Association for Information Systems

AIS Electronic Library (AISeL)

ACIS 2013 Proceedings | Australasian (ACIS)

2013

Adaptive Music Score Trainer for Visually Impaired in Sri Lanka

Kavindu Ranasinghe
University of Colombo School of Computing, kavindu.rana@gmail.com

Shyama Kumari
University of Colombo School of Computing, shyama3000@gmail.com

Dawpadee Kiriella
University of Colombo School of Computing, kiriella.dawpadee@gmail.com

Lakshman Jayaratne
University of Colombo School of Computing, klj@ucsc.cmb.ac.lk

Follow this and additional works at: https://aisel.aisnet.org/acis2013

Recommended Citation

Ranasinghe, Kavindu; Kumari, Shyama; Kiriella, Dawpadee; and Jayaratne, Lakshman, "Adaptive Music Score Trainer for Visually Impaired in Sri Lanka" (2013). *ACIS 2013 Proceedings*. 161.
https://aisel.aisnet.org/acis2013/161

This material is brought to you by the Australasian (ACIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in ACIS 2013 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Information Systems: Transforming the Future

24th Australasian Conference on Information Systems, 4-6 December 2013, Melbourne

Proudly sponsored by

ACIS 2013 Principal Sponsor

Adaptive Music Score Trainer for Visually Impaired in Sri Lanka

Kavindu Ranasinghe
Shyama Kumari
Dawpadee Kiriella
Lakshman Jayaratne
University of Colombo School of Computing
35, Reid Avenue,
Colombo 07, Sri Lanka
Email: klj@ucsc.cmb.ac.lk

Abstract

This paper is on a research, trying to solve the problem of how to assist visually impaired users in Sri Lanka to visualize and train Eastern music scores (notations) in a way that satisfies their music needs. Several researches have been done in Western music context to address this issue but fewer researches have been done in Eastern music context in this area. In this study, audio format was identified as the most feasible and effective readable format to read music notations for visually impaired. Therefore the research study has been steered to convert visual music notations (of Eastern music in Sri Lanka) into an audio output which gives visually impaired user a real feeling of reading visual notations as it is and assists to train on them. This can be further extended as a tool which assists visually impaired users in music score generation and music composition.

Keywords

visually impaired, Eastern music, music score/notation, singing synthesizing

INTRODUCTION

Education is a responsibility of all human beings. It is accepted as a need to be fulfilled regardless the disabilities or any other weakness of human being. Disabled does not mean inability or less talented, only differently able. Differently able users can be trained using appropriate tools and techniques similar to their other counterparts. This research is on music education of visually impaired. Aesthetic education has been a must in primary education as well as some aspects of secondary education. Taking into consideration visually impaired students, the easiest aesthetic subject which they can be followed is music as authors have disclosed via several surveys and observations.

World Health Organization (WHO 2012) has estimated that 285 million people are visually impaired worldwide while 39 million are blind and 246 have low vision. And also about 90% of visually impaired people of the world live in developing countries. There are number of visually impaired musicians who perform unbelievably. They were capable to overcome the barriers they had in learning music showing their extraordinary talent. World famous musicians such as Steve Wonder, Lemon Jefferson, Andrea Bocelli, Nobuyuk Tsujii and Sri Lankan musicians; Henry Caldera, Mekala Gamage, Hemapala Perera all are totally blind, but their compositions stood as milestones in history of music.

Sri Lanka has a disabled population of 247,711 out of the whole population (ICTA 2012). It is a significant percentage (2.7%) of the Sri Lankan population. Of the visually impaired children who start schooling only 24% make it up to the GCE Ordinary Level Examination and a mere 4% make it to the Advanced Level (ICTA 2012). In the schools which are specially allocated for visually impaired students approximately 99% of the visually impaired students follow music as their aesthetic subject as it is the only aesthetic subject they can follow comfortably. Among Sri Lankan visually impaired students approximately 95% follow Eastern music.

Schools of music have brought up the tradition of storing qualities of musical compositions in written formats. This has been practiced for many years and almost all the formal music education parties use these music notation systems in carrying out their teaching procedures. Purpose of music notation is to communicate the music created in composers mind to performer. Therefore music notation has become a major communication medium among musicians. Moreover music notations are compulsory in Sri Lankan local Eastern music syllabus and even in "Bhathkande" music exams.

In this context the contributions of this study are:

- A survey to identify the most critical problem which is faced by Sri Lankan visually impaired within the context of Eastern music

- A survey on IT literacy and computer affordability among target group since the proposing solution is ICT based
- Literature survey on available research prototypes and commercially available products in order to address the problem
- Proposing an innovative ICT based solution which is the most optimal way to solve the identified problem with available technology at present in a way that satisfies their music needs
- Analysis of the feasibility of implementing such a solution, its effectiveness and the challenges in implementing and developing the application

As a result of the initial questionnaire based survey of this study on what the major problems which are faced by visually impaired people in Sri Lanka in learning music are, the authors have revealed the problem of visualizing music notations.

