DOI 10.7603/s40601-013-0045-6

GSTF International Journal on Computing (JoC) Vol.3 No.4, April 2014

Music Training Interface for Visually Impaired through a Novel Approach to Optical Music Recognition

Dawpadee B. Kiriella, Shyama C. Kumari, Kavindu C. Ranasinghe, Lakshman Jayaratne

Received 27 Feb 2014 Accepted 10 Mar 2014

Abstract— **some inherited barriers which limits the human abilities can be surprisingly win through technology. This research focuses on defining a more reliable and a controllable interface for visually impaired people to read and study eastern music notations which are widely available in printed format. One of another concept behind was that differently-abled people should be assisted in a way which they can proceed interested tasks in an independent way. The research provide means to continue on researching the validity of using a controllable auditory interface instead using Braille music scripts converted with the help of 3rd parties. The research further summarizes the requirements aroused by the relevant users, design considerations, evaluation results on user feedbacks of proposed interface.**

Keywords-visually impaired, Eastern music, music score/notation, singing synthesizing, adaptive music score trainer, profile based approach, optical music recognition (OMR), Sri Lankan

I. INTRODUCTION

Music is a global language which every living being inherently get the ability to comprehend. An alphabet is also a kind of symbolic representation on simple sounds which human can produce. Each sound produced by a living being or nature has its own musical qualities. Music is heard and enjoyed by every human despite the level of capabilities they own. Studying and experimenting on this global language should not be restricted or disturbed due to any reason. The research builds up a foundation to ensure the independent accessibility of music notations created with symbolic representations of musical alphabet. Main focus in the research is about assisting the Sri Lankan visually impaired community to bring up their right to study on eastern music notations. This is a novel approach which explores on methods of converting printed eastern music notations to readable and controllable digitized output embedding the interface with qualities which sighted people experience when reading music notations. The research started its lifecycle with the name “Swarālōka” and will be referred in the paper from that name. The paper focuses more on the interface and user interaction aspect of “Swarālōka”.

Differently abled people have many unrevealed talents more than others. When interacting with them we have to stick to the belief that they are people whom do not lack intelligence [1]. They should be treated with respect and compassion. They may be slow learners but which does not mean that they cannot become experts in any field. It is believed that teaching music is a good way of enhancing the interest to learn in any other things also. That’s why a separate research field called music therapy has born. For differently abled people it is believed that one-on-one teaching is the most effective than one-on-many [1]. Usually humans feel uneasy when they are to perform in front of a crowd. This group which has different abilities would not like them to be judged or compared in front of a crowd. Therefore traditional classroom learning gives less result while more personalized teaching have the ability to put forward milestones in path of progression in learning. The teacher should adapt according to the way how his / her student learns. He/she should let the student know that the teacher is contented about the students learning pace [1] and should reinforce the student.

World Health Organization [2] 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 [3]. 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 [3]. 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

DOI: 10.5176/2251-3043_3.4.294

Published online: 27 May 2014

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:

- Literature Survey on
  - Capturing + recognizing music notations
  - Converting notations to readable formats
- Profile Based Approach to identify Eastern Music Notation Symbols
- Recorded audio based Swara singing synthesizing approach
- Summary on UI design considerations relevant to the context
- Incremental prototype designed for the research (deliverable)

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).

II. Related Work

In the research area of our research exists, the previous researches can be categorized into three categories as:

- Capturing and recognition of music notations
- Converting music notations into a readable format for visually impaired users
- Converting audio to notations (for training purposes)

Since this paper is focusing only in one of research components in our whole research, "capturing and recognition of music notations", only that part of our literature review is presented in this section.

The systems for music score recognition are called as OMR, and it is similar to Optical Character Recognition (OCR) except that OMR is used to recognize musical symbols instead of letters. OMR systems try to automatically recognize the main musical objects of a scanned music score and convert them into a suitable electronic format, such as a MIDI file, an audio waveform or ABC Notation [4]. Though OMR can be used for educational reasons to convert scores into Braille code for visually impaired people, to satisfy their music needs such as generating customized version of music exercises, these systems have been researched and applied only for Western music up to now. Because of the difference of the performance and the symbols of the music notations/scores in Eastern music rather than Western music to capture music scores in Eastern music the same mechanism in OMR cannot be used as it is. But the characteristics of OMR for Western music scores will guide in the process of identifying the most suitable mechanism to capture music scores in Eastern music as a part of our research focused on identifying an output format to represent Eastern music notation which visually impaired people can access effectively. Henceforth, three research approaches on improving the service of OMR will be discussed under this sub section.

According to [4] the aim of their study is to suggest novel methods, or modifications/ improvements of the available algorithms of OMR at the time of July, 2009. Johansen [4] has said that an OMR system consists typically of four main steps as; Pre-processing, Segmentation, Classification and Post-processing. In [5], they had being looked into each of these steps in detail but they have concentrated on identifying the main musical symbols, essential for playing the melody, and the other parts of a music score such as text, slurs, staff numbering are ignored by their system. And also according to them the last part of an OMR program usually consists of correcting classification errors by introducing musical rules and in their study that is only applied to correct wrongly classified pitched for accidentals.

Fujinaga [6] 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 [4] this study also has been limited to facilitate only for music scores in Western music. But, as this thesis [6] of doctor of philosophy has proposed a most completed research prototype to expand the quality OMR as an adaptive system, their research methodology will be a great support in extension of research prototype upto an adaptive system in the part of our research for recognizing the music scores to capture them in the context of Eastern music.

[7] has proposed a method called Description and Modification of Segmentation (DMOS) together with Enhanced Position Formalism (EPF) as an extension of the grammatical formalism for OMR. There, the authors of [7] have come up with another musical scores recognition system but for sheet music in Western context. Hence [7] is for Western music, directly it cannot be used for capturing music scores in our research. However, recognition of musical scores is one among three main research components in their research [7]. But much detailed facts rarely can be found though the concept has been disclosed. Since that it is difficult to use their [7] research idea in capturing of music scores (Eastern music) of our research.

The OMR system in [8] has been aimed toward the score

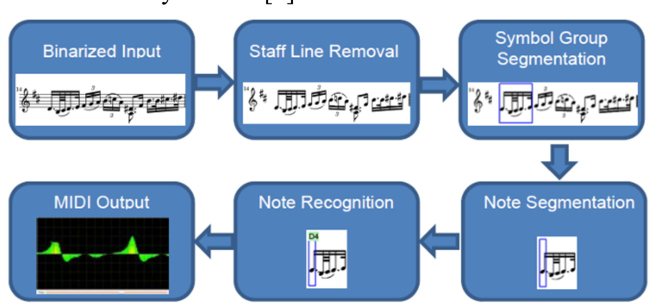

Figure 1. Sheet music reader image processing pipeline. [9]

images of the International Music Score Library Project (IMSLP). And that system mainly focuses on measures as the 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 isolated rigid symbols. They have reconciled these overlapping symbols by seeking non-overlapping variants of the composite symbols that best account for the pixel data. And the results of that system have been presented on a representative score from the IMSLP. Since the consideration measures of music scores in Eastern music is essential as well as Western music while capturing music scores, the focusing function of [7] will definitely guide to come up with an approach for implementation in the context of Eastern music too in sub functions of capturing scores such as "Isolated Chord Recognition[7]". So that is the significance of the study [7] in accordance with our research rather than other studies on OMR.

As demonstrated in [8] which has presented "A connected path approach for staff detection on a music score", the first challenge faced by an OMR system is staff line detection. And according to [8] it is the first task which dictates the possibility of success for the recognition of the music score. They have shown that in the case of handwritten music scores, the existing solutions except [8] are far from presenting satisfactory results. So in this study [8], a new algorithm for the automatic detection of staff lines in music scores was proposed. Since there is not a concept of staff lines on music scores in Eastern music, the research idea of the study [8] cannot be used that much in capturing music scores in the relevant part of our research.

