



International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A REVIEW PAPER ON TEXT-TO-SPEECH CONVERTOR

Sneha Tamboli¹, Pratiksha Raut¹, Lavkush Sategaonkar¹, Anjali Atram¹, Shubham Kawane¹, Prof. V. K. Barbudhe²

¹Student, Dept of E&TC Engineering, JCOET, Yavatmal, Maharashtra, India

²Assistant Professor, Dept of E&TC Engineering, JCOET, Yavatmal, Maharashtra, India

ABSTRACT

Speech is one of the oldest means of information to exchange ideas between humans. Text-to-speech (TTS) software project is a windows-based application that can read a text to the user. The software reads a file and pronounces it in its database. And then the program used reads an entire document for the user as assigned. The TTS software can be easily used to help people who are visually challenged or by people who don't want to look at their screen constantly and have to read the whole file with ease. The blind people cannot read on their own so this software can work as an assistant for them who would read out the given document. The proposed design will be acting as a connector to provide a one-way communication whereby the computer communicates with the user only by reading out textual documents or files for the purpose of quick response. The TTS model consists of two side or parts i.e., front-end and back-end. The front end has two works to do. Starting with the conversion of all the numbers and the symbols also the abbreviation into the written words. And the following process is called tokenization. The back-end mostly refers to the work like a synthesizer, which then converts the symbolic conversion into the sound. It sometimes also includes the pitch contour or the durations which in after gives that as a final output speech.

Keywords: TTS (Text-to-speech), symbolic, frontend, backend, database.

1. INTRODUCTION

Speech can be regressed as the composition of what to deliver and how to deliver. Speech conversion is a task which can change the styles of the speech, while retaining some of the linguistic contents. Vocal stimulation is found only in a very few kinds of the animals. But only humans can perform vocal stimulation can be found in some of the species like whales, birds, and bats; but their imitation is basically the stimulation of the sounds.

However, there are different levels of ability and complexities among the voice conversion applications which can be based on the machine learning. Among which some of the applications like Dragon Professional, Speech Matic's, Brains pro, Dictation and Microsoft Azure supports some of the natural way to communicate with other which acts as a barriers to communicate.

Some of the human organs such as vocal tracks and the glands have features which are being influenced by the parameters like human gender, educational belief and also the desired emotion. These three major factors can affect human voice, accent, pronunciation, tone and also device's volume. Also, transmission can affect the voice and distort the speech pattern. Today's speech technologies are commonly available for a limited but interesting range of task.

Text-to-Speech (TTS) is the ability of a computer to produce spoken words by converting text into speech. In other words, Text-to-Speech software is a speech synthesizer that vocalizes text in real time voice in a natural way. Text-to-Speech technology can be used in various areas:

Text to speech can be implemented in IVR systems to create an efficient self-service solution that improves customer satisfaction by informing and guiding callers while reducing cost. TTS can also be used in automated outbound call systems in order to provide information to customers without any need of agents.

2. CLASSIFICATION OF SPEECH RECOGNITION SYSTEM

Speech recognition of human speech has been a long intriguing issue among artificial intelligence and processing researchers. Speech recognition system can be classified in several different types. The challenges are described below briefly.

a) Types of speech utterance

- Isolated word:

Isolated word recognizer requires each word to speak on the bot. It accepts single word at a time.

- Connected word:
It is as same as isolated but it allows the utterance of mark content with a proper minimum pause.
 - Continuous Speech:
It allows the users to speak naturally and the computer will be having the content.
 - Spontaneous Speech:
This type of utterance allows the speech naturally inherit but it is not rehearsed.
- b) Types of speaker model:
- Speaker dependent models:
They are the types of specific speaker. They are easy to develop and also are very accurate but are not flexible.
 - Speaker independent models:
This type of systems is designed for the variety of speaker. This are difficult to develop and less accurate but very flexible.
- c) Types of Vocabulary:
- Small Vocabulary:
Single letter.
 - Medium Vocabulary:
Two or three letter.
 - Large Vocabulary:
More than three letters.

3. LITERATURE SURVEY

Recognition of noisy speech using Dynamic spectral sub band centroids (2004) IEEE Volume11, No. 2.

Author: Kuldip K. Paliwal

A procedure was proposed to construct the dynamic centroid feature vector that essentially embodies the transitional spectral information. It was demonstrated that in clean speech condition SSCs can produce performance comparable to that of MFCCs. Experiments were performed to compare SSCs with MFCCs for noisy speech recognition. The results showed that the centroids and the new dynamics SSC coefficients are more resilient to noise than the MFCC features

Text-to-Speech Synthesis (TTS) (2014) IJRIT, Volume 2, Issue 5.