Music notation is a set of signs/graphic symbols and the rules for representing music. The representation of music notation depends on the origin culture of the relevant art of music. Thus, music notation representation in Eastern music is different from Western music. In same type of music also, the way of visual presentation of music notation is different from country to country (i.e. In Eastern music, the visual presentation of music notation in Sri Lanka is different from the same one in India).

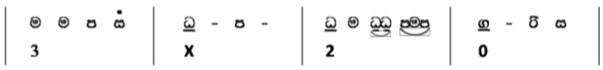

Figure 1: An example extract of common music notation (CMN) in Eastern Music using in Sri Lanka

Currently, the main methods which visually impaired people use to read music notation are music Braille and getting the assistance of a third party to get them read. There are specific standards for Eastern music Braille notations. Figure 2 shows the translation of the lead-in and first bar of the extract of CMN shown in Figure 1.

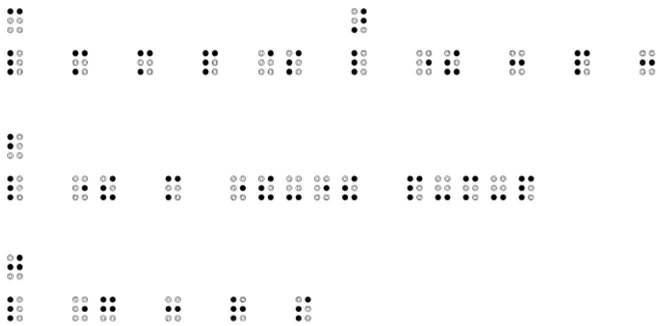

Figure 2: An example extract of Braille music notation

Currently, to read music notations by visually impaired, ordinary music notations have to be translated into Braille. There are many tools and researches carried out in order to generate Braille music notations for Western music, but it is hardly reported for Eastern music. Writing and reading Braille is exceedingly time consuming. A line in a music notation script consists of four columns as shown in Figure 2. But when writing them in Braille, users have to use one line per a column. Therefore, one line in text should be converted to four lines in Braille which make them difficult to grasp the notation quickly. Music notations require aligned wordings in writing and translating printed notations into Braille for reading. Therefore, it is hard to follow than converting normal text into Braille. Moreover Braille printouts require special and high cost hardware (Braille printers) which is not affordable for many users (Other devices such as Braille embossers, Braille displays are also highly expensive).

To follow music notations, visually impaired students have to first read the Braille notation and remember it and revise back when they play an instrument or sing it again. This leads to delays and disturbs the continuous flow of music. Furthermore, Braille documents have less durability and if one dot was missed, what is implied by a given notation will be totally different.

RELATED WORK

The previous researches done in this area can be categorized into three categories as:

- Capturing music notations
- Converting music notations into a readable format for visually impaired users

- Converting audio to notations (for training purposes)

This section describes them in detail.

Capturing Music Notations

Fujinaga (1996) has come up with a system called Adaptive Optical Music Recognition by proposing an adaptive software for the recognition of musical notation focusing to create a robust framework upon building a practical optical music recognizer. Since OMR is for Western music (Johansen 2009) this study also has been limited to facilitate only for music notations in Western music. But, Fujinaga (1996) proposed a research prototype to expand the quality OMR as an adaptive system. The research methodology of the study of Fujinaga (1996) was a great support in extension of research prototype up to an adaptive system in the part of our research for recognizing the music notations to capture them in the context of Eastern music.

The OMR system in the study of Raphael and Wang (2011) has been aimed toward the notation images of the International Music notation Library Project (IMSLP). This system mainly focuses on measures such as basic unit of recognition. They have identified candidate composite symbols in Western music (chords and beamed groups) using grammatically-formulated top-down model-based methods, while employing template matching to find out isolated rigid symbols. They have reconciled these overlapping symbols by seeking non-overlapping variants of the composite symbols those could best account for the pixel data. And the results of that system (Raphael and Wang 2011) have been presented on a representative notation from the IMSLP.

Ng and Khan (2012) have come up with a sheet music image processing application targeted towards beginners to provide an instructional aid. The application, "Automated Instructional Aid for Reading Sheet Music" accepts an image file of the sheet music, annotates all of the notes, and generates and plays a MIDI file of the song. Those authors mentioned that this application (Ng and Khan 2012) recognizes and classifies quarter, half, and whole notes with very high accuracy (97-100%) for digitized sheet music readily available in the internet.