According to [9], that study has come up with a sheet music image processing application targeted towards beginner music to provide an instructional aid for the novice musician. 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. Authors of the research paper have mentioned that this application [9] recognizes and classifies quarter, half, and whole notes with very high accuracy (97-100%) for digitized sheet music readily available on the internet by the paper publishing time, June, 2012. As mentioned earlier in the review of the study [8], since the music scores in Eastern music have not staff lines as in Western music, following the steps which comes after "Staff Line Removal [9]" in the "Sheet Music Reader Image Processing Pipeline [9]" as depicted in Figure 1 may support to achieve of capturing the music scores in Eastern music.

Genfang [10], [11], [12] with his partner researchers have come up with three research prototypes in 2003 to 2009 focusing reading digital images of sheet music. [10] has addressed the pattern recognition problem in OMR [5] by proposing a novel OMR algorithm based on mathematical morphology for Western music scores. According to the authors of [10], the recognition rate is 94% (using near 50 samples of musical score image). After [10], Genfang, Wenjun and Qiuqiu [11] have presented a more facilitated method which can be applied to all music score of Common Music Notation (CMN) in Western Music. At the same time (2009), a new algorithm for capturing to a rectangle box of each connected region in Gong-Che Notation (GCN) score image was presented by the study of [12] while GCN is a Chinese traditional music notation. [12] is focused on image segmentation based on an anticlockwise rotation extension algorithm which is more appropriate for capturing Chinese characters.

Among the researchers who is interested in the research area which covers capturing music scores there are ones [13] [14] who has tried to identify a solution for the problem of using digitally captured music (score) images while capturing music scores. Results of those researches can be used to get support in capturing music scores in our research because the soft copy of the music sheet which we are going to considering in our research prototype may be a digitally captured one using a digital camera as well as other digital formats.

[13] is such a research which has focused to reveal best practices for detail and color capture which are presented for creating an archival image containing all relevant data from the print source, based on commonly defined purposes of digital capture. They have been able to present options and recommendations for file formats for archival storage, web delivery and printing of musical materials.

[14] also has presented a "Fast Capture of Sheet Music for an Agile Digital Music Library". As the authors of [14] have mentioned, they have proposed their research prototype for an agile digital music library because a personal digital music library needs to be "agile", since it needs to make it easy to capture and index material on the fly.

In the paper [15], those researchers have introduced a web platform to create music lessons dynamically and collaboratively, with the assistance of a semi-automatic score annotation module called @-MUSE. To do so, they have described a methodology to design such a platform: Sign Management. Then, the paper has detailed its general architecture as an Iterative Sign Base System based on a common practice in musical learning: score annotation. And also this study has revealed various algorithms to generate relevant annotations on a score in order to explain it. These algorithms are based on the analysis of musical patterns difficulty. As [15] demonstrated, they are implemented within a module of @-MUSE called Score Analyzer. So before analyzing the music scores, they should be captured. But @-MUSE is also for Western music scores. Then, except getting some guidance from this research also, the part under capturing music scores of our research will not be able to use the same function of @-MUSE as it is, since we are focusing on only Eastern music.

If we focus on related works based on the conversion and interface, reliable interfaces are tactile, auditory, and a combination of tactile and auditory interfaces. Gustatory, olfactory or visual interfaces can be ignored in our context. Tactile interfaces always engage with a specific hardware component.

Weasel: A System for the Non-visual Presentation of Music Notation [16] is one of the most attention grabbing researches. But it does not only address the audio output, but also tactile output. Describing the Weasel research briefly; Weasel overlay consists with two main sections referred as bar area and control section. Once when the user presses on the bar area corresponding musical content is played. Circular symbols below the guidelines represent the presence of 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. Weasel had several output formats rather than restricting to one particular output format addressing the research problem in a better way. Basically Weasel research 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 has also considered on some successful solutions in order to overcome the limitations within Braille music representation and talking scores. The survey carried out parallel to the research highlights that existing methodologies used results with large amounts of information presented in a serial fashion which again limit the user to view more information at one glance. It provides a solution for finding out a way to filter unnecessary information. According to the researchers learner's reading speed is directly limited by the speed of the spoken description in talking scores even though prerequisite knowledge on Braille music is not needed. And addressing the problem in western music context seems less confused comparing with 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 converting process.

The major advantage of such a system is that user can get rid of serial information retrieval and can experience roughly an overall idea (through experiencing content within a music page using an interactive interface) on the music score more easily, precisely and quickly. The finding salso suggest that delivering more controllable output is also important. As the designers has concerned more on represent each and every information available in a manuscript notation, no information is lost as part of translation. PVC overlays and Intellikeys touchpad is something which can be studied in detail in order to get the use of that technology if required in our research. The control section in tactile interface which mostly used for output control purposes even can be replaced by a keyboard interface if most reachable keystrokes are wisely selected.

It is a known fact that, converting music scripts to readable format, require proper formatting of music score [17]. And in western music the music scores 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 pointed out. Since [17] suggests the methods to convert poorly formatted page in to a highly formatted page, in standardizing eastern music scripts, we can consider them as options where can get help from. Music XML format and Braille Music XML format are considered as most usable where we can use it for our research study since we are dealing with music notation conversions. [17] discusses points about BMR (Braille Music reader), BME (Braille music editor) and Resonare (Braille page formatter) to facilitate Braille music. Since we still have an idea to facilitate our system with the hand of Braille + Audio output conversions between Braille and digital musicmay be important.

Converting visual script in to speech/singing synthesis is another popular form of conversion. This is also the basic underlying theory that is being used in [16]. In the visually impaired music reading context there are considerable amount of researches carried out in converting visual script to singing. Here we focus on auditory interfaces.

Study [18] attempts to develop a framework which analyze and synthesize singing voice. In order to accomplish the goal it estimates source filter voice model parameters by considering both physical and expressive factors and models the dynamic behavior of these features over time using a Hidden Markov Model. Then while in progress of the research it describes the framework using mathematical algorithms and discusses the features of the framework.

The other application the [18] suggests for the framework is low-bitrate singing voice coding. And it claims that the previous researches which the study was built upon the compression advantage of encoding vowels using static templates. The topic itself may doesn't straight forward in our research project to be addressed since our project does not focus on singing lyrics of a song. But the theories discussed may be useful since still we plan to sing the notation. Hidden Markov Model is defined as "A statistical-tool used for modeling generative sequences characterized by a set of observable sequences." [19] Then this model can be used in our research 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.

Further elaborating the singing synthesizing process, we could surf on another important study based on the area. I.e. LYRICOS [20]. It is a system which uses data driven methods to model the phonetic information in the voice, resulting in an output that assumes the voice identity characteristics of recorded human vocalist and employs a high quality sinusoidal synthesis method. Voice data corpus used in the system is collected exhaustively by audio singings of 500 non sense words by trained vocalist. This approach provides support in synthesizing more natural singing than synthesized wave forms using algorithms. According to the target specified by the MIDI input, best inventory units are selected and the concatenation is done within the sinusoidal model frame work. These frameworks defined in LYRICOS exemplarily reveal methods to create synthesized singing output with high quality.

In deep experimenting the singing synthesizing field; the study [21], is discussing about proposing a score to singing synthesis system and as the method, score will be written in a score editor and saved in the MIDI format. As the study suggests Singing Voice Synthesis (SVS) become a trendy last few decades and it discusses about the urgency and significance of using SVS systems while describing the characteristics of a SVS system. At the same time it contrasts and compares speech and singing processes as well as the vocal music generation and instrumental music generation. The study discussed about the researches done to improve the natural quality of the sounds generated.

Since the system [21] is generating singing synthesis system researchers have discussed more about generating an audio out put through MIDI file. And since our research project concluded as singing output system, then using software like MBROLA and MIDI file formats and converting methods will be important to consider about. When comparing the [21] with our proposed research project the only and 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 concatenate synthesis in singing synthesize is discussed in [21] and the previous project failures happened in the past in synthesized singing and the reasons for failures. 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 a big issue, but still there can be such barriers have 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 project that's 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 consist of Sinhala diphone data bases. As an option for the above mentioned weaknesses of using existing diphone databases, the research study focuses in creating a new diphone database which we can use for Sinhala music notations.

In deep surfing score to audio conversion there are considerable amount of researches that being carried out with dynamic time wrapping to score to audio alignment. As both [22] and [23] suggests it is one of the best ways to minimize the error rate that can be occurred in converting score to audio.

