Speech Therapy System to Kannada Language

V.Udayashankara
Dept of IT, SJCE, Mysore, India
Email: v_udayashankara@sjce.ac.in

Swapna Havalgi
Dept of IT, SJCE Mysore.
Email: swapnah90@gmail.com

Abstract: **This paper presents an alternative communication technique to help people suffering from speech and language difficulties for various reasons. Electronic Speech synthesis is a process of generating human like speech from any text input to emulate human speaker. The objective of text to speech system is to convert an arbitrary Kannada text into its corresponding spoken waveform, using phoneme as basic unit for speech synthesis. A standard syllable level speech database consisting of 525 syllables is built for synthesizing naturally sounding speech. The main advantage of this system is the real time approach for conversion of entered text to corresponding speech. The initial and the final points of a speech waveform are determined using Maximum energy and zero crossing rate. The Unit selection based concatenation method is opted for syllable concatenation and the system is implemented using MATLAB**

Keywords: Text processing, Kannada, Maximum Energy, Zero Crossing rate, Unit selection Concatenation, Praat, speech synthesis, Unicode.

I. INTRODUCTION

Entering Kannada text from a user interface keyboard on a text document is a tedious process and converting the written words into naturally sounding speech is another most significant part of the Text to Speech system. The entire process of the system can be split into three stages i.e. Text entry, Text processing and speech processing. The Matlab R2014 software is used to develop this system.

Speech segments are significantly affected by co articulation, so if two speech segments that were not adjacent to each other are concatenated, there we can observe spectral or prosodic discontinuity. The Spectral discontinuities mainly occur when the formants at the concatenation point do not match. In synthetic speech even if each segment is very natural but if it contains large discontinuities in-between them, then the listener rates it as poor.

II. OVERVIEW OF KANNADA LANGUAGE

Kannada is the official and administrative language of the Karnataka state in India with 50 million native speakers. It is one of the scheduled languages of India. In contrast very small percentage of people use English as means of communication and due to the fact of low literacy rates in India, the conventional interfacing systems have become challenging. There is a huge need for technological development in linguistic languages such as Kannada.

In Kannada language there are 13 vowels and 34 consonants. It is phonetic in nature with one to one correspondence with written form and the spoken language. The phonemes are divided into Swaras (vowels) and vyanjanas (consonants) and these two togetherly constitute Varnamala (alphabet set).

Swaras are the independently existing letters also called as Vowels [2]. Sounds of vowels cannot be modified

ಅ ಆ ಇ ಈ ಉ ಊ ಋ ಎ ಏ ಐ ಒ ಓ ಔ

Vyanjanas dependent on Swaras to take independent form and also called as consonants. Consonants sounds can be modified by combining consonants with vowels.

ಕ ಖ ಗ ಘ ಙ, ಚ ಛ ಜ ಝ ಞ, ಟ ಠ ಡ ಢ ಣ, ತ ಥ ದ ಧ ನ, ಪ ಫ ಬ ಭ ಮ, ಯ ರ ಲ ವ ಶ ಷ ಸ ಹ ಳ

Vyanjanas can be classified as Vargeeya and Avargeeya [2]. Vargeeya vyanjanas are such as

ಯ್ ರ್ ಲ್ ವ್ ಶ್ ಷ್ ಸ್ ಹ್ ಳ್

And Avargeeya vyanjanas are such as

ಕ್ ಖ್ ಗ್ ಘ್ ಙ್, ಚ್ ಛ್ ಜ್ ಝ್ ಞ್, ಟ್ ಠ್ ಡ್ ಢ್ ಣ್, ತ್ ಥ್ ದ್ ಧ್ ನ್, ಪ್ ಫ್ ಬ್ ಭ್ ಮ್

By combining the above consonants with existing vowels Kannada alphabets called kagunitha are formed such as

ಕ್ + ಅ = ಕ; ಕ್ + ಆ = ಕಾ; ಕ್ + ಇ = ಕಿ; ಕ್ + ಈ = ಕೀ; ಕ್ + ಉ = ಕು; ಕ್ + ಊ = ಕೂ; ಕ್ + ಋ = ಕೃ; ಕ್ + ಎ = ಕೆ; ಕ್ + ಏ = ಕೇ; ಕ್ +ಐ = ಕೈ; ಕ್ + ಒ = ಕೊ; ಕ್ + ಓ = ಕೋ; ಕ್ + ಔ = ಕೌ;

The above basic rule is followed for the rest of all other consonants. The combination of vowels phoneme and consonants phoneme forms a syllable (Akshara).