Riley and Fujinaga (2002) as well as Ng and Khan (2012) have tried to identify a solution for the problem of using digitally captured music (notation) images while capturing music notations. Results of those researches can be used in capturing music notations in this research as the soft copy of the music sheet may be digitally captured using a digital camera or any other digital formats.

Converting music notations into a readable format for visually impaired

Weasel (Challis et al. 2000) is a system for the Non-visual Presentation of Music Notation and its overlay consists with two main sections referred as bar area and control section. Once the user presses on the bar area corresponding musical content is played. Circular symbols below the guidelines represent the dynamic information which consists in a manuscript made for sighted people and will be conveyed to the user through synthetic speech. The control section allows the user to select between different modes available. The modes allow users to navigate within the manuscript much more easily.

The survey in the paper of Weasel project (Challis et al. 2000) highlights that with Braille music; all the symbolic and written instructions are translated resulting, again, in large amounts of information presented in a serial fashion. Therefore, the research (Challis et al. 2000) has also considered some successful solutions to overcome the limitations within Braille music representation and talking notations. It highlights that existing methodologies used result with large amounts of information presented in a serial fashion which again limit the user to view the information at one glance. Further, it provides a solution for finding out a way to filter unnecessary information. The paper (Challis et al. 2000) describes that learner's reading speed is directly limited by the speed of the spoken description in talking notations even though prerequisite knowledge on Braille music is not needed. Therefore, addressing the problem in Western music context seems less confused compared to Eastern music context, since Western music is more structural than Eastern. The number of bars-to-a-line, lines-to a-page and graphical representation of the notations makes it much easier in the conversion process.

The paper done by Bertoni et al. (2003) reviewed available software to assist visually impaired people to read music notations and proposed a new web portal minimizing the weaknesses of existing tools. They have proposed a large digital archive for Braille music pieces to develop longer term access to Braille music and music literacy. That highlights the significance of music in visually impaired lives and challenges they face in learning music. Western music notations are standardized notations while Eastern music lacks of this feature. Those authors have listed down strengths and weaknesses of available music access assistive media (MIDI, Stave standard, Braille music) for visually impaired people. Among them it emphasizes the lack of self-evidence as one of the key weaknesses of Braille music. It summarizes converting methods of music into machine readable format while highlighting the methods of converting Braille music into digital music and vise-versa. That study mainly focused on addressing the problem of less efficient navigation of the blind reader compared to the sighted reader rather than facilitating higher amount of notations converting to Braille. They (Bertoni et al. 2003) have concluded converting music scripts to readable format requires proper formatting of music notation. In Western music the music notations itself encoded in standards and even for Braille music there are technologies to convert a less formatted Braille page in to a highly formatted Braille page(e.g. Resonare). But considering Eastern music in Sri Lankan context still there is no specific standard to be presented. In that case since we are addressing the conversion processes the significance of defining a standard is highly recommended. The study suggests the methods to convert poorly formatted page into a highly formatted page in standardizing Eastern music scripts.

The study (Kaprykowsky and Rodet 2006) connects with our system because the new algorithm "Short-Time Dynamic Time Warping" (STDTW) had evaluated in aligning musical notation and the time axis of audio performance of it. So the aligning results may be important for the current work in measuring the accuracy of the conversion music notation to audio. However, the problems in the study they have are, impossibility of identifying the exact tempo which is not constant and the notes played may be different from the written ones and it is difficult to understand the note playing in polyphonic music. With respect to these issues, we are addressing only monophonic music. Since our research focuses on vocal music but not instrumental music, the differences between audio file and written script must be minimum compared to instrumental music. But the problem of identifying the exact tempo is considered. Since the core of the study highly focused on mathematical algorithms, the opportunity of using the content to our research is very much limited. But the end product of the research (i.e. STDTW) may be worth in evaluating the project.

Kim (2003) has attempted to develop a framework which analyses and synthesizes singing voice. In order to accomplish the goal it estimates source filter voice model parameters by considering both physical and expressive factors. Further, it models the dynamic behaviour of these features over time using a Hidden Markov Model. Then, it describes the framework using mathematical algorithms and discusses the features of the framework. These are discussed with details below. The authors of that paper (Li et al. 2012) claim that the singing voice is one of the most complex instruments. This is a key point to consider in analysing the best output format for the current research project. It further discusses about singer identification techniques which can be used for our research if we plan to continue with adaptive trainer section. It suggests for the framework is low bit rate singing voice coding. As well as it claims that the previous researches on which the study was built upon the compression advantage of encoding vowels using static templates. However, the topic itself doesn't straight forward with respect to our research project since our project does not focus on singing lyrics of a song. But the theories discussed are useful for our research.