The research [23] presents a novel algorithm to align two sequences by time wrapping them optimally minimizing the weaknesses of Dynamic time wrapping those are cost and memory limitations using iterative approach. The evaluation is done between a score in MIDI format and instrumental performance of the score as the audio. The study [23] also something similar and connects with our research project because the new algorithm "Short-Time Dynamic TimeWarping" (STDTW) had evaluated in aligning musical score and the time axis of audio performance of it. So the aligning results may be important for us in measuring the accuracy of the conversion music score to audio. As 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 than written in the score and in polyphonic music it says difficult to understand the note playing. Relating these problems in to our research study, since we are addressing only monophonic music it can be assumed identifying the note playing can be not that difficult. And since our research study will be done to vocal music not to instrumental music, the differences between audio file and written script must be minimum compared to instrumental music. But still problem of identifying the exact tempo is valid to be considered.

Since the core of the study is highly focused on mathematical algorithms opportunity of usingthe content to our research is very limited. But the obviously the end product of the research (i.e. STDTW) may be worth at the evaluation of the project. The research study [24] is attempting to give a solution for the problem of reading a Chinese music score type which does not present the duration of each note in the music script (rhythmic immeasurability problem) by automatically interpreting it in staff. In order to achieve this they have first concluded 37 rhythmic patterns and then using HMM and mathematical calculations have been done.

One of the main problems they have faced in addressing the problem is different scores use different rules in music elements in Chinese music. But in Sri Lankan eastern music field there is an acceptance of a common interpretation of music elements like pitch, rhythm which doesn't raise this problem in our research study. And the study is focusing on converting text to another text. But in our project we are mainly focusing in text to audio conversion, but if the research study concludes Audio + Braille output, this study will be much more effective in converting notation to Braille. The study discusses that the interpretation has two parts rhythm and pitch (two basic characters of a music notation). Same as in Sri Lankan music the notation interprets pitch and thaal symbols.

One of the key points is that the paper describes is they have used the features of the notation like notes sequence (NS), numbers of notes (NN), pitch interval position and direction (PIDP) and their combinations: NS+NN, NN+PIDP, NS+PIDP, NS+NN+PIDP have used to develop the interpretation model. This is mainly used because the described notation structure does not give the measure of rhythm of the script. But in eastern music Sri Lankan context we don't have that problem but still we can use some of these techniques to develop the model.

Converting the visual score in to vibro-tactile is another common and effective methodology to give the sense of the score to visually impaired people. In this context, [25] is one of the most attention grabbing research that was being conducted. It is a well-known factor that if the visual score is being converted in to vibration then to sense the output it is obviously requires specific hardware. So in Sri Lankan context, with the economic aspects it is less feasible to spend much on specific hardware, but still the methodologies suggested remains worth to survey on.

The research [25] suggests a mechanism to read the music score with vibrations and as an extra feature Vibro-tactile Score Editor (Vib-Score Editor) that fully supports the vibro-tactile score with intuitive graphical user interface is also being developed. Vibro-tactile is being considered as one of the most efficient HCI mechanism and the researches of [25] tested the potential of common VibeTonz system and vibrotactile actuators such as piezoelectric actuators.

Apart from that the software tools like Hapticon Editor, Haptic Icon Prototyper, VibeTonz studio, and posVibEditor is being considerd.

The research suggests converting music scores in to particular vibration signals and same as Swarālōka does but output format only differs. Apart from that the research uses XML to store data. As both [25] and [16] suggest, if we utilize the vibro-tactile output format obviously the learning curve decreases since visually impaired people already practiced to read things by tactile (Braille reading). On the other hand the output format becomes comprehensive compared with audio format but time consuming. I.e. giving two different qualities of the sound simultaneouslywith vibro-tactile is bit confusing. But in audio output format it is not that confused and Swarālōka already does it. As [25] concludes, Vib-ScoreEditor has another metaphorical feature, vibrotactile clef (analogous to the musical clef), to make the process of composing vibrotactile patterns decoupled from the process of considering the signal-level characteristics of the vibration. This suggests that there can be some basic level music symbols that can be define a corresponding vibration signal and there can be some other which cannot be. And the mechanism to be followed to generate vibro-tactile output is confusing, as both [25] and [26] suggest. [27] study presents a detailed description of an evaluation carried out of a vibro-tactile pattern design using vibro-tactile score. One of the common characteristics that we have notices is all most all these researches carried out in vibro-tactile perspective is consist of a GUI whose purpose is questionable in the context of Swarālōka. Since Swarālōka is developed for visually impaired people it is not essential to focus seriously on GUI.

Apart from that the research [27] suggests two basic mechanisms named "Wave form editing" and "Score editing". Though the output format is a vibration, they first generate a wave form of that vibration signal. This point is critical to Swarālōka, since we always need to generate an audio signal as the output. Converting visual script to thermo-tactile output is another innovative method of presenting music notation to visually impaired people. In this discussed context there were significant researches carried out.

The study [28] is one of such researches which attempts to develop a thermo-score display dynamically changes the temperature of the instrument according to the frequency of the notes. The MIDI signals were converted to temperature using MIDI-to-temperature converters. To generate and control the temperature researchers have used Peltier device. And the device is placed on the instrument where the performer get the music signals through the instrument itself and using those bio-signals the performer generates the music with the instrument.

And the system discussed by [28] basically controls the time which performer plays a note. The system is mainly considered about keyboard playing and making one key hotter than the temperature the performer leads to pull finger by reflex action which makes the note plays short time. The research is important to us since it develops a non-visual output methodology. Basically this system is generated to use in instrumental performance in music which always requires the music instrument to sense the music notes. But in our research project we are not only focusing on instrumental music we want to facilitate the vocal music even (singing the music notes as in eastern tradition).

The study highlights an important point that is it claims a performer always conducts music information processing which implies the feedback the performer gets both from the audience and the instrument affects to the performance. The paper discusses it in detail with figures and the theory will be important to us to decide whether our final product should give the performer instant sense of the notation while playing or should there be a time gap between the notation reading and performance. I.e. will it be much more successful if we produce a system to convert the music notation in to readable format and make the performer memorize it and then make him to perform?

The instrument discussed in the study make the frequently used keys hotter than others. In order to do that it first creates a chroma-profile (chart representation of frequency of pitch notations computed from MIDI data). This chroma-file generation may be useful in our research too since whatever the output format of our research project having a quick summary of the frequency of music note occurrence in the script must be important information. The paper discusses further more details that can be given to the performer using the device. It suggests with the same device giving the audiences excitement. This feature is not much critical to our research study, but as it is discussed in the paper it may be a redundant feature to the device. It depends on the performer. Since there can be performers whose performance may be less when user excitement goes up, if so even the average performance even may not achieved as expected through the device.

One another important factor highlighted by the literature was that auditory interfaces are more appropriate when temporal tasks are engaged and combination of auditory-tactile signals does not come up with significance enhancement in communicating information.

III. Methodology

To address our research question, Swarālōka delivers a non-visual presentation of sheet music which can be easily grasped by visually impaired users.

Audio format has been identified as the most optimal readable format (discovered as a research finding through the interviews, observations and surveys conducted) among other possible non-visual output formats such as hepatic/kinesthetic, gustatory and olfactory. Hence, the output file format of Swarālōka is audio format.