Author: Okpala Izunn

Text-to-Speech synthesis is a technology that provides a means of converting written text from a speech. The models run on JAVA platform and methodology used were object-oriented analysis and development methodology. With Text-to-Speech synthesis, one can medicate on the capabilities of same as like the handicapped individuals. Actually, in these models it's never been that easy to use Text-to-Speech synthesis at just one click and computer will speak text aloud in a clear and natural soothing voice.

A Communication system for the disabled with emotional synthetic speech produced by rules. (1991) ICA Volume 1

Author: Iain R. Murray, John L. Arnott, Norman ALM and Alan F. Newell.

A system for producing synthesis speech while incorporates vocal emotion effects has been developed. A range of common emotions can be simulated by the TTS system. The system which was made runs on a standard laptop PC, and was enable non vocal persons to express a range of emotions via a high-quality speech synthesizer. And also, conversational speech acts and speaking them with appropriate vocal emotion were developed.

A focus on codemixing and codeswitching in Tamil speech -to-text. (2020). IEEE 2020 8th International conference in Software Engineering Research and Innovation. May 20, 2020.

Authors: Dheenesh Pubadi, Ayush Basandri, Ahmed Mashat, Ishan Gandhi.

This paper was an attempt to develop an application that converts Tamil languages speech to Tamil text, with a view to encourage usage and indirectly ensure the preservation of a classical language, the application converts spoken Tamil to text without autocorrection. The research maintains that it is very much important to maintain the utilization of Tamil language via technology to help in preservation of one of the oldest surviving languages in the world from ancient times. The system is extendable to any other of the languages just by changing the language rules, intonations and the database. This research work also emphasized on the indigenous design considerations for such applications.

Speech to text and text to speech recognition systems (2018) IOSR, Volume 20, Issue 2.

Author: Ayushi Trivedi, Navya Pant, Pinal Shah, Supriya Agrawal

Most of the application find the use of function such as articulatory and acoustics based speech recognition, conversion from speech signals to text signals and from text to synthetic speech signals, language translation amongst various other. In this paper different techniques and algorithms were applied to achieve the mentioned function abilities. Hybrid machine translation is widely used due to its inoculation of advantages of both rule-based as well as statistical machine. It makes sure that there is a creation of syntactically connected and grammatically correct text while also taking care of smoothness in a text, fast learning ability, data acquisitions which are a parts of SMT.

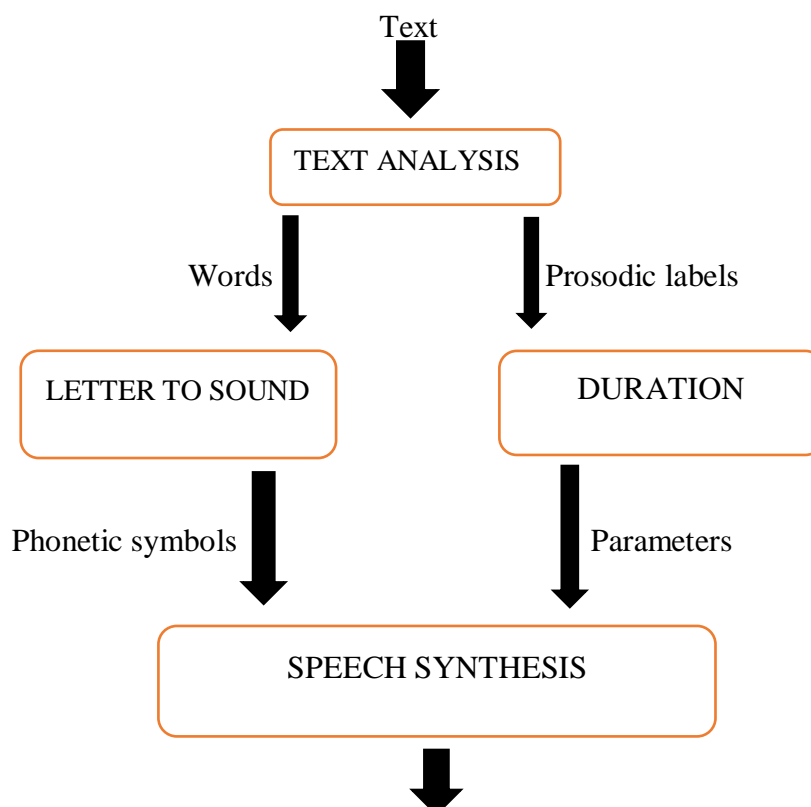
A Robust Isolated Automatic Speech Recognition Systems by using Machine Learning. (August 2019) IJITEE ISSN:2278-3075, Volume-8 Issue-10.

