

Text to Speech Conversion in Punjabi-A Review

Priya* and Amandeep Kaur Gahier**

ABSTRACT

Speech is the essential method for communication between individuals. Speech is regularly dependent upon natural speech's concatenation i.e. units which are obtained from a sentence or word. The synthesis of concatenative speech has turned out to be exceptionally well known in these years because of its enhanced affectability to unit context over easier predecessors. Speech synthesis has been worked on for quite a few years from now. Latest progress in speech synthesis has delivered synthesizers having very high comprehensibility. In the process of speech synthesis, the accurateness of information extraction is pivotal in creating synthesized speech having very high quality. Speech synthesis incorporates the algorithmic change of text data which is input to waveforms of speech. The characterization of Speech Synthesizers is done by the technique utilized for synthesis, encoding and storage of the speech. The method of synthesis is controlled by the size of vocabulary as every single conceivable utterances of the language which should be modelled. It is a properly confirmed statement that TTS (text-to-speech) synthesizers intended for utilization in a confined domain dependably perform superior to their counterparts which are of general purpose. The outline of such universally useful synthesizers is complex by the concept that the sound output requires to be very near to natural speech. In this paper, I have presented a survey on the various techniques of text to speech synthesis system.

Keywords: text to speech, NLP, TTS, ASR

1. INTRODUCTION

In the field of sciences, machines are used to perform various computational tasks. Such as machine learning, optical scheduling in cloud computing, transaction and bio-inspired systems. Now a machine needs to take input from human or environment in order to perform its task. Old machines involved Command Line Interfaces which were a good start towards development of computational technologies, but due to the user unfriendly environment, Graphical User Interfaces came into the picture [1]. GUI's are successfully deployed in various web based systems as well as desktop applications. But the exponentially rising technologies has arisen a need to develop systems which are even friendlier than GUI's.

That system is made in artificial intelligence. That is intelligence exhibited by machines or software. It is also the name of the academic field of study which studies how to create computers and computer software that are capable of intelligent behaviour [1].

In today's life everybody needs that the present PC frameworks ought to act like people and turned out to be easy to understand. Indeed, even a number of the considerable research scholars had longed for including the machines in each aspect of human life. With the development of the computing machines power their applications in modern and daily life are additionally rising. The system of text to speech has numerous applications that are valuable in our life, for example, for video conferencing, peopling with

* Research Scholar Computer Science and Engineering Lovely Professional University Phagwara, Punjab, India,
Email: priyamahi94@yahoo.in

** Assistant Professor Computer Science and Engineering Lovely Professional University Phagwara, Punjab, India,
Email: amandeep.19340@lpu.co.in

perusing trouble, telecommunications and multimedia, for visual guides procedure, Elementary and Applied Research, programs of screen reader for the visually impaired and educational state. [3]

As of late, utilization of PCs in speech recognition and speech synthesis has turned into an essential zone of study among computer scientists and speech. The essential inspirations are to furnish clients with a neighbourly vocal interface with the PC and to permit individuals with specific impediment, (for example, visual impairment) to utilize the PC. Speech is the essential method for communication between individuals. The language Punjabi is being discussed by around 104 million people groups in Pakistan, India, and different nations with Punjabi migrants. The language is being composed in script Gurmukhi in Indian Punjab, though in script Shahmukhi in Pakistani Punjab. Punjabi is a language of syllables, hence waveform output is produced by linking the sounds of phoneme set apart in document of sound recorded. Phoneme sounds in various settings have been checked and put away in the database of speech to get normal sounding combined speech. [5]