The basic functionalities of the system are:

- Convert any eastern music script in Sri Lankan context into the relevant audio clip
- Facilitate user with navigating forward backward or any point through the music script to hear necessary part of audio
- Generate an audio file with several audio layers and user should be able to enable and disable different layers of the audio file and listen to them in a preferred way
- Allow user to save notes as user preference and get rid of time consumes re-processing
- Ability of controlling the tempo of the output format and return back to default tempo (while navigating)

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 (create the input soft copy) of a sheet music (hard copy)

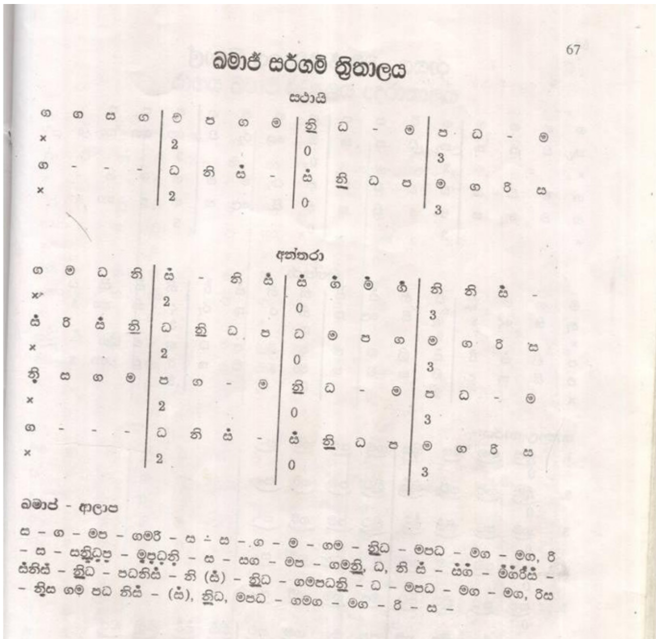

Figure 2. Original Image source.

- Phase 1: Zoning and Segmentation of the image of sheet music
- Phase 2: Adaptive OMR - Recognition of atomic symbols
- Phase3: Converting into the readable format (audio format)

A. *Capturing the image (create the input soft copy) of a sheet music (hard copy)*

Inserting a softcopy of a sheet music (as depicted in Figure 2) is supported by the system through a scanner or camera which can deliver a digital photograph. In user perspective, visually impaired users are more familiar with taking digital images with scanner. So, in this research that method was used.

B. *Zoning and Segmentation of the image of sheet music*

Overall process of symbol recognition (Figure 3) can be considered as two parts as zoning, segmentation and OCR (adaptive OMR) processing. When image was fed into the system, at the pre-process stage image will be enhanced to a state which each interested feature appears with more exaggeration. Image will be filtered by a threshold value and converted to a binary format which each of the pixel would have either black or white. The input image may be a skewed, noised image. Vertical lines of the table structures will be recognized and used in de-skewing the image in order to correct orientation as showed. Each part of the table will be recognized separately. For a single session, tabled zones are detected and saved separately in a temp location as separate files. Eroded versions of the same images are also saved separately. Eroded images are used to infer the boundaries of the rows through vertical projection.

One of the most successful methods in zoning was using differently processed clone of the image (Figure 4) to detect blobs and to extract the region of interest. To detect the rows and characters together with all important symbols, we have to erode the image. Otherwise some of the symbols which help in

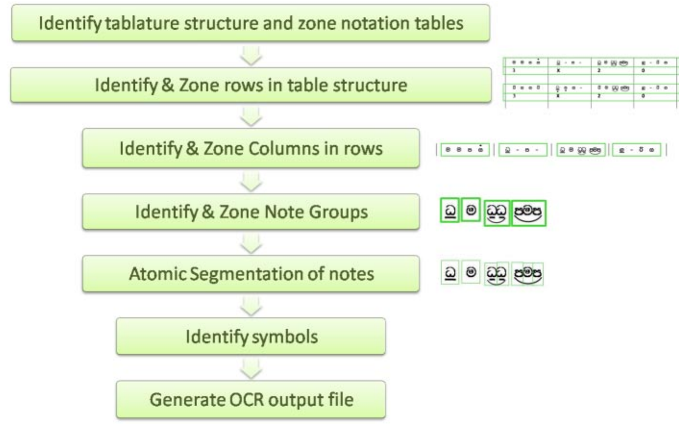

Figure 3. Overall Process of Symbol Recognition

creating composite notation information accurately will not be extracted correctly. But eroding the image contaminates the original source and may result in providing false and incorrect information. Therefore a clone of original source is used to erode and detect blob regions. The original source is used to extract information using detected coordinates. These UIs are only available in developer testing mode. Each row in a table will be then saved as separate images in a temporary location. These images relevant to rows of the music notation tables will be the final image version for composite symbol blob detection.

C. *Adaptive OMR: Recognition of atomic symbols*

Recognition of atomic symbols (blobs) is done using adaptive OMR which is newly introduced in this research. Our OMR has been developed on top of a novel approach called profile based recognition approach (this is described in detail under fourth section in this paper) which is a research finding of ours.

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 will contain 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.

D. *Converting to the Readable Format (Audio Format)*

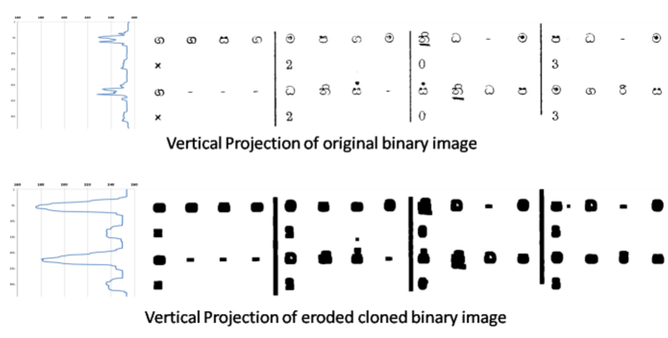

Figure 4. Original binary image and cloned binary image.

The first level prototype was developed in a way which user can get an idea on the output format of the solution. 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, 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. Layers in the next level prototype also were the same and can be illustrated as depicted by Figure 5.

Actual implementation design was not that easy as we have to keep the synchronization between the layers. OMR output is used to create intermediary files for thaal and and notes and the whole conversion process depends these two intermediary files. If we consider the basic layers in output format, audio clips for each note/swara and thaalakshara (played by tabla) will be kept in a sound bank. We observed that when one note sung for more than one maatraas, only the vowel sound is sung in the stretched maatrasas implied by Figure 6. Therefore this effect of stretched maatras can be successfully replicated using two audio clips recorded with initial consonant sound and stretched vowel sound. As each note has different unique frequencies every note has to be recorded with these two morphs. Having fixed length audio chunks have many benefits. It helps to easily synchronize between layers. When more than one sound played simultaneously, there should not be a delay in start points. If we used lengthy audios or lengthy data lines start point delays and next sound clip starting delay will be significant. But when small chunks of audios used, these delays become inconsiderable even though they are present. Looping lengthy audios also does not work due to the same problem.

According to the number of maatraas in a line and thaal symbol sequence, relevant thaal is determined. Thaalakshara sequence owns to each thaal (for initial research model, Dadara, Jap thaal and Theenthal are considered) is kept in separate properties files. In the final step, chunks of audio in the sound bank relevant to the finalized thaalakshara sequence and note sequences will be added to separate data-lines and

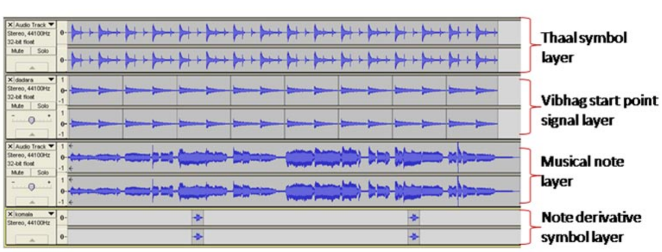

Figure 5. Layers in Prototype output

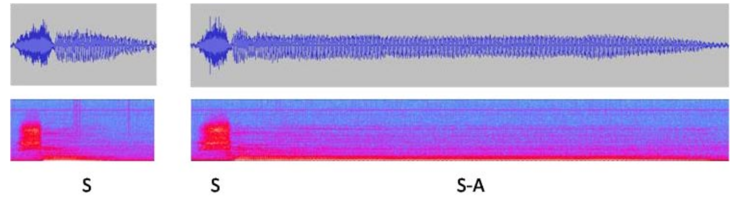

Figure 6. Replicating stretched maatra

played concurrently. This will be a demonstration of the input visual script of Eastern music notation as it is.

