Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 6.017

IJCSMC, Vol. 6, Issue. 4, April 2017, pg.40 – 44

TEXT RECOGNIZER AND DOCUMENT READER IN VARIOUS LANGUAGES AND ACCENTS

Ajith U[1], Aravind N[2], Deva Prasad L[3], Swaminathan B[4]

[1]Computer Science and Engineering & Rajalakshmi Engineering College, India
[2]Computer Science and Engineering & Rajalakshmi Engineering College, India
[3]Computer Science and Engineering & Rajalakshmi Engineering College, India
[4]Professor, Computer Science and Engineering & Rajalakshmi Engineering College, India

[1] ajith9026.u@gmail.com; [2] aravind.loke333@gmail.com; [3] devaspark7@gmail.com

Abstract -- The goal of this application is used to convert the text into the speech. Speech is the primary means of communication between people. During synthesis very small segments of recorded human speech are concatenated together to produce the synthesized speech. The quality of a speech synthesizer is judged by its similarity to the human voice and by its ability to be understood. A text to speech synthesizer is a computer based system that should be able to read any text aloud and it also allows people with visual impairments and reading disabilities to listen to written works on a home computer. Many computer operating systems have included speech synthesizers since the early 1990s. Recent progress in speech synthesis has produced synthesizers with very high intelligibility. The benefit of the proposed system is that the user can also control the variation of the voice output. Hence speech synthesis task is to develop a machine having a clear, usual sounding voice for conveying information to a user in a preferred accent, language, and voice.

Keywords -- Text to Speech, Speech synthesis, Voice Output, Prosody, Text Analysis.

I. INTRODUCTION

Speech is the most broadly and frequently used method of communication between humans. Intended for real communication, the clearness of speech and accent are the important part to bring the message correctly. The synthetic making of human speech is called Speech Synthesis. The word Synthesis is defined by the vocabulary as fusion or mixing. Speech synthesis is a process of automatic generation of speech by machines or computers. Naturalness defines the output speech sounds alike to human speech and intelligibility is which the output language sound is agreed. In general, alphabets are collected as word, and then words are formed as a sentences or text for a meaningful message These systems are only applicable when a limited vocabulary is required, and when sentences to be pronounced have a very restricted structure, as in the case for the

announcement of arrivals of. In the context of text to speech synthesis, it is impossible to record and store all the words of the language. It is thus more suitable to define text to speech as the automatic production of speech. There are different types of synthesis for delivering speech.

A. Concatenation synthesis

Concatenation synthesis is based on the concatenation of segments of the recorded speech. Generally, concatenation synthesis produces the most natural-sounding synthesized speech. However, differences between natural variations in speech and the nature of the automated techniques for segmenting the waveforms sometimes result in audible glitches in the output.

B. Unit selection synthesis

Unit selection synthesis uses large databases of recorded speech. During database creation, each recorded utterance is segmented into some or all of the following: individual phones, diphones, words, phrases, and sentences. Typically, the division into segments is done using a specially modified speech recognizer set to a "forced alignment" mode with some manual correction afterward, using visual representations such as the waveform and spectrogram. An index of the units in the speech database is then created based on the segmentation and acoustic parameters like the fundamental frequency pitch, duration, position in the syllable, and neighbouring phones. At run time, the desired target utterance is created by determining the best chain of candidate units from the database (unit selection). This is achieved using a weighted decision tree. Unit selection provides the greatest naturalness, because it applies only a small amount of digital signal processing to the recorded speech. Digital Signal Processing often makes recorded speech sound less natural, although some systems use a small amount of signal processing at the point of concatenation to smooth the waveform. The output from the best unit-selection systems is often indistinguishable from real human voices, especially in contexts for which the ext to speech system has been tuned. However, maximum naturalness typically require unit-selection speech databases to be very large, in some systems ranging into the gigabytes of recorded data, representing dozens of hours of speech. Recently, researchers have proposed various automated methods to detect unnatural segments in unit-selection speech synthesis systems.

C. Diphone Synthesis

Diphone synthesis uses a minimal speech database containing all the diphones (sound-to-sound transitions) occurring in a language. The number of diphones depends on the phonetics of the language: for example, Spanish has about 800 diphones, and German about 2500. In diphone synthesis, only one example of each diphone is contained in the speech database. At runtime, the target prosody of a sentence is superimposed on these minimal units by means of digital signal processing techniques such as linear predictive coding or more recent techniques such as pitch modification in the source domain using discrete cosine transform Diphone synthesis suffers from the sonic glitches of concatenation synthesis and the robotic-sounding nature of formant synthesis, and has few of the advantages of either approach other than small size. As such, its use in commercial applications is declining although it continues to be used in research because there are a number of freely available software implementations.

