Malayalam Text-to-Speech

Priyanka Jose , Govindaru V

Abstract— **The computer synthesis of natural speech is an objective for both engineers and linguists that would provide many useful applications to human-computer interaction. This paper explores text to speech system with Optical Character Recognition (OCR), where spoken utterances are automatically produced from text. Text to speech system with OCR is yet to develop for Indian languages, especially for Malayalam. Therefore we attempt to develop a text-to-speech for Malayalam using by tools available in matlab. Some facts of the current state technology are illustrated and the final section will explain the authors approach to the field of voice synthesis. The output generated by the proposed system to have very closeness to natural human voices.**

***Index Terms* - text-to-speech, ocr, character recognition, concatenative synthesis, Unicode, ASCII code**

I. Introduction

Malayalam is a Dravidian language, spoken in the south west part of India. It is the official language of Kerala State and Lakshadweep Union Territory. In India there are about 50 million speakers of Malayalam. Another 500,000 people who speak Malayalam are residing outside India.

Speech is the major source for communication in all stages. Text to speech (TTS) synthesis with OCR is a complex combination of language processing and signal processing. Automatic conversion of text to speech system is useful for many commercial and humanitarian applications. Such as:

A. Reading Aid for Blind People

The visually impaired can benefit tremendously from text to speech technology. TTS software would enable input text to be generated to spoken words [3].

B. Talking Aid for Vocally Handicapped People

People those who have lost the ability to speak but can still hear, can use a type writer or similar interface has the potential to TTS to provide themselves with a voice[3].

C. Training and educational aid

Speech has several advantages over written language. Virtual teachers contributing to a distance learning course, for example, could teaching on-line tutorials. This can be particularly advantageous in situation, where the presence of a real teacher can be embarrassing for the student, as has been noted for sufferers of dyslexia[2].

D. Remote access to online information

Any written information that is stored online, for example electronic mail, news items can all be accessed aurally by means of speech synthesizer

II. HISTORY of TTS

Developments in electronic signal processing resulted in development/research in the field of machine to create human voice. In 1779, the Danish scientist Christian Kratzensteim, built model of human vocal tract, that could produce five vocal sound, they are [a:], [e:], [i:], [o:] and [u:]. This was the first invention related to TTS. In 1950, the first computer based speech synthesize system was created. TTS was firstly developed for English language in Japan. In 1991, the Ministry of Communication & IT (MCIT) started a program called The Technology Developed for Indian Languages (TDIL), for building technology solution for Indian Languages. This was the turning point of Malayalam TTS.

III. MALAYALAM SCRIPT

Malayalam now consists of 53 letters including vowels and consonants. The character set consist of 13 vowels, 2 left vowel sign, 7 right vowel sign, some appear on both side of the Conj\consonant 30 commonly used conjuncts, 36 consonants and vowel signs[10]. The orthographic representation of speech sounds for Malayalam language is the Aksharas, which are the basic unit of the writing system.

IV. OCR SYSTEM

In recent years OCR system has received considerable attention because of the tremendous need for digitization of printed document. The goal of OCR is to classify optical patterns corresponding to alphanumeric or other characters. An OCR system for printed text documents in Malayalam, segments the scanned document images into text line words and further characters. The scanned image of a printed Malayalam text is the input to the system and the output is the editable computer file containing the text data in the printed page. Segmentation and feature extraction are the most important phases involved in the system. There are many OCR systems available for handling English documents, however there are many not reported effort for Indian languages.

V. SPEECH SYNTHESIS

Manuscript received May 10, 2013

Priyanka jose, Electronics and Communications Department, Mahatma Gandhi University College of Engineering. Thodupuza, India.

Dr. V. Govindaru, Research & Development Department, Centre for Development of Imaging Technology, Trivandrum,India.

The conversion of an arbitrary given text into a spoken waveform is the main objective of a TTS system. Synthesized speech can be generated from the corresponding pieces of recording speech that store in the database.

A. Input

Indian language script is stored in digital computers in UNICODE, ISCII, ASCII and in transliteration process of various fonts. The input text could be available in Unicode font is synthesis by the engine. Unicode is a computing industry standard for the purpose of encoding, representation and handling of text in world's writing system.

B. Speech Generation

To synthesis the acoustic wave form is the objective of the speech generation component. Speech generation has been attempted by the corresponding recorded sound file by segregating words, sentence and paragraphs.