Separate signals will also be played to convey derivatives of notes. A harp like sound from the relevant frequency of derivative note will be played in the background in order to distinguish between shuddha (natural) and vikruthi (sharp / flat) notes. These signal audios also have the same length relevant to duration of a maatra.Start points of vibhag (sound of thaalam-pota arranged according to thaal symbols) also will be played if the layer is enabled and again have the same length audio chunks. Sometimes novice users may find it difficult to identify notes in higher octave and lower octave. If higher and lower octave help has been enabled notes in higher octave will be played from speaker in right hand side only and the notes in lower octave will be played from left hand side speaker only while middle octave notes will be played from both speakers. As a requirement came from the users another note information narration layer has been provided. This allows the users to get information about each note through a narrated voice. Basic conversion process is depicted by Figure 7.

Design had to be disciplined with some specific considerations. Introductory sessions will be 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 will 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 be given through simple keyboard strokes. User will need initial guidance in getting used to key–function mapping. Introductory sessions for users are explained here. The users are not required to have a sound knowledge on music to follow this research solution. As the solution is using distinct audio symbols to represent semiotics in a music notation, users need to have knowledge on the usage of those symbols as a pre-requisite. The introductory session should fill this knowledge gap in a way which user does not feel as if he/she is learning a new language. Auditory symbols used to represent derivatives and ornamentation is explained here. User should be able to easily distinguish between auditory symbols used to represent derivatives of music notes. If more than one derivative effect on same note, the overlapping and composite or collective effect should still be sharp enough to distinguish and identify each auditory symbol separately. Auditory symbols used to represent derivatives of music notes should be carefully selected considering on the meaningfulness which they will offer when listened or heard. This fact is also true for the auditory symbols used to represent music ornamentation.

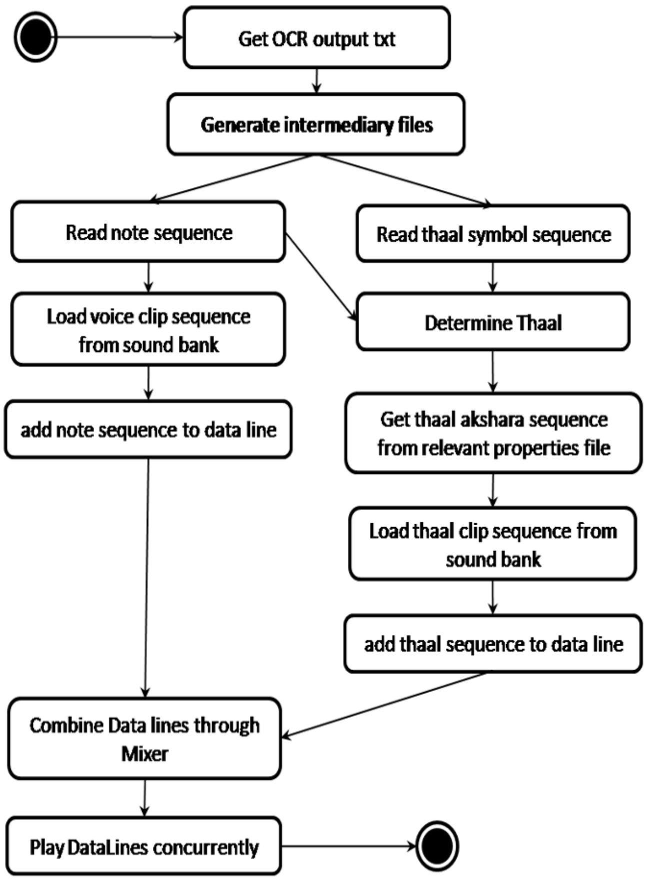

Figure 7. Converting intermediary files to output format

Following are some of the considerations taken into account when selecting representative auditory symbols. Considering Thaal symbols, symbolic auditory chunk which represent main thaal position in a row should different from the chunk used to represent rest maatras. A consistency in the use of these symbolic auditory chunks will help the user to get in to track easily without confusions.

Considering transitions, while user navigates through a music notation (jumping from one point to another, move back to last column or row, move forward to last column or row, switch between verses or scripts) audio effects and clips should be used in a way to imply status changes. One of the major fact in this navigable approach is that user is capable to go back to the previous state easily. User is always given the information about the row that is being played through a narration to make sure that he/she will not get lost the cognition of the point he/she is, in the middle of a script.

Conceptual cognition created in users mind should not violate. As the system is fully focused on eastern music notations, the conceptual cognition or image created in users mind when interacting with the prototype (listening, hearing, following readable format of music notation generated in the research solution and even in undergoing introductory sessions or user manual instructions) should not disturb the feelings aligned with eastern culture and peacefulness intrinsically inherited to eastern music.

Controls should be given through simple keyboard strokes which is limited for not more than 8-10 keys (next [Fast Forward], back [Rewind], select [in], deselect [out], Mark Start, Mark Stop [escape], Option Menu, move level up, move level down). Though getting used to key – function mapping requires initial guidance; it will be more accurate and less ambiguous compared to enabling voice commands.

All the instructions and audio signals were recorded and kept within a separate sound repository for instructions. Relevant instructions were played depending on the context and key event triggered by the user. User is given 3 modes to navigate within the notation script. They are NbN : Note by Note, VbV : Vibhag by Vibhag and RvR : Row by Row. User was given privileges to enable and disable signal layers in default play back mode which is RbR. In NbN mode thaal symbol layer and vibhag separation layer are kept automatically disabled. Left and right arrow key strokes are used for forward and backward navigation.

In implementation, the intermediary files are read using IntermediaryReader class. To map audio files in sound bank NaadaJanaka, ThaalaJanaka, ThalamJanaka and VikrutiJanaka classes are used. To add the audio file sequence defined in intermediary files Sequencer classes are used.

IV. PROFILE BASED APPROACH

Our newly introduced adaptive OMR is one of the major contributions of our research in this research area. It has been developed on top of profile based approach for symbol recognition which is discovered by us using the source reference of a web cam Sudoku solver program by Banko [29].

In our OMR, after detecting blobs, characters on the image of sheet music, we considered some qualities specific for our research problem. In our domain we only have a few numbers of characters. In the set of composite notation symbols which has to be identified after segmentation can be minimized when signs used to indicate Saptak and derivative of music note is identified separately. Left part will be just seven characters. For such a limited domain we can use simplified character recognition algorithms. The qualities we observed in the characters to be identified are as follows.

- There is less number of characters to be identified.
- All the notation symbols (which are letters in Sinhala alphabet) have a specific shape; all of them had curves with specific shape.

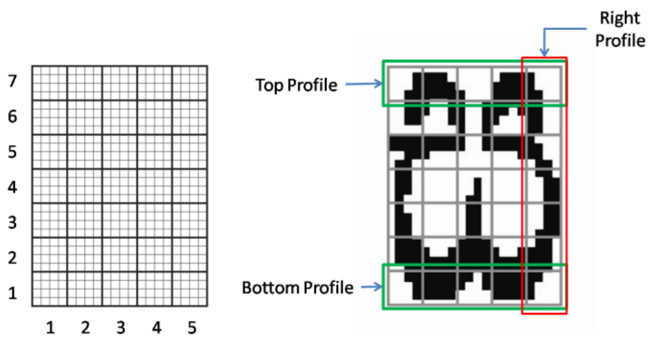

Figure 8. Basic profiles in a normalized sub image

- Each specific shape can be categorized into distinguishable classes by using corner profile shapes (Specifically top profile, bottom profile and right profile).
- Horizontal profiles can be divided into five specific regions.
- Vertical profiles can be divided into seven specific regions.

From the preprocessed raster images we observed that when considering profiles, we can avoid exhaustive matching of each and every pixel. Only profile pixels at the corners need to be examined after normalization. Avoiding exhaustive comparison of each and every pixel enhances the performance of the process. Each sub image within a blob is resized to 20 * 28 pixel image. Three basic corner profiles and one optional horizontal profile are detected and calculated in recognition process. They are bottom profile, top profile and right profile as illustrated in Figure 8. Figure 9 illustrates how identification process flows.

