PRINTED BOOK TO AUDIO BOOK CONVERTER FOR VISUALLY IMPAIRED

Ajinkya Domale, Bhimsen Padalkar, Raj Parekh, M.A. Joshi
College of Engineering, Pune
National Highway 4, Shivaji Nagar, Pune, Maharashtra - 411005

Abstract

Visually impaired people are dependent solely on Braille books & audio recordings provided by NGOs. Owing to many constraints in above two approaches blind people can't have book of their choice. The presented work will provide them an opportunity to have an audio-book of their choice in English or Marathi language of any printed book having English, Marathi or Braille script. Printed text from textbook having English, Marathi or Braille script will be taken as input in the form of an image which will be converted into plain editable text with the help of Optical Character Recognition (OCR). This plain text will be then fed to Text to Speech (TTS) converter which will generate the audio output file in English or Marathi language corresponding to the input text image script. Printed book to audio book converter has been successfully implemented and satisfactory results were obtained.

1. Introduction

Visually impaired people can't learn as we people normally do. The only way to learn for them is through Braille script which is a tactile writing system used by blind & visually impaired. An A4 size page of a normal textbook is equivalent to 4 pages in Braille script. Due to this fact Braille script books are large, heavy, take more time to read and are not easily available as compared to normal textbooks. Therefore, Braille books are not produced for higher studies where book size is considerably large. Owing to these constraints many social organizations voluntarily come forward and convert normal textbooks into audio books by manually recording which requires significant amount of manpower & time investment.

So, the proposed converter is intended to do the same task which will eliminate the cumbersome procedure of printing large size Braille script books & manual recording of normal textbooks. This converter will open new doors of Knowledge Sea for visually impaired as they can listen to any book of their choice in English & Marathi language.

1.1 Technical Background

The very first system employing both OCR & TTS was designed by Bernard Conan Pobiak [1], he has built ADJUSTABLE ACCESS ELECTRONIC BOOKS, The Reader Project of Washington, D.C.

Since, then there have been tremendous amounts of work done regarding both OCR & TTS for English language. Many papers have been published employing various techniques which significantly improved English language OCR & TTS systems.

Also, there are many products available in the market in the form of software as well as hardware. Products such as ABBYY FineReader [2], Readiris [3], etc. are available in the market for English language OCR and products such as NaturalReader [4], AT & T Natural Voices [5], etc. serves the purpose of TTS system for English language in the form of software. But they are very costly, there prices range from $100 to $500, making them unaffordable for the visually impaired students for which presented work is dedicated. There are other free versions as well available over the internet but they don't give sufficient accuracy in case of English OCR and clear perceptibility in case of English TTS.

There is another category of products which are available in the market such as KNFB Reader [6], Optelec ClearReader+ [7], etc. which provides the complete system of English OCR & TTS in the form of hardware employing stand-alone camera and a processing unit. But, these systems are way too costly, there costs range from $2500 to $3500. Visually impaired people, at least in India can't afford such systems.

Thorough literature review of English OCR & TTS gave encouragement to build a system which can effectively convert a printed book in English language to its audio book by utilizing the existing hardware like personal computer and scanners which are readily available in blind schools and colleges, thus making it cost effective and easy to use.

Although, the OCR technology has made a lot of advancements in scripts like Persian, Latin and Arabic etc., there have been a considerably limited number of studies on Devanagari script character recognition which is used in Marathi language. At

978-0-7695-5146-3/13 $26.00 © 2013 IEEE
DOI 10.1109/TIIEC.2013.27

present, several organizations have started work on Indian languages OCR. Ministry of Information Technology, Government of India, has initiated a Technology Development on Indian Languages (TDIL) project under which OCR system development for most of the important Indian language scripts have been taken up by different labs and academic institutions.