C. Methodology

A quick review of literature shows that following Malayalam speech engines are available with some advantages and limitations. They are:

I) E-Speak: It is originally knows as Speak and written for Acorn/RISC_O computers starting in 1995. Espeak is speech synthesizer software for English and other languages including Malayalam. It uses a formant synthesis method. Advantages: read more than 50 languages at a stretch, provides complete speech support for Orca, size is 2MB, provide support to online learning of visually challenged. Limitations: native speakers not involved in development, the Malayalam phonemes used at present is not perfectly legible to comprehend the spoken text.

Swaram: a joint project of Kerala State IT Mission, Society for Promotion of Alternative Computing and Employment (SPACE) and designed by INSIGHT. It can be used for listening any written work in Malayalam. Any type of file that support Unicode format can run on this software. Advantages: native speakers involved in development, smaller size 3-4MB.

Limitations: Malayalam speakers not perfectly legible.

I) ML-TTS: It works with both Windows and Linux. ML-TTS was developed through the effort of IIT Madras, IIT Hyderabad, C-DAC Trivandrum and Mumbai.

Advantages: native speakers involved in development, human voice, legible.

Limitations: bigger size 2GB, less mobility, slow and without speed control.

Dhvani: this system has been developed by Simputer trust, headed by Dr. Ramesh Hariharan at Indian Institute of Science, Bangalore in year 2000. Using of various tools available in Matlab solves above said limitations of speech engines already developed; especially in controlling of talking speed.

VI. PROPOSED SYSTEM

The proposed system in figure.1 has two sections: OCR system and TTS system.

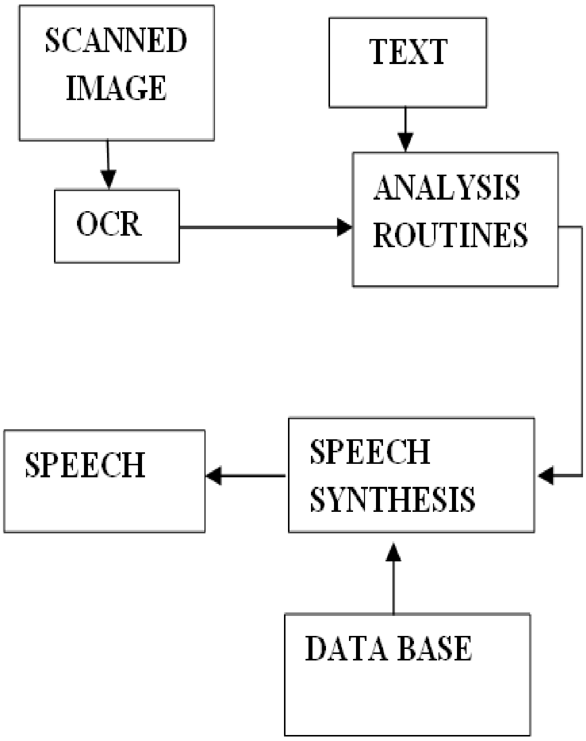

Fig.1 Block diagram of proposed system

First discuss about the OCR system partially developed by C-DIT Trivandrum. The process of OCR involves several steps including pre-processing, feature extraction and post-processing(Fig.2).

A. Scanning

Text digitization is a process to convert the image into proper digital image. This can be performed either by a flat-bed scanner or a hand-held scanner. Scanned image has a resolution level typically 300-1000 dot per inch for better accuracy of text extraction and saves it in preferably TIF,JPG and GIF format.

B. Pre-processing

Pre-processing consists of a number of preliminary steps to make the raw data usable for recognizer[9]. Firstly the scanned image is converted to gray scale image by binarization method. sometimes skew detection and correction method is necessary to digitized image to make text lines horizontal. The noise free image is passed to the segmentation step, where the image is segmented in to characters. Various segmentation processes are explained in[9]. It is the most important aspect of pre-processing stage.