D. Domain-specific synthesis

Domain-specific synthesis concatenates pre-recorded words and phrases to create complete utterances. It is used in applications where the variety of texts the system will output is limited to a particular domain, like transit schedule announcements or weather reports. The technology is very simple to implement, and has been in commercial use for a long time, in devices like talking clocks and calculators. The level of naturalness of these systems can be very high because the variety of sentence types is limited, and they closely match the prosody and intonation of the original recordings. Because these systems are limited by the words and phrases in their databases, they are not general-purpose and can only synthesize the combinations of words and phrases with which they have been pre-programmed. The blending of words within naturally spoken language however can still cause problems unless the many variations are taken into account. For example, in non-rhotic dialects of English the ***"r"*** in words like ***"clear"*** /ˈklɪə/ is usually only pronounced when the following word has a vowel as its first letter (e.g. ***"clear out"*** is realized as /ˌklɪəɾˈʌʊt/). Likewise in French, many final consonants become no longer silent if followed by a word that begins with a vowel, an effect called liaison. This alternation cannot be reproduced by a simple word-concatenation system, which would require additional complexity to be context-sensitive.

II. OVERALL ARCHITECTURE

A text-to-speech system is composed of two parts a front-end and back-end.The front-end has two major tasks.

A. Text Normalization

It converts raw text containing symbols like numbers and abbreviations into the equivalent of written-out words. This process is also known as pre-processing, or tokenization.

B. Text to Phoneme

The front-end then assigns phonetic transcriptions to each word, and divides and marks the text into prosodic units, like phrases, clauses, and sentences. The process of assigning phonetic transcriptions to words is also known as grapheme-to-phoneme conversion.

C. Linguistic Information

Phonetic transcriptions and prosody information together make up the symbolic linguistic representation that is output by the front-end.

D. Synthesis

The back-end—often referred to as the synthesizer—then converts the symbolic linguistic representation into sound. In certain systems, this part includes the computation of the target prosody (pitch contour, phoneme durations), which is then imposed on the output speech.