One of the earliest efforts made in the direction of recognition of printed Devanagari characters were made by R.M.K. Sinha [8], [9]. He presented a syntactic pattern analysis system and its applications for Devanagari script in his doctoral thesis [8]. The first complete OCR system development of printed Devanagari is perhaps due to Palit and Chaudhuri [10] as well as Pal and Chaudhuri [11]. A team comprising Prof. B. B. Chaudhuri, U. Pal, M.Mitra, and U. Garain of Indian Statistical Institute, Kolkata, developed the first commercial level product for printed Devanagari OCR. The same technology has been transferred to Centre for Development for the Advance Computing (CDAC) in 2001 for commercialization and is marketed as "Chitrankan" which do not give satisfactory results and the upcoming versions is costly, it would cost approximately Rs. 10,000 per single licensed copy.

In the context of building TTS systems in Indian languages such as Marathi language, there has considerable amount of efforts and works in the past. These works can be divided into parametric ([12] [13] [14] [15] [16]) and unit selection synthesis ([17] [18] [19]).

Parametric Synthesis: The idea here is to derive a set of parameters such as formant frequencies, linear prediction coefficients for each phone in the language and develop co-articulation rules for joining of two phone segments. This type of synthesis technique needs a rich understanding of relation of acoustic properties of sounds (phones/syllables) and their dynamics. Such knowledge is typically acquired by observing several examples of sounds in different possible contexts.

Unit Selection Synthesis: This technique relies on using a set of pre-recorded sub-word units. For each sub-word such as diaphone or syllable, there are many than one pre- recorded examples. Units such as diaphone or syllables are preferred in this approach as they inherently capture the co-articulation between phones. The speech databases used in unit selection synthesis already encapsulate prosody associated with different contexts at the word, phrase and sentence levels. During synthesis time, a unit is selected such that it matches with the input specification as well as with the others units in the sequence.

Significant contribution in the field of TTS for Marathi and other Indian languages is done by a consortium consisting of five institutions in India to build TTS systems in six Indian languages including Marathi [20]. Similar efforts were taken by IIIT-H to provide free speech databases of some Indian languages including Marathi [21].

Dhvani[22] is a text to speech system designed for Indian Languages. It utilizes the concatinative approach giving a fair perceptible output but still it sounds more robotic.

Concatinative or unit selection synthesis based approach doesn't give the required perceptibility of the generated speech. So, this approach is not that effective. Parametric synthesis based approach takes a lot of time and sounds artificial even though it is more accurate as compared to the selection synthesis based approach. There is software named NVDA (Non Visual Desktop Access) which is now equipped with support for Marathi language and it is much better than the previous approaches, it makes use of Festival Speech Synthesis technique [23]. Being more reliable and perceptible TTS system, NVDA is used along with MATLAB to generate the speech corresponding to Marathi text in the presented work.

Attempts have been made to optically recognize embossed Braille using various methods. In 1993 & 94, Mennens and his team designed an optical recognition system which recognized Braille writing. It used scanner for digitizing the Braille document [24-25].

A grid was constructed consisting of horizontal and vertical lines that run through all the dots, and then verification of dots present in the line intersection points. The method proved to be good, when the embossing was in good condition. Effort was also made to recognize Braille embossing on both sides of a single sheet in one scanning, but the deformation in the dot grid alignment was not handled.

In 1999 Ng and his team approached the problem using boundary detection techniques to translate Braille into English or Chinese [26]. The paper presented an automatic system to recognize the Braille pages and convert the Braille documents into English/Chinese text for editing. They separated embossing on two sides of a sheet into two standard templates. Using both single side and double sided Braille pages as the specimen; the system has been proved to be 100% and 97% accurate respectively. The system took the advantage of regular spacing between Braille dots

within a cell, and the regular spacing between cells. The paper did not discuss anything about grid deformation. The capturing device used in this experiment was a digital camera, which was placed directly above the Braille page.

In 2001, Murray and Dais designed a handheld device which handles the scanning as well as the translation [27-28]. Since the user is in control of the scanning orientation, and only a small segment is scanned at each instance, grid deformation is not a major concern and a simpler algorithm was used to yield efficient, real-time translation of Braille characters.

In 2005, Nestro Falcon and his team developed further more efficient techniques for Braille writing recognition using Image processing Techniques [29]. The paper presented the development of BrailLector, a system able to speak from Braille writing. By means of dynamic thresholding, adaptive Braille grid, recovery dots techniques and TTS software (Text-To-Speech). BraileLector is a robust application with innovative thresholding and Braille grid creation algorithms which detects and read Braille characters with 99.9% of correct symbols and error variance below 0.012.