Hidden Markov Model is defined as a statistical tool used for modelling generative sequences characterized by a set of observable sequence (Gales and Young 2008). Therefore, this model can be used to confirm the accuracy of the final product. Whatever the output readable format would be, it could be considered as a sequence of characters (audio or tactile characters sequence which gives the same meaning of the music notation). Then, that sequence may be possible to model over the original sequence of music notes in the script using HMM.

Kyritsi et al. (2007) have discussed about proposing a score (notation) to singing synthesis system and as the method, notation will be written in a notation editor and saved in the MIDI format. As that study suggests Singing Voice Synthesis (SVS) has become a trend during the last few decades and it discusses about the urgency and significance of using SVS systems while describing the characteristics of a SVS system. Further, it contrasts and compares speech and singing processes as well as the vocal music generation and instrumental music generation. When comparing that study (Kyritsi et al. 2007) with our proposed system the key different of the output format is this study proposed a lyrics singing and our project is proposed to notation singing.

The use of concatenative synthesis in singing synthesize is discussed and the previous project failures in synthesized singing and the reasons for failures are also presented in that study (Kyritsi et al. 2007). As an example Swedish use of MBROLA output was not much natural because their diphone database was derived from spoken language. Since our research is not about lyrics singing it may not be an issue. But still there can be such barriers to search for a better output. And it suggests if we use such software to produce the output then first we have to check the availability of particular database of the particular language (in our research this is Sinhala language) According to the study; most of the diphone databases are available in male voice. That itself suggests as a drawback of using available databases. MBROLA project doesn't contain Sinhala diphone database. As an option for the above mentioned weaknesses of using existing diphone databases, the research study focuses on creating a new diphone database which can be used for Sinhala music notations. In future works the paper suggests about creating a database with both male and female voices with differentiations of formants, pitch, timbre, and vibrato. This suggestion is applicable for our research study too.

Converting Audio to Notations (for training purposes)

Conceptually a music trainer can be defined as one who gives a genuine feedback on a musical performance by analysing it. According to Study on Software-Based Extraction of Objective Parameters from Music Performances by Lerch (2008) Music Performance Analysis (MPA) aims at studying the performance of a musical score rather than the score itself. It deals with the observation, extraction, description, interpretation and modelling of music performance parameters as well as the analysis of attributes and characteristics of the generation and perception of music performance. According to him four classes of acoustical parameters that can be used for the description or characterization of music performances have been identified. They (Lerch 2008) are:

- Tempo and timing: global or local tempo and its variation, rubato, or expressive timing, subtle variation of note lengths in phrases, articulation of tones
- Velocity, loudness or intensity: musical dynamics, crescendo and diminuendo, accents, tremolo
- Pitch: temperament, tuning frequency, expressive intonation, vibrato, glissando
- Timbre: sound quality and its variation resulting from instrumentation and instrument-specific properties such as bow positioning (string instruments)

Various researches has gone through experimenting various methods in extracting above parameters from a music performance, analysing them in order to come up with conclusions on the performance, filtering out separate performances from a polyphonic music, discriminating between vocals and non-vocals.

The research by Simon Dixon, Austrian Research Institute for Artificial Intelligence on the Computer Recognition of Solo Piano Music (Dixon 2000) suggests a new approach to extract specific parameters from a performance of a piano music. The system is closely related to the task of automatic transcription but other than calculating which notes were played (pitch) and when they were played (timing) it focuses also on how loud they were played (velocity).

According to the study Standard signal processing techniques based on the short time Fourier transform are used to create a time-frequency representation of the signal, and adaptive peak-picking and pattern matching algorithms are employed to find the musical notes. The data processing stages consists with down sampling the audio, creating time-frequency graph, converting the graph to a power spectrum to identify peaks, isolate peaks to create time-frequency atoms, and derive musical notes using those frequency tracks.

Some of the Key design methods used in the research (Dixon 2000) are highlighted as entry paths which can be used in Music Performance Analysis. The study reveals that down-sampling the signal to a 12 kHz sampling rate decreases the processing time of subsequent stages in system data processing and also states that the use of a lower bandwidth signal does not appear to affect results adversely. In order to obtain a more accurate estimate of frequency, one extension of the research uses the rate of change of phase in the FFT filter bank channels, rather than the centre frequency of the channels. These methods can be used in identifying the performance of the user in prototype which was created in our research. As in most of the researches this has also used synthesized test data for evaluations.