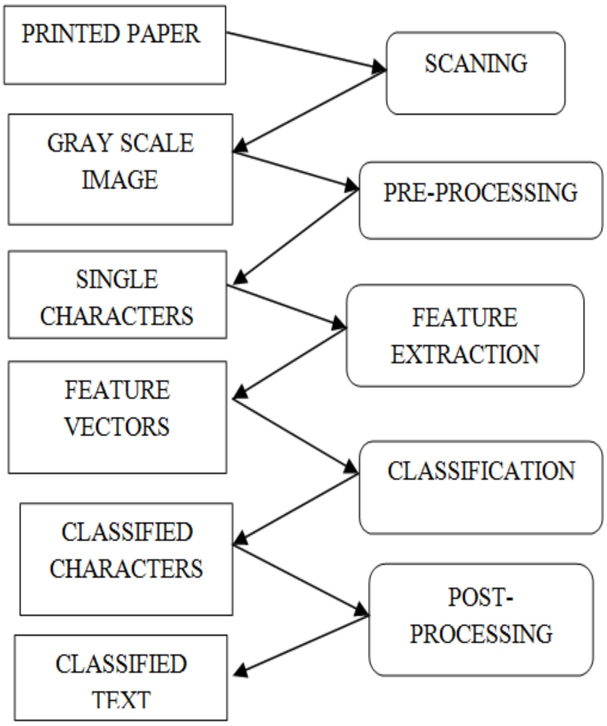

Fig.2 Steps of OCR

C. Feature extraction and Classification

All characters will be divided into geometric elements like lines, arc and circles and compare the combination of these elements with stored combination of known characters[9]. Common feature extraction and classification method is explained in [10].

D. post-processing

Remaining step is post-processing in reorganization. It include spell checking, error checking and text editing etc, when the recognized character does not match with the original one or cannot be recognized from the original one.

Second section explains about TTS system. Generating technologies for synthetic speech waveforms are, Formant synthesis, Articulatory synthesis and Concatenative synthesis. Formant synthesis seeks to mimic human speech by artificially creating the movements of formants. Formants are the resonant regions exhibited by the vocal tract[2]. In formant synthesis, dynamically changing formant frequencies and bandwidths bare no relationship with the articulatory specifications of the vocal tract is the main disadvantage. Thus articulatory synthesis attempt to modeling the geometry of human vocal tract that would recreate a specific spectram, and has a relation to the human vocal system. But in articulatory synthesis, the major disadvantage is articulatory ambiguity. From the above analysis it is convinced that concatenative synthesis is well for speech wave generation.

E. concatenative synthesis

The concatenation of segments of recorded speech is the concatenative synthesis. For a given text, the wave form segments are stored in a database are joined based on some joining rules. This was the type of method employed by the UK telephone network's speaking clock, introduced in 1936[2]. It is the easiest way to produce natural intelligible and natural sounding synthetic speech by connecting the prerecorded natural utterances. To find correct unit length is the most important aspects in concatenative synthesis. Unit selection concatenative synthesizers utilize extremely large speech corpuses. It is necessary to select the segments with minimum of joints, (Fig.1).

F. Database generation

The proposed system is maintained with a database for both Audio and text file. In mat lab, ASCII values and corresponding wavsounds are stored in matrix format, ascii values in the first raw and wave sounds in second raw. As per the ascii values length it has been sorted in ascending order to reduce error rate(Table 1). Turning of speak speed in accordance with native speakers is the major issue concern in acceptance of output.

TABLE 1

Input Text	**Ascii Value**	**Speech Out**
1256= 1+2+5+6	49 0 50 0 55 0 54 0	Onnu Randu Anju Aaru
കോഴി=ക+ോ+ഴ+ി	21 13 75 13 52 13 63 13	കോഴി
ആന=ആ+ന	6 13 40 13	ആന

The TTS system is illustrated in fig.1 consists of a set of analysis derived from computational linguistics. The system identifies words or numbers in the given text, and splits into syllables. The syllables are converted to corresponding Ascii values. The speech is generated by the concatenating coded speech segments.

And the same letter which has two sounds is identified by the conditions.
For eg:

നനഞ്ഞു=ന+ന+ഞ്ഞു

Here first na() and second na() has different pronounciation. First letter is the dental nasal and second letter is the aiveolar nasal as shown in figure.3.