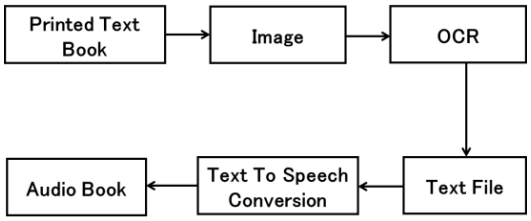
An effort to recognize Arabic Braille recognition was found in the paper presented by Abdul Malik and his team in 2007 [30]. The algorithm was developed to recognize an image of embossed Arabic Braille and then convert it to text. It aimed at building fully functional Optical Arabic Braille Recognition system. The conversion of Braille to text was complicated because two or more cells would represent a single symbol, and sometimes a single cell would represent one text symbol or two or more symbols. The algorithm also tested for variations of Braille documents; skewed reversed or worn-out.

Detecting the Dots of the Braille character has been developed by Amany Al and his team in 2008 [31]. The core of the proposed method was the use of stability of threshold with Beta distribution to initiate the process of thresholds estimation. Segmented Braille image is then used to form a grid that contains recto dots and another one that contains verso dots. Using the segmented image, Braille dots composing character on both single sided and double sided documents are automatically identified from those grids with good accuracy.

Saad D Al-Shamma and Sami Fathi [32] presented Image processing technique in 2010 to convert Arabic Braille into equivalent Arabic and also to voice.

To conclude, no software has been successfully developed to convert Braille of Indian languages into its equivalent. The main reason behind this is the complexity involved in writing Indian languages. Most of them adopt syllable writing, and hence becomes a complex task to convert. In the presented work we have successfully converted Braille Marathi into its equivalent normal Marathi text with a very good accuracy.

1.2 Proposed solution



An attempt to provide a printed book to audio book converter for English, Marathi and Braille (English & Marathi) script with minimal hardware requirement has been performed successfully. Also, a new algorithm has been devised for Braille character recognition which almost gives 100 % accuracy.

1.3 Organization of the report

The remainder of the paper is organized as: section 2 describes the proposed solution; the details of implementation are given in section 4. The results are presented in section 4. Finally, conclusion is given in section 5.

2. Proposed Solution

The whole project has been implemented in MATLAB environment. Image quality should be considerably well to obtain the specified accuracy. Accuracy will be less for degraded images.

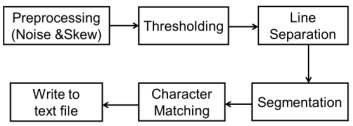
3. Implementation

3.1 Hardware Implementation

Full software implementation is done. Hardware implementation can be done by using Texas Instrument's Floating Point DSP TMS320C6713B as it provides the integrated digital audio interface and has been highly optimized for audio, making it a good choice as the end output is audio only.

3.2 Software Implementation

English OCR Flow Chart



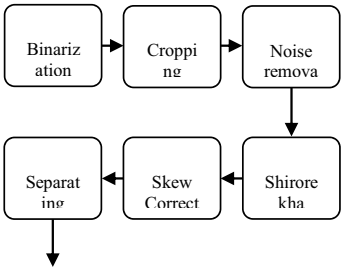
English TTS Algorithm

Text-to-Speech is implemented by using the Microsoft Win32 Speech API (SAPI). The Speech Application Programming Interface or SAPI is an API developed by Microsoft to allow the use of speech recognition and speech synthesis within Windows applications. It is shipped as part of the Windows OS itself. Therefore, it can be easily used on any genuine Microsoft Windows without having any copyright infringement issues. The main motto behind using Microsoft SAPI is its excellent quality and it can also be very easily interfaced with MATLAB to obtain the required output speech file. This saves the lots of hustle of building a completely new TTS system from scratch. Following algorithm is employed in MATLAB while building the TTS system using Windows SAPI.

