Designing, development and implementation of Text to Speech algorithm for Gujarati text using concatenative methodology

Prof. JJ Kothari[1], Dr. CK Kumbharana[2]

[1]Associate Professor, Shri M. P. Shah College of Commerce, Jamnagar.
[2]Associate Professor & Head, Department of Compute Science, Saurashtra University, Rajkot

Abstract- Speech is the most natural and major form of communication between humans. Since last three decades humans are trying to create computer that can understand and talk like human. Compared to English and other European languages like French, Spanish etc., much research is not done in Indian languages. There is less work done particularly in Gujarati. This paper describes Designing, development and implementation of a concatenate based Text to Speech algorithm for Speech Synthesis in Gujarati language. When the researcher tries to develop certain recognition system, they require certain previously stored data i.e. database for respective recognition system. Concatenative Synthesis described in this paper uses database of prerecorded Gujarati phonemes and concatenates them to produce sound. It also describes concatenative method which boosts up data matching and speech generation process.

Index Terms- Text to speech, Concatenative Synthesis, Gujarati consonants and vowels, TTS IDE.

I. Introduction

Text-to-speech (TTS) systems are software that convert natural language text into synthesized speech [1] . Text-To-Speech (TTS) synthesis system has a wide range of applications in every day life like public speaking, Listening aid, Screen reader etc.. In order to make the computer systems more interactive and helpful to the users, especially physically and visibly impaired and illiterate masses, the TTS synthesis systems are in great demand for the Indian languages. In this paper concatenate approach is used to generate TTS engine. It uses phoneme corpus database [2] containing all basic symbols, numbers and 'Barakhadi' characters made using combination of Gujarati Consonants and vowels. The concatenative approach is based on the small pieces of recorded speech. In this approach, to prepare "speech database", the small pieces are either cut from the recordings or recorded directly and then stored. Then, as per shown in figure 1, at the synthesis phase, units selected from the speech database are concatenated and, the resulting speech signal is synthesized as output.

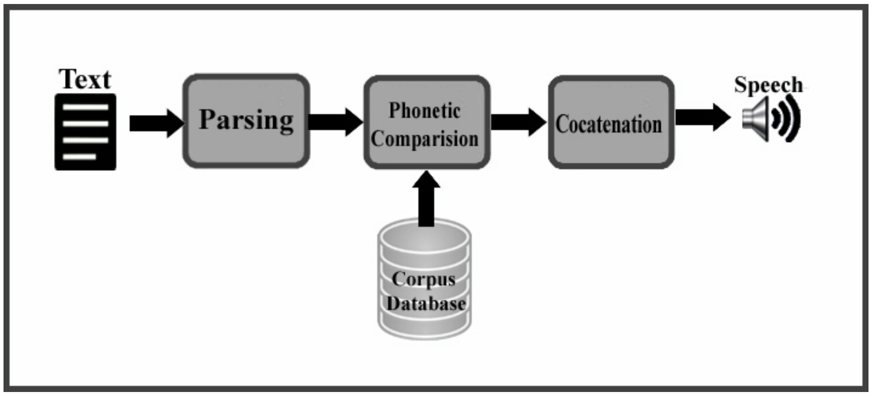


Figure 1: Concatenative TTS Synthesis System Design

II. CURRENT RESEARCH IN GUJARATI TTS SYNTHESIS

dhvani is a text to speech system designed for Indian Languages. The aim of this project is to ensure that literacy and knowledge of English are not essential for using a Computer[3]. Currently *dhvani* is capable of generating intelligible speech for the following Indian Languages . Bengali, Gujarati, Hindi, Kannada, Malayalam, Marathi, Oriya, Panjabi, Tamil, Telugu, Pashto (experimental). Dhvani works in GNU/Linux platforms as of now.

Tanvina B.Patel and their team from DAIICT have developed TTS. For their project, TTS voice in Gujarati is built using festival frame work and HTS frame work[4].

SAFA (Screen Access For All) Reader is a program developed by National Association For The Blind, New Delhi [5]. It is a screen reader. SAFA can detect the text language on the fly and calls the relevant TTS for speaking it. The latest version of SAFA is supporting following languages: Hindi, English, Sanskrit, Tamil, Marathi, Bengali, Nepali, Gujarati, Kannada and Telugu.

eSpeak is a compact, multi-language, open source text-to-speech synthesizer. This version is a SAPI5 compatible Windows speech engine which should work with screen readers such as Jaws, NVDA, and Window-Eyes. There is also a version of eSpeak which can be run as a command-line program.