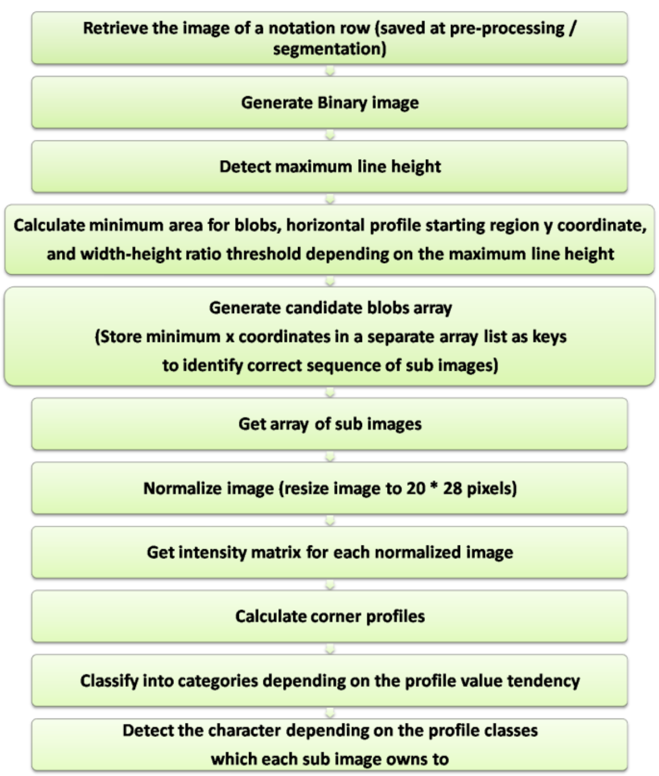

Figure 9. Character identification process flow

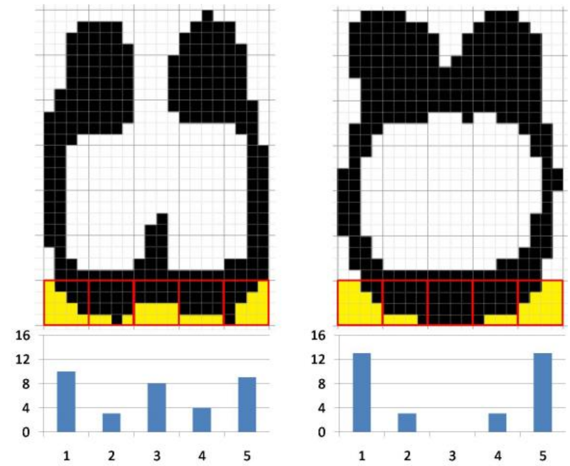

Figure 10. White pixel density distribution in bottom profile

Calculating corner profiles, identifying characters and deriving intermediary files each sub image is 20 * 28 pixels in scale and therefore horizontally pixels can be divided into 5 columns each column having 4 pixels width. Each 4 rows in image are considered as a single row. Therefore we have 7 conceptual rows which each row has 4 pixels height. The region which will be used in calculating top profile is the first conceptual row (4 * 20). This row has 5 columns, each column having 4 * 4 pixels square. White pixel density is calculated per each 5 squares as depicted in Figure 10. This is done in a specific way. Counting white pixels (having intensity value 255) always started from top to bottom in top profile analysis. Once when a black (having intensity value 0) pixel detected counting for that column will stop and cursor will move onto next column. Therefore only the white pixels outside the characters are counted. Resulting array will have corresponding white pixel count for each 5 squares. Fluctuations in these white pixel density values are used to detect number of curves in corner profiles. If one curve detected, profile is considered as owns to "Class A". If more than one symmetric curve detected, profile is considered as owns to "Class B". The profiles which have more than one asymmetric curve own to "Class C" and all the rest profiles will be categorized under "Class D".

Rest basic profiles (bottom profile and right profile) are calculated in same way. The only difference in bottom profile calculating is starting to count from bottom of the columns. In right profile even though we have 7 rows (7 * 4 pixels) only the 5 rows in the middle are considered omitting the uppermost and lowest row. In right profile calculating method, counting white pixels starts from the rightmost pixels and are counted row by row. For horizontal profile (which is an optional profile) actual white pixel counts are calculated and the specific conceptual row is inferred depending on the maximum line height. In the array of values relevant to horizontal profile, only the 3 squares (4 * 4 pixels) in the middle are considered. If the value in Left Square or Right Square does not deviate significantly from the value in the middle pixel profile is considered as owns to "Class G". Else the profile owns to class "X".

When resolving characters relevant profile conditions are checked according to the most optimized sequence. Therefore it becomes not necessary to calculate all profiles for each

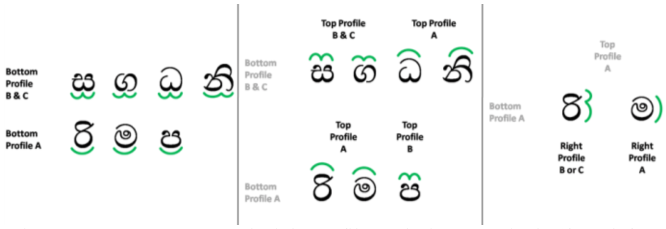

Figure 11. Bottom, Top and Right profile analysis respectively (from left to right)

character. As presented in Figure 11 (first left illustration), bottom profile is the one to be considered in first level. Figure 11 (middle illsutration) implies how the candidate images are further classified using top profile. Right profile (Figure 11: right illstration) and Horizontal Profile (Figure 12) are used to finalize the classification.

Figure 13 explains how these profiles are used to uniquely identify each character. While going through developer testing, threshold values are fine tuned and adjusted in a way which suits to all general scenarios addressed in the research. Apart from the basic profile those characters which needed to be treated differently are verified using width-height ratio and the optional profile.

After the composite symbols have been recognized, intermediary files are created by going through an OCR process. The two intermediary files created for music notes and Thaal symbols will be saved and user can directly use saved data when they use the system next time.

User will not have to bother on any of these steps as system will only convey the final outcome of the process. Developer testing mode and temporary image files created at the middle of the process allowed the developers to examine what actually happens when various types of test data has been inserted as input images.

V. Research Evaluation

Functional testing of the Swarālōka was done in three stages. The OMR was tested using images of sheet music in different sequences and frequencies. Mostly existing fonts (both font type and size), “IskolaPota” (a UNICODE font) and FMAbhaya (an ASCII font) insizes of 11, 18 and 44 considered. Accuracy level for each character was calculated in testing.

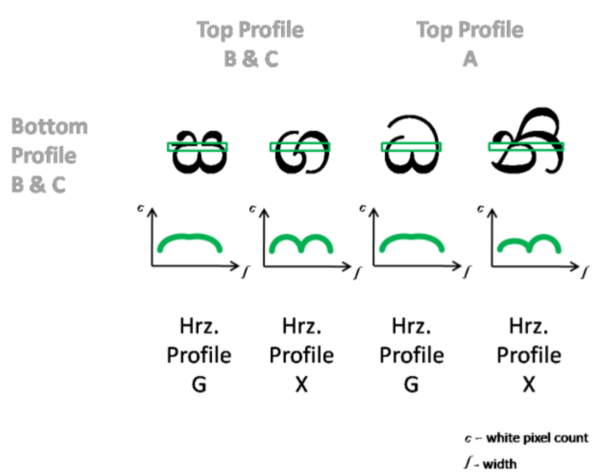

Figure 12. Horizontal profile analysis

Character	Bottom Profile	Top Profile	Right Profie	Horizontal Profie
ස	B	B		G
රි	A	A	B	
ග	B	B		X
ම	A	A	A	
ප	A	B		
ධ	B	A		G
නි	B	A		X

Figure 13. Identifying characters uniquely using profiles

Other than functional testing, usability testing was carried out with the participation of 16 students from Blind School, Ratmalane, Sri Lanka. The testing environment was set up at the computer laboratory in the school. Eastern music notations containing “Sargam” of Raag “Bilawal”, “Bhupali” and “Khamaj” was used in testing and “Khamaj” was taken in order to examine how students can identify derivatives of music notation through the auditory output in the prototype. 10 Female students and 6 Male students participated in testing who are in the age range of 12 yrs. – 16 yrs. All of them were given the music notation to be written in Braille while one of the teachers is reading and were asked to read it again using the written notation. This is to replicate the environment in manual system they are using. Same notation was given to the system and students were asked to repeat the notation while navigating it using the system. To avoid the influence that can be occurred by getting used to the notation, 8 of the students were allowed to engage in manual process first and rest 8 were asked to engage with the prototype first. All students who were selected studied both IT and Eastern Music. 10 of the students were given “Sargam of Raag Bilawal”(Jap Thaal) having 4 lines of 10 notes (40 notes). They used slate and stylus in writing with Braille. 4 of the students were given “Sargam of Raag Khamaj”(Theen Thaal) having 2 lines of 16 notes (32 notes) and 3 were given “Sargam of Raag Bhupali”(Theen Thaal) having 2 lines of 16 notes (32 notes). Time consumed in manual process and when using the system was calculated. Other than that depending on how the students have recognized the notes error rates were calculated for each test.