1) Text-to-Speech synthesizes speech from string TXT, and speaks it. The audio format is mono, 16 bit, 16 KHz by default. Function WAV = TTS(TXT) does not vocalize but output to the variable WAV.
2) Function TTS (TXT,VOICE) uses the specific voice and by default voice is
"Microsoft Anna - English (United States)".
3) Function TTS (...,PACE) set the pace of speech to PACE. PACE ranges from -10 (slowest) to 10 (fastest). Default value is 0.
4) Function TTS(...,FS) set the sampling rate of the speech to FS KHz. FS must be one of the following: 8000, 11025, 12000, 16000, 22050, 24000, 32000, 44100 & 48000. Default value is set to 16000.

Once the digitized text is obtained after running English OCR on scanned image, it is fed to the TTS converter which produces speech corresponding to text file in .wav audio file format.

Marathi OCR Flow Chart



Marathi TTS Algorithm

First of all TTS system for Marathi language is built using the database made available by CDAC under NVDA screen reader software. It is done by using concatenative approach. In this approach, all the consonants, vowels and numbers in Marathi language are pre-recorded. Then depending upon the word to be uttered speech corresponding to the letters in the word is retrieved from the database, assembled and concatenated to form the whole word. Once the speeches corresponding to the whole word are concatenated; prosody & phonetic rules are applied to the concatenated speech which tries to make the utterance more natural. Even after applying these rules, the utterances were not sounding natural as they should be. Thus, making it difficult to figure out what's being uttered. Dhvani – a TTS system for Indian Languages utilizes similar approach, but it sounds robotic being a concatenative based TTS system. The biggest problem with such system is that one can't hear and grasp whatever is being uttered over a longer duration of time. Therefore this approach is scarped as desired results were not getting generated using this approach.

Again, building a completely new system having more perceptibility from scratch would have taken much time. There is significant progress done in the field of building a more perceptible system for Marathi languages by CDAC, JAWS, NVDA, etc. But, among all NVDA is the best in terms clear perceptibility as it has the database of whole words from Marathi language instead of characters making it more natural TTS system. NonVisual Desktop Access (NVDA) is a free and open source screen reader for the Microsoft Windows operating system. Therefore, this system can be used as a platform without anya copyright infringement issues to build TTS system for the presented work which integrates NVDA with MATLAB. Following algorithm is employed in MATLAB while building the TTS system using NVDA.

1) Text file obtained after running OCR for Marathi language is opened in MATLAB using 'winopen' function.
2) Executable file corresponding to NVDA (Non Visual Desktop Access) screen reader is executed using 'winopen' function in MATLAB which turns on the screen reader.
3) Till the time NVDA gets loaded in the system, an audio recorder is created in MATLAB at back-end which will record whatever being uttered by the NVDA screen reader.
4) The audio recorder created in MATLAB records the sound coming directly from sound card of the computer. This is done by using the 'Stereo Mix' option available in Windows OS under recording devices.
5) Once NVDA screen reader gets loaded in the system, it starts uttering whatever is displayed on the screen in clear & perceptible Marathi language. As, the text file in Marathi to be uttered is already opened on the screen, the NVDA screen reader starts uttering the opened text file in Marathi language.

6) The audio recorder function created in MATLAB records all the uttered sound and stores it in .wav audio file format.

Braille OCR Algorithm

1) First of all database is created which is different for English & Marathi Braille scripts.

2) Then Binarization of Braille Script Image is performed.

3) After that segmentation is performed this involves separating lines, words & letters using horizontal and vertical projection profiles.

4) Once a line is separated, all the pixels in each dot are clustered together by using dilation. So the dots become more prominent.

5) After dilation, each dot is represented by one pixel only. Here, all the dots in one column are aligned vertically so that they constitute only one pixel column. Same is done for the dots in one row (horizontal alignment).

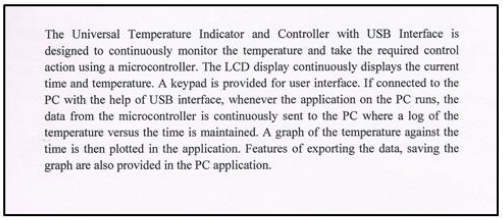
6) Once the line is separated, the words can be separated using the distances between consecutive columns of letters.

