Shodhika Bhashini
Multilingual Search and TTS System

Dr.K.V.N.Sunitha
Professor& HOD, CSE Dept.,
G.Narayanamma Institute of Technology & Science
Shaikpet, Hyderabad

M.Charan Teja
IV/IV CSE ,I semester
SreeNidhi Institute of Technology & Science
Ghatkesar, Hyderabad

Abstract— This paper presents a multilingual search engine and text to speech system (TTS) developed to help hands-busy and visually impaired/aged people. Wide spread use of computers and Internet for the last two decades has made any information available in fraction of seconds. People of all ages prefer web searching than books. This resulted in development of very powerful search engines. But they are not that user friendly when it comes to searching using regional language and producing the required information in a convenient way to disabled/aged/hands-busy people, particularly for Indian languages. The work presented here is an effort to overcome these limitations of search engine. Shodhika Bhashini not only searches the information in a user friendly way but also speaks out the information. This multilingual TTS and search engine is developed for communicating in Telugu, English & Hindi language. Same can be extended for other south Indian languages.

Keywords- Search engine, TTS, Syllable, Transilaration, segmentation

I. INTRODUCTION *(HEADING 1)*

Though there are many search engines and TTS systems available independently[1,3]**,** Multi Lingual TTS with search engines are not available in market. To use search engine in languages like Telugu or Hindi , no virtual keyboard is available . The proposed system Shodhika Bhashini is a solution that overcomes all the problems of existing TTS and search engines. This multilingual TTS is for communicating in Telugu, English & Hindi language. The input is given by virtual keypad which is transcribed and segmented into syllables. The corresponding speech units of syllables from speech corpus are concatenated into one signal[2,5,6,9]. The text input is not restricted to particular words or sentences, it can grow applicably. Most of the work is done in English language[8], and effort is made in other Indian languages. As Telugu, Hindi being the popular languages spoken in different parts of the country, the government of India has taken interest in developing the work in these languages.

II. SYSTEM ARCHITECTURE

Modules of the proposed system are: Text Normalization, Text Syllabification, Syllable Extraction and Concatenation.

The system is developed to take text input and produce its equivalent speech output [1,4,7] or searches the text in Web. The input size is not restricted to one word or sentence. The context level diagram of the system is shown in Figure 1.

The design of the system mainly concentrates on accepting the data from the user then transcribing it into an orthographic form which is understood by the system. The input text is then made into syllables to extract the corresponding speech units from the speech corpora. The system is designed into different modules like text normalization, text syllabification, syllable extraction and concatenation.

III. SYSTEM IMPLEMENTATION

Before you begin to format your paper, first write and save the content as a separate text file. Keep your text and graphic files separate until after the text has been formatted and styled. Do not use hard tabs, and limit use of hard returns to only one return at the end of a paragraph. Do not add any kind of pagination anywhere in the paper. Do not number text heads-the template will do that for you.

Finally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling and grammar:

A. User Interface

Shodhika Bhashini is web based system with a user interface to accept the input and play the concatenated speech files and also to search for a text in Telugu, English and Hindi. The interface is developed using ASP.net with Visual Studio 2008 Frame Work. A virtual keyboard for Telugu font and another for entering Hindi text is designed to enter the text input. The given text is converted to WX notation and displayed in the text box. The interface also provides a facility to read from files and to search from World Wide Web.