Hidden Markov Models (HMM) (Cont and Schnell 2004) are used in analysis of audio with speech content. This extended observation model which is based on Gaussian Mixture Models, is trainable using a learning algorithm called automatic discriminative training to obtain better audio to score real-time alignment for musical applications. The research reveals that the new model has proven its viability in a concert performance, and its flexibility in an automatic accompaniment application for ballads. That study (Cont and Schnell 2004) also states that the probability density functions (PDFs) are good candidates for an automatic learning algorithm.

Approach to labelling a large amount of training data for vocal/non-vocal discrimination in musical audio with the minimum amount of human labour is revealed by the dissertation “Automatic labelling of training data for singing voice detection in musical audio” (Lee and Cremer 2000). This research use symbolic music files like MIDI files as a tool to find vocal/non-vocal boundaries in real recordings.

EXPERIMENTAL SETUP

The all researches/studies discussed under the previous section have been developed for Western music. As identified via the critical review, to overcome the problem, we are proposing the system, “Adaptive Music Score Trainer” which delivers a non-visual presentation of music notations of Eastern music. This can be easily grasped by visually impaired users. Through the interviews, observations and surveys conducted, the output format of the system was decided as audio format. We revealed that the audio output format is much more convenience, flexible and familiar to the visually impaired user compared with the other non-visual output formats such as hepatic/kinaesthetic, gustatory and olfactory.

The basic functionalities of the system are:

- Facilitate users in converting any script of Eastern music notations used in Sri Lankan context into readable output format (synthesized singing of corresponding "Swara" and background play of corresponding "Thaal") which visually impaired students can grasp each and every point of the music notation more easily
- Facilitate users in navigating through the auditory output created in a similar way which a sighted music student is capable when they are using a printed music notation
- Generate an audio output with several audio signal layers to represent or signal the users about ornamentation and "Vikiriti" qualities of music notes / "Swara" and user can enable or disable any layer depending on the necessity
- Facilitate users in saving processed music notations if he/she prefers and get rid of time consumes for re-processing
- Facilitate users in changing the tempo of the output format accordingly and playing back in low("Vilambha Laya"), medium("Madhya Laya"), fast("Dhrutha Laya") tempos without violating the nature of the auditory composition
- Facilitate users in navigating through the output audio note by note and getting details about the qualities of each note in speech if they prefer
- Assists users to train (vocal) a selected music notation

The first level prototype was developed in a way the user can get an idea about the output format of the solution to gain constructive ideas from domain experts and users. It also acted as a successful method in conveying the conceptual idea more precisely to the target group. It was an audio which composed by merging four layers of audio tracks. Each layer represented an isolated symbolic layer identified in an Eastern music script. Musical note layer, "Thaal" symbol layer, "Vibhag" start point signal layer, and Note derivative symbol layer are the layers combined to produce the effective outcome. Musical note layer was a recorded singing and is expected to be synthesized in the actual solution. "Thaal" symbol layer was generated using FL Studio software. Note derivative symbol layer were composed using distinct audio chunks which inherit unique qualities to imply the relevant derivative type ("Komala"/flat, "Theevra"/sharp) or ornamentation (kan, meand, bend) of a musical note it represents at the correct occurrence of the relevant musical note in music notation time line. "Vibhag" start point signal layer contained harp rings which dims at start points of "Vibhags"/columns. Audocity was used to merge tracks and come up with a single audio track. Each sample was created as mono channel and dual channel output in order to identify the better alternative which users can grasp more easily.

First level prototype audio samples are designed in a way which even sighted user can understand the content of the music notation which it represents. Therefore, at the very first phase they were verified to be evaluated by sighted. As this initial evaluation would give birth to ideas and feedbacks which were lately contribute a well fine-tuned prototype beforehand.

The prototype level system should go to the real user testing with minimum errors and with higher quality. If we are to get all the weaknesses in the prototype or requirements clarified from the beginning, the user might get exhausted and they would also have to participate in many testing sessions. Therefore, a separate test data set was created to examine the effectiveness of using each auditory symbol.

The proposed solution involves a framework with a set of processes and tools/ techniques. This framework is described as phases in the fourth section of this paper, and in each phase the tools and techniques that are used for the process are described.

Ideally when a visual script of music notations is found by a visually impaired user, the following process is carried out in the system with a pre-requisite stage and three main phases as follows.

- Pre-requisite: Capturing the image of a script of music notation (create input soft copy) from the hard copy
- Phase 1: Identification of symbols on the input image of the visual script
- Phase 2: Converting to the readable format (audio format)
- Phase 3: Converting user performance (audio) into comparable formats (for training purposes)

Capturing the Image of a Script of Music Notation (Create Input Soft Copy) from the Hard Copy

This initial phase is a pre-requisite. Acquisition of image with a considerable resolution has to be done using a scanner or camera which fits the recommended quality requirements.