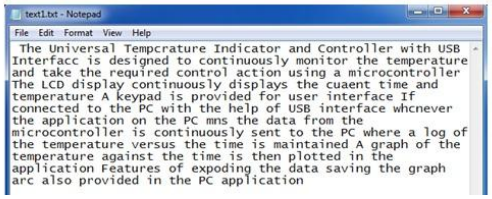
7) Letters can be separated using the distances between consecutive columns of letters.

8) The code assigned to the letter is compared to the codes in the database of that class assigned to it. Then applying the rules of Braille script, corresponding letter is returned.

4. Results

English OCR


The Universal Temperature Indicator and Controller with USB Interface is designed to continuously monitor the temperature and take the required control action using a microcontroller. The LCD display continuously displays the current time and temperature. A keypad is provided for user interface. If connected to the PC with the help of USB interface, whenever the application on the PC runs, the data from the microcontroller is continuously sent to the PC where a log of the temperature versus the time is maintained. A graph of the temperature against the time is then plotted in the application. Features of exporting the data, saving the graph are also provided in the PC application.

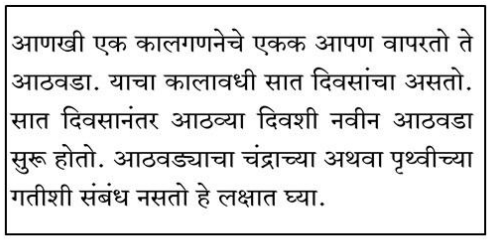


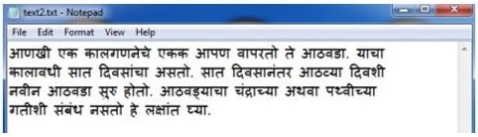
Words in the input text image: 111

Incorrectly recognized words: 7

Accuracy: 93.7 %

Marathi OCR


आणखी एक कालगणनेचे एकक आपण वापरतो ते आठवडा. याचा कालावधी सात दिवसांचा असतो. सात दिवसानंतर आठव्या दिवशी नवीन आठवडा सुरू होतो. आठवड्याचा चंद्राच्या अथवा पृथ्वीच्या गतीशी संबंध नसतो हे लक्षात घ्या.

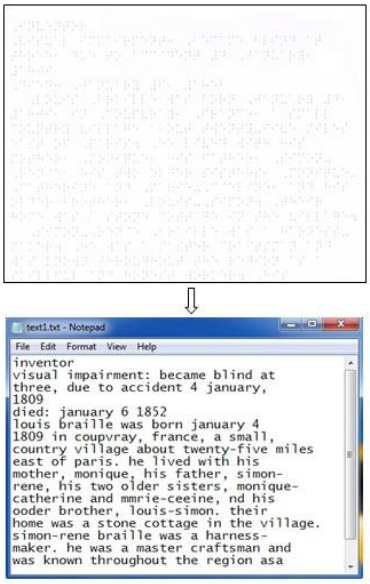


Words in the input text image: 95

Incorrectly recognized words: 2

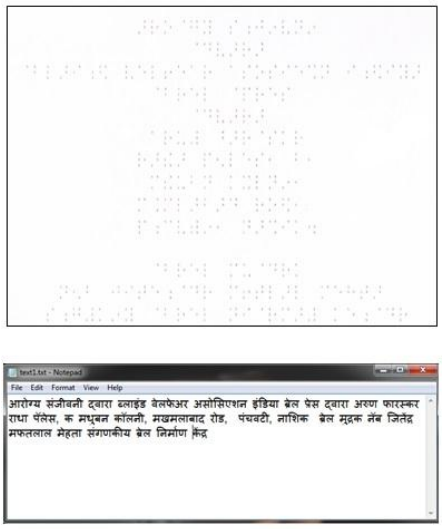
Accuracy: 97.89 %

English Braille OCR



Accuracy = 99 %

Marathi Braille OCR



Accuracy = 99

English language OCR system has an accuracy of above 93 %, Marathi language OCR system has an accuracy of above 90 % and Braille script OCR system of both English & Marathi language has an accuracy of around 99 %. The accuracy may decline depending on the quality of the scanned documents. Windows SAPI for English language TTS & NVDA for Marathi language TTS have successfully integrated with MATLAB to obtain an audio file corresponding to the digitized text returned by the OCR system on the scanned image.

