Music Training Interface for Visually Impaired with a Universally Applicable OMR Engine

DB Kiriella#, KC Ranasinghe and L Jayaratne
University of Colombo School of Computing, Colombo 7, Sri Lanka
#dbk@ucsc.cmb.ac.lk

Abstract*— Assistive technologies specifically built for the visually impaired have not been able to cater to all their requirements. Visually impaired need third party assistance to convert visual Eastern music notation scripts to formats read able to them. Translation error rate remains high even with human assistance since music Braille form becomes more difficult to follow especially for complex notations. This research focuses on recognizing Eastern music notation scripts and translating them to an auditory output for the users. The main goal is to assist the visually impaired to independently visualize and train music notations. Optical Music Recognition (OMR) engine brought forward in this research consists of a pre-processor, regions-detector, recognizer and a post-processor. Pre-processor captures images of notation scripts de-skewing and through binaries. Regions-detector identifies tabulated segments, rows, columns, notation groups and atomic notation symbols. Recognizer provides shape definitions for each symbol and recognizes the best match for given language. The OMR also provides an adaptable API which developers can use to initially train a new set of symbols. It is capable of providing descriptions called shape definitions for new symbols. Mapping shape definitions to corresponding music notation symbols can be done through a configuration file. The intermediary file produced by the Recognizer is further analysed by the post-processor which refines the notation sequence depending on a matrix on probabilities of one note following another. This also suggests most appropriate substitutions for missing or noisy symbols. Recognized notation sequence is then converted to an auditory form. The OMR engine performs with 94.2% accuracy rate for Sinhala Eastern music notations while revealing successful adaptations for Hindi and English language symbols. 81% of the users accept that this type of an interface is more convenient for them compared to the existing method.*

Keywords*— Optical Music Recognition, Visually Impaired, Assistive Technologies*

I. INTRODUCTION

There are number of visually impaired musicians who perform unbelievably. They were capable to overcome the barriers they had in learning music, amazing us about their extraordinary talent. Observations even tempt us to think whether they were gifted with more capabilities required by a musician. World famous musicians such as Steve Wonder, Lemon Jefferson, Andrea Bocelli, Nobuyuk Tsujii, Henry Caldera, Hemapala Perera all are totally blind, but their compositions stood as milestones in history of music. Most of their biographies and explanations given by them reveal that their paths to succession were not fallen towards smooth and easy roads but brimmed with many leaps and bounds.

Among lists of difficulties they found in learning music, mostly highlighted were reading and regenerating music notations. Requirement for a solution to overcome this barrier was what they highly requested in interviews we carried out. Hence Swarālōka (Kiriella et al, 2014) has given a local solution for Sri Lankans; here the proposed research is focused on a universal solution targeting visually impaired music lovers worldwide. The research addresses the problem of how to assist visually impaired people to independently visualize and train music scores (notations) universally available, in a way that satisfies their music needs. Researchers are trying to reach the goal of assisting visually impaired people to independently study, train and experiment on printed music notation scripts universally available, avoiding regional music scripting barriers achieving the two major objectives stated below:

- Come up with a universal OMR engine which facilitates all Eastern music genres
- Come up with the most feasible converter which converts the output of the universal OMR engine into the most convenient readable format for the visually impaired regardless of the scripted natural language

II. RELATED WORK

The systems for music score recognition are called as OMR systems and they 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 (Musical Instrument Digital Interface file), an audio waveform or ABC Notation (Johansen, 2009). Johansen

