



ISSN 2278 – 0211 (Online)

Speech Synthesis Using Android

Shailesh S. Sangle

Assistant Professor, Department of Information Technology
MCT's Rajiv Gandhi Institute of Technology, Mumbai, India

Nilesh M. Patil

Assistant Professor, Department of Information Technology
MCT's Rajiv Gandhi Institute of Technology, Mumbai, India

Abstract:

Speech Synthesis is one of the most leading application areas in natural language processing (NLP). This is also known as “Text-To-Speech” (TTS) and is mainly the capability of the device to “speak” text of different languages. This application acts as an interface between two different representations of information, namely text and speech, to perform effective communication between two parties. Our main objective is to make an application of speech synthesis for Android based mobile phones. We have developed an application on the Android environment and the voice conversion libraries provided by Android environment are used. The application developed is user friendly and reliable and effective communication is performed.

Keywords: NLP, TTS, Android, OS, SLR

1. Introduction

Speech Synthesis is one of the major applications of NLP. We have developed an application using the Android operating system. Android is the open source OS developed by Google and is widely used within several types of embedded and mobile platforms, including mobile phones and tablets. Our work basically consists of three different aspects. First aspect is to convert English text to English speech. Second aspect is conversion of regional language text to regional voice. The third and most important aspect is the integration of the presented system on android environment. The android environment is the most common and the popular platform used in mobile devices so that the application can be attached to a mobile phone or the system so that the effective communication will be performed.

2. Text to Speech Conversion

Our system consists of preprocessor, text analyzer, morphological analyzer, contextual analyzer, syntactic prosodic parser, letter to sound module and prosody generator. A preprocessor check for the correct syntax of the sentences and splits them into list of individual words. Text analyzer identifies numbers, abbreviations, and idioms and transforms them into full text as and when required. A morphological analyzer performs task to propose all possible part of speech categories for each word taken individually, on the basis of their spelling. Inflected, derived and compound words are decomposed into their elementary graphemes units by simple regular grammars exploiting lexicons of stems and affixes. The contextual analyzer module considers words in their context, which allows it to reduce the list of their possible part of speech of neighboring words. Finally a syntactic parser examines the remaining search space and finds the text structure which more closely relates to its expected prosodic realization.

In this application we used an algorithmic approach to perform the TTS conversion. Speech synthesis is the artificial production of human speech. It converts normal language text into speech. A TTS engine converts written text to a phonemic representation and then converts the phonemic representation to waveforms that can be output as sound. A TTS engine is composed of front end and back end. At the earlier stage the preprocessing is done on input text. Front end is responsible for preprocessing by converting raw text (containing symbols like numbers and abbreviations) into equivalent of written out words. This process is also called normalization or tokenization. After the input text is split to the individual words, classification of the word is done. The front end assigns phonetic transcriptions to each word, divides and marks the text into prosodic units, likes phrases, clauses and sentences known as text to phoneme conversion. Once the phonetic equivalent is obtained, the next work is to connect it with the lookup library to identify the voice representation of that specific word. Phonetic transcriptions and prosody information together make up the sign language

recognition (SLR) that is output by the front end. At the final stage, the library connected to produce the person specific voice. Back end converts the SLR into sound.

3. Related Work

Er. Sheilly Padda and Er. Nidhi have discussed the text to speech conversion for Punjabi (Gurmukhi) language^[1]. The paper also discusses various issues which were found when converting text to speech. Eyob B. Kaise proposed algorithms and methods that address critical issues in developing a general Amharic text to speech synthesizer^[2]. Aidan Kehoe proposes a number of guidelines to assist in the creation and testing of help material that may be presented to users via speech synthesis engines^[3]. Erik Blankenship describes handicapped accessible text to speech markup software developed for poetry and performance^[4].

4. Proposed Work

The following steps were performed to develop the application.

