**TEXT**

**TO**

**SPEECH SYNTHESIS**

**INTRODUCTION**

Digital speech processing plays a vital role in modern speech communication research and applications. The fundamental purpose of speech is communication; it means transmission of message between human and machine. Text to speech system (TTS) converts text into voice using a speech synthesizer . It is the artificial production of human speech. A computer system used for this purpose is called a speech synthesizer, and that is implemented in both software and hardware form like ARM7 microcontroller which converts Text to Speech and Speech to Text. A text-to-speech (TTS) system converts normal language text into US & UK English accents. This synthetic speech cannot be understood by a person with average communication skill in English language. Text to speech system processes are significantly different from live human speech production. Live human speech production depends on complex fluid mechanics dependent on changes in lung pressure and vocal tract constrictions .The objective of a text to speech system is to convert an arbitrary given text into a corresponding spoken waveform. Text processing and speech generation are two main components of a text to speech system. The objective of the text processing component is to process the given input text and produce appropriate sequence of phonemic units. These phonemic units are realized by the speech generation component either by synthesis from parameters or by selection of a unit from a large speech corpus. For natural sounding speech synthesis, it is essential that the text processing component produce an appropriate sequence of phonemic units corresponding to an arbitrary input text.

**SPEECH SYNTHESIS**

Speech synthesis is the artificial production of human speech. A computer system used for this purpose is called a speech computer or speech synthesizer, and can be implemented in software or hardware products. A text-to-speech (TTS) system converts normal language text into speech; other systems render symbolic linguistic representation like phonetic transcriptions into speech.

Synthesized speech can be created by concatenating pieces of recorded speech that are stored in a database. Systems differ in the size of the stored speech units; a system that stores phones or diphone provides the largest output range, but may lack clarity. For specific usage domains, the storage of entire words or sentences allows for high-quality output. Alternatively, a synthesizer can incorporate a model of the vocal tract and other human voice characteristics to create a completely "synthetic" voice output.

A text-to-speech system is composed of two parts a front-end and a back-end. The front-end has two major tasks. First, it converts raw text containing symbols like numbers and abbreviations into the equivalent of written words. This process is often called text normalization, pre-processing, or tokenization. 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 called text-to-phoneme or grapheme-to-phoneme conversion. Phonetic transcriptions and prosody information together make up the symbolic linguistic representation that is output by the front-end. 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.

**TYPES OF SPEECH SYNTHESIS**

The two primary technologies generating synthetic speech waveforms are concatenative synthesis and formantsynthesis. Each technology has strengths and weaknesses, and the intended uses of a synthesis system will typically determine which approach is used.

1. **CONCATENATIVE SYNTHESIS**

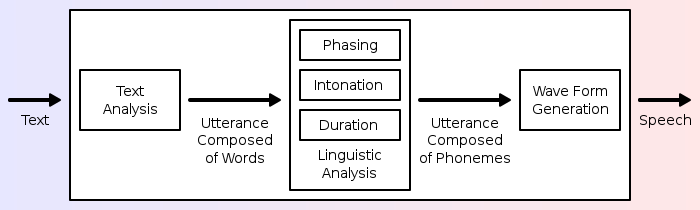
Concatenative synthesis is based on the concatenation (or stringing together) of segments of recorded speech. Generally, concatenative synthesis produces the most natural-sounding synthesized speech. However, differences between natural variations in speech and the nature of the automated techniques for segmenting the waveforms sometimes result in audible glitches in the output. There are three main sub-types of concatenative synthesis.

1. Unit Selection Synthesis
2. Diphone Synthesis
3. Domain Specific Synthesis
4. **FORMANT SYNTHESIS**

This method is sometimes called rules-based synthesis; however, many concatenative systems also have rules-based components. Many systems based on formant synthesis technology generate artificial, robotic-sounding speech that would never be mistaken for human speech. However, maximum naturalness is not always the goal of a speech synthesis system, and formant synthesis systems have advantages over concatenative systems. Formant-synthesized speech can be reliably intelligible, even at very high speeds, avoiding the acoustic glitches that commonly plague concatenative systems. High-speed synthesized speech is used by the visually impaired to quickly navigate computers using a screen reader. Formant synthesizers are usually smaller programs than concatenative systems because they do not have a database of speech samples. They can therefore be used in embedded systems, where memory and microprocessor power are especially limited.