(2009) has said that an OMR system consists typically of four main steps as; Pre-processing, Segmentation, Classification and Post-processing. These are the main steps in most of the character recognition systems as well. Fujinaga (1996) has come up with a system called Adaptive Optical Music Recognition by proposing adaptive software for the recognition of musical notation focusing to create a robust framework upon building a practical optical music recognizer. Raphael et al. (2011) has proposed a method called DMOS (Description and Modification of Segmentation) together with EPF (Enhanced Position Formalism) as an extension of the grammatical formalism for OMR. And that system mainly focuses on measures as the basic unit of recognition. Since music notation also can be considered as a language, grammatical formalism discussed by Raphael can be used to strengthen and enhance accuracy rate in post-processing. Depending on the probability of one note following another, grammatical rules can be defined to validate the recognized notation sequences. According to Genfang et al. (2003), that study has come up with an 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. Even though MIDI is a standard auditory output format, for a novice music learner midi output will not be enough to grasp all information embedded within a script. Genfang (2003, 2009) with his partner researchers have come up with three research prototypes in 2003 to 2009 focusing reading digital images of sheet music. They (Genfang et al., 2003) have presented a more facilitated method which can be applied to all music score of CMN (Common Music Notation) in Western Music. At the same time (2009), a new algorithm for capturing to a rectangle box of each connected region in GCN (Gong-Che Notation) score image was presented by the study of (Ng et al., 2012) while GCN is a Chinese traditional music notation. Riley et al. (2002) and Lobb et al. (2005) have tried to identify a solution for the problem of using digitally captured music (score) images while capturing music scores. Almost all of these researches are based on Western Music Manuscript notations. Western music manuscripts notations are well structured and symbols do not depend on language. Therefore it is a more favourable format when defining a universal OMR compared to Eastern music notation scripts. A Web Cam Sudoku solver program by Banko (2011) gave the idea of their own OCR (Optical Character Recognition) which uses profile features of each numeric character in the recognition algorithm. In their future works they had mentioned about post processing process which will be used to enhance the accuracy rate of identifying "6" and "9". In implementation of Swarālōka OMR (Kiriella et al, 2014), the symbol identification approach of Banko (2011) was helpful in defining a method to recognize atomic symbols. Modelling a universal OMR engine needed more deep literature survey studying all possible aspects of symbol identification in printed music scripts. Especially the findings by Fujinaga and Raphael provide more insight and valuable directions to come up with a more generalized and robust recognition framework.

III. METHODOLOGY AND EXPERIMENTAL DESIGN

In Swarālōka (Kiriella et al, 2014) (one of our previous researches) has already proved success of the research approach in inventing a music training interface which can be handled by visually impaired users independently. This initial research only supported Eastern music scripts with characters in the Sinhala alphabet. In the initial research we observed following main aspects:

- Effective use of universal OMR engines on interfaces for visually impaired are still at research stage
- Any alphabet owns to any language consists of a writable symbol set. Therefore the shapes are constrained with rules (orientation, writing direction, pauses in writing parts of a single letter)
- Apart from the graphical representation, music symbols can be defined using set of rules depending on how they are written
- Collection of rule sets can be used to define all symbols for a specific music notation
- Even though the Eastern music script notations differ depending on the language used, the auditory implication is almost same
- Sinhala is a strong language which has skilfully filtered out the units of sound which human voice can generate and have potential to be used as a primary source when defining the units of sound
- Definition of a music notation sequence best fits to a format in temporal domain compared to a visual script in spatial domain. Therefore, an auditory music training interface in temporal domain would be more successful
-

Since it is proved that music is a universal communication and entertainment medium, the research was built upon the hypothesis; if the conversion to a universally comprehensive output is possible for Eastern music scripts in Sinhala, it should be possible for all the other

Eastern music scripts in other languages when OMR is capable to accept rule sets adjusted for specific language. In this novel study, the main three research components are the universal OMR engine, the generalised script converter and the music score training interface

.