In OMR evaluation, test results of the profile based OMR algorithm prove that the method can affirm 85.71% of overall accuracy when considered both Unicode and ASCII fonts. Only for Unicode font (Iskola pota) it gives 95.23% accuracy and for ASCII font considered (FMAbhaya) it gives a 71.42% overall accuracy. Accuracy level of the method used in the research can be enhanced by adding more rules and post-processing process to match against a probability matrix which defines the probability of one music note after the other. Different font sizes (11, 18, and 44) were considered in calculating the accuracy percentage.

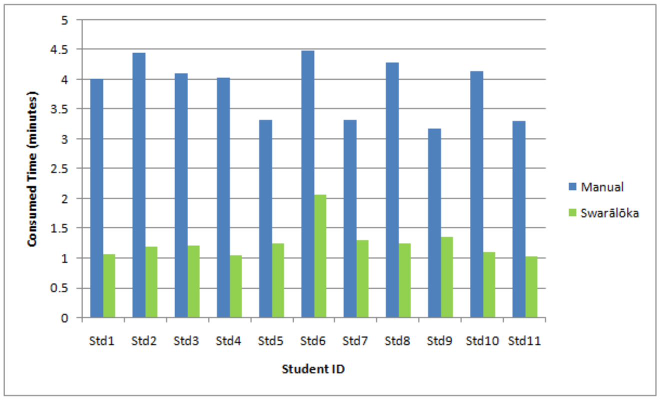

Figure 14. Consumed time comparison

Through the results obtained on success rate of recognizing each character and overall accuracy level it is evident that this profile based character recognition approach can be used successfully when the number of characters are limited in the domain of a research.

In evaluating user samples, participants covered representatives from more than one age category as only a fair sample selection could 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.

Figure 14 shows time the graph with time comparison between manual method (Braille) and the proposed system for a selected script of music notations ("Bilawal Sargam"). According to the final results of user evaluation against time factor as shown in Table I, the proposed system saves users time about three times than manual method.

In version 2.0 (v 2.0) of the system users were given an option which ambiguous notations can be verified using a speech description given on each music note. And those notes which were heard with little difference which make the students difficult to distinguish were re-recorded having more exaggeration of pronunciation. Therefore accuracy level could be enhanced. We can derive that automated script recognition process (v 2.0) supports to enhance the accuracy that can be occurred in translation process because users can easily go a step backward than in manual process. The results also suggest that one of the critical factors in delivering a synthesized singing of notation is the pronunciation of the syllable. Table II shows error rate comparison between the manual method (Braille) and the proposed system.

TABLE I. CONSUMED TIME COMPARISON

Used Method	Consumed Average Time
Manual method (Braille)	3min 7sec
Swarālōka	1min 31sec

min = minutes, sec = seconds

TABLE II. ERROR RATE COMPARISON

Student ID	Manual Method %	Swarālōka V 1.0 %	Swarālōka V 2.0 %
Std1	2.50	15.00	0.00
Std2	5.00	10.00	0.00
Std3	2.50	7.50	0.00
Std4	0.00	5.00	0.00
Std5	5.00	15.00	0.00
Std6	0.00	2.50	0.00
Std7	22.50	12.50	0.00
Std8	0.00	10.00	0.00
Std9	0.00	12.50	0.00
Std10	0.00	0.00	0.00
Std11	0.00	15.00	0.00
Std12	0.00	25.00	0.00
Std13	0.00	15.63	0.00
Std14	0.00	28.13	0.00
Std15	3.13	9.38	0.00
Std16	0.00	10.00	0.00

V 1.0 = version 1.0, V 2.0 = version 2.0

As a summary feedback of all students on usefulness of Swarālōka compared to manual method (Braille) in studying and training on Eastern musical scores is depicted in Table III.

Students were very enthusiastic about using the system to study music notations. The feedback given by them confirms that they are very comfortable with the novel way proposed by the proposed system. All the test results evidently suggest that this type of a navigable and controllable auditory output can be used more efficiently to assist visually impaired students to study on music notations (Compared to Braille music notation scripts). Visually impaired students can grasp auditory outputs more comfortably and it is a more convenient way to remember things. Success rate of identifying signals given to indicate derivatives of music notations proves that concurrent auditory signals in same frequency but different timbre can be recognized more quickly and easily in a controlled environment (less noisy) compared to several tactile symbols denoting all information of a particular music note. As overall time taken to music notation script translation into Braille with the help of a third-party can be reduced in a significantly using the proposed solution it also helps to enhance the speed of learning of the visually impaired students as method is more customized for them compared to traditional method.

TABLE III. OVERALL USER FEEDBACK

Feedback	Users Given %
Swarālōka is more useful than manual method (Braille)	81
Manual method (Braille) is more useful than Swarālōka	13
Usefulness of both methods is same	6

VI. Discussion

Discussing the limitations of the research, the following can be mentioned. The current research prototype only supports printed music notations written in standard tabular format (handwritten scripts are not supported). OMR function has not been tested against all the existing font faces. Exceptional notations other than standard notations used in applied music still cannot be identified through the research prototype. Swarālōka does not assist to translate lyrics embedded in notation scripts. Some of the complex music ornamentation symbols rarely seen in the scripts are not taken into consideration. But since the system can be easily extended with more signal layers, the limitation is avoidable for some extent. The research prototype has been only designed for visually impaired people those who do not have multiple disabilities (hearing disabilities).

VII. Conclusion and Future Work

The research objectives have been successfully achieved throughout the research via our research contributions as follows.

Most convenient format that visually impaired people can use to understand Eastern music scripts is a controllable audio format (Compared to Braille music scripts). In segmenting notation rows, having one image to detect region of interest (ROI) and another to extract content avoids loosing information signs owns to composite symbols. Profiles are used mostly in post-processing in character recognition, but the implemented solution proves that it can be used as basic recognition process when there is limited number of characters. Audio chunks in a sound bank with a fixed duration can be used to synthesize more naturalistic rhythmic singing compared to *Digital signal processing* (DSP). Accurate pronunciation of syllables becomes critical in a synthesized Swara singing to distinguish between each unique syllable. A Swara singing which stretched up to more than one Matras can be successfully replicated combining audio chunks relevant to consonant sound and vowel sound stretched. Concurrent auditory signals can be used to convey composite information more effectively (compared to special tactile signs). Assistive technology (in our research context) becomes more usable when general hardware is used which is affordable and accessible. Capturing user inputs through frequently used keyboard strokes and easily identifiable keys allows maintaining a greater accuracy and less effort in initial training. Concurrent auditory signals can be used to convey composite information more effectively (compared to special tactile signs). Capturing user inputs through frequently used keyboard strokes and easily identifiable keys allows maintaining a greater accuracy and less effort in initial training. Assistive technology (in our research context) becomes more usable when general hardware is used which is affordable and accessible.

Swarālōka has allowed the following mentioned future work in order to feed the research area more and more. Post processing of music notation symbol recognition can be further enhanced by using a well tested probability matrix previews probability of one note following another (which can be used like a lexicon which helps to enhance word level accuracy of a recognized set of characters). By comparing a data set which includes notes own to each Raag against the frequency of notes (in descending and ascending order) on recognized script we can enhance the accuracy level further while providing more intelligent information about notation scripts to the users. Research prototype can be enhanced with functionality which assists users to memorize the music notations and have a feedback on its own singing. Assistance in capturing music notation scripts will help the users to be more independent and convenient. A new definition language for music scripts has been put forward through the research having the potential of extending as an assistive music composer.