2. NATURAL LANGUAGE PROCESSING

NLP (Natural Language Processing) is the most straightforwardly science related to handling human (natural) language. It gets from computer science field since PCs, or some other units of handling, are the objective gadgets utilized to finish such preparing. This portrayal reacts essentially to the "Preparing" constituent part in NLP. What makes NLP unique in relation to some other preparing related action is the field of use: the languages of human. They manage more information related perspectives in this way necessitating the backing of abilities of learning by the content processors. Finally, it could be communicated that majority of the mechanized tasks which are of NLP and also related to NLP could be enclosed by the more broad idea of Machine Learning, plainly identified with computer science, which thinks about any subject identifying with the utilization of PCs for the surmising of the significant elements of the frameworks of examination. Since the specific field of analysis is natural language, these specific techniques of learning are of imperative premium, in light of the fact that somehow, we people make utilization of this sort of language as our fundamental method for reasoning and communication, intrinsically. In the event that generally a languagewhich is formal was to be concentrated on (e.g., a programming language), there would be no motivation to make utilization of such approaches of learning in light of the fact that the issues of logic and construction destined to the formalism of that kind of language would as of now be predefined or known.

Regular uses of complete NLP at high level would bargain exclusively with text information (at the system's input and output) for example, machine translation, question answering, text summarization or text classification. At the point when technologies of speech and also other domains ought to be considered, regardless of NLP strategies allude solely to the textual synthesis or analysis of the applications. Either ASR (Automatic Speech Recognition) or TTS (Text to Speech) synthesis need a trustful module of NLP since content information dependably shows up some place in the handling chain. TTS creates utterances of speech from an input content whereas ASR delivers such content from an utterance of speech which is given as input. [2]

3. TEXT TO SPEECH SYNTHESIS

Speech is the most imperative type of correspondence in regular life. In any case, the reliance of human PC cooperation on images and written text makes the utilization of PCs inconceivable for illiterate masses and physically and visually impaired people. TTS (Text-to-speech synthesis) helps researchers of speech processing to follow up on this issue by synthesizing speech (in languages local e.g. Punjabi, Hindi, Tamil and so forth.) from written content as in mobile phones, browsers and so forth. The synthesis of speech can be done by basically three techniques: Articulatory synthesis, Concatenativesynthesis and Formant synthesis.

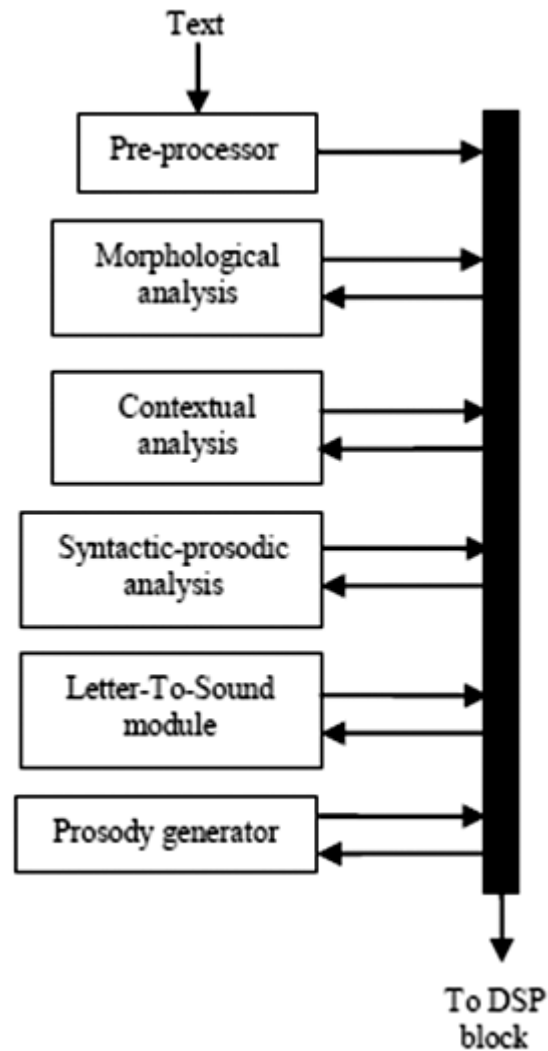


Figure 1: Block diagram of Natural Language Processing module [5]

Articulatory synthesis tries to display the system of production of human speech (particularly vocal tract framework, different articulators' viz. jaw, tongue, lip and so on.) and articulatory procedures straightforwardly. In any case, it is additionally the most troublesome technique to actualize because of absence of information of the complex organs of articulation of human.