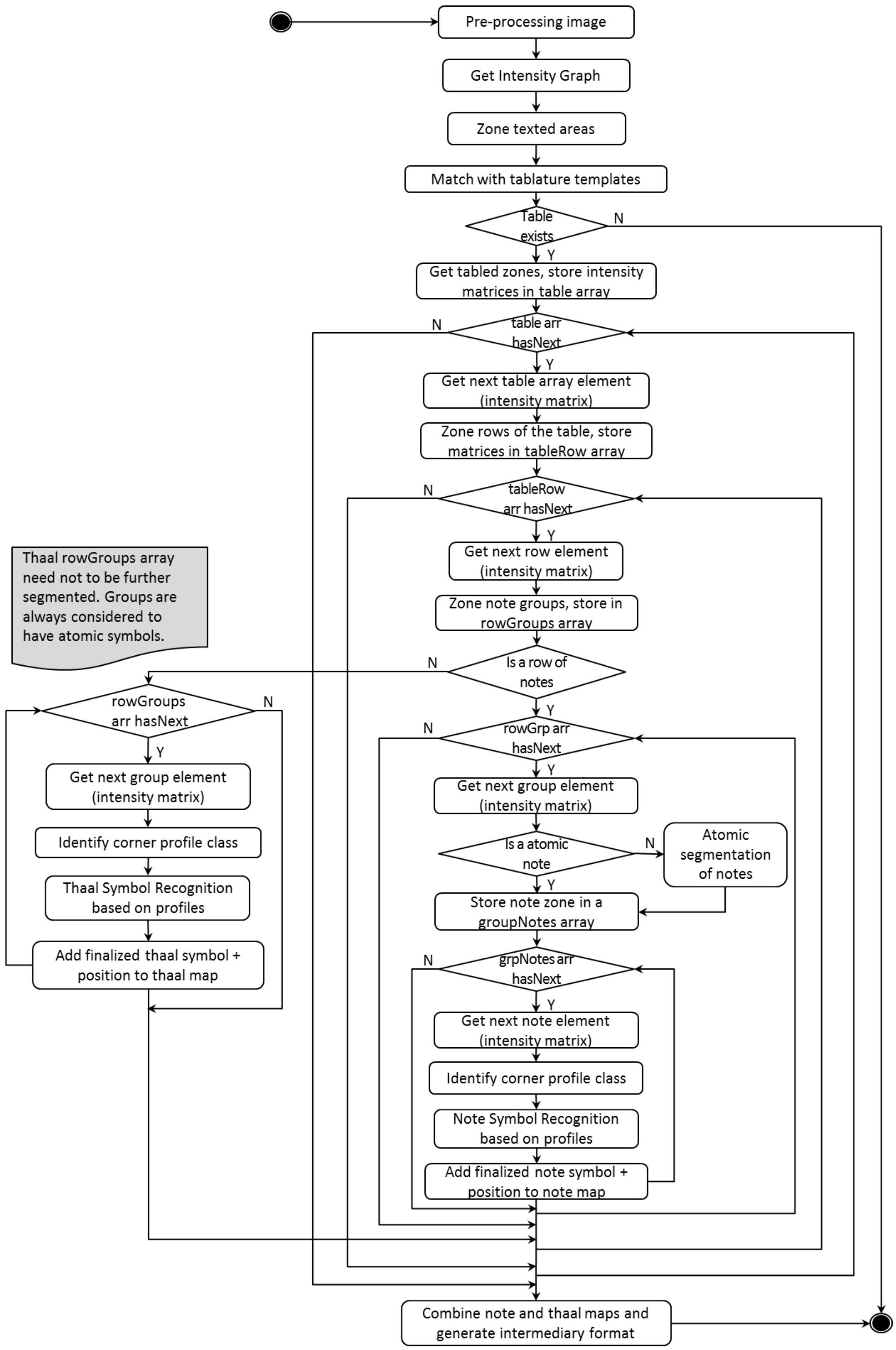

Figure 1. The flow diagram for the process of detection, segmentation and recognition of music note

A. The Universal OMR

Eastern music notations have several ways to write music symbols which slightly deviate from one way to the other regardless of the scripted natural language. Therefore when preparing music notations for the conversion phase, it should traverse through a common standard intermediary format. The Vibhags/columns are sometimes separated using an angled vertical line which lies only in a row/line of notation and placed whenever a separation needed. Those vertical lines are not fused together across the notation rows like bar lines which separate columns in commonly used music notation tablature structures. Therefore in zoning and segmentation rows of music notes should be identified initially and then split into columns. At the phase which atomic symbols are identified letter which denotes a musical note together with any marks used to represent derivative, ornamentation or Saptak should be treated as a composite symbol. Allowing segmentation to another level will increase the complexity and will become costly while processing. The flow diagram for the process of detection, segmentation and recognition of music notes is shown in Figure 1.

Recognizer comes up with shape definitions for each symbol and recognizes best matches for a given language. The OMR also provides an adaptable API which developers can use to initially train a new set of symbols by defining and adjusting rule sets for symbols in the language to be trained. It is capable to come up with descriptions called shape definitions for the new symbols. Mapping shape definitions to corresponding music notation symbols can be done through a configuration file. An intermediary file produced by the Recognizer is further analysed by the post-processor which refines the notation sequence depending on a matrix on probabilities of one note following another. This also suggests most appropriate substitutions for missing or noisy symbols. As mentioned in the related work, pre-defined grammatical formalism constraints on the note sequences can be used to further validate the recognized sequences and enhance the accuracy rate.

B. The Generalised Script Converter

The inputs for this process are the intermediary files generated in the previous phase and note group sequence map. These intermediary scripts are devoid of the qualities of the characters of scripted natural language in scores. While proceeding with the conversion into the audio output, two intermediary files for note and Thaal sequences will be 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 to represent all qualities of one note, more than one character will be used. Each note will be followed by a coefficient to represent its own duration. Rounded value of sum of these co-efficient will reveal the number of Maatras in the line. Intermediary Thaal file will include the sequence of Thaal symbols extracted from the OMR output file.

Whole conversion process depends on 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. 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 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. 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 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 should be easily distinguishable.