III. BUILDING KANNADA SPEECH DATABASE

During the process of speech synthesis the particular speech file has to be fetched, concatenated and processed by using a suitable algorithm [2],[3] and given as a sound output. A speech database is a collection of sound files placed within a single folder. So creating an error free database is of at most importance. For this purpose utility software compatible for windows operating system called as the Praat [2], is used for recording of each phoneme.

978-1-5090-1025-7/16/$31.00 ©2016 IEEE

The above task is performed to create a database consisting of 525 syllables. Each syllable is recorded at sampling frequency of 44 KHz represented with 16 bit.

The following gives the procedural steps for recording of speech waveform to build the database using Praat software:
Step 1: Open the Praat utility software and select the record mono sound option from New in menu bar.
Step 2: Select 44100 Hz sampling frequency radio button and then press record option to start recording.
Step 3: Utter the required phoneme/syllable
Step 4: Once recording is done, press the stop button and play to listen the recorded sound. If it is correctly recorded select the option save to list.
Step 5: Use view and edit option to see the particular waveform and carefully select the required portion of the waveform containing the phoneme and save with .wav file format.
From the above procedural steps the waveforms of 'ಕ' 'ಮ' 'ಲ' are shown in Figure 1,2,3 respectively.

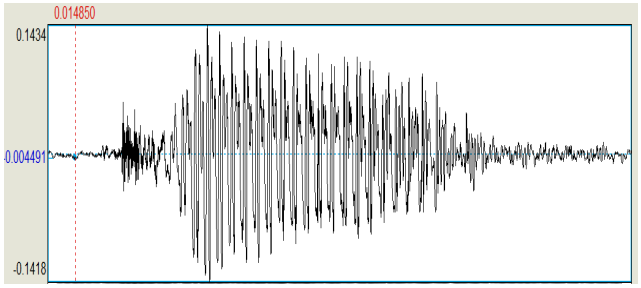

Figure 1.wave form of consonant 'ಕ'

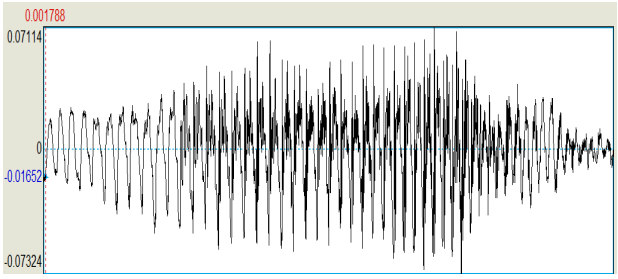

Figure 2.Wave form of consonant 'ಮ'

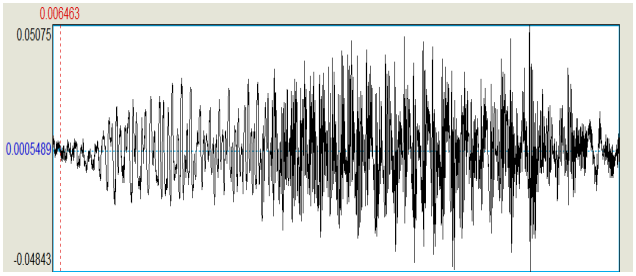

Figure 3.Wave form generated for consonant 'ಲ'

III. METHODOLOGY

Text to speech system (TTS) is to convert orthographic text into acoustic signal. The entire process involved in Text to Speech conversion system is divided into three stages such as Text Entry, Text processing and Speech processing [3]. This section briefs about building the TTS system as shown in Figure 4.

