



An English Text to Speech Conversion System

Kaladharan N

Assistant Professor, Dept. of Electrical Engineering,
Annamalai University, Tamilnadu, India

Abstract-This paper is demonstrating to convert the international language English text into speech sign. The exchange of text to speech is made by the speech synthesizer. Speech synthesis is the imitation technique of human speech. Text handling and speech generation are two main mechanisms of text to speech system. In a text to speech system, spoken words are automatically formed from text. The extreme vital talents of a synthesized speech are genuineness and fluency. Text to speech system will support in saving the information from websites and documents that contain information in different languages. Database formation, character recognition and text to speech conversion are the essential phases in text to speech analysis.

Key words: Text to Speech system, Speech synthesis, Synthesis tool, C#, .Net

I. INTRODUCTION

Speech is the most broadly and frequently used method of communication between humans [1]. Intended for real communication, the clearness of speech and accent are the important part to bring the message correctly [2]. The synthetic making of human speech is called Speech Synthesis [3]. The word 'Synthesis' is defined by the vocabulary as 'fusion' or 'mixing' [4]. Speech synthesis is a process of automatic generation of speech by machines or computers. 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 [5]. The significant assets of speech synthesis systems are naturalness and articulation. Naturalness defines the output speech sounds alike to human speech and intelligibility is which the output language sound is agreed [6]. In general, alphabets are collected as word, and then words are formed as a sentences or text for a meaningful message. A text to speech (tts) method converts standard human language text into speech. This conversion involves text processing like text study and finding, linguistic theory, models of speech construction, text ruling, phonetic inquiry and linearization, prosodic making and pitch, acoustic-phonetic characterization of language, acoustic processing and speech generation processes [7]. The TTS system can be used to read text from emails, SMSs, web pages, news, articles, blogs, talking books and toys, games, man machine communications etc. This paper presents a system to design a Text to Speech conversion.

This paper is structured as follows, section II-Literature review, and section III-Techniques of speech synthesis, section IV- Methodology, section V-Synthesis tool design, section VI-Implementation, section VII-Results and section VIII-Conclusion.

II. LITERATURE REVIEW

Several systems has been developed which translate one language to another and that convert text to speech. Some of them which deed as inspiration for my system are as following:

The system that can convert the English sentence into corresponding Punjabi sentence, as well as read the text audibly. The system is capable to be used as a web application or desktop application, hence making it more flexible [7]. The phonemes of the Hindi language can be used as the simple element for speech synthesis. Speech database for Hindi Verbal will be developed by using phoneme. The Text to Speech System for Hindi using English Language is able to express a loud Hindi word which is typed in English [8]. Malayalam is a phonetic language having a written form that has direct communication to the spoken form. Input text from an image to a speech synthesis system consists of a character recognizer and TTS system [9]. Yoruba is one of the three major languages in Nigeria. Tone languages, such as Yorùbá and some others are different from languages that have no tone, example is originate in languages like English and French. Clearly a Yoruba Text-to- Speech System is a feasible and a realizable task [10]. Gujarati TTS system using phoneme concatenative methodology is developed by scholars. It focuses primarily on the process of creating a voice for a concatenative Text-To-Speech system, having specific standard output voice to sound like the target voice. [11]. Tamil is one among the Dravidian languages in India. Problems in developing tamil TTS include understanding tamil phonetics, database conception of tamil language, syllable level concatenation, complexity of the language etc. A record has been created from the various domain words and syllables and create a Tamil TTS [12]. Sanskrit with support of Hindi, will communicate the text to speech for delivering the speech output from the Sanskrit and/or Hindi text of the books. System would be accepting various Sanskrit texts, their Hindi clarifications and recite it out using the module covering speech synthesis. Since both the languages has their base script as 'Devnagri', also they are phonetic in nature they do not require separate TTS [13]. The system accepts Kannada text as input which requires normalization. It then produces the

consistent Kannada text which is accepted to the TTS to produce equivalent speech output by understanding the corresponding speech file from the speech database [14].

III. TECHNIQUES OF SPEECH SYNTHESIS

There are different ways to perform speech synthesis:

Concatenative synthesis based on the concatenation of segments of recorded speech. It involves two phases, viz., the offline phase and the online phase. Offline phase includes pre-processing, segmentation and pitch marking. Online phase includes text analysis and synthesis.

Articulatory synthesis is based on physical models of the human speech making system. It uses mechanical and acoustic models of speech production to synthesize speech. This technique basically uses five articulatory parameters like area of lip opening, tongue blade, glottal area, nasal cavities, vocal tract and so on [10].