III. GUJARATI CHARACTERS FEATURE

Gujarati (U]HZFTL) is an Indo-Aryan language spoken by the people of Gujarat. It is a derived from Old Western Rajasthani which is the ancestor of modern Gujarati and Rajasthani. Gujarati is one of the 22 official languages and 14 regional languages of India. It is officially recognized in the state of Gujarat, India [6]. It is spoken by over 46 million people all over the world.

Gujarati Script is based on abugida system rather than the alphabet system commonly used for European languages [7]. A character in Indian language scripts is close to a syllable and can be typically of the form: C*V*N, where C is a consonant, V is a vowel and N is anusvAra, visarha, jivhAmUllya etc. There is fairly good correspondence between what is written and what is spoken [8].

IV. GUJARATI CONSONANTS / VOWELS

Gujarati language is phonetic in nature. The grapheme to phoneme mapping is linear. Gujarati language has its own set of Vowels and Consonants. The Vowels and Consonants of Gujarati are listed in Table 1 and 2.

Table 1: Consonants in Gujarati Language.

S	B	U	3	R	K	H	Ü	8	9	0	-
6	T	Y	N	n	G	5	O	A	E	D	I
Z	,	J	;	X	Ø	C	/	Ù	7		

Table 2: Vowels and Diacritic with S in Gujarati Language

V	VF	>	.	p	é	V[	V{	VM	VÁ	V\	Vo
S	SF	lS	SL	S]	S}	S[	S{	SM	SÁ	S\	So

V. TTS MODULE

A TTS process can be dividing into two modules: Natural Language Processing (NLP) and Digital Signal Processing (DSP). The NLP module takes Text as an input and gives a normalized phonetic sentence. These phonetic sentences are the input for DSP module [9] which is responsible for generating the corresponding possible natural speech.

Concatenative TTS (CTTS) synthesis approach is used here for solving the text-to-speech (TTS) paradigm. Recorded speech feature segments, which may be words, phonemes, or even sub-phonemes, are used in this method. In this approach, speech is generated by concatenating the best compatible segments according to certain concatenation rules [10]. Speech generated by this approach inherently possesses natural quality. However, its quality depends on the size of the recorded database, as high-quality CTTS needs an extensive database.

Researchers have developed text to speech model which is shown in figure 2. It includes entire process containing master database creation, Phoneme recording and TTS Engine creation.

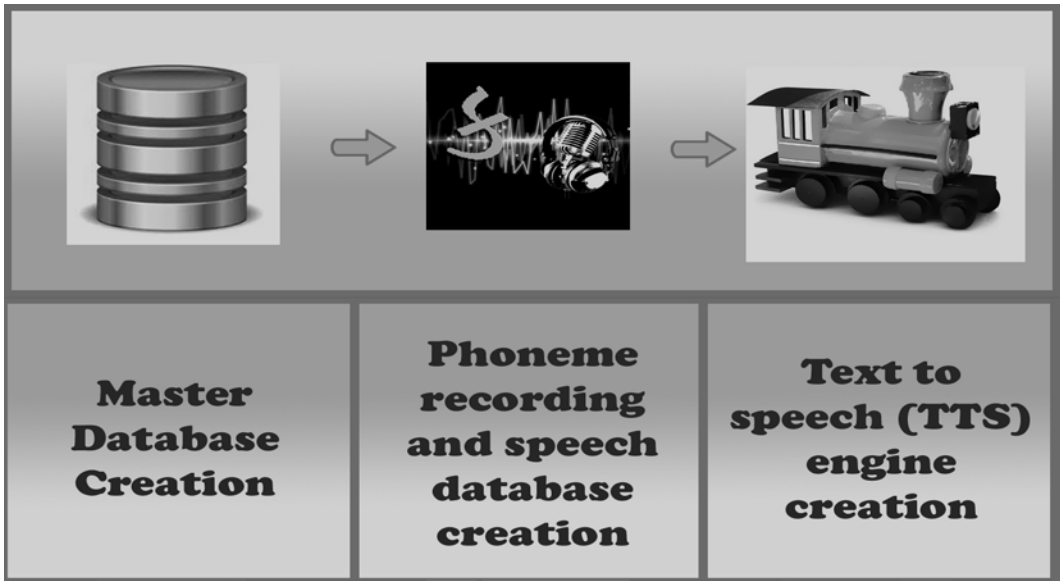

Master Database Creation
Phoneme recording and speech database creation
Text to speech (TTS) engine creation

Figure 2: Text to speech model

A. Master Database Creation

This project focuses primarily on the process of creating a voice for a concatenative Text-To-Speech system own standard output voice to sound more like the target voice. In this methodology, text entered by user is compared with master database. Master database is created using base tables shown in table 3. Each base tables (base table 1 to base table 5) contains syllabus with its ASCII code.

