ijecs open access www.ijecs.in

International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 3 Issue 9 September 2014 Page No. 8020-8025

Punjabi Speech Synthesis System for Android Mobile Phones

Jagmeet Kaur[1], Parminder Singh[2]

[1]Department of Computer Science and Engineering, SGI, Farrukh Nagar, Haryana, India
jagmeet2704@gmail.com
[2]Department of Computer Science and Engineering, G.N.D.E.C, Ludhiana, Punjab, India
parminder2u@gmail.com

Abstract: *Mobile phone usage is approximately 3.5 times more than the usage of Personal Computers. Android has the biggest share among all Smartphone Operating systems like Symbian, Windows etc. because it has very few restrictions for the developers to develop an application. Text to speech synthesis is one application which reads the written text aloud. TTS systems on Android are available for many languages but not for Punjabi. Our present work is to develop a Punjabi text to speech synthesizer that can produce an output speech on a mobile device. While porting this TTS system to a resource limited device like mobile phones, some practical aspects like application size and processing time are considered. The Concatenative Speech Synthesis technique has been used which uses the Phonemes as the smallest single units for concatenation.*

Keywords: *Android Operating System, Concatenative Synthesis, Phonemes, Speech Database.*

1. INTRODUCTION

In the past decade, mobile device have made rapid progresses. A few years back an open source platform Android is being popular. An Android mobile phone also provide the Text to Speech synthesis with vocal interface for the users to allow them to read their text or email aloud on their mobile screen, hence reducing the use of visual modality. This TTS application also helps the users to read the text while driving, jogging etc. [2]. Android TTS application is helpful for users with visual disabilities and illiterate masses [1].

Many speech synthesizers are already available for various languages, but building a speech corpus for other languages than that of English speech corpus is a different task [3]. The problem of porting a TTS system to mobile devices is limited storage and processing power. Gopi et al. develop Text to speech synthesizer (TTS) for Android platform. Authors use ESNOLA (Epoch Synchronous Non Overlap Add) based concatenative speech synthesizer technique with Partnemes as the smallest unit for implementing the TTS for the Malayalam language [4]. Ahlawat and Dahiya developed English and Hindi TTS engines on an Android environment. For Hindi TTS, authors present a two layer process by first getting the input text in Hindi language and then map this whole Hindi data into English language and then generate the output speech. They use phonemes of the English language are used as the smallest units for the concatenation [5]. Mhamunkar et al. present an application of Android the speech to text and text to speech conversion for searching a word meaning through

voice. The application accepts the word in the voice then searches the word in its mobile dictionary and generates the meaning of the word as the synthetic speech output [6]. A Bengali Text to speech synthesizer for the Android operating system is developed by Mukherjee and Mandal using ESNOLA (Epoch Synchronous Non Overlap Add) based Concatenative Speech synthesis technique with Partnemes as the smallest units for the concatenation [2]. Saychum et al. use the hidden Markov model (HMM) based speech unit for a bilingual TTS System on Android operating system which converts the text into Thai language and English language separately and then plays an audio file for that [7]. The application text to speech synthesis on Android mobile phones for people who are blind or visually impaired is discussed by Shaik et al. [8]. An African text to speech engine is developed in the mobile platform for E-learning by Roux et al. The authors use HMM technique to read a portion of text aloud and hence generating the output speech [9]. A Thai speech synthesizer on a mobile device based on Flite, a unit concatenation synthesizer is implemented by Wongpatikaseree et al. [1]. Singh and Lehal developed Punjabi TTS systems which produced reasonably acceptable synthesis output on PC [10]. However, it has not yet been implemented for resource-limited devices such as mobiles. The goal of our research is to develop a Punjabi text-to-speech synthesizer that can produce the synthesized output speech in a real-time on a mobile device. The Concatenative speech synthesizer technique has been used in order to get the two qualities of the output synthesized speech: intelligibility and naturalness. Phonemes of the Punjabi language have been used as the basic unit. Punjabi speech database is developed for the Punjabi language containing the valid phonemes (combination of vowels (V) and consonant-vowel combination (CV)) and their corresponding sound filenames are stored against them. Sounds for these V and CV combinations are recorded in wave file. The input text is first segmented into Punjabi phonemes, then these phonemes are searched in the database and corresponding filename are retrieved. These files are then searched in the resources folder and played [11].