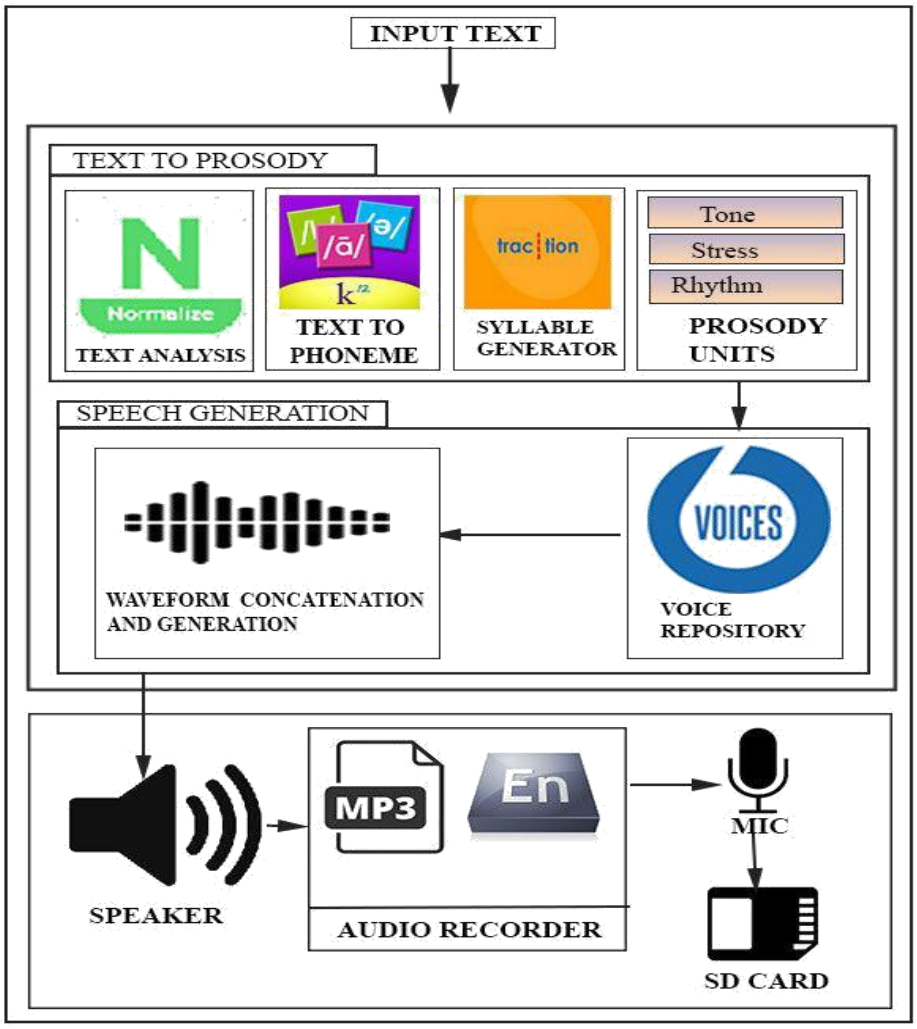

Fig. 1 Overall Architecture

III. **IMPLEMENTATION**

The application is developed on the basis of three modules.

1) Text Input

This module fetch the input from the user in the form of text. The input characters can be in-form of direct entry of text in specified field or the user can upload a text document to the application from the mobile device. The system also has a file browser to navigate easily to the folder where the document is located. The user has also the facility of reviewing the pre -processed output.

2) Voice Output

This module is the core for processing the fetched input. In simple the user input the doc or text into the application, and the output voice language must be chosen and On click, the fetched input is parsed and analyzed. This analyzed text is converted into the normalized form after which the normalized text is transformed into a phonetic representation and this in turn assigned into its equivalent waveform .This process of text to speech/voice conversion is called as synthesis.

3) Audio Storage

Audio storage module performs the saving of output audio in mobile device. At first the output audio is processed by the audio manager and a audio file is created which in turn can be stored in device storage or in external storage devices.

IV. **CONCLUSION**

User can easily access the application. Portable application. People can use this application from anywhere. If the user wants to communicate with others, can easily give the text and convert it into the specified natural languages. So, the user can able to convey his message to others. In this way, This application makes the user communication easier and reduces the difficulties of the user.

ACKNOWLEDGEMENT

We have taken efforts in this application. However, it would not have been possible without the kind support and help of our institution which provided data for this project. We highly indebted to Swaminathan for his guidance and constant supervision as well as for providing necessary information regarding the application and also his support in completing the project. The authors are thankful and gratefully acknowledge all reviewers for their valuable suggestions for enriching the quality of the paper.

REFERENCES

[1]. Itunuoluwa Isewon, Jelili Oyelade, and Olufunke Oladipupo. ”Design and Implementation of Text To Speech Conversion for Visually Impaired People ”. International Journal of Applied Information Systems Volume, April 2014. covenantuniversity.edu.ng/content/download/23746/161028/file/ijais14-451143.pdf

[2]. Kaladharan N. “An English Text to Speech Conversion System”. International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 10, October-2015. https://www.ijarcsse.com/docs/papers/Volume_5/10_October2015/V5I10-0228.pdf

[3]. Hay Mar Htun, Theingi Zin, and Hla Myo Tun. “Text To Speech Conversion Using Different Speech Synthesis”. International Journal Of Scientific & Technology Research Volume 4, Issue 07, July 2015. http://www.ijstr.org/final-print/july2015/Text-To-Speech-Conversion-Using-Different-Speech-Synthesis.pdf

[4]. Tapas Kumar Patra, Biplab Patra, and Puspanjali Mohapatra. "Text to Speech Conversion with Phonematic Concatenation". International Journal of Electronics Communication and Computer Technology, Volume 2 Issue 5 (September 2012) . http://www.ijecct.org/v2n5/223_0205M30.pdf.

[5]. Prachi Khilari and Bhope V. P. "A Review On Speech To Text Conversion Methods". International Journal of Advanced Research in Computer Engineering & Technology, Volume 4 Issue 7, July 2015. http://ijarcet.org/wpcontent/uploads/Ijarcet-Volume 4-Issue-7-3067-3072.pdf

[6]. Jisha Gopinath, Aravind, Pooja Chandran, and Saranya. "Text to Speech Conversion System using OCR" International Journal of Emerging Technology and Advanced Engineering, Volume 5, Issue 1, January 2015. www.ijetae.com/files/Volume5Issue1/IJETAE_0115_62.pdf