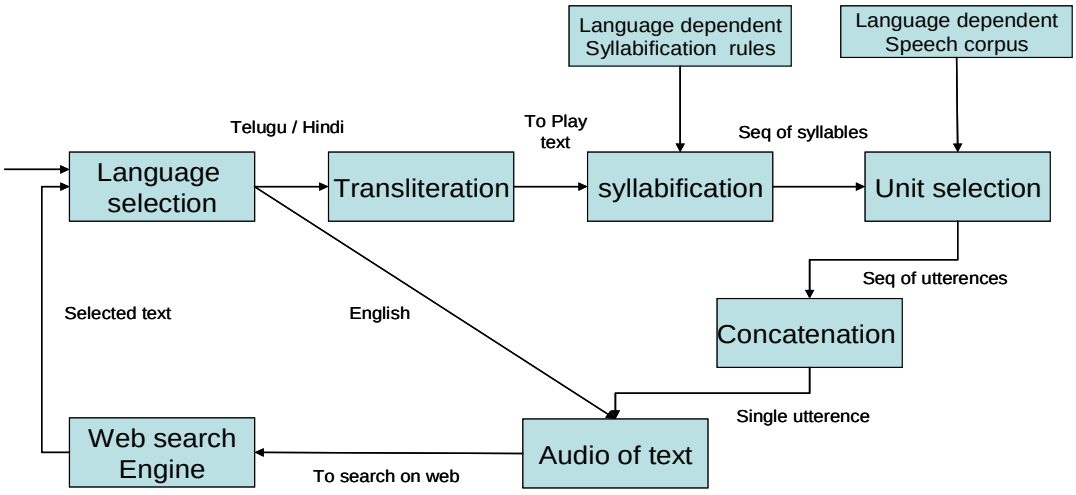

Figure 1. Context Level Diagram of the system

The virtual keyboard consists of two panels, one for vowels and another for consonants. The vowel panel consists of primary vowels and the vowel modifiers and the second panel consists of consonants. The character set for consonants in Telugu and Hindi are more complex and peculiar in their function. Their character signs have often three or more than three distinct shapes depending on whether they are used as standalone characters called Primaries or Base consonants, or when used with a vowel other than the inherent vowel /a/ functioning as a hanger or a pure consonant, or when used as a constituent of a conjunct called as secondary or dependent consonant characters called as *ottulu.* Stress button is used for consonant modifier. The interface consist of three buttons one to speak the input text that is given, one to clear the contents of the input text box and one button to stop speech output, one button for Transliteration and one button for syllabification.

B. Darabase Creation

The speech corpora used for the Shodhika Bhashini search and TTS system is created using the speech utterance of three telugu stories and two Hindi articles. All the sentences are recorded with a sampling frequency of 22500Hz in a noise free environment. The syllables are identified and segmented and stored in the database with an identifier to retrieve it

The syllable speech units are stored using the context dependent structure where the information regarding to the syllable to its left, right and the position of the syllable in the word.

```
Struct syllable {
                char * Syllable_next;
                char * Syllable_prev;
                int    Syllable_pos;
                int    Syllable_id;
        }
```

Segmentation of Speech Segmenting the speech signal according to the phonetic transcription is a fundamental task in any voice activated system . One of the major reasons for considering syllable as a basic unit for TTS systems is its better representational and durational stability relative to the phoneme.

It is demonstrated that segmentation at syllable-like units followed by isolated style recognition of continuous speech performs well..Using PRAAT tool speech signal is segmented as shown in figure 2.

C. Phase 1 : Transliteration

The raw text is the input that is given by the user in Telugu language, which is transcribed into an orthographic form using one of the standard forms which the system can understand. To transcribe the raw text WX notation is used where each Telugu alphabet is represented using this notation.

This is language specific procedure. Here the procedure is explained for telugu and the same procedure is used for Hindi language. The user is allowed to enter Telugu text or browse a Telugu file. The word or contents of the file are converted into character array. The conversion is done character by character. The character position is represented by i, whose value initially is 0.