Formant synthesis is also known as rule based synthesis, makes the acoustic speech data fully through rules on the acoustic correlates of the numerous speech sounds. Formant synthesized speech is consistently clear, even at very high speeds.

Domain specific synthesis concatenates prerecorded words and phrases to create complete utterances. It is used in applications like transit schedule announcements or weather reports, talking clocks and calculators. Where the variety of texts the system will output is limited to a particular domain [15]. The technology is very simple to implement.

Unit selection synthesis is the dominant synthesis technique in text to speech. Unit selection synthesis uses large databases of recorded speech.

Diphone synthesis is most popular method used for creating a synthetic voice from recordings or samples of a particular person [16]. In diphone synthesis, the strength of speech depends on expression or sentence and the model used for prosody. It uses a nominal speech database.

HMM based synthesis, the speech parameters like frequency spectrum, essential frequency and interval are statistically modeled and speech is generated by using HMM based on supreme probability principle [17].

Sinusoidal synthesis uses a harmonic model and decomposes each frame into a set of harmonics of an estimated fundamental frequency. The fundamental parameters like amplitudes, frequencies and phases are changed by care the same spectral envelope [18]. The simple idea is to model every significant spectral component as a sinusoid.

Corpus based speech Synthesis, most state of the art speech synthesis systems which are able to produce more normal speech are generalization of the concatenative combination which is based on dynamic selection of units are based on large quantities of speech data [19].

IV. METHODOLOGY

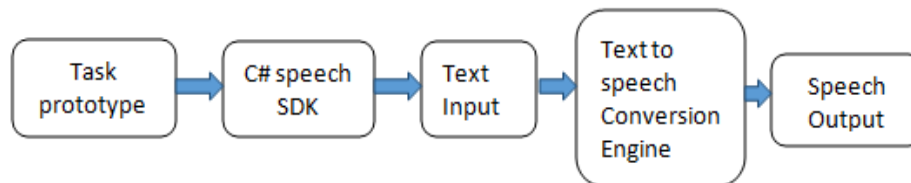


Fig. 1. Block diagram of typical text-to speech system

Task prototype block pacts the design of the scheme of text to speech like buttons, copyreaders and so on. The c# speech software development kit block is used to compile the desired program or code module. Text input block is used to feed the data text from the database file or directly to the editor. Synthesizer block is used to convert the text to speech, as per the input text. Speech output is used to deliver the sound of the corresponding text in to desired manner. This method is built on Microsoft platform. Speech synthesizer tool is designed in Microsoft.NET 4.5 framework using C# Programming in Microsoft Visual Studio 2012 Environment.

V. SYNTHESIS TOOL DESIGN

Synthesis tool is ingenious in Microsoft .NET Framework 4.5 SDK using Microsoft Visual C# object oriented language in Microsoft Visual Studio 2012 environment.

The .NET Framework is an expansion platform for building apps for Windows, Windows Phone, Windows Server, web service and Microsoft Azure. It consists of the common language runtime and the .NET Framework class library, which includes classes, interfaces, and value types that support a wide range of technologies. The .NET Framework delivers a managed execution environment, simplified growth and utilization, and integration with a variety of programming languages, including Visual Basic and Visual C#. The documentation for the .NET Framework includes an extensive class library reference, pace by pace procedures, conceptual overviews and information about trials, compilers, and command-line tools. The .NET Framework 4.5 delivers better performance, reliability, and security.

Integrated development environment (IDE) is a software application that provides complete amenities to computer programmers for software developments. An IDE normally consists of a source code editor, build automation tools and a debugger. Microsoft Visual Studio is an integrated development environment from Microsoft. It is used to develop console and graphical user interface applications along with Windows Forms applications, web sites, web applications, and web services in both built-in code together with managed code for all platforms supported by Microsoft Windows, Windows Mobile Windows CE, .NET Framework, .NET Compact Framework and Microsoft Silverlight .

Microsoft Visual C#, spelled C sharp, is a programming language that is designed for constructing a variety of applications that run on the .NET Framework. C# is simple, powerful, type safe, and object-oriented. C# is similar to java programming. Visual Studio supports Visual C# with a full featured code editor, compiler, designers, project templates, code wizards, a potent and easy to use debugger, and other tools.

The *System.Speech.Synthesis* namespace contains classes for initializing and configuring a speech synthesis engine, for creating prompts, for generating speech, for replying to events, and for modifying voice characteristics. It is a simple Windows application and shows how to use text to speech.

5.1. Computer system configuration:

Microsoft windows 8.1,8 GB RAM, 1TB HD, Realtek HD audio sound card, 2.1 Speaker

5.2. Necessary Softwares required:

Microsoft Visual studio 2012, Microsoft .NET 4.5 Framework, Microsoft WordPad/ Notepad.