Concatenative speech synthesis systems can synthesize sound speech which is more natural and high quality however keeping in mind the end goal for synthesizing speech with different characteristics of voice, for example, emotions, styles of speaker, individualities of speaker and so on., a lot of memory and corpus of speech is required as put away essential units of speech (like diphones, syllables and so forth.) are connected to shape sequence of word utilizing dictionary of pronunciation.

Formant synthesis depends on the guidelines which portray the thunderous frequencies of the vocal tract. The formant strategy utilizes the source-channel model of production of speech, where speech is displayed by parameters of the filter model. The formant synthesis which is rule-based can deliver quality speech which sounds unnatural, as it is hard to assess the model of vocal tract and parameters of source. [6]

Research in T-T-S is a field which is multi-disciplinary: from acoustic phonetics (production and perception of speech) over morphology (pronunciation) and syntax (parts of speech, grammar), to processing

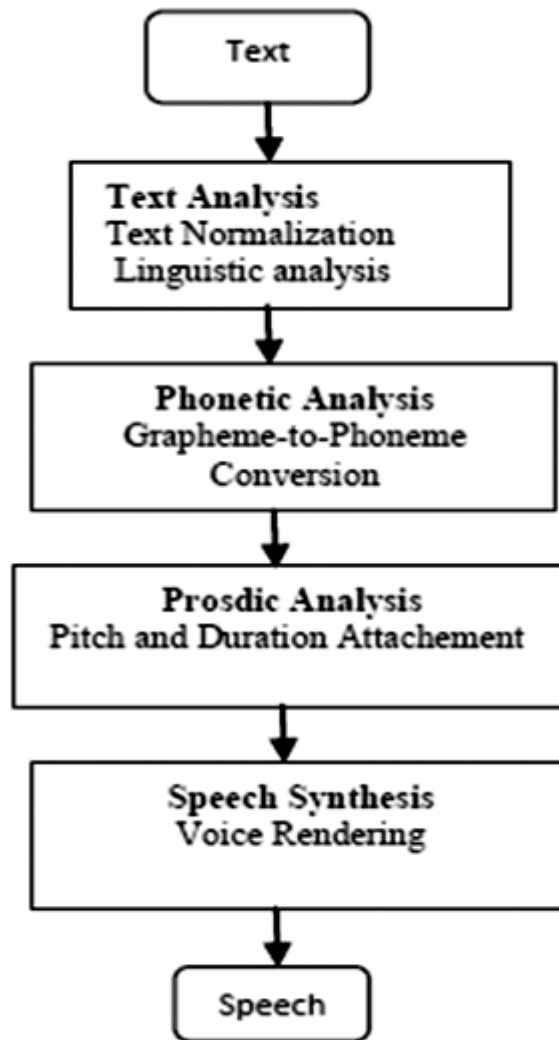


Figure 2: Block diagram of TTS [7]

of speech signal (synthesis). There are numerous stages of processing in T-T-S system: the text front –end analyses and standardizes the text which is incoming, makes conceivable pronunciations for every word in context, and produces prosody (intonation, rhythm, melody, emotions) of the sentence to be spoken. For T-T-S systems evaluation, three parameters need to be assessed: naturalness, intelligibility and accuracy. [7]

The procedure of transformation of text into speech extensively contains two phases: 1) Text analysis and 2) generation of signal of speech.

Text analysis comprises of standardization of the content wherein the symbols and numbers get to be abbreviations and words are supplanted by their entire words or expressions and so forth. The most difficult undertaking in the block of analysis of text is the linguistic analysis which implies syntactic and semantic examination and goes for comprehension the connection of the content. The factual techniques are utilized to locate the most plausible importance of the expressions. This is critical in light of the fact that the word pronunciation may rely on upon its importance and on the setting.

