music.

5. CONCLUSIONS

Audio recording and the sheet music are two major representations of music. They emphasize musical characteristics from two different but related perspectives. Bridging the semantic gap and showing the extent of similarity between these two representations have recently become an emerging research topic with many potential applications. In this work, we introduce a new musical representation and analysis system Musicians Aid to address this issue. Musicians Aid solves the previous problem by analyzing individual audio recording and further comparing the results with the data collected from the corresponding sheet music. The proposed system consists of two major functional modules: (1). The *sound analyzer* computes the correctness of input song in

	Songs Played by Guitar Player	Songs Generated by Computer
Total Number of Songs Analyzed	100	100
Absolute Correctness (AC)	90.3%	98.9%
Scaled Correctness (SC)	98.3%	99.2%
Number of Songs Perfectly Created	90	99

Table 2. Comparison of the music analysis performance of the Musicians Aid system with real songs played by an amateur guitar player and digital songs completely generated by a computer program.

terms of the similarity to the corresponding sheet music and reproduce the hard copy of the original sheet music; and (2). The *sound generator* creates digital music based on the user textual input and provides the playback of the computer-generated music. We employ both quantitative and qualitative experiments to show the efficacy of the developed system.

The Musicians Aid is successful in some practical scenarios. However, we believe the system has some limitations at this point. First, the system currently can only identify the musical notes in the form of quarter note due to its weakness on timing analysis. In the future, the system should be able to identify the duration of a musical note by implementing time stamps. Moreover, the system can not effectively communicate with users regarding the mistakes they made during playing. Ideally, the system will be able to visualize these mistakes by highlighting the incorrect notes and displaying a text box explaining the mistake to users. Finally, we plan to extend the application over the web to assist users in playing new songs, learning to read sheet music, and developing a better understanding of tones, which helps creating a convenient environment for amateur musicians and their music instructor.

6. ACKNOWLEDGMENTS

This work was supported by the Frostburg State University President's Experiential Learning Enhancement Fund Program (PELEF) and by a FSU Foundation Opportunity Grant (# 30566).

REFERENCES

- [1] Pan, X. and Free, K., "Interactive real-time media streaming with reliable communication," *SPIE Symposium on Electronic Imaging* (2014).
- [2] Pan, X., Zhang, X., and Lyu, S., "Detecting splicing in digital audios using local noise level estimation," *IEEE International Conference on Acoustics, Speech and Signal Processing* (2012).
- [3] Pan, X. and Lyu, S., "Region duplication detection using image feature matching," *IEEE Transactions on Information Forensics and Security (TIFS)* **5**(4), 857–867 (2010).
- [4] Fu, Z., Lu, G., Ting, K. M., and Zhang, D., "A survey of audio-based music classification and annotation," *IEEE Transactions on Multimedia* **13**(2), 303–319 (2011).
- [5] Casey, M., Veltkamp, R., Goto, M., Leman, M., Rhodes, C., and Slaney, M., "Content-based music information retrieval: Current directions and future challenges," *Proceedings of the IEEE*, 668–696 (2008).
- [6] Logan, B. and Salomon, A., "A music similarity function based on signal analysis," *IEEE International Conference on Multimedia and Expo* (2001).

- [7] Aucouturier, J.-J. and Pachet, F., “Music similarity measures: What’s the use?,” *International Conference on Music Information Retrieval* (2002).
- [8] Aucouturier, J. J. and Pachet, F., “Finding Songs That Sound The Same,” *IEEE Workshop on Model Based Processing and Coding of Audio* (2002).
- [9] Pampalk, E., Dixon, S., and Widmer, G., “On the evaluation of perceptual similarity measures for music,” *International Conference on Music Information Retrieval* (2003).
- [10] Berenzweig, A., Logan, B., Ellis, D. P. W., and Whitman, B., “A large-scale evaluation of acoustic and subjective music similarity measures,” *International Conference on Music Information Retrieval* (2003).
- [11] ho Tsai, W., ming Yu, H., and min Wang, H., “A query-by-example technique for retrieving popular songs with similar melodies,” *International Conference on Music Information Retrieval* (2005).
- [12] Jang, J. R. and Lee, H., “A general framework of progressive filtering and its application to query by singing/humming,” *IEEE Transactions on Audio, Speech and Language Processing* **16**(2), 350–358 (2008).
- [13] Unal, E., Chew, E., Georgiou, P. G., and Narayanan, S. S., “Challenging uncertainty in query by humming systems: A fingerprinting approach,” *IEEE Transactions on Audio, Speech and Language Processing* **16**(2), 359–371 (2008).
- [14] Serrà, J., Gómez, E., Herrera, P., and Serra, X., “Chroma binary similarity and local alignment applied to cover song identification,” *IEEE Transactions on Audio, Speech and Language Processing* **16**(6), 1138–1151 (2008).
- [15] Ravuri, S. V. and Ellis, D. P. W., “Cover song detection: From high scores to general classification,” *IEEE International Conference on Acoustics, Speech, and Signal Processing* (2010).
- [16] Fremerey, C., Mller, M., Kurth, F., and Clausen, M., “Automatic mapping of scanned sheet music to audio recordings,” *International Conference on Music Information Retrieval* (2008).
- [17] Fremerey, C., Clausen, M., Ewert, S., and Mller, M., “Sheet music-audio identification,” *International Conference on Music Information Retrieval* (2009).
- [18] Müller, M., Goto, M., and Schedl, M., eds., [*Multimodal Music Processing*], vol. 3 of *Dagstuhl Follow-Ups*, Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, Germany (2012).
- [19] Şentürk, S., Gulati, S., and Serra, X., “Score informed tonic identification for makam music of turkey,” *International Conference on Music Information Retrieval* (2013).
- [20] Şentürk, S., Holzapfel, A., and Serra, X., “Linking scores and audio recordings in makam music of turkey,” *Journal of New Music Research* **43**, 34–52 (2014).
- [21] Rabiner, L. R., “On the Use of Autocorrelation Analysis for Pitch Detection,” *IEEE Trans. on Acoustics, Speech, and Signal Processing* (1), 24–33 (1977).
- [22] Jorgensen, O., [*Tuning: containing the perfection of eighteenth-century temperament, the lost art of nineteenth-century temperament, and the science of equal temperament, complete with instructions for aural and electronic tuning*], Michigan State University Press (1991).
- [23] Audacity, “Audacity(R) program.” <http://audacity.sourceforge.net>.