VI. IMPLEMENTATION

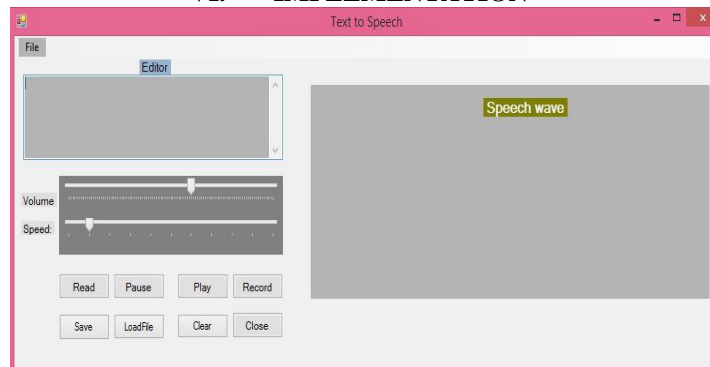


Fig. 2. Implementation of the Text to speech system

- Read the text from the editor
- Clear the editor
- Load the text (.txt) file from the folder
- Editor text is to save in the folder as (*.txt)
- Record the speech as (*.wav), after the Read
- Play the text, when it is pause
- Pause the speech, when it is read and play
- Close the application

- Editor is used to type the text from the keyboard or load from the folder.
- Speech Wave editor is to plot the recorded speech.
- Volume key is used to increase the speech sound to audible.
- Speed is the key to manage the speech quickness

VII. RESULTS

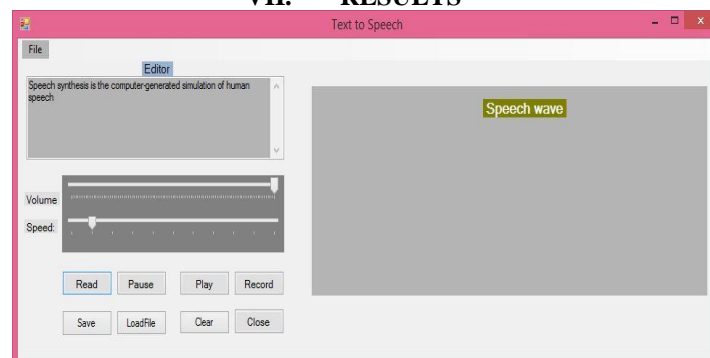


Fig. 3. Text to speech system Input through editor

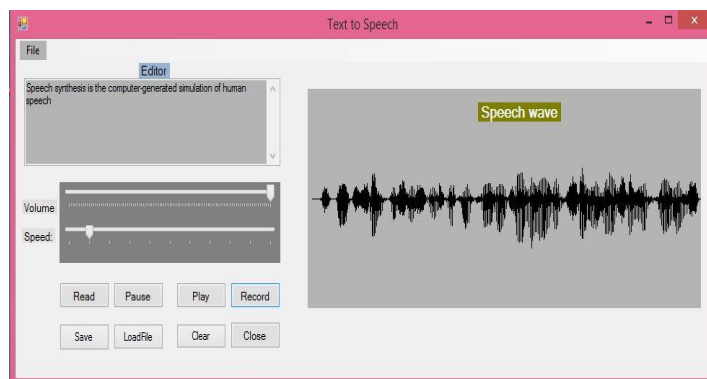


Fig. 4. Text-to-speech system output for 'Sentences'

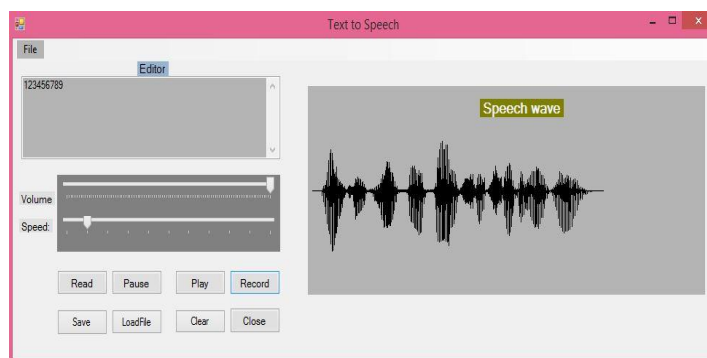


Fig. 5. Text-to-speech system output for 'Numbers'

VIII. CONCLUSION

The planned system gives a very simple method for text to speech conversion. Text inputs like the alphabets, sentences, words and numbers are given to the system. Text to speech conversion is achieved and receive better results which is audible and perfect. The output speech of the corresponding sentence and numbers is showed in the right side screen. Even though many text to speech system is available in the tts field, the .net framework system contributes a satisfactory results. Hence this system is very much used in the web applications, email readings, mobile applications and so on for making an intelligent speaking system.