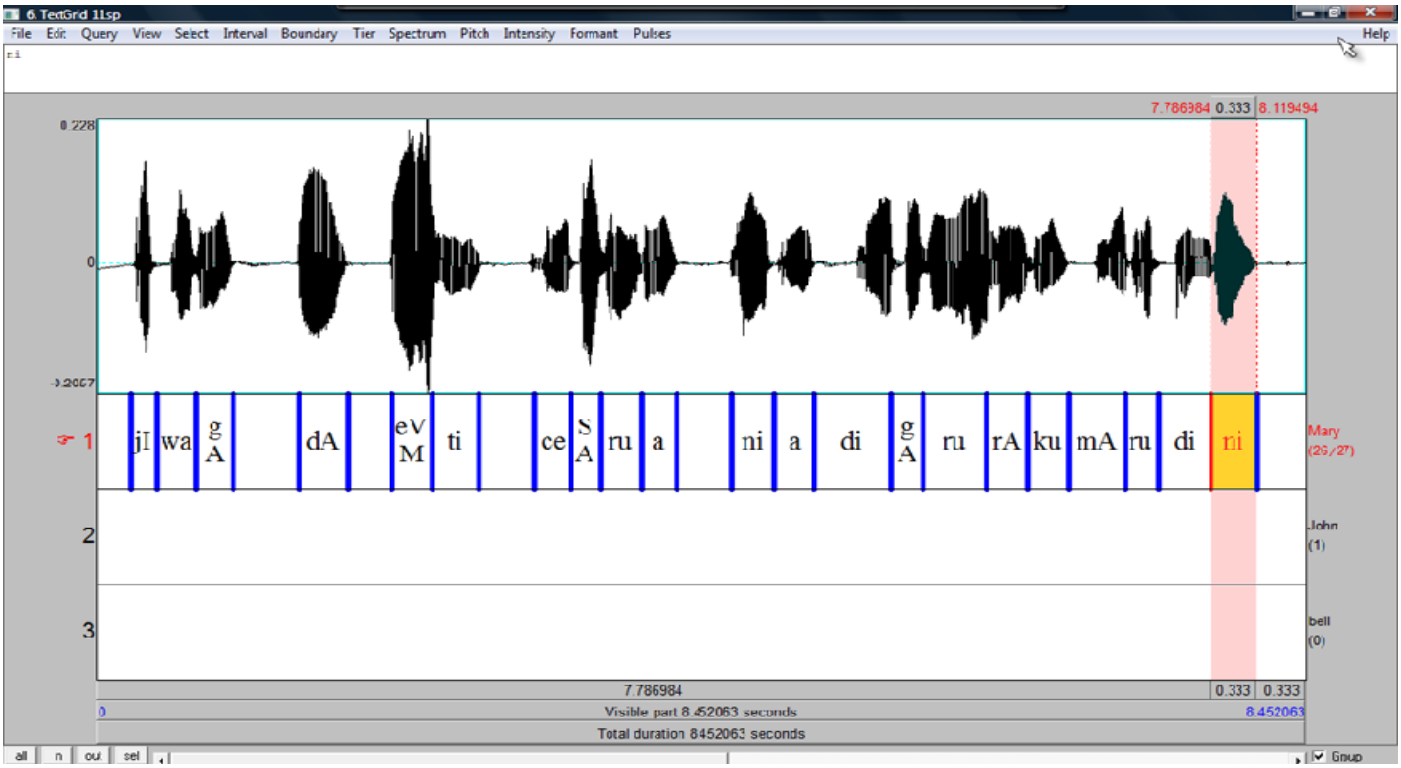

Figure 2.: Segmentation of recorded speech

Conversion procedure, which is iterative till all the characters are converted, is as follows:

- If the Unicode of the character is between 0C15 and 0C39 (క to హ), English representation corresponding to the Unicode is retrieved and added to English text.
- If the Unicode of the character is between 0C3E and 0C4C (ా to ౌ), the last letter from the English text is removed. English representation corresponding to the Unicode is retrieved and added to English text.
- If Unicode of the character is 0C4D (stress mark ్), the last letter from the English text is removed.
- Else, the character encountered is copied into English text.
- i is incremented 1

D. *Phase 2 :Text Syllabification*

Text syllabification function each of the normalized words and break them into syllables and arrange it according to the sequence of the syllables based on Telugu phonological rules. Syllable structure is represented as C*VC* in most of Indian languages like Telugu, Tamil, Hindi. Etc. The syllables in Telugu language can exist as vowel alone or as CV, VC, CVC. CCVC. The syllabification function reads the normalized text input and first identifies the consonants and vowels, using phoneme list which consists of all the phonemes of Telugu language. The second part of syllabification is to identify the syllables and the following algorithm is designed to perform consonant – vowel identification and then identifies the syllables. The consonant – vowel identification part reads the input text and places the result into file1, this is taken as input and the syllables are identified and placed into the file2.