1. Articulatory Synthesis
2. HMM Based Synthesis
3. Sinewave Synthesis

**TEXT TO SPEECH SYSTEM**

****

**LITERATURE SURVEY**

Review on Text-To-Speech Synthesizer

8 AUGUST 2015

Suhas R. Mache

Research Fellow, Dept. Of Computer Science & IT, Dr. B.A.M. University Aurangabad, India

Manasi R. Baheti ,

C. Namrata Mahender

Asst. Professor, Dept. Of Computer Science & IT, Dr. B.A.M. University Aurangabad, India

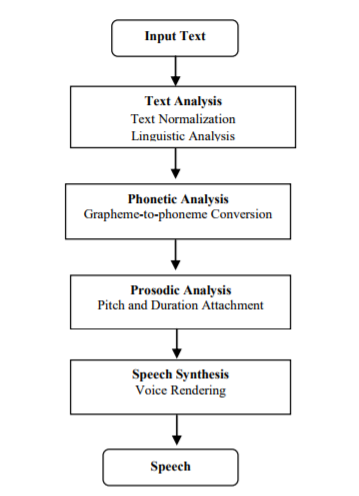
ABSTRACT

In this paper we try to explore various Text-to-Speech synthesis techniques developed and implemented by several speech synthesis researchers and research groups over the world during the last decades. Few Text-to-Speech synthesizer software and hardware’s are mentioned with its features. Today’s most popular device smart phone has capabilities to read a text and e-book aloud. A continues research is going on speech smoother with the help of one of the key approaches i.e. statistical parametric approach to speech synthesis. This paper aims to give an overview of speech synthesis in Indian languages, summarizes and compares the characteristics of various synthesis techniques used.

BLOCK DIAGRAM FOR

TEXT

TO SPEECH SYNTHESIS



**TTS SYSTEM’S MAIN PHASES**

1. Text processing -

A text-to-speech system the input text is first analyzed, normalized and transcribed into a phonetic or some other linguistic representation. Text processing components deals with low level processing issues such as sentence segmentation and word segmentation

Document structure detection - The document structure can be detected by interpreting punctuation mark and paragraph formatting.

Text normalization - The text normalization handles abbreviation and acronyms. The goal of normalization is to match the text e.g. Dr could be rendered as doctor. Proper normalization makes the good output.

Linguistic analysis - Linguistic analysis includes a morphological analysis for proper word pronunciation and syntactic analysis to facilitate accenting and phrasing to handle ambiguities in written text

1. Speech generation –

The speech generation component processes to generate the speech by using parameters as Phonetic analysis-

It focuses on the phone level within each word. Each phone is tagged with information about what sound to produce and how to produce it means style and emphasis.

Grapheme to phoneme conversion: Exact pronunciation of each word of the input sentences is determined.

Homograph disambiguation: Figuring out whether input sentence use the present tense or past tense version of the word. To identify a word tense system depends on dictionary.

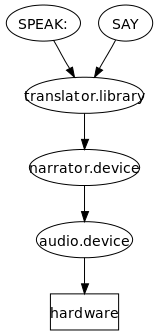
Prosodic analysis - The analysis of prosody is important because it gives basis for marking prosodic effect around our utterance plans i.e. phonological prosodic processing and later to arrive at suitable rendering strategies for the marked prosody i.e. phonetic prosodic processing.

There are two approaches in the prosody.

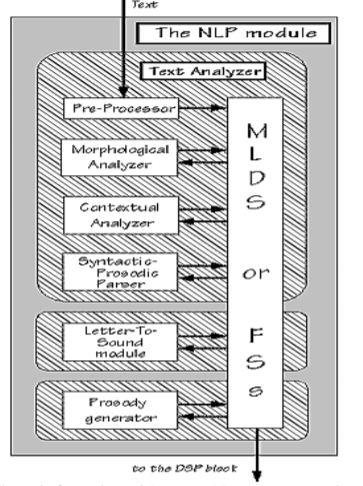
Create an abstract descriptive system which characterizes observations of the behavior of the parameters of prosody within the acoustic signal (fundamental frequency movement, intensity changes and duration movement) and promote the system to a symbolic phonological role.

Create a phonological system which can be used to input to process which eventually result in an acoustic signal jugged by listeners to have a proper prosody.

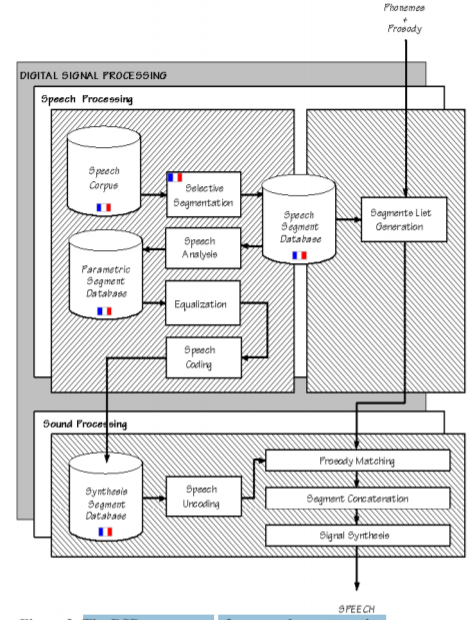
**BASIC ALGORITHM FOR TEXT TO SPEECH TRANSLATION**