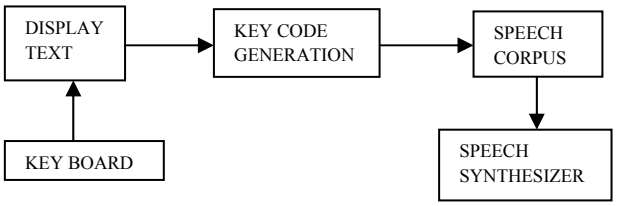

Figure 4. Block Diagram of Text to Speech Synthesis

A. Text Entry

It is the foremost step in building TTS system [3]. Input is provided by manual entry from keyboard. The advantage of using keyboard to enter the text is to make user to feel comfortable and easy to type and to do so some of the settings are to be made with the personal computers or the laptops using windows as operating system, so that it can allow entering the Unicode from a general keyboard.

The settings required are as follows:
In Control panel >> Go to Language and Region >> Select Keyboards and Language option->click on Change Keyboards->click on ADD button and in the drop down box select the language as Kannada (India)->click on keyboard within that check option Kannada.

After following the above instructions open the blank text document and in the right corner of the tool bar a symbol as EN can be seen. Click on it and select option KD Kannada (India). Now the normal general QWERTY keyboard works as Kannada keyboard.

B. Text Processing

It is an essential stage of TTS system. The main purpose of this stage is to generate the Unicode values from the entered text which is further required for mapping onto speech database for speech synthesis.

Table1. Independent Vowel

Alphabets	Unicode	Decimal Equivalent
ಅ	0C85	3205
ಆ	0C86	3206
ಇ	0C87	3207
ಈ	0C88	3208

ಉ	0C89	3209
ಊ	0C8A	3210
ಋ	0C8B	3211
ಎ	0C8E	3214
ಏ	0C8F	3215
ಐ	0C90	3216
ಒ	0C92	3218
ಓ	0C93	3219
ಔ	0C94	3220

Table2. Consonants

Alphabets	Unicode	Decimal Equivalent
ಕ	0C95	3221
ಖ	0C96	3222
ಗ	0C97	3223
ಘ	0C98	3224
ಙ	0C99	3225
ಚ	0C9A	3226
ಛ	0C9B	3227
ಜ	0C9C	3228
ಝ	0C9D	3229
ಞ	0C9E	3230
ಟ	0C9F	3231
ಠ	0CA0	3232
ಡ	0CA1	3233
ಢ	0CA2	3234
ಣ	0CA3	3235
ತ	0CA4	3236
ಥ	0CA5	3237
ದ	0CA6	3238
ಧ	0CA7	3239
ನ	0CA8	3240
ಪ	0CAA	3242
ಫ	0CAB	3243
ಬ	0CAC	3244
ಭ	0CAD	3245
ಮ	0CAE	3246
ಯ	0CAF	3247
ರ	0CB0	3248
ಲ	0CB2	3250
ವ	0CB5	3253
ಶ	0CB6	3254
ಷ	0CB7	3255
ಸ	0CB8	3256
ಹ	0CB9	3257
ಳ	0CB3	3251

The program reads the entered character one by one and generates corresponding key code i.e. Unicode value of each character. This program is implemented in MATLAB software and the algorithm is described below:

1. From the keyboard the characters are entered on a text document.
2. The text document is first opened by giving the path of the text file to a variable and stored in the form of bytes.
3. These bytes are read and converted into unsigned 8 bit integers.
4. The value obtained with respect to each entered phoneme will be in the decimal equivalent which is converted to a Hex value that gives the Unicode value as shown in the Table 1. For example, if the entered character is **ಅ** it is directly mapped on to a Table 1 containing vowels and its decimal equivalent is 3205. Then it is converted to Hex value and stores it as 0C85.
5. Similarly if the entered character is from a consonant group say **ಕ,** it is mapped to Table 2 and its decimal value is 3221.Then it is converted to Hex value as 0C95 and it is stored.
6. After each Unicode value generated, a set of delimiters will occur automatically before the next corresponding Unicode value is stored. For example if the entered text is **ಉದಯ** the corresponding Unicode values along with delimiters are E0 B2 **89** E0 B2 **A6** E0 B2 **AF**.
7. And even if there is a space in between two characters or if a dependent vowel sign character is to be read, a particular delimiter pattern containing a set of three values will automatically occur before the next character is read. For example, if the entered text is with spaces in between say **ಉ ದ ಯ** E0 B2 **89** 20 E0 B2 **A6** 20 E0 B2 **AF** and for characters with dependent vowel signs say **ಅರುಣ** E0 B2 **85** E0 B2 (**B0** E0 B3 **81)** E0 B2**A3.**