C. The Music Score Training Assisting Interface

Among the design considerations used, influence of music teaching and training methods are significant in developing the interface for end users. Capturing the user performance accurately, focusing on one-to-one teaching, encouraging the user 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. Figure 2 depicts the process in this phase as a flow diagram. Monophonic User performance is captured and depending on the frequency graphs intermediary files are created for performance analysis. Depending on the successful matches, feedback is given to the user with a replay.

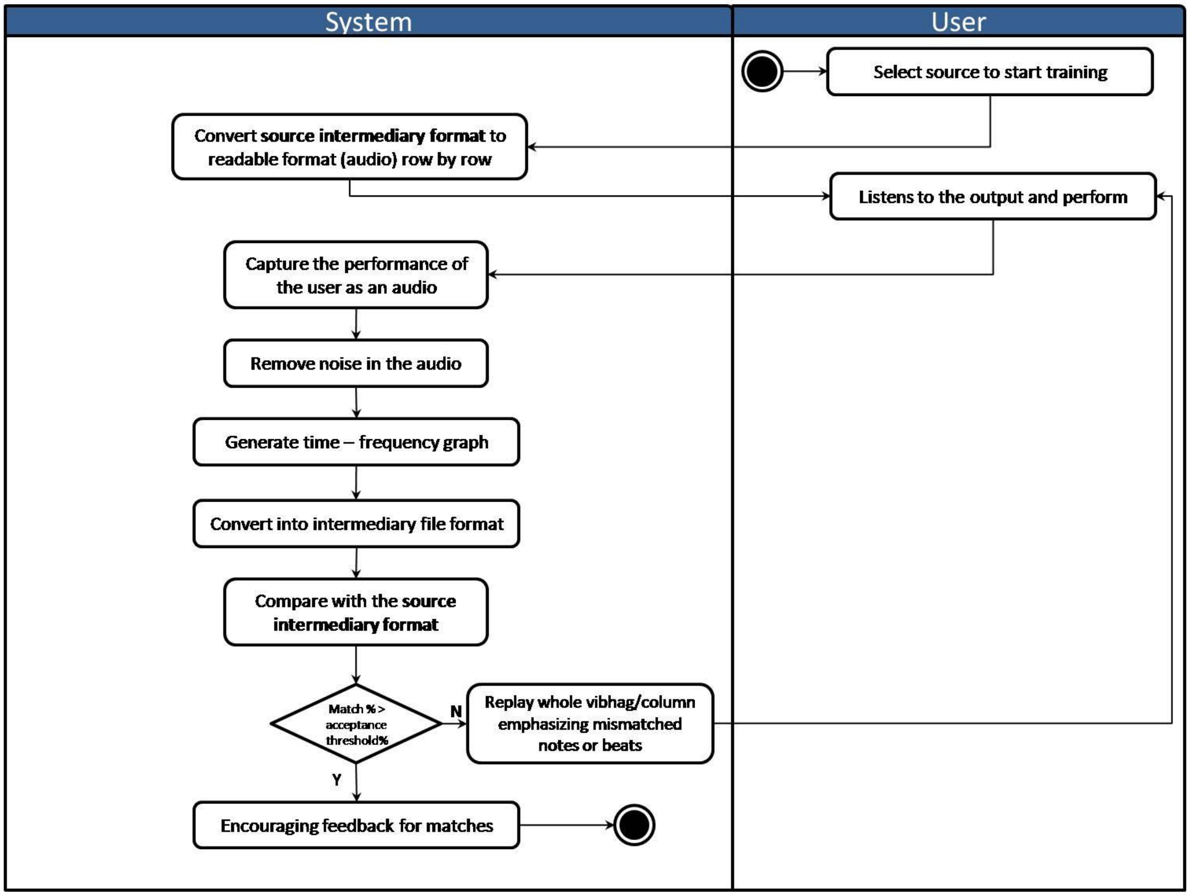

Figure 2. The flow diagram for the process of music score training assisting interface

IV. RESULTS AND DISCUSSION

Implemented prototype and the research findings were evaluated and verified through real experiments with selected visually impaired students willing to learn music and the teachers who are willing to assist these students. 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.

The effective time taken by a visually impaired student or total human hours (together with an involvement of a teacher) spent to capture "n" (n should be a pre-defined reasonable value for similar type of experiment series) number of music score lines using the proposed solution and existing way will be analysed and compared. Comparing comments by the user on the ease of use of the developed system would provide necessary means to draw conclusions on the research findings.

The effective time taken by a visually impaired student or total human hours (together with an involvement of a teacher) spent to capture n (n should be a pre-defined reasonable value for similar type of experiment series) number of lines in a music notation using the proposed solution and existing way. Comparing comments by the user in ease of use would also provide necessary means to come up with a conclusion on research findings.