ACKNOWLEDGMENT

The author thanks the authorities of Annamalai University for providing the required facilities in order to complete this portion of work and personally thanks to his colleague Dr.NJ.Vinothkumar and Dr.T.Shankar to give an inspiration to do this project. And also the author would like to express gratitude to his cousin Mr.G.Kirubashankar.M.Sc, for spending his precious time to finish this project successfully.

REFERENCES

- [1] Kaladharan N, "A Study of Speech Recognition" International Journal of Innovative Research in Computer and Communication Engineering, Vol. 3, Issue 9, September 2015
- [2] Mukta Gahlawat, et al, "Natural Speech Synthesizer for Blind Persons Using Hybrid Approach" Procedia Computer Science Volume 41, 2014, Pages 83–88
- [3] S.M.K.Chaitanya, et,al, "Text to Speech Conversion on Intel Atom Processor, International Journal of Engineering Research-Online, Vol.3, Issue.3, 2015
- [4] Jashin Marina, et al, "Design of Gujarati Text-to-Speech System", International Journal of Research, Volume 02 Issue 05 May 2015
- [5] Helal Uddin Mullah, "A Comparative Study of Different Text-to-Speech Synthesis Techniques", International Journal of Scientific & Engineering Research, Volume 6, Issue 6, June-2015
- [6] R.Shantha Selva Kumari, et, al, "Conversion of English Text-to-Speech (TTS) Using Indian Speech Signal" International Journal of Scientific Engineering and Technology, Volume No.4 Issue No.8, pp: 447-450
- [7] Prabhsimran Singh et.al, "A Text to Speech (TTS) System with English to Punjabi Conversion", November 2014
- [8] Kaveri Kamble , Ramesh Kagalkar, " A Review: Translation of Text to Speech Conversion for Hindi Language" International Journal of Science and Research (IJSR), Volume 3 Issue 11, November 2014
- [9] Aiswarya TS,et.al , " Malayalam Text To Speech Synthesis with Emotions", International Journal of Advance Research in Computer Science and Management Studies Volume 3, Issue 5, May 2015
- [10] O.O. Adeyemo and A. Idowu, "Development and integration of Text to Speech Usability Interface for Visually Impaired Users in Yoruba language", African Journal of Computing & ICT, Vol 8. No. 1 – March, 2015

- [11] Prof. JJ Kothari, Dr. CK Kumbharana, “Designing, development and implementation of Text to Speech algorithm for Gujarati text using concatenative methodology”, International Journal of Scientific and Research Publications, Volume 5, Issue 9, September 2015
- [12] J.Sangeetha.et.al, “Text to Speech Synthesis System For Tamil”, International Journal of Emerging Technology and Advanced Engineering, Volume 3, Special Issue 1, January 2013
- [13] Piyush Mishra,et.al , “ Research Proposal Paper on Sanskrit Voice Engine: Convert Text-to-Audio in Sanskrit/Hindi”, International Journal of Computer Applications (0975 – 8887) Volume 70– No.26, May 2013
- [14] Jagadish S Kallimani et.al , “Normalization of Non Standard Words for Kannada Speech Synthesis”, International Journal of Advances in Computer Science and Technology , 1(1), November-December 2012, 21-26
- [15] Shruti Gupta et al, “Comparative study of text to speech system for Indian language”, International journal of advances in computing and information technology, April 2012
- [16] Abhishek srivatsava ,et al, “An Intelligent Text to Speech System for Windows based Systems and Mobile Devices”, International Journal of Computer Applications (0975 – 8887) Volume 90 – No 16, March 2014
- [17] S.Saraswathi,et,al “ Design of Multilingual Speech Synthesis System”, Intelligent Information Management, 2010, 2, 58-64
- [18] A. Indumathi ,et,al, “Survey On Speech Synthesis”, Signal Processing: An International Journal (SPIJ), Volume (6): Issue (5): 2012
- [19] Kiruthiga S, et.al, “Annotating Speech Corpus for Prosody Modeling in Indian Language Text to Speech Systems”, IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 1, No 1, January 2012
- [20] <http://www.informatix.net/Technologies/SoftwareTechnologies/tabid/90/Default.aspx>
- [21] [https://msdn.microsoft.com/en-us/library/system.speech.synthesis\(v=vs.110\).aspx](https://msdn.microsoft.com/en-us/library/system.speech.synthesis(v=vs.110).aspx)