2. ANDROID ARCHITECTURE

Google on 5th November 2007 launched the mobile platform called Android for mobile devices like PDA, net books and smart phones [12]. Android is an open source OS based on Linux kernel which acts as an abstraction layer between the hardware and the software stack [13]. All applications are written in Java Programming language and Eclipse is the IDE for developing Android apps. Google created Dalvik as the virtual machine environment for mobile devices for compiling the projects, each application runs on its own VM, not on the Java VM [14]. Since embedded systems have the constraints of application size and processing time, so Dalvik Virtual Machine (DVM) is the one important feature which is optimized for low memory requirement [12]. The architecture of Android operating system is divided into four layers: the first layer is Application layer, the second layer is Application framework, the third layer is divided in two sub layers: libraries and Android Runtime, and the last layer is Linux Kernel. So in total there are five layers [14]. Main features of Android operating system are:

- Free use and adaptation of operating system to manufactures of mobile devices.
- Optimized use of memory with DVM.
- High quality of audio visual content.
- Quick and easy development of applications using development tools and rich database of software libraries [15].

3. METHODOLOGY

The methodology followed for the project is as follows:

a) Concatenative synthesis technique is used to get the naturalness quality of the synthetic speech which involves taking real recorded speech; cutting it into segments and concatenating these sound segments back together during synthesis to produce the desired output speech.

b) The user enters the Punjabi text into the textbox which is divided into words. This phase first analyzes the positions of vowels and consonants in a word and then segments the words into phonemes as a vowel and consonant vowel combination.

c) The database of this Punjabi TTS consists of two fields: word and filename. The word field contains the phoneme and the filename field contains the sound file. The database preparation consists of the selection of sentences containing all the vowels, consonants and their combinations, recording of these sentences and finally marking of the phoneme sounds in these recorded sound files. SQLite DBMS is used to store the Punjabi phonemes and their corresponding sound files.

d) Now the phoneme is searched in the database and retrieves the filename, which is further searched in the application resources for its corresponding sound wave file to get played. At the end, these phoneme sounds are concatenated to generate the sound corresponding to the input text.

4. DEVELOPING PUNJABI TTS SYSTEM

4.1 DATABASE PREPARATION

Following are the steps which are followed for the preparation of Gurmukhi Punjabi speech database:

4.1.1 PUNJABI PHONEME SELECTION

For the development of this TTS system, Phonemes are selected as the basic unit of concatenation. The reason for selecting phonemes as basic speech units is that, any word can be made using phonemes while keeping the database relatively smaller than any other method [10][11]. In Punjabi language there are two types of phonemes V and CV gave rise to 380 phonemes with non-nasalized vowels and 380 phonemes with nasalized vowels, resulting total 722 valid Punjabi phonemes, where V and C represent vowel and consonant respectively as shown in TABLE 1.

4.1.2 TEXT FOR RECORDING

For the analysis of phonemes, a carefully selection of unbiased Punjabi Corpus was made, having nearly four million total words. For labeling the phoneme sounds, we have selected the words having all the consonants, vowels and their valid combinations for recording.

TABLE 1. VALID AND INVALID PHONEMES

Phoneme Type	**V (Non-Nasalized)**	**V (Nasalized)**	**CV (Non-Nasalized)**	**CV (Nasalized)**	**Total**
No. of Phonemes	10	10	380	380	760
Invalid Phonemes	0	0	7	51	58
Valid Phonemes	10	10	373	329	722
Examples	ਆ	ਆਂ	ਜੀ [ਜ(C) +ਈ(V)]	ਮਾਂ [ਮ(C) + ਆਂ(V)]	

4.1.3 RECORDING WORDS

The selected words have been recorded by a native female speaker of Punjabi. The speech quality depends upon the quality of the recorded sound and hence, sound quality of extracted speech units from this recorded sound. So, a professional female speaker of Punjabi is selected for recording. The recording has been done in the studio with the following characteristics:

Sampling Rate: 44100 Hz, Bit Depth: 16 bit, Channels: Mono.