Identification of Symbols on the Input Image of the Visual Script

In this 1st phase of the proposed system, pre-processing of image and normalization is done. The relevant symbols (Eastern music notations on the visual script) on the input image (captured image) are identified with the pre-defined knowledge the system has with a specific Optical Character Recognition (OCR) engine, developed and trained for the all possible symbols in Eastern music notations in Sri Lanka. The output file of this process is a single text file. The steps of music symbol identification are shown in Figure 3.

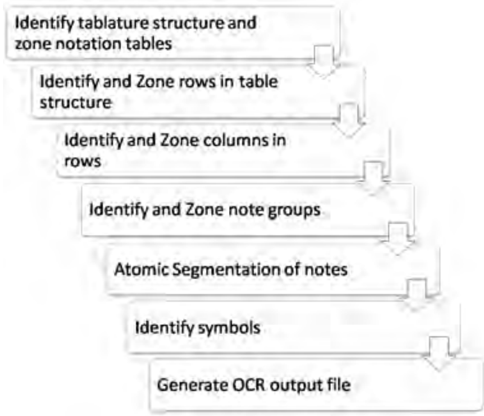

Figure 3: Symbol Identification Process

Segmentation of columns and symbol groups are done using the horizontal projection of grey levels as illustrated in Figure 4.

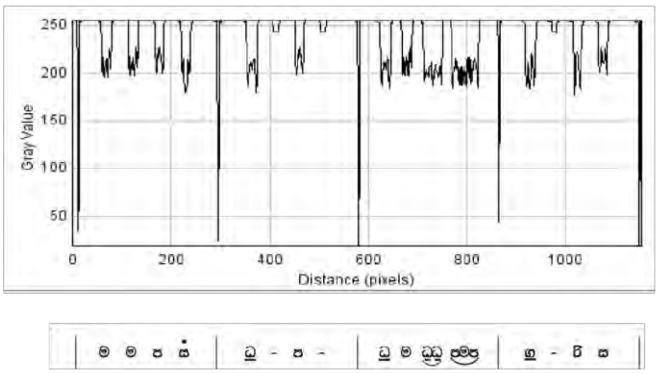

Figure 4: Segmentation of columns and symbol groups using horizontal projection

Average width of non-grouped notes and central peaks in horizontal projection of note groups (Consider frequency matrix of greyed pixels laid horizontally in a note group and curved line. Curved line matrix is substituted from note group matrix to isolate note symbols. Central peaks are identified in the resulting matrix) are used to separate atomic notes from a group of notes. Identification is done using height, width ratio and pixel ratio of atomic composite (including signs used to represent derivatives of music notes. e.g. “Vikriti Swara”) symbols.

After identification of atomic notes, ambiguous sequences are verified in the post processing, against a probability matrix created revealing the probability of occurring of one note followed by another. A sequence id of note groups and their corresponding cardinalities are kept in a map. Output text file contains note sequences and “Thaal” symbol sequences, with delimiters to separate “Vibhags”. For initial research model, only “Sargam” notations and notations which only include “Swara” (no lyric lines) are considered.

Converting to the Readable Format (Audio Format)

The inputs for this process are the text file generated in previous step and note group sequence map. While proceeding the conversion to audio output, two intermediary files for note and “Thaal” sequences are generated from the text file.

In the note sequence of intermediary note file, attributes of each note such as note name, "Saptak", derivative of note and ornamentation are represented using distinct symbols. Therefore, more than one character is used to represent all qualities of one note. Each note is followed by a coefficient to represent its own duration. Rounded value of sum of these co-efficient reveals the number of "Maatras" in the line as showed in Figure 5. Intermediary "Thaal" file includes the sequence of "Thaal" symbols extracted from the OCR output file.

සසස	0.3S, 0.3S, 0.3S
සස	0.5S, 0.5S
ස	S
ස -	S, S-A
ස - -	S, S-A, S-A

Figure 5: Coefficients representing the duration of notes

Audio clips for each note/"Swara" and "Thaal Akshara" (played by tabla) are kept in a sound bank. We observed that when one note sung for more than one "Maatraas", all the "Maatras" except the first one has the singing of only the vowel sound as implied by Figure 6.