Algorithm for Syllabification

- The Read the input text which is in WX notation.
- Label the characters of the normalized text as consonants and vowels using the following rules
  - Any consonant except(y, H, M) followed by y is a single consonant, label it as C
  - Any consonant except (y, r, l, lY, lYY) followed by r is taken as single consonant
  - Consonants like (k, c, t, w, p, g, j, d, x, b, m, R, S, s) followed by l is taken as single consonant.
  - Consonant like (k, c, t, w, p, g, j, d, x, b, R, S, s, r) followed by v is taken as a single consonant.
  - Label the remaining as Vowel (V) or Consonant(C) depending on the set to which it belongs.
  - Store the attribute of the word in terms of (C*VC*)* in file1.

- For each word in the normalized text get its label attribute from file1.
    - If the first character is a C then the associate it to the nearest Vowel on the right.
    - If the last character is a C then associate it to the nearest Vowel on the left Check.
    - If sequences correspond to VV then break is as V-V.
    - Else If sequence correspond to VCV then break is as V-CV.
    - Else If sequence correspond to VCCV then break is as VC-CV.
    - Else If sequence correspond to VCCCV then break is as VC-CCV.
    - The strings separated by – are identified as syllable units.
- repeat
    - Store the text in syllable form in file2 for synthesis process.

Example 1: Telugu origin word: ruXrAksha (WX representation)

Consonant-Vowel identification:CVCVCCCV (<Xr> is taken as 1 Consonant)

Syllabification: CV- CVC-CV-CV-CV

Example 2: English loan word: yApil 'Apple'

Consonant-Vowel identification: CVCVC

Syllabification:CV-CVC

E. *Phase 3 : Syllable Extraction and Concatenation*

This module will receive a sequence of syllables that has been properly arranged according to the raw text. Based on the list of syllable, Syllable Extraction module will search for speech units in the speech corpus. The context of the syllable to be searched in speech corpus is given weights according to next, previous and the position of the syllable in the word. For context corresponding to next is given a weightage of 4, previous is given 2 and the matching of position in the word is given 1. This phase reads the input from file2 and performs the search for the syllables and performs the concatenation according to input syllables and plays the speech files as a single sound file.

- Read the syllabified text.
- Compute the weight for each Syllable using the following table

Table 1: weights for syllables

Next	Previous	Position	Weight
Y	Y	Y	7
Y	Y	N	6
Y	N	Y	5
Y	N	N	4
N	Y	Y	3
N	Y	N	2
N	N	Y	1

    - If the search syllable matches with next, previous and position context in speech corpus, it is given weight of 7 and so on.
- The syllable which gives maximum weight is selected for synthesis.
- If the next context of the syllable is space or end mark of sentence then include the silence unit depending on the unit.
- Concatenate all the units selected and generate a single wave file which has the utterance of the given text.
- Play the wave file.

Examples showing how the synthesizer will select the speech syllables.

Synthesizer:

na-mas-kA-ra-mu (for the word namaskAramu)

ru-XrAk-sha (for the word ruXrAKsha)

yA-pil (for the word yApil)

F. *Phase 4 : Searching in Web*

This module gets the keyword in Unicode format, transcribes the text and retrieves the related web page dynamically.

IV. RESULTS

The execution of Shodhika Bhashini – Multi Lingual Search and TTS System is explained in detail with the help of the interface designed for the system. The interface is designed with minimum complexity for the user in understanding the interface and giving the text input for generation of speech. For

the generation of speech, the database is created with 5000 syllables. The utterance of these stories and the article are collected in a noise free environment. These utterances are segmented into words and again segmented at micro level of syllable and are labeled automatically. Testing is done for generating speech for the text where the text is framed with collection of known vocabulary and unknown vocabulary. The following figures3,4, shows the working process of the system.