Author: Sunanda Mendiratta, Neelam Turk, Dipali Bansal.

The paper covered architecture of ASR that helps in getting ideas about basic stages of speech recognition system. Also, the techniques of machine learning are used in the model. And artificial neural networks are also covered. The work is done by using the support of vector machines and artificial networks is also covered. The translation of spoken words into respective written scripts is done by speech recognition and language of speech is identified using Automated Speed Recognition (ASR) system. The work shows that traditional classifier results can be further improved by doing hybridization of it with other optimization algorithms.

4. PROPOSED WORK

The development of an software is not an easy or one day task. It requires a lot of time, discussion and knowledge where the real need of the software is analysed. The software will be first tested for feasibility then requirements were specified and analyzed. The Iterative waterfall model will be used in order to provide feedback and make necessary changes even after the completion of a module. Block diagram is given below:



Different python libraries which will be used in project

a) Tkinter:

Tkinter is a python binding to the Tk GUI toolkit. It is the standard python interface to use the TK GUI toolkit. Tkinter is included with standard Linux, Microsoft windows and MAC OS installs of python. Tkinter is implemented as a python wrapper around a complete TCL interpreter embedded in the python interpreter.

This Python framework provides an interface to the Tk toolkit and works as a very thin object-oriented layer on top of Tk. The Tk toolkit is a cross-platform collection of 'graphical control elements, that is widgets, for building application interfaces.

b) Pyttsx 3:

Pyttsx is a good text to speech conversion library in python but it was written only for python2. However, one library gTTS which works perfectly in python3 but it needs internet connections to work smoothly since it relies on google to get the audio data.

5. CONCLUSION

Text-to-speech conversion software project is windows-based application that reads a text file to the user. The software reads a text file or the entered text or the image selected and associated pronunciation in its temporary database.

Text-to-speech converter is like a recent software project that allows even the visually challenged to read and understand various type of documents by their own. The user can just choose the different types of modes and the write the sentences or provide a document then the user can easily explain what he wants to say. So, this software is not just an advancement towards the future development but also a boon for the people those who cannot speak and see.

REFERENCES

- [1] A Robust Isolated Automatic Speech Recognition System using Machine Learning. (August 2019) IJITEE ISSN:2278-3075, Volume-8 Issue-10. Sunanda Mndiratta, Neelam Turk, Dipali Bansal.
- [2] Text-to-Speech Synthesis (TTS). (May 2014) IJRIT, Volume 2, Issue 5. Okpala Izunna, Oluigbo Ikenna.
- [3] A communication system for the disabled with emotional synthetic speech produced by rule. (Sept 1991) ISCA Volume 1. Iain R.Murray. John Arnott.
- [4] Recognition of noisy speed using dynamic spectral sub-band centroids. (March 2004). IEEE Volume11, No. 2. Kuldip K.Paliwal
- [5] A focus on codemixing and codeswitching in Tamil speech to text (2020).IEEE 2020 8th international conference in Software Engineering Research and Innovation. May 20, 2020. Dheenesh Pubadi, Ayush Basandri, Ahmed Mashat, Ishan Gandhi.
- [6] Speech to Text and text to speech recognition systems. (April 2018) IOSR, Volume 20, Issue 2. Ayushi Trivedi, Navya Pant, Pinal Shah.
- [7] G. K. K. Sanjivani S. Bhabad, "An Overview of Technical Progress in Speech Recognition," International Journal of Advanced Research in Computer Science and Software Engineering, 2013.
- [8] P. K. Kurzekar, "A Comparative Study of Feature Extraction Techniques for Speech Recognition System," International Journal of Innovative Research in Science, Engineering and Technology, 2014.
- [9] N. K. P. K. Bhupinder Singh, "Speech Recognition with Hidden Markov Model: A Review," International Journal of Advanced Research in Computer Science and Software Engineering, 2012.
- [10] S. R. Mache, "Review on Text-To-Speech Synthesizer," International Journal of Advanced Research in Computer and Communication Engineering, 2015.
- [11] Ayushi Trivedi, Navya Pant, Pinal Shah, Simran Sonik and Supriya Agrawal, "Speech to text and text to speech recognition systems-Areview," IOSR Journal of Computer Engineering (IOSR-JCE), 2018.
- [12] Kaveri Kamble, Ramesh Kagalkar, "A Review: Translation of Text to Speech Conversation for Hindi Language," International Journal of Science and Research (IJSR) , 2012.
- [13] Mark Gales and Steve Young, The Application of Hidden Markov Models in Speech Recognition, Foundations and Trends in Signal Processing, 2008.