5. Conclusions

Printed book to audio book converter has been successfully implemented for English, Marathi & Braille script. A new algorithm for English & Marathi Braille script has been implemented successfully which is almost 100 % accurate. Such system will greatly help illiterate visually impaired people and people working along with them. OCR systems can't have a perfect 100 percent accuracy. So, different algorithms which can improve the accuracy of OCR for all three scripts can be employed in the future. In the presented work, OCR for Marathi & OCR for Braille Marathi are implemented. In future, more Indian regional languages can be included.
Improvements in the output voice can be made by changing the prosody rules & Indian accent database can be used which will make the English TTS system sound more natural having Indian accent. In future both these TTS systems for English & Marathi languages can be made stand-alone systems which are independent of any third party TTS platforms. Whole work can be converted into C language code. So that it can be easily dumped into any DSP processor which will serve as a stand-alone device for converting printed book into audio book. In such stand-alone device either a scanner or High Definition camera can be interfaced with DSP processor which will acquire images of printed book pages.

References

[1] Bernard Conan Pobiak - ADJUSTABLE ACCESS ELECTRONIC BOOKS, The Reader Project of Washington, D.C., 1992 IEEE
[2] http://finereader.abbyy.com/
[3] http://www.irislink.com/c2-2115-189/Readiris-14--OCR-Software--Scan--Convert-- Manage-your-Documents-.aspx
[4] http://www.naturalreaders.com/
[5] http://www.naturalvoices.att.com/
[6] http://www.knfbreader.com/
[7] http://www.optelec.com/en_US/optelec-clear-reader
[8] R. M. K. Sinha, "A Syntactic pattern analysis system and its application to Devanagari script recognitio," Ph.D. Thesis, Dept. Elect. Eng., Indian Institute of Technology, Kanpur, India, 1973.
[9] R. M. K. Sinha and H. Mahabala, "Machine recognition of Devanagari script" IEEE Trans. Syst. Man Cybern., vol. 9, no. 8, pp. 435–441, Aug.1979.
[10] S. Palit and B. B. Chaudhuri, "A feature-based scheme for the machine recognition of printed