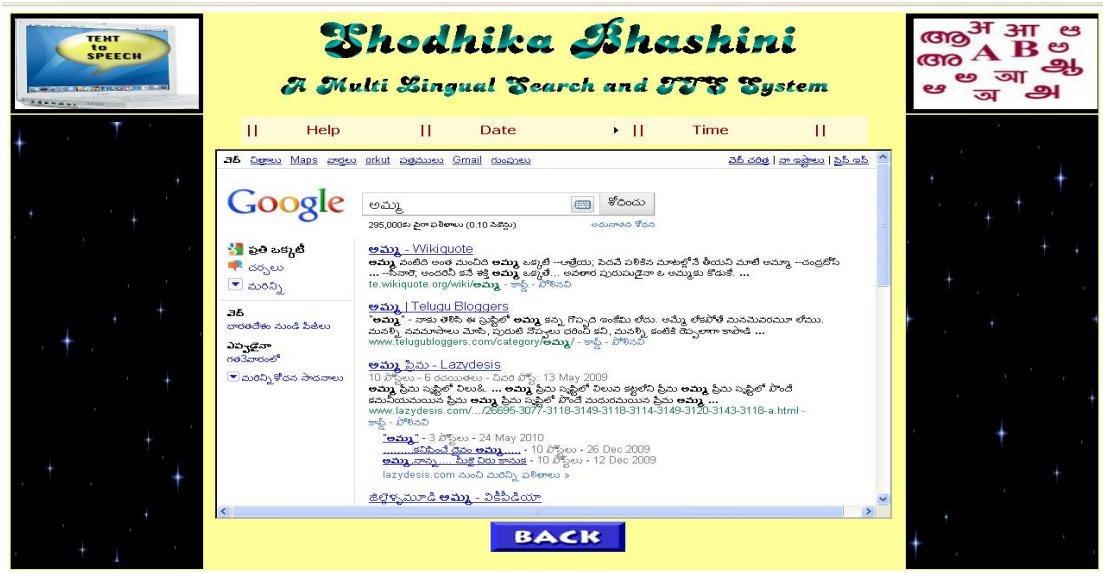

Figure 3: Shows the output of the Telugu word search.