Table 3: The structure of Gujarati syllables and required speech records

Base table	Syllabus Structure	syllables	Required	Total
1	Half Consonant	É b u r ß ^ t y w g % a e d i < j : x Q ? 1 ` å	24	24
2	Barakhadi	S SF lS SL S] S} S[S{ SM SÁ S\ So Likewise Barakhadi for B U 3 R K H h 8 9 0 - 6 T Y N W G 5 O A E D I Z , J ; X Ø C / Ù 7	34X12	408
3	Digits	_ ! ¼ # $ %F & * ( )	10	10
4	Sp. Single Characters	k ~ ¢ ® ± ½ È Ð Ò Ó Ú Ô Þ ª æ ç è ê ë ô õ ü lü\ ý	24	24
5	Special Characters Barakhadi	V + Ç Ë Ì Ý à â ã ä	10X12	120
				586

B. Phoneme Recording

Entire system is developed in Microsoft Visual studio 10 in .NET environment. C# and SQL are used as Front End and Back End tools respectively. 586 phonemes of entire master table described in table 3 are recorded in following format by native Gujarati speaker.

- File format: Wav
- Recording Device: Microphone
- Sample rate : 44100
- Channels: 16 bit Stereo

The recorded sound files are then named and stored by using the phoneme name itself. For example the sound file of /ka/ (S) is named ka.wav. All the sound files recorded are named and stored in the similar way.

C. Text to Speech Engine Creation

TTS methodology is explained here in conceptual part presenting algorithm and Implementation part as TTS IDE (Integrated Development Environment).

Following is the Algorithm describing entire text to speech conversion process:

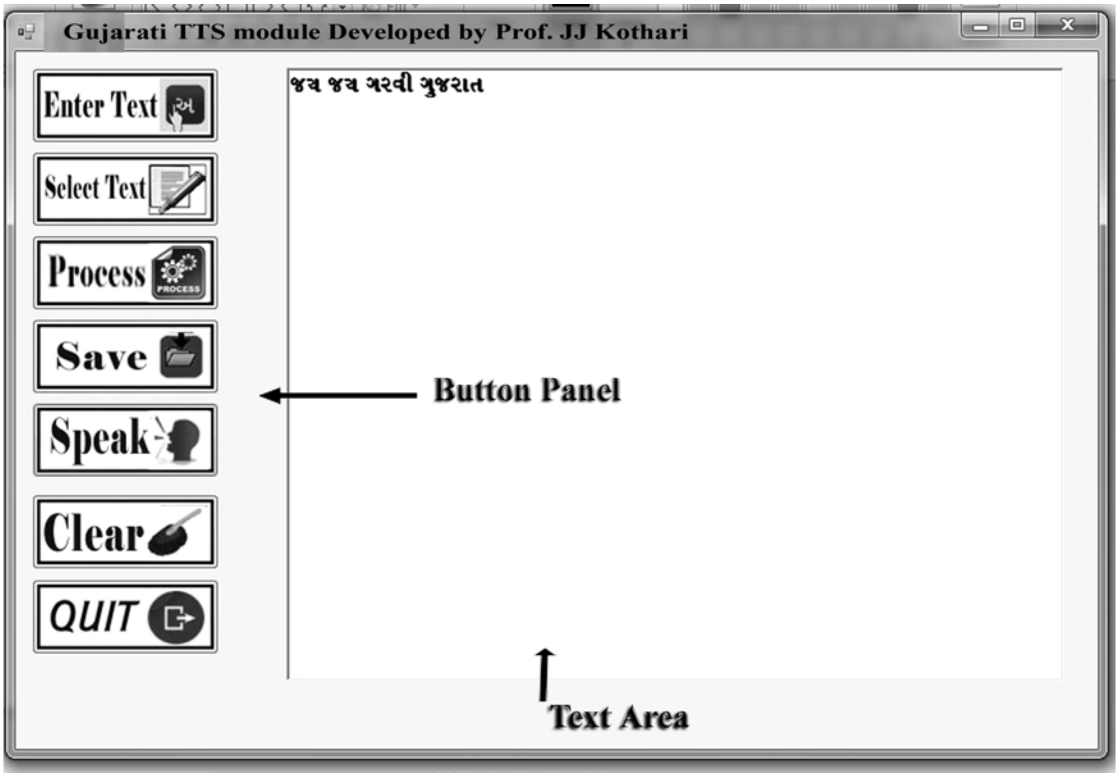
```
Step 1:  Enter text say message
Step 2:  Find length of message say N
Step 3:  I=1 as character to be extracted from message
Step 4:  j=1 as new character position during next phoneme finding process
Step 5:  k=1 as position to put next phoneme to phoneme array say myphones
Step 6: Initialize next phone as blank string
Step 7: Extract I th character from message and add it to phone
Step 8: Find match of phone from phoneme database
Step 9:  If found  then add 1 to I,
Step 10: If I <= N then go to step 8
Step 11: Remove last character from phone.
Step 12: Add phone as k th element of myphoneme array.
Step 13: make phone to blank string
Step 14: If I > N then go to step 17
Step 15: Go to step 8
Step 16: Initialize I as 1
Step 17: Take I th element of myphones array.
Step 18: Copy corresponding .wav file to SpeakIt .wav file
Step 19: Repeat step 20  to 22  till last element in myphoneme array. Then move to step 23
Step 20: Add 1 to i
Step 21: Take I th element of myphones array.
Step 22: Concatenate it at end of SpeakIt .wav file.
Step 23: Play speakIt .wav file
Step 24: End of process.
```