4.1.4 LABELING OF THE PHONEMES

The next phase is to label the Phoneme sounds in the recorded sound file. This phase of labeling the phoneme sounds in the database is very crucial and time consuming task and needs to be done very carefully, because the naturalness of the synthetic speech produced by the TTS system depends upon how exactly the phoneme boundaries have been marked. For this purpose the sound editing software- *Sonic Foundry Sound Forge 10.0* has been used. The phoneme sounds have been labeled manually one by one, after carefully listening and analyzing the word sounds. The phoneme boundaries have been marked and noted down [11].

4.1.5 CREATION OF PUNJABI SPEECH DATABASE

The Punjabi speech database is an important part, which it is optimized for the high quality TTS system. In order to obtain the naturalness of our application, we have used the concatenative technique for combining the sound files of the phonemes. As this TTS system is to be developed for portable devices on Android operating system, so a very light weight SQLite database is used. The database of our TTS system has two fields: word and filename. The first field contains phoneme itself, second field contains the sound filename. In total 578 phonemes and their sound files have been recorded and the size of our database is 13.1 MB.

4.2 TEXT NORMALIZATION

The Punjabi text normalization module consists of processing the input text before passing the text for the TTS conversion. The input text is a raw text containing the abbreviations, special symbols and numeric values are first searched and then replaced with their expanded form, so that they are spoken in full word form. The entered text is then segmented into words [1].

4.3 WORD SEGMENTATION INTO PHONEMES

Since Phoneme is the basic unit of concatenation, so it is necessary to first segment the Punjabi input text into words and then the words are segmented into phonemes which are stored in the database. For example, the word "ਜਗਮੀਤ" will be segmented into five phonemes: ਜ (ਜ ਅ) + ਗ (ਗ ਅ) + ਮੀ (ਮ ਈ) + ਤ (ਤ ਅ).

4.4 SEARCHING PHONEMES AND CONCATENATION

This phase is responsible for searching the corresponding phoneme in the database and retrieving the sound filename. Then the corresponding sound file is searched in the application resources, loaded into the memory. If the search is successful for that particular phoneme, the corresponding sound file is returned from database. If there is no entry for that particular phoneme in the database, then this will be skipped as invalid phoneme. At the end, these phoneme sounds are concatenated to generate the sound corresponding to the input text.