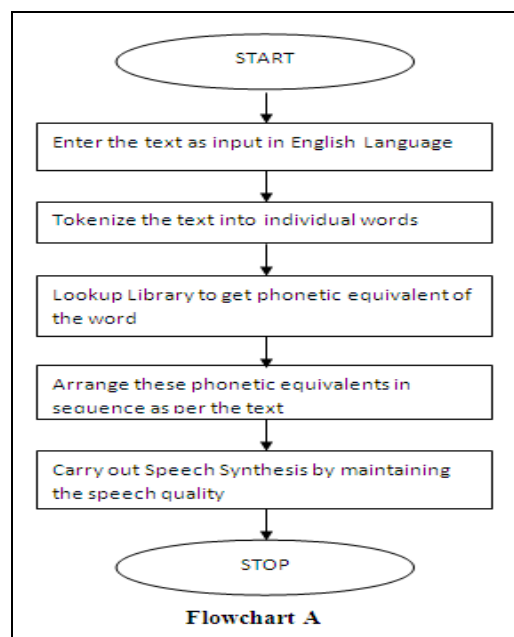
- To get the natural quality in synthetic speech we adopted concatenative speech synthesis techniques.
- For speech synthesis, phonemes of the English language were used as the basic ingredients.
- Using these phonemes, speech database for English language was developed.
- The input text was then separated into English phonemes.
- Phonemes were searched in the database and corresponding phoneme sounds were concatenated to generate synthesized output speech.

We developed this application to provide an efficient language translator in mobile phones which will provide hand-held device users with the advantage of instantaneous and non-mediated translation from one human language to another. Two way communications is possible between the users with minimum time lag. The communication is performed from English text to English voice and the Hindi text to English form and then to Hindi speech. However, a person can understand a sentence only if it is pronounced correctly. But still there are gaps in pronouncing in mobile computing. So this application has come up with a better and user understandable pronunciation mechanism. Current speech recognition API's are only capable of recognizing a single word. This application will enhance the speech recognition to recognize sentences. Next one is the homophone detection. A homophone is a word that is pronounced the same but differs in meaning (example, to too two). The speech recognition engine will be able to detect those words according to the sentence.

5. Algorithm

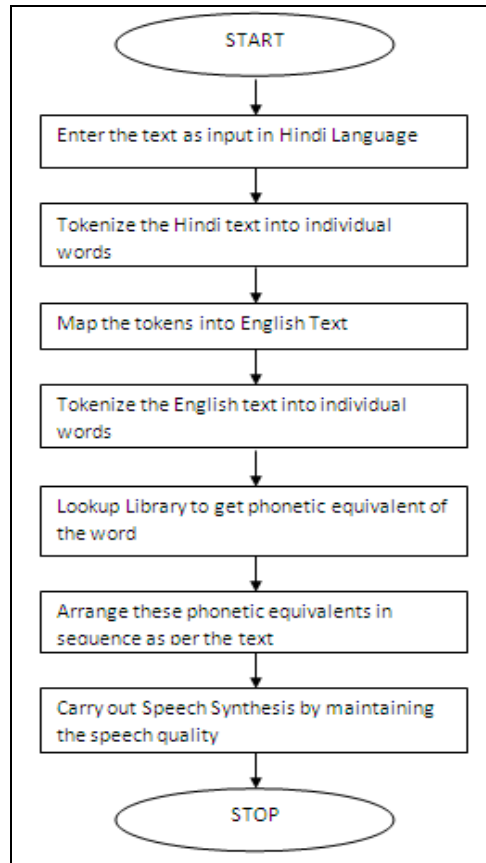
5.1. English Text to English Speech

The application of speech synthesis is developed in Android 4.04. The following procedure was carried to convert the English text to English speech as shown in the flowchart A. First we took the text in English language as the input. By means of lexical analyzer, we split that text into individual words. Then we searched in the library for an equivalent phonetics of those individual words. After that as per the text in English, we arrange this phonetics. Then the corresponding phoneme sounds were concatenated to generate synthesized output speech.



5.2. Hindi Text to Hindi Speech

The application of speech synthesis is developed in Android 4.04. The following procedure was carried to convert the Hindi text to Hindi speech as shown in the flowchart B. First we took the text in Hindi language as the input. By means of lexical analyzer, we split that text into individual words. Then we map these tokens into English language. By means of lexical analyzer again, we split that text into individual words. Then we searched in the library for an equivalent phonetics of those individual words. After that as per the text in English, we arrange this phonetics. Then the corresponding phoneme sounds were concatenated to generate synthesized output speech.