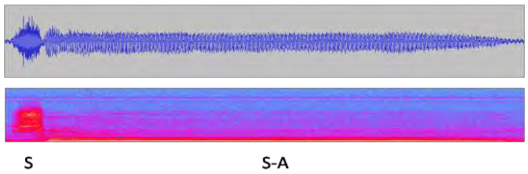

Figure 6: "S" note sung for more than one maatraas

According to the number of "Maatraas" in a line and "Thaal" symbol sequence, relevant "Thaal" is determined. "Thaal Akshara" sequence owns to each "Thaal" (for initial research model, Dadara and Teental are considered) is kept in separate properties files. In the final step, chunks of audio in the sound bank relevant to the finalized "Thaal Akshara" sequence and note sequences are added to separate data-lines and played concurrently. This is a demonstration of the input visual script of Eastern music notation as it is. Separate signals are also played to convey derivatives of notes and start points of "Vibhag" (sound of "Thaalam-pota" arranged according to "Thaal" symbols) if the layers are enabled. Abstract flow of conversion (to readable format) process is illustrated in Figure 7.

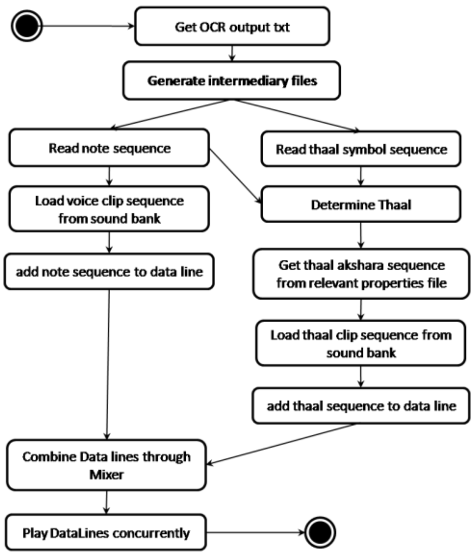

Figure 7: Abstract flow of conversion (to readable format) process

Design had to be disciplined with some specific considerations. Introductory sessions are provided to communicate the distinct audio symbols used by the prototype to represent semiotics in music notation. User should be able to easily distinguish between auditory symbols used to represent derivatives of music notes and they should be carefully chosen considering on the meaningfulness which they offer when listened or heard. While user navigates through a music notation audio effects and clips should be used in a way to imply status changes. As the system is fully focused on Eastern music notations, the conceptual cognition or image created in users mind when interacting with the prototype should not disturb the feelings aligned with Eastern culture and peacefulness intrinsically inherited to Eastern music. Controls should be allowed to give by simple keyboard strokes. Therefore, the user is needed an initial training in getting used to key – function mapping.

Converting audio to notations (for training purposes)

This functionality is bit complex in implementing. The user first selects the composition which he/she wants to get trained. Note intermediary file is already available for these processed compositions and for comparison purposes an audio DNA file is also created. Even though the audio output played by the system consist many extra layers, only the main audio layer which consists singing synthesizing of note sequences are considered in training process. The output audio is played row by row and the user also performs only the corresponding row at a time. Starting points of the user's performance are detected using the point which user breaks the silence immediately after a given beep sound. The time – frequency graph of the captured audio chunk performed by the user is converted to standard notes intermediary file format after undergoing a pre-processing to remove noise. For additional comparisons, an audio DNA file is created for the "Swara" / note singing performed by the user. These files are considered as representations of actual performance and the initial files created by the system at the beginning of the training are considered as representations of expected performance. This reverse process will only be helpful to detect the accuracy of the frequencies of the notes sung, aligned with the corresponding beat. Actual and expected performances are compared using the generated files. A "Maatra"/ a single beat is considered as a frame and mismatches higher than a specific threshold percentages are communicated to the users through a playback highlighting the frames which user needs improvements.

Among the design considerations relevant to training assistance sessions, influence of music teaching and training methods are significant. Capturing the student performance accurately, focusing on one-to-one teaching, encouraging the student each and every time, even in frequent failures but avoiding loss of self control, allowing him/her feel that the system is always satisfied on what he/she performs are some of the major facts among them.

EVALUATION

Final prototype and the research findings was evaluated and verified through real experiments with selected visually impaired students and teachers willing to learn music. The prototype and the methodologies used to evaluate the test results should be clever enough to proceed despite the computer literacy level of the selected users. In evaluating test samples, participants should cover representatives from wide range of age categories as only a fair sample selection can grant enough support to derive acceptable conclusions and inferences through statistical experiments.

Evaluation of actual and expected output format quality by musical expertise was used to evaluate functionality of research model. Main evaluation criteria used in evaluating the effectiveness of the solution from users perspective was error rate, success rate and time spent for an evaluation test. The effective time taken by a visually impaired student to capture and revise or translate "n" (n should be a pre-defined reasonable value for similar type of experiment series) number of music notation lines using the proposed solution and existing way was analysed and compared. Error rates also were calculated for notations selected from different music compositions (depending on density in presence of "Vikriti Swara") and in different environmental conditions (Class rooms, noisy environment, individual participation). Analysing comments by the user on the ease of use of the developed model would also provide necessary means to draw conclusions on the research findings and further development requirements.