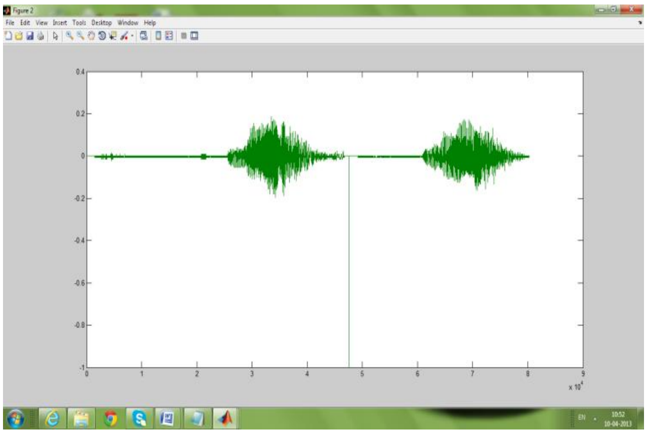

Fig.3 Different sound for na(ന)

The output wave file is modulated by a modulation technique. A parameter of a sound or audio signal called carrier, is varied systematically, the signal is said to be modulated. Full modulation or 100% modulation refers to the maximum permissible level of the system.

G.Naturalness in speech

Text to speech is the artificial production of sound. This sound can be created by the recorded speech segments that are stored in database. The storage of entire letters allows for high quality output. Mainly the quality of the speech synthesizer is analyzed by its similarity to the human voice and its ability to understand.

VII. Conclusion

In this paper we discussed about Malayalam TTS synthesizer. It is observed that the development of TTS in Indian languages is a difficult task, especially for Malayalam, in which same letters is pronounced in multiple ways. Success of Malayalam TTS depends not only on addressing of above said issue but also in corporating of regional variation in speaking of Malayalam.

Acknowledgment

It is with great sense of satisfaction that I present this paper. First of all, I thank God Almighty for leading me talent for undertakes this venture. I immensely thankful to Dr. B. Lethakumary, Head of Electronics and Communications Department, Mahatma Gandhi University College of Engineering, Thodupuzha, for giving me the opportunity and permission to undertake this work. I am thankful to Dr. V. Govindaru, Head of R&D Department, C-DIT, Trivandrum, for his valuable suggestions and guidance throughout my work. I also express my sincere gratitude to all my teachers and staff members of C-DIT, Trivandrum for their timely help and suggestions.

References

[1] S.D. Shirbahadurkar and D. S. Bormane, "*Speech synthesizer using concatenative synthesis strategy for Marathy language*" International journal of recent trends in engineering. Vol 2, November 2009.

[2] Alan O Cinneide, David Dorran and Mikel Gaiza, "*A brief introduction to speech synthesis and voice modification*" Sound Electric 2007, November 2007.

[3] D. H. Klatt, " Review of text to speech conversion for English." Journal of acoustical society of America. Vol -82. PP-(737-793), 1987.

[4] J. Andrew Hunt and W. Alan Black, " Unit selection in a concatenative speech synthesis system using a large speech database." IEEE international conference on acoustics, speech and signal processing, Atlanta, Georgia.

[5] Arun Soman, S. Sachin Kumar, V. K. Hemanth, M. Sabarimalai Manikandan, and K. P. Soman, "*Corpus driven Malayalam text to speech synthesis for interactive voice response system*" International journal of computer application, vol-29, no.4, September 2011.

[6] Haowen Jiang, "*Malayalam- A grammatical sketch and a text*" Dept of Linguistic, Rice University, April 2010.

[7] R. K. Sunil Kumar and N. K. Narayanan "*Malayalam speech phoneme recognition using zero crossing information of speech signal and artificial neural network*".

[8] Stephen Weiss, *"Speech synthesis with hidden marcov models*" Speech processing group, ETH Zurich 2007.

[9] Farjana Yeasmin Omee, Shiam Shabbir Himel and Md. Abu Naser Bikas. "A Complete Workflow forDevelopment of Bangla OCR", in International Journal of computer Application, vol-29. No-9, May 2011.

[10] Bindu Philip and R. D. Sudhaker Samuel, "An Efficient OCR for Printed Malayalam Text Using Novel Segmentation Algorithm and SVM Classifier", in International Journal of Recent Trends in Engineering, vol-1, May 2009.

Priyanka Jose ia a post graduate engineering student in Applied Electronics at Mahatma Gandhi University College of Engineering, Thodupuzha, India.She completed her graduation in Electronics and Communication Engineering.

Dr. Govindaru V is working as Head of Research and Development Division in C-DIT, Thiruvananthapuram. India. He did his Ph.d from ISEC. Banglaore, India.