Devanagari script" in Pattern Recognition, Image Processing and Computer Vision, P. P. Das and B. N. Chatterjee, Eds. New Delhi, India: Narosa Publishing House, 1995, pp. 163–168.
[11] U. Pal and B. B. Chaudhuri, "Printed Devanagari script OCR system" Vivek, vol. 10, pp. 12–24, 1997.
[12] P. V. S. Rao and R. B. Thosar, "A programming system for studies in speech synthesis," IEEE Trans. Acoust., Speech and Signal Processing, vol. 22, no. 3, pp. 217–225, 1974.
[13] P. Bhaskararao and S. Mathew, "Phonemic transcription rules for text-to-speech synthesis of Hindi," in In: Sinha, R.M.K. (ed) Proc. Computer Processing of Asian Languages-2. New Delhi: Tata McGraw Hill., 1992.
[14] B. Yegnanarayana, S. Rajendran, V. R. Ramachandran, and A. S. Madhukumar, "Significance of knowledge sources for a text-tospeech system for Indian languages," Sadhana, pp. 147–169,1994.
[15] S. S. Agrawal, "Synthesizing Hindi speech using Klsyn and Hlsyn for natural sounding," in Workshop on Spoken Language Processing, TIFR, Mumbai, 2003.
[16] X. A. Furtado and A. Sen, "Synthesis of unlimited speech in Indian languages using formant-based rules," Sadhana, vol. 21, no. 3, pp. 345–362, 1996.
[17] K. Prahallad, A. W. Black, R. Kumar, and R. Sangal, "Experiments with unit selection speech databases for Indian languages," in in Proceedings of National Seminar on Language Technology Tools: Implementations of Telugu, Hyderabad, India, 2003.
[18] K. Bali, A. G. Ramakrishnan, P. P. Talukdar, and S. K. Nemela, "Tools for the development of a Hindi speech synthesis system," in 5th ISCA Speech Synthesis Workshop, Pittsburgh, USA, 2004.
[19] S. Thomas, M. Nageshwar Rao, H. A. Murthy, and C. S. Ramalingam, "Natural sounding speech based on syllable-like units," in EUSIPCO, Florence, Italy, 2006.
[20] Hema A. Murthy, Ashwin Bellur, Vinodh Viswanath, Badri Narayanan, Anila Susan, G Kasthuri, K. Sreenivasa Rao, Sudhamay Maity, N. P. Narendra, Ramu Reddy, Krishnendu Ghosh, K. G. Sulochana, E. L. Abhilash, T. Sajini, M. Sasikumar, Bira Chandra Singh, Pranaw Kumar, P. Vijayaditya, E. Veera Raghavendra, Kishore Prahallad,"Building Unit Selection Speech Synthesis in Indian Languages: An Initiative by an Indian Consortium"
[21] Kishore Prahallad, E.Naresh Kumar, Venkatesh Keri, S.Rajendran, Alan W Black," The IIIT-H Indic Speech Databases".
[22] http://dhvani.sourceforge.net/
[23] Anila Susan Kurian, Badri Narayan, Nagarajan Madasamy, Ashwin Bellur, Raghava Krishnan, Kasthuri G., Vinodh M.V., Hema A. Murthy, "Indian Language Screen Readers and Syllable Based Festival Text-to-Speech Synthesis System", Proceedings of the 2nd Workshop on Speech and Language Processing for Assistive Technologies, pages 63–72, Edinburgh, Scotland, UK, July 30, 2011.
[24] Jan Mennens "Optical recognition of Braille writing", IEEE 1993. pp 428-431.
[25] Jan Mennens, LUC Va Tichelen, Guido Francois and Jan J Engelen. Optical Recognition of Braille writing using Standard Equipment. IEEE Trnsactions on Rehabilitation Engineering Vol 2. No.4 December 1994.
[26] C. Ng, V.Ng and Y.Lau, Regular feature extraction for recognition of Braille. Third International conference on computational Intelligence and Multimedia Applications, 1999. ICCIMA '99 Proceedings, Page 302—306,1999.
[27] I. Murray and T. Dias. A portable device for optically recognizing braille - part i: hardware development. In The Seventh Australian and New Zealand Intelligent Information Systems Conference 2001, pages 129–134, 2001.
[28] I. Murray and T. Dias. A portable device for optically recognizing braille - part ii: software development. In The Seventh Australian and New Zealand Intelligent Information Systems Conference 2001, pages 141–146, 2001.
[29] Nestor Falcon, Carlos M. Travieso, Jestis B. Alonso and Miguel A Ferrer. Image Processing Techniques for Braille writing Recognition. EUROCAST 2005, LNCS 3643.
[30]Abdul Malik Al-Salman, Yosef ALOHAI, Mohammed ALKanhal and Abdulla AIRajith. An Arabic Optical Braille Recognition System. ICTA 07 April 12-14 Hammamet, Tunisia.
[31] Amany AL-Saleh, Ali EL-Zaart, and Abdul Malik AISalman. ICCHP 2008 LNCS 5105. Dot Detection of Optical Braille Images for Braille Cells Recogniton.
[32] Saad D. Al-Shamma and Sami Fathi. "Arabic Braille Recognition and Transcription into Text and Voice. 2010 " 5th Cairo International Biomedical Engineering Conference Cairo, Egypt, December 16-18 2010.

7. Acknowledgements

We would like to thank Mr. K. Raman Shankar, Director - National Association for the Blind (NAB), Mumbai for providing us the required guidance regarding Braille script. We thank Mrs. Dhanavanti Vinay Hardikar, Bal Bharati, Pune (Maharashtra State Bureau of Textbook Production & Curriculum Research) for providing us the guidance in the Devanagari script. Last, but not the least, we express sincere appreciation to all our friends, colleagues especially Santosh Bhase & Vishal Bajare from Nowrosjee Wadia College, Pune who are visually impaired and helped us a lot in the due course of our project regarding testing and suggestions.