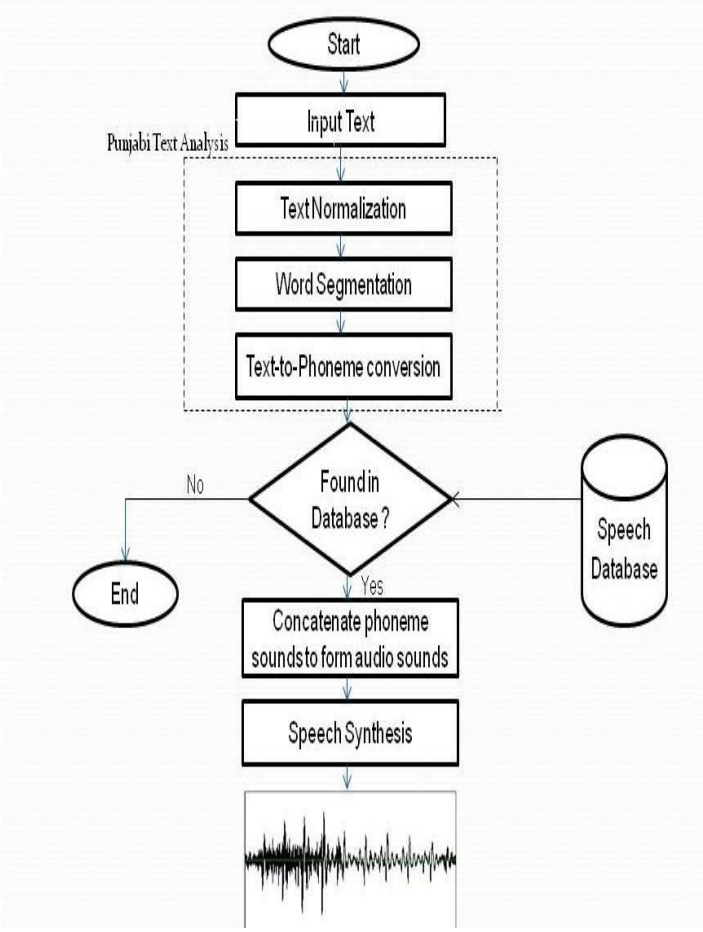
4.5 PORTING TTS TO ANDROID PLATFORM

The final phase of our TTS application is to port the TTS system from the desktop to the Android platform. 'Gurumukhi' font has been introduced in the application to render the typed Punjabi text. With the help of Gurumukhi font, for every key pressed from mobile its corresponding Punjabi letter is displayed in the input text field. This text is then converted into the Unicode values. These Unicode values are the inputs for the TTS engine.

5. IMPLEMENTING PUNJABI TTS SYSTEM ON ANDROID

The minimum system specification for implementing Punjabi TTS on Android is Android OS version 2.2 with 512MB RAM. Android application should be efficient because they will run on mobile device with limited computing power, storage and constrained battery life. The size of our application is 10.3 MB. On Android device keeping the database connections opened and occupying the memory all the time is very expensive. So database connections are closed as soon as the sound filename is retrieved and the memory is released as soon as the sound file is finished playing. The TTS system has the functionality that the user can generate Punjabi speech sound by typing the Punjabi

word in an English alphabet format. The input text types in an English alphabet can be written in the textbox. The 'Play Sound' button when clicked generates the speech file corresponding to the text and plays the audio file generated. The application can be distributed to the end user, and can be made available on the Android devices by connecting to a PC through the USB port and Bluetooth. Fig. 1 illustrates the Gurmukhi Speech Synthesis system and Fig. 2 illustrates the Interface for our TTS application.



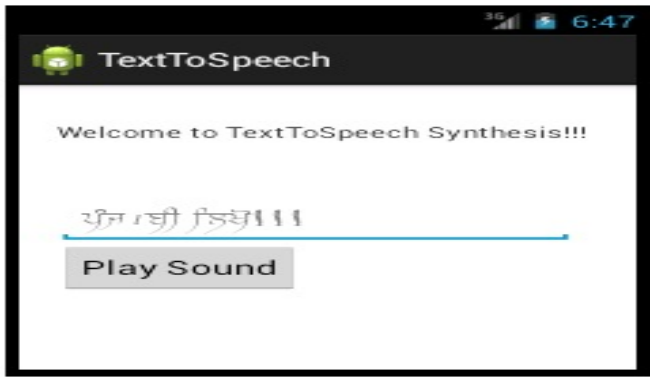


Figure 1. Flow chart of Punjabi TTS

Figure 2. Interface of TTS application

6. CONCLUSION

In this paper, we describe our implementation of a Punjabi speech synthesizer on an Android OS mobile device. Our aim is to develop a TTS (Text-to-Speech) application that can produce an output speech in almost real time on the Android based smart phones. Our development of TTS system is based on principles of concatenation using phonemes as the speech unit. The size of the speech database is 13.1 MB with 578 phonemes sounds extracted from the recorded sound file and the size of this Punjabi TTS application is 10.3 MB. The developed system shows good results for segmentation of words into phonemes and can segment the text of any length to the phonemes. This application mainly caters the need of visually weak people who cannot read. This application will help them in knowing what is written. It also enables the ill-literate people, who don't know how to read, in getting the information which is in written form. It also helps the book-lovers who like to read while travelling. Sometimes they cannot read while travelling due to some reasons. They can simply plug in the ear phones and can listen the text. It also helps the people in learning Punjabi language. Although these are some advantages of this presented work, but still there are gaps in pronunciation mechanism which can be improved by improving the quality of the sound file using DSP techniques.