A TTS IDE is developed for entering text and retrieves it in sound form. It is shown in figure 3 which contains different components as per described below.

IDE contains two parts:

[a] Text area: this is the area where user enters Gujarati text which is to be converted into speech. '*mansi.ttf*' font is used for text entry. This font is developed by Dr. CK Kumbharana for his research work regarding speech synthesis [11].

[b] Button panel: This area contains different buttons helping text to speech conversion. Function of each buttons is explained below by taking assumption that Gujarati word "ZFHF" is entered in text area.



જય જય ગરવી ગુજરાત

Figure 3: IDE of TTS Module

Enter Text: Work area becomes enabled to enter Gujarati text.

Select Text: This option allows to select portion of entered text.

Process: First it makes text area disable. Now text cannot be modified or selected. It executes 3 steps as follow:

Step 1: Converts each character in ASCII form.
i.e. ZFHF → Z F H F → 090 046 072 046
Step 2: Matches and finds corresponding phonemes from master database.
i.e. 090 046 072 046 → 090046 072046 → ra ja → ra.wav ja.wav

Representation of extracted wav files ra.wav and ja.wav are shown in figure 4a and 4b.

Step 3 Stores each wav files in an array.

Save: Concatenates wav files stored in array and puts resulting wav file in a folder. i.e. 1.wav

Here 1.wav is resulting wav file after concatenating ra.wav and ja.wav. Its wav representation is shown in figure 4c.

Speak: Plays resulting wav file. It will produce vocal form of entered text.

Clear: Clears text entered in text area.

Quit: To exit from TTS Module.

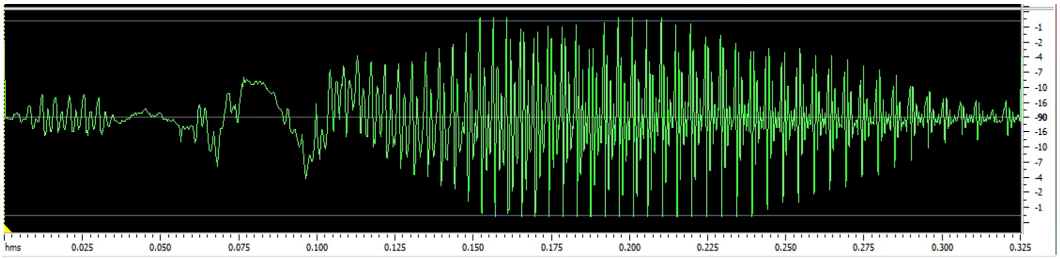


Figure 4a: wave diagram showing the phoneme 'ZF'

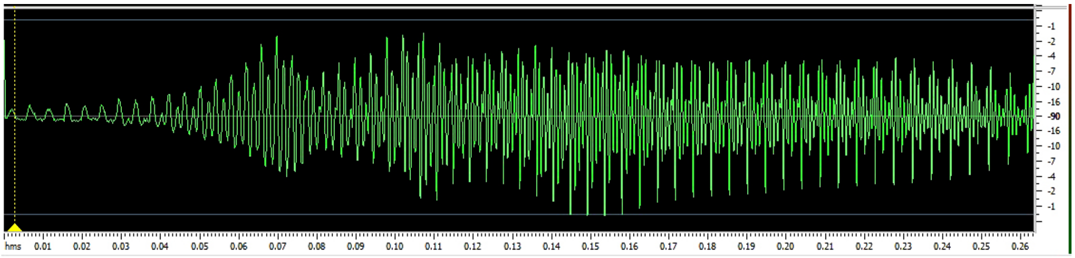


Figure 4b: wave diagram showing the phoneme 'HF'