Functional evaluation of the training process was done by re-feeding a "Swara" / music note singing synthesized by the prototype. As the system have an intermediary note sequence file for each of its outputs, it can be compared with the intermediary note sequence file generated when same audio was given to the system at training process. Accuracy and error rates was analysed for different music notations owns to selected sample compositions which can represent various natures of music notes.

CONCLUSION

Concluding the paper we are clear that the adaptive music score trainer for visually impaired users in Sri Lanka is proposed via this research is able to address the identified problem under the most appropriate methodology. As the authors we believe that this will be a turning point of the world of music since with this invention, the doors will be opened for more new and talented people to enter the Sri Lankan Eastern music world. As well as not only the people who familiar with Eastern music, but also the music lovers all around the world regardless their disabilities, will be able to enjoy the taste of Sri Lankan Eastern music nourished with the colourful thoughts born in their minds. This adaptive music score trainer of visually impaired can be extended towards a music score generator which assist user to create a visual script of music notations from an audio input.

REFERENCES

Bertoni, G., Nicotra, G., and Quatraro, A. 2003. "Access to music by blind people."

Cont, A., and Schnell, N. 2004. "Training IRCAM ' s Score Follower"

Challis, B.P., and Edwards, A.D.N. 2000. "Weasel: A System for the Non-visual Presentation of Music Notation," *Computers Helping People with Special Needs: Proceedings of ICCHP*, January, pp 113-120.

Dixon, S. 2000. "On the Computer Recognition of Solo Piano Music."

Fujinaga, I. 1996. *Adaptive Optical Music Recognition*. Applied Image, Ink.

Gales, M., and Young, S. 2008. "Time Flies When You're Having Fun: Cognitive Absorption and Beliefs About Information Technology Usage The Application of Hidden Markov Models in Speech Recognition," *Foundations and Trends* (1:3), pp 195-304.

ICTA. 2012. "Digital Inclusion: Equality of Opportunity to the Visually Handicapped through ICT." Retrieved 15 July, 2013, from http://nanasala.org/icta-handicapped.htm

Johansen, L.S. 2009. "Optical Music Recognition," *Thesis for the Degree of Master of Science in Modelling and Data Analysis*, July, pp 37-81.

Kaprykowsky, H., and Rodet, X. 2006 "Musical Alignment Using Globally Optimal Short-Time Dynamic Time Warping Short-Time Dynamic Time Warping."

Kim, Y.E. 2003. "A Framework for Parametric Singing Voice Analysis/Synthesis," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, October.

Kyritsi, V., Georgaki, A., and Kouroupetroglou, G. 2007. "A Score-to-Singing Voice Synthesis System for Greek Language," *Proceedings of the Inter. Computer Music Conference*, August.

Lee, K., and Cremer, M. 2000. "Automatic Labeling of Training Data for Singing Voice Detection in Musical Audio."

Lerch, A. 2008. "Software-Based Extraction of Objective Parameters from Music Performances."

Li, R., Ding, Y., Li, W., and Bi, M. 2012. "Automatic Interpretation of Chinese Traditional Musical Notation Using Conditional Random Field," *International Symposium on Computer Music Modelling and Retrieval*, June.

Ng, A., and Khan, A. 2011. "New Approaches to Optical Music Recognition Automated Instructional Aid for Reading Sheet Music."

Raphael, C., and Wang, J. 2011. "New approaches To Optical Music Recognition," *International Society for Music Information Retrieval Conference*, pp 305-310.

Riley, J., and Fujinaga, I. 2002. "Recommended Best Practices for Digital Image Capture of Musical Scores."

WHO. 2012. "Visual Impairment and Blindness." Retrieved 15 July, 2013, from http://www.who.int/mediacentre/factsheets/fs282/en/

ACKNOWLEDGEMENTS

Late Prof. Wimal Weerakkody, Faculty of Arts, University of Peradeniya and Ms. Shantha Shyamalee, Eastern music teacher, De Soysa Maha Vidyalaya, Moratuwa are reminded with respect and gratefulness for their guidance and support given. Without the great support of students of School for the Blind, the Ceylon School for the Deaf and Blind, Ratmalana it might be very hard to carry out the surveys, so they are also gratefully reminded.

COPYRIGHT

Kavindu Ranasinghe, Shyama Kumari, Dawpadee Kiriella, Lakshman Jayaratne © 2013. The authors assign to ACIS and educational and non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The authors also grant a non-exclusive licence to ACIS to publish this document in full in the Conference Papers and Proceedings. Those documents may be published on the World Wide Web, CD-ROM, in printed form, and on mirror sites on the World Wide Web. Any other usage is prohibited without the express permission of the authors.