REFERENCES

[1] K. Wongpatikaseree, A. Ratikan, A. Thangthai, A. Chotimongkol and C. Nattee, *A Real-time Thai Speech Synthesizer on a Mobile Device*, Proc. 8th IEEE Conference on Symposium on Natural Language Processing (SNLP), Bangkok, 2009, 42-47.

[2] S. Mukherjee and S.K.D. Mandal, *A Bengali Speech synthesizer on Android OS*, Proc. 1st ACM Workshop on Speech and Multimodal Interaction in Assistive Environments (SMIAE), Jeju, Republic of Korea, 2012, 43-46.

[3] S. Kiruthiga and K. Krishnamoorthy, *Annotating Speech Corpus for Prosody Modeling in Indian Language Text to Speech Systems*, International Journal of Computer Science Issues (IJCSI), 9(1), 2012, 104-108.

[4] A. Gopi, P. Shobana, T. Sajini and V.K. Bhadran, *Implementation of Malayalam Text to Speech Using Concatenative Based TTS for Android Platform*, Proc. IEEE International Conference on Control Communication and Computing (ICCC), Thiruvananthapuram, Kerala, India, 2013, 184-189.

[5] S. Ahlawat and R. Dahiya, *A Novel Approach of Text To Speech Conversion Under Android Environment*, International Journal of Computer Science & Management Studies (IJCSMS), 13(5), 2013, 189-193.

[6] P.V. Mhamunkar, K.S. Bansode and L.S. Naik, *Android Application to get Word Meaning through Voice*, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), 2(2), 2013, 572-575.

[7] S. Saychum, A. Thangthai, P. Janjoi, N. Thatphithakkul, C. Wutiwiwatchai, P. Lamsrichan and T. Kobayashi, *A Bi-lingual Thai-English TTS System on Android Mobile Devices*, Proc. 9th IEEE International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTICON), Phetchaburi, 2012, 1-4.

[8] A.S. Shaik, G. Hossain and M. Yeasin, *Design, Development and Performance Evaluation of Reconfigured Mobile Android Phone for People Who are Blind or Visually Impaired*, Proc. 28th ACM International Conference on Design of Communication, SIGDOC, S.Carlos, 2010, 159-166.

[9] J.C. Roux, P.E. Scholtz, D. Klop, C. Povlsen, B. Jongejan and A. Magnusdottir, *Incorporating Speech Synthesis in the Development of a Mobile Platform for E-learning*, Proc. 7th International Conference on Language Resources and Evaluation (LREC), Valletta, Malta, 2010, 1507-1510.

[10] P. Singh and G.S. Lehal, *Text-To-Speech Synthesis System for Punjabi Language*, International Conference on Multidisciplinary Information Science & Technologies (INSciT), Merida, Spain, 2006, 388-392.

[11] S. Luthra and P. Singh, *Punjabi Speech Generation System based on Phonemes*, International Journal of Computer Applications (IJCA), 49(13), 2012, 40-44.

[12] N. Gandhewar and R. Sheikh, *Google Android: An Emerging Software Platform for Mobile Devices*, International Journal on Computer Science and Engineering (IJCSE), NCICT Special Issue, ISSN: 0975-3397, 2010, 12-17.

[13] J. Liu and J. Yu, *Research on Development of Android Applications*, Proc. 4th IEEE International Conference on Intelligent Networks and Intelligent Systems (ICINIS), Kunming, China, 2011, 69-72.

[14] S. Primorac and M. Russo, *Android application for sending SMS messages with speech recognition interface*, Proc. 35th IEEE International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO, Opatija, Croatia, 2012, 1763-1767.

[15] B.R. Reddy and E. Mahender, *Speech to Text Conversion using Android Platform*, International Journal of Engineering Research and Applications (IJERA), 3(1), 2013, 253-258.