6. Results

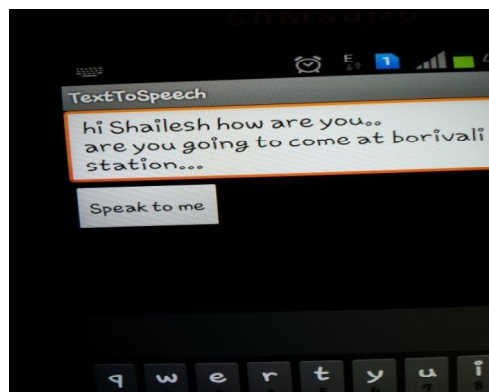


Figure (a)

Figure (a) above shows the textbox in Android mobile where the text in English language is given as the input.

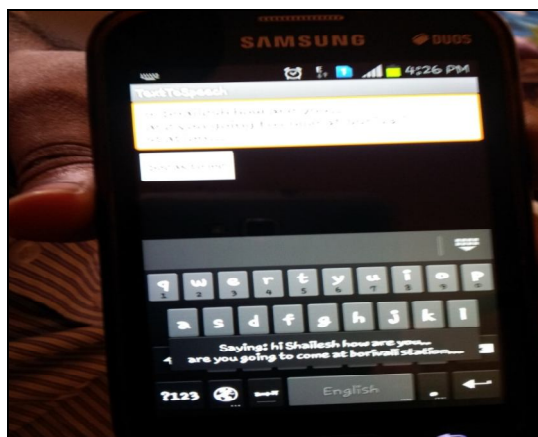


Figure (b)

Figure (b) above shows screen in Android mobile when speak to me button is clicked and the audio output is given in English for text given in the text box.

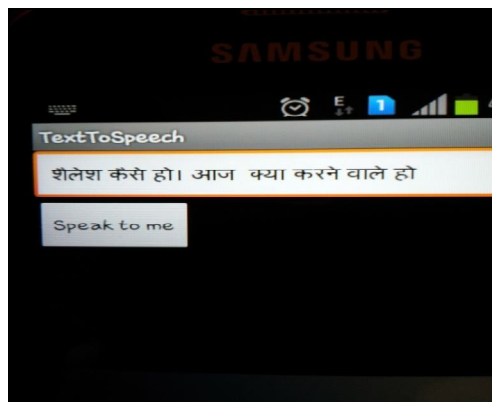


Figure (c)

Figure (c) above shows the textbox in Android mobile where the text in Hindi language is given as the input.

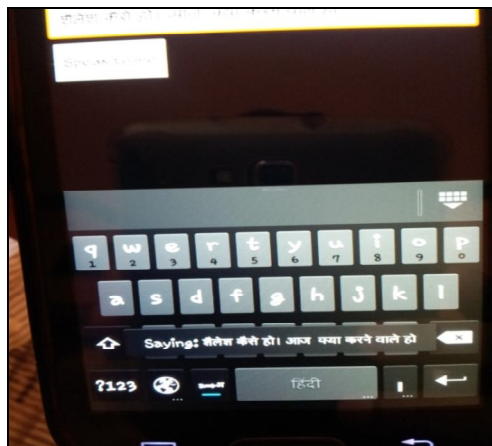


Figure (d)

Figure (d) above shows screen in Android mobile when speak to me button is clicked and the audio output is given in Hindi for text given in the text box.

7. Conclusion and Future Scope

We have developed an application of speech synthesis on the Android environment. The application developed is user friendly and reliable and effective communication is performed. This system can be a solution to the problems of various individuals in their busy life and especially for the people with low vision or reading disabilities as it would help them to listen to their emails while relaxing,

listen eBooks, study for exams by listening to notes. The proposed work has been done for the English and Hindi language. This work can also be done for the other regional languages such as Tamil, Gujarati, etc. We can also integrate a person voice with the system.

8. References

1. Er. Sheilly Padda, Er. Nidhi; "A Step towards Making an Effective Text to speech Conversion System", International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622, Vol. 2, Issue 2, Mar-Apr 2012, pp.1242-1244
2. Eyob B. Kaise; "Concatenative Speech Synthesis for Amharic using Unit Selection Method", MEDES'12, October 25-31, 2012, Addisababa, Ethiopia.
3. Aidan Kehoe, "Designing Help Topics for Use with Text to Speech", SIGDIC'06, October 18- 20, 2006, Myrtle Beach, South Carolina, USA.
4. Erik Blankinship, "Tools for Expressive Text to Speech Markup", UIST'01 Orlando FLO, November 11-14, 2001