Acknowledgment

Late Prof. D. P. M. 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 and usability testing in order to carry out successful research evaluation. At the same time Mr. Amila Indrajith, Music teacher, School for the Blind, The Ceylon School for the Deaf and Blind, Ratmalana, Ms. Padmakumari Karunatilake, Mr. Sarath Kumara IT teachers, School for the Blind, The Ceylon School for the Deaf and Blind, Ratmalana, Mr. Asoka Weerawardena, IT Instructor for the visually impaired, University of Colombo and Ms. Aruni Samarasinghe, Eastern music teacher, Kanuwana Vidyalaya, Ja-Ela will also be reminded with gratitude for their kind assistance. Mr. B. R. Dassanayake, Dean, Faculty of Music, University of the Visual and Performing Arts, Ms. Sandya Kodduruarachi, the principal - School for the Blind, the Ceylon School for the Deaf and Blind, Ratmalana, Ms. Indu Bandara, Director, Census Division, Department of Census and Statistics and Ms. Mekhala Gamage, famous vocalist also helped a lot to make this effort a success.

References

[1] M. B. Lewis, "Teaching Students with Learning Disabilities and Physical Impairments," 2003 [Online]. Available: http://www.marthabeth.com/learning_disabled_students.html [Accessed: 26-Jan-2014].

[2] "WHO | Visual impairment and blindness."

[3] "Digital Inclusion: Equality of Opportunity to the Visually Handicapped through ICT." [Online]. Available: http://nanasala.org/icta-handicapped.htm [Accessed: 26-Jan-2014].

[4] S. Johansen, "Optical Music Recognition," no. July, 2009.

[5] I. Fujinaga, "Adaptive Optical Music Recognition".

[6] B. Couasnon, "Dmos: A generic document recognition method, application to an automatic generator of musical scores, mathematical formulae and table structures recognition systems." Document Analysis and Recognition, 2001. Proceedings. Sixth International Conference on. IEEE, 2001.

[7] C. Raphael and J. Wang, "New approaches to optical music recognition," no. Ismir, pp. 305–310, 2011.

[8] J. S. Cardoso, A. Capela, A. Rebelo, and C. Guedes, "A connected path approach for staff detection on a music score," pp. 1005–1008, 2008.

[9] C. Genfang, and X. Shunren, "The study and prototype system of printed music recognition." *Neural Networks and Signal Processing, 2003. Proceedings of the 2003 International Conference* on. Vol. 2. IEEE, 2003.

[10] C. Genfang, Z. Wenjun, and W. Qiuqiu, "Pick-up the musical information from digital musical score based on mathematical morphology and music notation." *Education Technology and Computer Science, 2009. ETCS'09. First International Workshop* on. Vol. 1. IEEE, 2009.

[11] C. Genfang, and Z. Wenjun, "Chinese Gong-Che notation musical score image segmentation based on an anticlockwise rotation extension algorithm." *Computational Intelligence and Software Engineering, 2009. CiSE 2009. International Conference* on. IEEE, 2009.
[12] A. Ng, “Automated Instructional Aid for Reading Sheet Music.”
[13] J. Riley, “Recommended Best Practices for Digital Image Capture of Musical Scores.”
[14] R. Lobb and D. Bainbridge, “FAST CAPTURE OF SHEET MUSIC FOR AN AGILE DIGITAL MUSIC LIBRARY,” pp. 145–152.
[15] V. Sébastien, D. Sébastien, and N. Conruyt, “Constituting a Musical Sign Base through Score Analysis and Annotation,” vol. 4, no. 3, pp. 386–398, 2011.
[16] B. P. Challis, A. D. N. Edwards, and C. Science, “Weasel: A System for the Non-visualPresentation of Music Notation,” pp. 113–120, 2000.
[17] “Weasel: Access to Music Notation for Blind Peaple.” [Online]. Available: https://cultivateint.org/issue8/weasel. [Accessed: 24-Apr-2013].
[18] G. Bertoni, G. Nicotra, and A. Quatraro, “Access to music by blind people.”
[19] A. St, “A FRAMEWORK FOR PARAMETRIC SINGING VOICE ANALYSIS / SYNTHESISYoungmooE . Kim,” 2003.
[20] N. Ramanathan, “Applications of Hidden Markov Models,” 2006.
[21] M. W. Macon, L. J. Link, J. Oliverio, M. A. Clements, E. B. George, "ConcatenationBased MIDI to Singing Voice Synthesis", 1997.
[22] A. Georgaki, “A SCORE-TO-SINGING VOICE SYNTHESIS SYSTEM,” no. August,pp. 27–31, 2007.
[23] H. Kaprykowsky and X. Rodet, “Musical Alignment Using Globally Optimal Short-TimeDynamic Time Warping Short-Time Dynamic Time Warping.”
[24] H. Kaprykowsky, and X. Rodet, "Globally optimal short-time dynamic time warping,application to score to audio alignment." Acoustics, Speech and Signal Processing, 2006.ICASSP 2006 Proceedings. 2006 IEEE International Conference on. Vol. 5. IEEE, 2006.
[25] R. Li, Y. Ding, W. Li, and M. Bi, “Automatic Interpretation of Chinese Traditional MusicalNotation Using Conditional Random Field,” no. 60933004, pp. 19–22.
[26] J. Lee, JonghyunRyu, and S. Choi, "Vibrotactile score: A score metaphor for designingvibrotactile patterns." EuroHaptics conference, 2009 and Symposium on Haptic Interfacesfor Virtual Environment and Teleoperator Systems. World Haptics 2009. ThirdJoint. IEEE, 2009.
[27] J. Lee, JonghyunRyu, and S. Choi. "Graphical authoring tools for vibrotactile patterns."EuroHaptics conference, 2009 and Symposium on Haptic Interfaces for Virtual Environmentand Teleoperator Systems. World Haptics 2009. Third Joint. IEEE, 2009.
[28] J. Lee and S. Choi. "Evaluation of vibrotactile pattern design using vibrotactile score."Haptics Symposium (HAPTICS), 2012 IEEE. IEEE, 2012.
[29] B. Banko, “Realtime Webcam Sudoku Solver,” [Online]. Available: http://www.codeproject.com/Articles/238114/Realtime-Webcam-Sudoku-Solver [Accessed: 26-Jan-2014].

Dawpadee Kiriella is currently a fourth year Information and Communication Technology undergraduate at University of Colombo School of Computing (UCSC), Sri Lanka. Her research interests include Image Processing, Human Computer Interaction, Character Recognition, Data Mining, Database Technology, Bio-Informatics, Information Retrieval, Multimedia Computing and Audio Signal Processing.

Shyama Kumari is currently a fourth year Information and Communication Technology undergraduate at University of Colombo School of Computing (UCSC), Sri Lanka. Her research interests include Image Processing, Audio signal processing, HCI, Multimedia Computing, Character recognition.

Kavindu Ranasinghe is currently a fourth year Information and Communication Technology undergraduate at University of Colombo School of Computing (UCSC), Sri Lanka. His research interests include Image Processing, Audio Signal Processing, Character Recognition, Human Computer Interaction, Multimedia Computing and Music Information Retrieval.

Dr Lakshman Jayaratne - (Ph.D. (UWS), B.Sc.(SL), MACS, MCS(SL), MIEEE) obtained his B.Sc (Hons) in Computer Science from the University of Colombo, Sri Lanka in 1992. He obtained his PhD degree in Information Technology in 2006 from the University of Western Sydney, Sydney, Australia. He is working as a Senior Lecturer at the University of Colombo School of Computing (UCSC), University of Colombo. He has wide experience in actively engaging in IT consultancies for public and private sector organizations in Sri Lanka. His research interest includes Multimedia Information Management, Multimedia Databases, Intelligent Human-Web Interaction, Web Information Management and Retrieval, and Web Search Optimization. Also his research interest include Audio Music Monitoring for Radio Broadcasting.