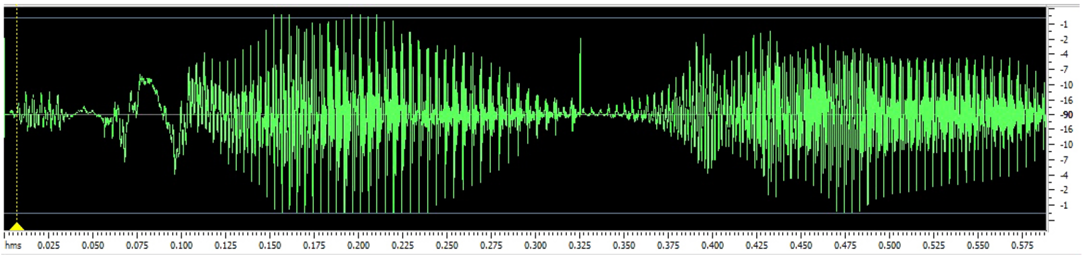


Figure 4c: Wav diagram showing how the word 'ZFHF' (King) was concatenated

Based on concatenation in figure 4c (stringing together) of segments of recorded syllables in figure 4a and 4b, stringing of syllables occurred when the text being supplied was linked to its corresponding syllables in the sound database. The pronunciation of the word 'ZFHF' (king) was achieved by concatenated the two syllables involved (ZF and HF) with the correct tone on each syllable, the sound of each syllable were fetched from the sound database and then stringed them together which then produced distinct sound.

VI. CONCLUSION

Algorithm and implementation of Gujarati TTS system using phoneme concatenative methodology is developed by researchers. This project focuses primarily on the process of creating a voice for a concatenative Text-To-Speech system, or altering the TTS systems own standard output voice to sound more like the target voice. The synthesized speech produced through this model is reasonably natural. About 80% of basic Phonemes are covered in database corpus which makes mismatch ratio very less. Phoneme matching is based on ASCII code combination. Researcher have observed the efficiency of this approach for Gujarati language and found that the performance of this approach is better. The generated speech shows distortion at the concatenation point of two syllables. If this distortion is significant then it would loose the naturalness. Future work will mainly focus on improving the naturalness [12] of the synthesizer.

REFERENCES

[1] J. Allen, M. S. Hunnicutt, D. H. Klatt, R. C. Armstrong, and D. B. Pisoni, From text to speech: The MITalk system. Cambridge University Press, 1987.

[2] http://research.ijcaonline.org/volume117/number19/pxc3903311.pdf

[3] http://dhvani.sourceforge.net/

[4] Intranet.daiict.ac.in:8085/DonlabTTS/

[5] http://safa-reader.software.informer.com/2.0/

[6] Ramani, S., Chandrasekar, R., Anjaneyulu, K.S.R. (eds.): KBCS 1989. LNCS, vol. 444. Springer, Heidelberg (1990)

[7] G. Cardona & B. Suthar. "'Gujarati'- In the Indo-Aryan Languages", 722-765. G. Cardona & D. Jain (eds). Routledge (2007).

[8] Ravi D J and Sudarshan Patilkulkarni, "A Novel Approach to Develop Speech Database for Kannada Text-to Speech System", Int. J. on Recent Trends in Engineering & Technology, Vol. 05, No. 01, 2011.

[9] Atal B. S and Hanauer Suzanne L., "Speech analysis and synthesis by linear prediction of the speech wave", The journal of acoustic society of America, 1971, pp 637-655.

[10] A. Hunt and A. Black, "Unit selection in a concatenative speech synthesis system using a large speech database," in Proc. IEEE Int. Conf. Acoust. Speech Process., Munchen, Germany, 1996, vol. 1, pp. 373–376.

[11] Kumbharana CK, "Speech Pattern Recognition for Speech To Text Conversion", Thesis, Nov. 2007, 113-140,

[12] Carlson, R., & Nord, L.“Vowel dynamics in a text-to-speech system - some considerations”. In Proceedings Eurospeech '93 (pp. 1911-1914). Berlin, 1993.

AUTHORS

First Author - Prof. JJ Kothari, Associate Professor, Shri M. P. Shah College of Commerce, Jamnagar. Research Fellow, Department of computer science, Saurashtra University, Rajkot. Email: jitendrajkothari@gmail.com

Second Author – Dr. CK Kumbharana , Associate Professor & Head, Department of Computer Science, Saurashtra University, Rajkot, Email: ckkumbharana@yahoo.com

Correspondence Author- Prof. JJ Kothari
Email: jitendrajkothari@gmail.com