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 will be error rate, success rate and time spent for an evaluation test.

Usability testing was carried out with the participation of 16 students from Blind School, Ratmalana, 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. 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.

The OMR engine performs with 94.2% accuracy rate for Sinhala Eastern music notations while revealing successful adaptations for Hindi and English language symbols. Compared to a MIDI format, the output is strong enough to convey all the information embedded within a script since it implies the exact quality and name of the notations. Composite information is also implied using concurrent audio layers and is a feature which is absent in existing methods. 81% of the randomly selected sample group of participants among the target user group accepted that this type of an interface is more convenient for them compared to existing method.

V. CONCLUSION AND FUTURE WORK

The research study is to be extended to come up with another assistive function of OMR engine for identifying missing symbols of damaged printed music scripts. Moreover we expect to enhance the research prototype in order to facilitate the users who use music scores scripted in Tamil language. Further studies are going on Japanese and Chinese scripts too since they also have some similar music qualities in Sri Lankan and Indian Eastern music.

As the authors we believe that this proposed universally accessible music training interface 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 world of music without any genre barrier in Eastern music. 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 almost all the Eastern music genres nourished with the colourful thoughts born in their minds.

In fact, steering the research towards will be not only a help for those who are in need but will help these differently-abled people to illuminate their world with the colours of music.

ACKNOWLEDGMENT

The research is carried out at the University Of Colombo School Of Computing, University of Colombo, Sri Lanka and granted by the National Centre for Advanced Studies in Humanities and Social Sciences (NCAS), Sri Lanka. At the research experimental setup evaluation, the support of the School for the Blind, the Ceylon School for the Deaf and Blind, and Ratmalana, Sri Lanka should be gratefully acknowledged.

REFERENCES

Banko B (2011) Realtime Webcam Sudoku Solver [Online]. Available: http://www.codeproject.com/Articles/238114/Realtime-Webcam-Sudoku-Solver [Accessed: 14-May-2015].

Fujinaga I (1996) Adaptive Optical Music Recognition".

Genfang C, and Shunren X (2003) 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*.

Genfang C, Wenjun Z, and Qiuqiu W (2009) 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*.

Genfang C, and Wenjun Z (2009) Chinese Gong-Che notation musical score image segmentation based on an anticlockwise rotation extension algorithm, *Computational Intelligence and Software Engineering (CISE) 2009. International Conference on. IEEE.*

Johansen S (2009) Optical Music Recognition.

Kiriella DB, Kumari SC, Ranasinghe KC, *et al* (2014) Music training interface for visually impaired through a novel approach to optical music recognition, *GSTF International Journal on Computing (JoC)*, vol. 3, no. 4, 90-103.

Lobb R and Bainbridge D (2005) Fast capture of sheet music for an agile digital music library, 145–152.

Ng A, and Khan A (2012) Automated instructional aid for reading sheet music.

Raphael C, and Wang J (2011) New approaches to optical music recognition, *ISMIR*, 305–310.

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

BIOGRAPHY OF AUTHORS

Dawpadee B. Kiriella obtained her B. Sc. in Information and Communication Technology from the University of Colombo, School of Computing (UCSC) in 2014. She is currently working as an Assistant Lecturer at UCSC. While following a M. Phil. at the same University, she is engaging in researches on developing a universally applicable optical music recognition engine and also on adaptable interfaces related to assistive technologies. Since July 2014, she is also working as a co-investigator of the research project of FaceID - Computer based Forensic Facial Reconstruction for Sri Lanka using 3D Graphics. Her major research interest includes Image Processing, Computer Vision, Human Computer Interaction, Artificial Intelligence, and Audio Signal Processing

Kavindu C. Ranasinghe obtained his B. Sc. in Information and Communication Technology from the University of Colombo School of Computing (UCSC), University of Colombo, Sri Lanka in 2014. He is currently working as a Software Engineer at Cambio Software Engineering (Pvt.) Ltd. While following a M. Phil. at the UCSC, he is researching on synthesized singing as a mediator in musical communication. His major research interest includes Image Processing and Computer Vision, Audio Signal Processing, Character Recognition, Human Computer Interaction, Multimedia Computing and Music Information Retrieval.

Dr. Lakshman Jayaratne 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. At Present he is working as a Research Advisor to Ministry of Defense, 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 includes Audio Music Monitoring for Radio Broadcasting and Computational Approach to Train on Music Notations for Visually Impaired in Sri Lanka.