A Simultaneous Text To Speech Conversion System

[1]Vandana Chandola, [2]Pooja Jain, [3]Abhishek Mendon, [4]Dilip Dalgade

[1]Vandana Chandola, IT Department, RGIT, Mumbai

[2]Pooja Jain, IT Department, RGIT, Mumbai

[3]Abhishek Mendon, IT Department, RGIT, Mumbai

[4]Dilip Dalgade, IT Department, RGIT, Mumbai

Abstract— **Digital Technology has been flourishing since long and with the introduction of smart phones, it is now one of the most competitive technologies in the world. Smart phones can now provide us with such services in our day-to-day lives that could only be described as privileges in the past. With Simultaneous Text-to-Speech Conversion System, we aim to present an Android app that receives input from the user in the form of text and converts it into speech simultaneously, i.e. the time lag between the input and the output is non-existent. Working on this idea, we intend to help people who lack the power of speech by assisting them communicate with others effortlessly via this app.**

***Index Terms*— Android application, Speech Synthesizer, Text-to-Speech**

I. INTRODUCTION

Text-To-Speech (TTS) system is a mobile-based system that is capable of reading out aloud any text that the user asks it to. Despite seemingly many similarities between this TTS system and any other speaking machine, there is a very basic difference that our proposed system aims at achieving automatic generation of new statements. Here, the latter are systems that need not apply a wide variety of vocabulary in its functionality, that is, usually only a few hundred or thousand words. The sentences spoken out loud here normally have a pre-defined and strict structure. For instance, announcements made in buses and trains. The Text-to-Speech Conversion system that we propose is very different in the sense that automatic generation of speech is to be achieved through a grapheme-to-phoneme (that is text-to-speech) conversion system to deal with the vast and varied vocabulary used by the system.

Although the task seems quite easy to perform from a human's perspective, it is difficult to implement it to a computer. The human being can easily pronounce a set of unknown words or sentences because he already is familiar with the rules of the spoken language. However, the same cannot be said for a computer. It is quite impossible for a reading system to process language and speech synthesis as we humans do. Despite, the leaps of technological advancements made in the field of Artificial Intelligence and Signal Processing, the computer is still too far to match human beings in this respect.

The computer systems consists of two major modules – NLP and DSP, that are responsible for the natural language processing and digital signal processing respectively. The NLP is responsible for the translation to sound along with the appropriate pitch and tone. Whereas, the DSP processes the signals received and converts them into speech. But this processing has a few limitations such as the text spoken often lacks in the emotional aspect when compared to a human.

II. EXISTING SYSTEM

The existing system often puts up insufficiencies regarding the user-friendliness of the UI. The punctuations and abbreviations used can often be misunderstood and mispronounced by the system. For instance, the punctuations used in the abbreviation "B.E.I.T." can be falsely interpreted as the completion of a sentence. The 'Voice Response Systems' are only efficient with a limited vocabulary, because they just concatenate different words or phrases and form a sentence. Also, the pronunciation of the sentence follows a very strict and pre-defined structure.

In the context of TTS synthesis, it is nearly impossible to record and store all the words in a language, of multiple languages. It is therefore more appropriate to describe Text-To-Speech Synthesis as the automatic generation of speech, through a grapheme-to-phoneme translation of the sentences to utter.

III. ISSUES WITH EXISTING SYSTEM

The difficulties in speech generation is almost completely language dependent and showcases many issues. For instance,

ISSN: 2278 – 1323

in English and some other languages (except Hindi, Gujarati, etc) the speech synthesis is complicated as in this language the written words do not always relate to its pronunciation, that is more often than not, there is a remarkable difference between how a word is written and pronounced. To counter this issue, an extremely huge set of all the rules and their corresponding exceptions is required to generate correct pronunciation and manner of speaking for generated speech. Text preprocessing is often a very complicated task and includes various language dependent issues. Digits and numbers must be expanded into full words in some cases only. For example in English, the number 1997 is spoken as "nineteen-ninety-seven" (in case of year) or "one-thousand nine-hundred and ninety-seven "(in case of measure). Dates
and fractional are also different. 2/19 can be expressed as "two-nineteenths" (in case of fraction) or "February nineteenth" (in case of date). Some words, also known as homographs, also create the most difficult issues in TTS synthesis. Homographs are words that have the same spelling but have different meanings and usually different pronunciation. For example: "bow". The word "bow" is pronounced differently in different cases : "Bow before the king!" and "He borrowed the bow from the warrior". Therefore semantic information regarding the word is necessary to produce the correct pronunciation.

Speech generation along with the production of proper tone, stress, pitch and duration is going to be one of the most challenging tasks for years. All these features together are called the prosodic features as they deal with the patterns of rhythm and sound.

The dictionary-based method is fast and accurate, but is an absolute failure if a new word is introduced which is not present in its dictionary of words. As the size of the dictionary increases, so does the memory requirement of the system. On the other hand, the rule-based approach is capable of working on any given input, but more the number of irregular word spellings, more is the complexity of rules.