Phonetic Analysis changes over the orthographical symbols into phonological ones by utilizing an alphabet which is phonetic one. For e.g. the alphabet of the *International Phonetic Association* contains symbols of phoneme, their marks which are diacritical and different symbols associated to their pronunciation, other alphabets which are phonetic, for example, SAMPA (*Speech Assesment Methods-Phonetic Alphabet*), *Arpabet* and *Worldbetare* accessible.

Prosody is an idea which contains the speech rhythm, intonation and patterns of speech. At the perceptual level, instinctive nature in speech is credited to specific properties of the signal of speech identified with discernable changes in syllabic length, loudness and pitch by and large known as prosody. Acoustically, these progressions relate to the varieties in the basic frequency (F0), duration and amplitude of units of speech.

Speech Synthesis block at long last produces the signal of speech. This can be accomplished either taking into account parametric representation, in which realizations of phoneme are delivered by machine, or by choosing units of speech from a database. The subsequent short speech units are together joined to create the final signal of speech. [7]

4. NEED OF TEXT TO SPEECH SYSTEM

The field of application of TTS is quickly growing whilst the nature of systems of TTS is likewise steadily increasing. The systems of Speech synthesis are additionally turning out to be more moderate for common clients, which make these systems more appropriate for use in everyday life. A few uses of TTS are:

- Aid to handicapped,
- Talking toys and books,
- Education and Games,
- Multimedia and Telecommunication,
- Man-machine communication,
- An email which is voice enabled
- Vocal monitoring.

5. CHALLENGES IN TEXT TO SPEECH SYSTEM

Speech synthesis has been steadily created over the late decades and it has been incorporated into a numerous new applications. Creating system of speech synthesis is a confounded procedure and, it incorporates the accompanying critical difficulties.

- Advancement of TTS frameworks necessitate information about production of human speech and about languages which are being produced.
- The real implementation of a complete functional system needs good skills of software.
- Majority of systems of TTS do not generate semantic representations of their input text, subsequently, different heuristic techniques are utilized to figure the correct approach for disambiguating homographs, such as looking at neighboring words and utilizing statistics about recurrence of event.
- The most critical characteristics of a system of speech synthesis are intelligibility and naturalness. Intelligibility is the simplicity with which the understanding of output is done whereas naturalness portrays how intently the output sounds as a speech of human. The perfect synthesizer of speech ought to be both intelligible and natural. [8]

6. RELATED WORK

Lots of work has been done in the area of text to speech synthesis system. Little work in this area is given as follows:

<i>S. No.</i>	<i>Title of Paper</i>	<i>Authors</i>	<i>Contribution of Paper</i>
1.	Punjabi Speech Synthesis System for Android Mobile Phones [9]	Jagmeet Kaur, Parminder Singh	This paper presented a Punjabi text to speech synthesizer that can generate output speech on mobile device. Few aspects such as processing time and size of application are taken into account.
2.	Android Based Punjabi TTS System [10]	Hardeep, Parminder Singh	This paper examined the strategy which is utilized for developing synthesis system of TTS for content of Punjabi. The method of concatenative speech synthesis with phonemes is utilized as fundamental concatenation units.
3.	Punjabi Text-To-Speech Synthesis System [11]	Parminder Singh, Gurpreet Singh Lehal	This paper portrayed the Punjabi text to speech synthesis system. Concatenative method has been utilized and syllables have been stated as good selection of unit of speech for databases of speech for many languages.
4.	Improved System for Converting Text into Speech for Punjabi Language [12]	Ramanpreet Kaur, Dharamveer Sharma	This paper discussed few enhancements in the text to speech synthesis system which is format based for input text of Punjabi. Analysis of eSpeak for Punjabi is done and then faults are accurate by utilizing eSpeakedit.
5.	A Technique for Speech Synthesis of Gurmukhi embedded in LATEX document [13]	Jaswinder Singh, Tirath Singh, Sarabjit Singh, Brahmjit Kaur	In this paper, mapping of Gurmukhi script which is embedded in Latex document utilizing Pandey package for the respective audio files of Punjabi is done. Parsing is used in this process.