**Operations of the natural Language processing module of a TTS synthesizer**

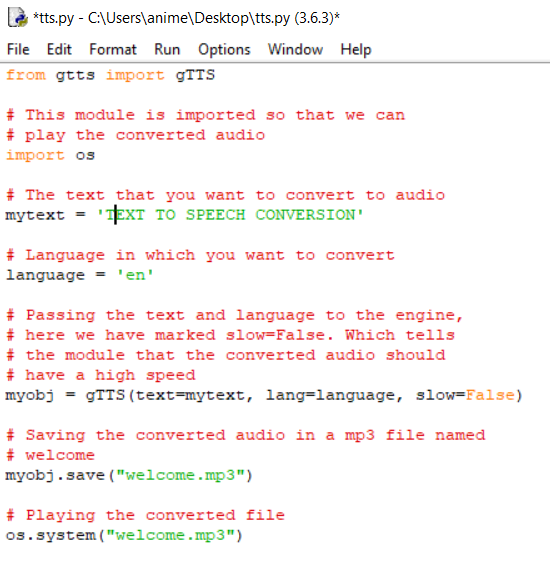
****

**The DSP component of a general concatenation based synthesizer**

****

**ANALYSIS OF PERFORMANCE**

Used the ‘gTTS’ library in python to achieve the desired output in form the form of speech in the specific language required.



Achieved the normal Text To Speech conversion using the above python’s gTTs library and got the desired output to whatever text mentioned in the specified landuage.

**CONCLUSION**

Text-to-Speech synthesizer has steadily developed from last few decades to gain the current shape. The three basic methods of speech synthesis are Articulator, Formant and Concatenative synthesis used in various synthesizers. Many new applications are developed, but intelligibility and comprehensibility of synthetic speech has not reached the acceptable level. Even in India some research organizations are also working on Text-to-Speech in regional languages like Marathi, Hindi, Telugu, Punjabi, Kannada etc. But all these systems do not replicate the human natural speech. There is huge scope in improvement in the synthesis to achieve a high level of natural and emotion aspect.

**Future Scopes**

Text to speech synthesis is a rapidly growing aspect of computer technology and is increasingly playing a more important role in the way we interact with the system and interfaces across a variety of platforms. We have identified the various operations and processes involved in text to speech synthesis. We have also developed a very simple and attractive graphical user interface which allows the user to type in his/her text provided in the text field in the application. Our system interfaces with a text to speech engine developed for American English.

With the use of the code and the library mentioned above in python we have achieved the normal text to speech conversion in the specified language. In future we would like to achieve using the similar library the conversion of text in any language to any speech in any other language and also try to record any desired voice and obtain the speech conversion sound in the recorded desired voice.

REFERENCES

**[1] Archana Balyan, S.S. Agrwal and Amita Dev, Speech Synthesis: Review, IJERT, ISSN 2278-0181 Vol. 2 (2013) p. 57 – 75.**

**[2] D.D. Pande, M. Praveen Kumar, A Smart Device for People with Disabilities using ARM7, IJERT, ISSN 2278-0181 Vol.3(2014) p. 614 – 618.**

**[3] J.O. Onaolap, F.E. Idachaba, J. Badejo, T. Odu and O.I. Adu, in Proc. of the World Congress on Engineering, (London, UK. 2014).**

**[4] Alistair Conkie, Thomas Okken, Yeon-Jun Kim, Giuseppe Di Fabbrizio, Building Text-To-Speech Voices in the Cloud, in Proc. AT&T Labs Research, Park Avenue, Florham Park, NJ- USA).**

**[5] Mark Tatham and Katherine Morton, Developments in Speech Synthesis (John Wiley & Sons, Ltd. ISBN: 0-470-85538-X, 2005).**

**[6] A. Indumati and Dr. E. Chandra, Speech processing –An Overview, Int. J. of Engg. Sci. and Tech., Vol. 4, (2012) p. 2853-2860.**

**[7] Mattingly I. G., Speech Synthesis for Phonetic and Phonological Models, T.A. Sebeok (Ed.) Current Trends in Linguistics, Vol. 12, (1974) p. 2451-2487**

**[8] Klatt Dennis, Review of Text-to-Speech Conversion for English, J. of the Acoustical Soc. of America, Vol. 3, (1987) p. 737-793.**

**[9] Schroeder M., A Brief History of Synthetic Speech, J. Speech Communication, Vol. 13, (1993) p. 231-237.**

**[10] Allen, John, Hunnicutt, Sharon, and Dennis Klatt, Text To Speech, The MITTALK System (Cambridge: Cambridge University Press, 1987).**

**LIST OF ABBREVIATIONS FLOWCHARTS USED**

1. **FLOWCHARTS**
2. **TTS System**
3. **Block Diagram Of Speech Synthesis**
4. **Algorithm For Speech Translation**
5. **Operation of NLP In TTS Synthesizer**
6. **DSP Component Of Analyzer**
7. **Python Library Code Used for TTS translation.**
8. **ABBREVIATIONS**
9. **TTS – text to speech**
10. **NLP- natural language processing**
11. **DSP- digital signal processing**
12. **ARM- Advanced RISC Machines**
13. **RISC- reduced instruction set computer**
14. **HMM- Hidden Markov Model**