IV. PROPOSED SYSTEM

With this program, one can convert a text file, or something you type in the program's text area, into WAV audio so the information it contains can also be heard. We can also transfer the text in voice format to other devices. It will have a reasoning process before producing the output. The new system's speech rate and pitch of the voice will be adjustable.

V. DESIGN

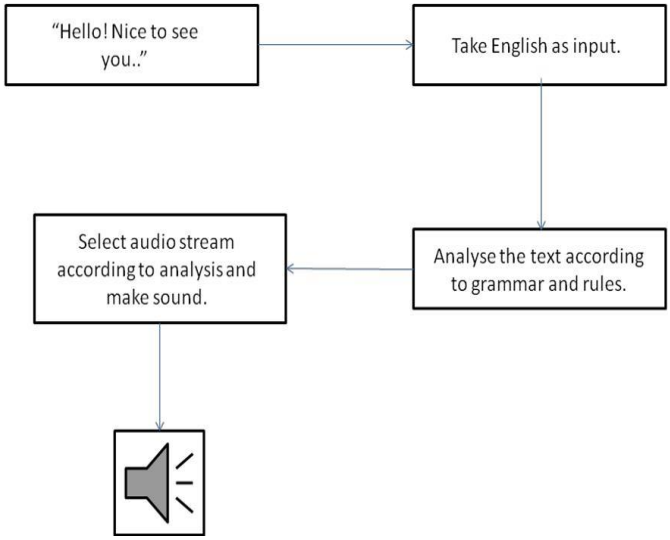

Fig 1: Block Diagram of Proposed System

We intend to use Google Text-To-Speech libraries and make a Text-To-Speech Synthesizer which would have a higher level of comfort while in use. We are making an android application keeping ease of use as our priority. The application will be portable and can be used by anyone and everyone. We aim to design the application in such a way that it can be used quickly as soon as the need arrives. It will basically allow you to convert your text input into speech format. For the same, we will be using TextToSpeech class provided by android.

In order to use this class, we will have to instantiate an object of this class. It provides us with various methods like addSpeech(String text, String filename), getLanguage(), isSpeaking(), setPitch(float pitch), setSpeechRate(float speechRate), shutdown(), stop() etc. We will be using some of these methods to implement our idea of the app. The basic procedure for text to speech conversion involves grapheme to phoneme conversion. It is all about transforming the input text into a form which can be easily spoken by the system. At the minimum, this contains the normalization of the text so that numbers and symbols become words, abbreviations are replaced by their corresponding whole words or phrases, and so on.

VI. CONCLUSION

To summarize, we have discussed the topics relevant to the development of TTS systems .The text to speech conversion may seem effective and efficient to its users if it produces natural speech and by making several modifications to it. This system is useful for deaf and dumb people to interact with other people from society. Text-to-Speech Synthesis is a crucial research and application area. Since the idea of this

ISSN: 2278 – 1323

app has hit us we have researched and planned how our application will be implemented and used. The aim seems achievable after a detailed research on the existing system and then analyzing and planning the system to be implemented. Our aim is to provide user convenience which will be achieved with high level of performance. The project plan and methodologies are apt for the successful completion of the project. The existing system is studied properly and so the proposed will surely eliminate the existing problems. We can visualize the project completion and successful deployment of the same.

REFERENCES

[1] Mrs. S. D. Suryawanshi, Mrs. R. R. Itkarkar, Mr. D. T. Mane, "High Quality Text to Speech Synthesizer using Phonetic Integration" , International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE) Volume 3, Issue 2, February 2014

[2] J. Sangeetha,S. Jothilakshmi, S. Sindhuja, V. Ramalingam, "Text to Speech synthesis system for Tamil", International Conference on Information Systems and Computing (ICISC-2013), INDIA.

[3] Jia *Member, IEEE,* Shen Zhang, Fanbo Meng, Yongxin Wang, and Lianhong Cai*, Member, IEEE,"* Emotional Audio-Visual Speech Synthesis Based on PAD", IEEE transactions on audio, speech, and language processing, vol. 19, no. 3, march 2011.

[4] n.swetha 2k..anuradha," Text-to-speech conversion", International Journal of Advanced Trends in Computer Science and Engineering, Vol.2 , No.6, Pages : 269-278 (2013).

[5]] Swati Ahlawat, Rajiv Dahiya," A Novel Approach of Text to Speech Conversion Under Android Environment", IJCSMS International Journal of Computer Science & Management Studies, Vol. 13, Issue 05, July 2013.

[6] D.Sasirekha, E.Chandra," Text to Speech: A Simple Tutorial", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-1, March 2012.

[7] Miss. Priyanka V. Mhamunkar, Mr. Krishna S. Bansode and Prof. Laxman S. Naik," Android Application to get Word Meaning through Voice", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 2, Issue 2, February.

ISSN: 2278 – 1323