7. MOTIVATION

In the course of the most recent couple of years there has been an awesome improvement of the speech quality which is generated with text to speech. Numerous individuals surmise that generated speech which is synthetic as it is additionally known as sounds like robots from more established motion pictures. The reality is however that a few voices verging on recorded speech which is like sound. Before the development of this system, the text to speech system was not developed for words, numbers, and integers to convert into Punjabi language. In the system or software developed before, to develop this system firstly we have to record the sound like .wave files in human voice then these files were used to convert the number into speech.

The research is based on following objectives:

1. Detailed understanding of NLP and text to speech conversions.
2. Study of various techniques designed and implemented by researchers in this field.
3. Understanding the computation systems implemented in Punjabi language.
4. Design a Text-to-Speech tool which is able to teach a Layman Punjabi language by converting Punjabi alphabets in their speech wave form.

8. CONCLUSION & FUTURE SCOPE

An initial step towards this end, a study of the artificial technologies in which there is study of tasks of natural language processing has done. These tasks include speech recognition and speech synthesis. The

speech recognition is used for changing speech into text. It uses an automated speech recognition system. Then speech synthesis can convert the text to speech. It can convert with the help of some different tools such as Dialog system. The survey of various techniques of text to speech synthesis system has been presented in this paper. It is found that many systems for Punjabi language to speech conversion are yet to be formulated. So, it is proposed to develop a system which is able to guide laymen about Punjabi language. It will pronounce the Punjabi letter (dictation set of thing). It has been found that this kind of system is not developed it.

REFERENCES

- [1] C. Science, "Text to Speech Synthesis for Numerals into Punjabi language," no. July, 2013.
- [2] Alexandre Trilla, "Natural Language Processing techniques in Text-To-Speech synthesis and Automatic Speech Recognition", Department de Technologies Media, Spain, January 2009.
- [3] Er. Sheilly Padda, Ms. Rupinderdeep Kaur, Er. Nidhi, "Architecture and Implementation of Punjabi Text to Speech System Using Transcriptions Concept", International Journal of Engineering Research and Development, Vol. 1, Issue 5, June 2012.
- [4] J. O. Onaolapo, F. E. Idachaba, J. Badejo, T. Odu, O. I. Adu, "A Simplified Overview of Text-To-Speech Synthesis", Proceedings of the World Congress on Engineering, Vol. 1, July 2014.
- [5] Simmi Luthra, Parminder Singh, "Punjabi Speech Generation System based on Phonemes", International Journal of Computer Applications, Vol. 49, No. 13, July 2012.
- [6] Divya Bansal, Ankita Goel, Khushneet Jindal, "Punjabi Speech Synthesis System using HTK", International Journal of Information Sciences and Techniques (IJIST), Vol. 2, No. 4, July 2012.
- [7] Archana Balyan, S. S. Agarwal, Amita Dev, "Speech Synthesis: A Review", International Journal of Engineering Research & Technology, Vol. 2, Issue 6, June 2013.
- [8] Shruti Gupta, Parteek Kumar, "Comparative study of text to speech system for Indian language", International Journal of Advances in Computing and Information Technology, pp. 199-209, ISSN 2277-9140, April 2012.
- [9] Jagmeet Kaur, Parminder Singh, "Punjabi Speech Synthesis System for Android Mobile Phones", International Journal of Engineering and Computer Science, Vol. 3, Issue 9, September 2014.
- [10] Hardeep, Parminder Singh, "Android Based Punjabi TTS System", International Journal of Computer Science Trends and Technology, Vol. 3, Issue 3, May-June 2015.
- [11] Parminder Singh, Gurpreet Singh Lehal, "Punjabi Text to Speech Synthesis System", Proceedings of Coling 2012, pp. 409-416, December 2012.
- [12] Research Journal of Engineering and Technology (IRJET), Vol. 3, Issue 4, April 2016.
- [13] Jaswinder Singh, Tirath Singh, Sarabjit Singh, Brahmjit Singh, "A Technique for Speech Synthesis of Gurmukhi Embedded in LATEX document", International Journal of Science and Research (IJSR), Vol. 5, Issue 2, February 2016.