Figure 4: Loading Telugu text from file

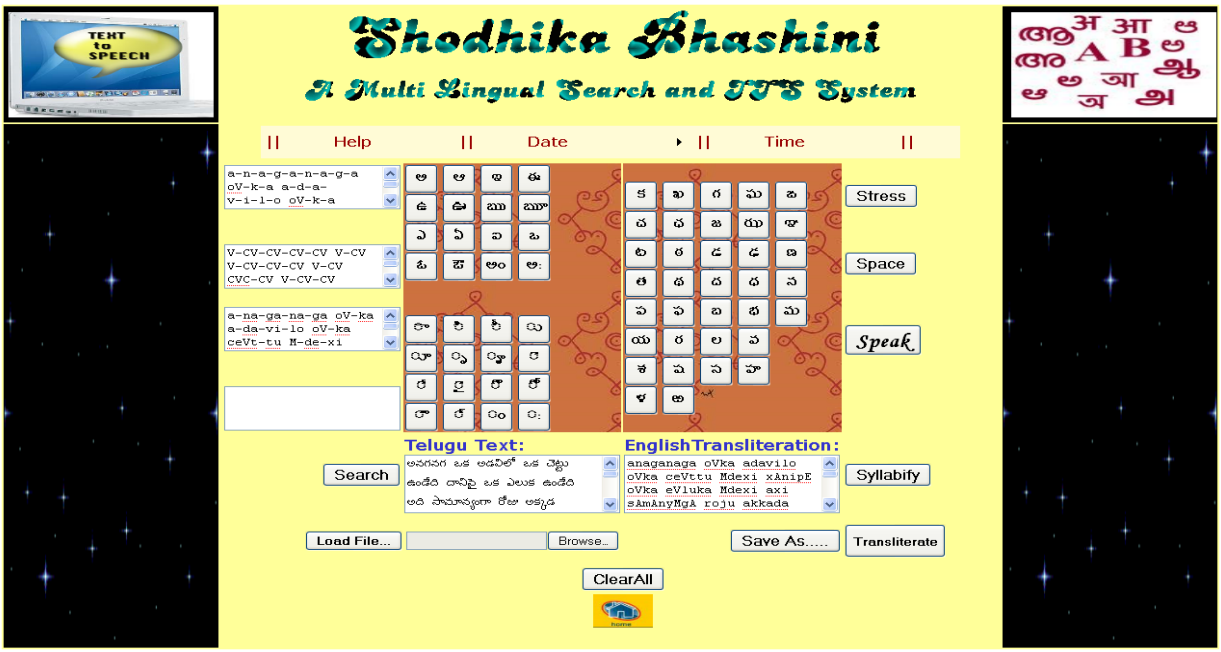

Figure 5: Telugu text loaded from file, its transliteration, and syllabification.

Figure 6: Displays the web pages for English word ‘India’.

Figure 7: Shows the output of the Hindi word search.

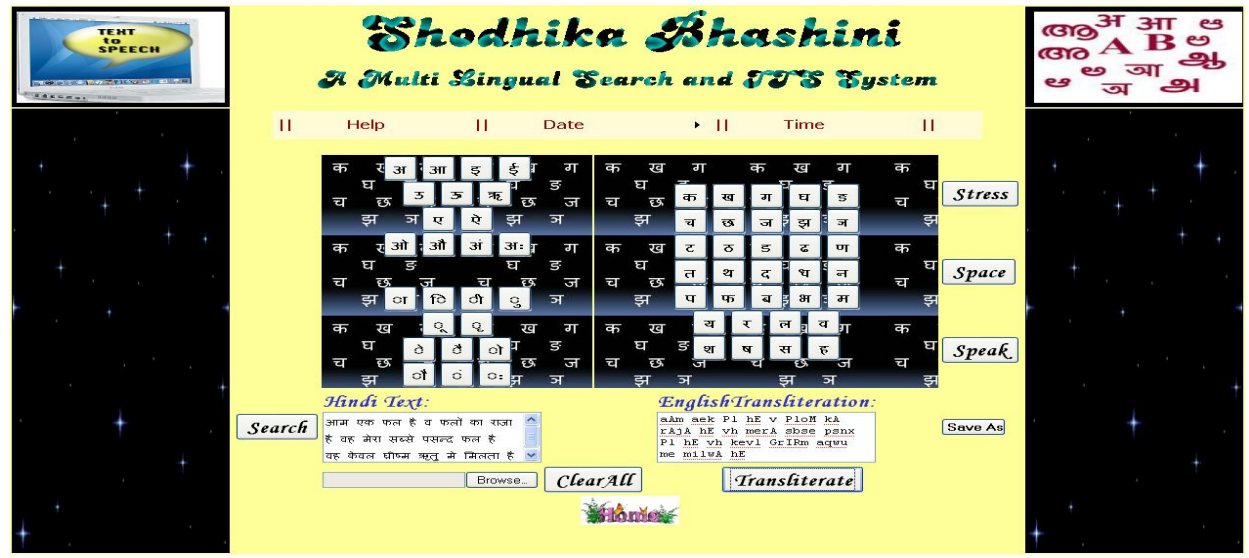

Figure 8: Hindi text loaded from file and its transliteration

V CONCLUSION AND FUTURE SCOPE

The main objective of any TTS system is to generate speech that is close to the natural voice, and is achieved in this system "Shodhika Bhashini – Multilingual Search and TTS system". This system generates speech output for the given input text using context dependent concatenative synthesis with syllable as a basic unit. The system provides two modes of entering text. Virtual keyboard is provided with Telugu, Hindi phoneme set which enables the user to enter the text in their native language. File interaction mode is provided where the user can directly open a file and convert to speech. This system can synthesize for input that is not included in the vocabulary using closest possible match. Speech generated for such vocabulary is close to naturalness. This also searches the web for the given text in respective languages.