C. *Speech Processing*

This stage is the progression of the text processing phase and involves mapping of each Unicode value with the saved .wav file and further proceedings contain silence removal and concatenation of those audio file.
Mapping involves extracting of a particular audio file with the same name of the Unicode value that is generated

during the text processing. The picked audio file is passed through a silence removal algorithm with which we could able to find the starting and ending point of the desired audio waveform by suppressing all other waveforms generated by noise and pauses created while uttering and recording of each phoneme. The process involves calculation of short time energy as in (1) and zero crossing rates as in (2)
Short Time Energy

$$E_n = x^2(n) * h(n) = \sum_{m=-\infty}^{\infty} x^2(m)h(n-m) \quad (1)$$

Where E_n indicates energy and $x(n)$ is the signal and $h(n)$ is the gain of the system

Zero Crossing Rate: It is the rate at which the signal changes from positive to negative or back is given by the equation (2)

$$Z_n = \sum_{m=-\infty}^{\infty} |sgn[x(m)] - sgn[x(m-1)]| w(n-m) \quad (2)$$

Where Z_n indicates zero crossing rate and $w(n)$ indicates window function

The following algorithm describes the extraction of initial and final point of the desired waveform

1. Check the first crossing of ITU and then for crossing of ITL
2. ITU crossing indicates the beginning and end points of the speech signal
3. Then on Zero cross, move backward and forward from N1 and N2 respectively.
4. If the threshold IZCT is crossed more than three times, move backward to first time crossing which gives the boundary.
5. Similarly if threshold is crossed more than three times, move forward to first crossing which gives the boundary.

The algorithm removes the unwanted waveforms and makes it ready for concatenation process. For concatenation and speech synthesis, unit selection based concatenation technique is used to get the speech output [3]

IV. RESULTS

This paper is worked towards the people suffering from utterance problems and for those who face difficulty in pronouncing the phonemes properly.
The results of this work are evaluated by conducting intelligibility test for 50 subjects considering 25 male and 25 female at All India Institute of Speech and Hearing. In this test, 100 Kannada words are chosen randomly and subjects are allowed to intelligibility test. Some of the words which are used for intelligibility test are tabulated in Table 3.

Table 3 Randomly Chosen Words

ಕಮಲ
ಅನುಪಮ
ಉದಯ
ಕಿಟಕಿ
ಚಮಚ
ನವಿಲು
ಭಜನೆ
ಡಮರುಗ
ಹವಳ
ಔಷದಿ

The subjects are able to listen and understand all the hundred words. Hence the system accuracy is 100 percent for the above words.

V. CONCLUSION

Text to Speech System is developed for speech therapy is productively implemented in real time using MATLAB programming language. The prosody mismatch is minimized by effectively selecting the starting and end points of the speech waveform using maximum energy and zero crossing rates from the syllable level standard database. The quality of the synthesized speech is evaluated by performing the intelligibility test. As per the results, cent percent accuracy is obtained. The proposed work with little modification can be extended for most of the other Indian languages.

REFERENCES

[1] D.J.Ravi and Sudarshan Patilkulkarni "Text-To-Speech Synthesis System for Kannada Language," International Journal of Advanced Research in Computer Science, Vol. 2, Jan-Feb 2011, pp 298-304.

[2] D.J.Ravi and Sudarshan Patil Kulkarni "A novel approach to build Speech Database for Kannada TTS system," International Journal on Recent Trends in Engineering & Technology, Vol. 05, Mar 2011

[3] Shreekanth.T, Udayashankara.V, Arun Kumar.C "An Unit Selection based Hindi Text To Speech Synthesis System Using Syllable as a Basic Unit," IOSR Journal of VLSI and Signal Processing , Volume 4, Jul-Aug 2014 pp 49-57