The naturalness of the system can be improved by adding speech samples for all possible syllables with varied context into database. There is need to develop algorithms that can concatenate two speech units so that the coupling effect is not observed by human ear. Corpus-based Text-to-Speech (TTS) enables us to dramatically improve the naturalness of

synthetic speech over that of rule-based TTS. However, so far no general-purpose TTS has been developed that can consistently synthesize sufficiently natural speech. Furthermore, the flexibility of TTS is still inadequate. This system can be easily extended to any Indian language which is syllabic in nature.

Text normalization also can be added to this to improve the coverage like abbreviations A.D, BC etc. This system can be deployed in mobile phones so that it will be convenient for use as it becomes PC free. Here only 2500 syllables are trained. The corpus size also can be extended further to cover all syllables.

REFERENCES

[1] E. J. Yannakoudakis and P. J. Hutton, Speech Synthesis and Recognition Systems, Ellis Horwood Limited, 1987.

[2] Murtaza Bulu, Shrikanth S. Narayanan, and Ann K. Syrdal, "Expressive speech synthesis using a concatenative synthesizer," in Proceedings of International Conference on Spoken Language Processing, sept 2002.

[3] M. H. O'Malley, "Text-to-Speech conversion Technology," IEEE Computer Journal, vol. 23, no. 8, pp. 17–23, Aug 1990.

[4] X. Huang, A.Acero and H.-W. Hon, "Spoken Language Processing A Guide to Theory, Algorithm and System Development", New Jersey: Prentice Hall, 2001.

[5] Digital Processing of Speech Signals by L.R. Rabiner and R.W. Schafer.

[6] L.R. Rabiner and B.H. Juang. Fundamental of Speech Recognition Pearson Education

[7] Speech and Language Processing by Daniel Jurafsky & James H.Martin.

[8] S. Lemmetty, "A Review of Speech Synthesis Technology", Master Thesis, Department of Electrical and Communication Engineering, Helsinki University of Technology, Helsinki, Finland, March 1999.

[9] A.W. Black and N. Campbell, "Optimising Selection of Units from Speech Databases for Concatenative Synthesis", Proceeding Eurospeech '95, Madrid, Spain, September 1995, pp. 581 – 584.

[10] A.M. Zeki and N. Azizah, "A Speech Synthesizer for Malay Language", NationalConference on Research and Development in Computer Science, Selangor, Malaysia, October 2001.

Authors Profile:

M.Charan Teja is studying IV/IV CSE at Srinidhi Institute of Science & Technology,Hyderabad.He has carried out this work under the guidance of Dr K.V.N.Sunitha.

Dr K.V.N.Sunitha did her B.Tech **ECE** from Nagarjuna University, M.Tech **CS** from REC Warangal and Ph.D in from JNTU in 2006. She has 19 years of Teaching Experience, worked at various engineering colleges - G. Pulla Reddy Engineering College Kurnool, KSRM College of Engineering Cuddapah, Vasavi College of Engg., Hyderabad & DVR College of Engineering, and Hyderabad. She has been working as HOD CSE Dept in G.Narayanamma Institute of Technology and Science from the **inception of the CSE Dept since 2001**. She is a recipient of Academic Excellence award by GNITS in 2004. She has received "Best computer engineering Teacher award "by ISTE (Indian society for Technical education) in Feb 2008.Her biography is included in Marquis who is who in the world ,28th edition 2011.She has published more than 45 papers in International & National Journals and conferences. She is guiding 15 PhDs. She is a reviewer for International Journala"JATIT",. She is fellow of Institute of engineers, Sr member for International association of Computer Science and information technology and life member of many technical associations like CSI, IEEE, and ACM. She has authored two books, "Programming in UNIX and Compiler design"- BS Publications & "Formal Languages & Automta Theory" -Tata Mc Graw Hill.. Now she is working in the areas Natural language processing, Speech recognition, Network Security. She is advisory board member for ACE Engg College. She is BOS member for G.Pulla Reddy Engg College., Kurnool.