A Phonetic Study for Constructing a Database of Gujarati Characters for Speech Synthesis of Gujarati Text

Prof. JJ Kothari
Associate Professor,
Shri M. P. Shah College of Commerce, Jamnagar.
Research Fellow, Department of
Computer Science, Saurashtra University, Rajkot.

CK Kumbharana, Ph.D
Associate Professor & Head,
Department of Compute Science,
Saurashtra University,
Rajkot

ABSTRACT
Authors are under procedure to develop Gujarati Text to speech(TTS) synthesis system using concatenation of Gujarati phonemes for unrestricted input Gujarati text. In this methodology, Gujarati phonemes are needed to record and store as speech database. This paper describes a flexible procedure to develop speech database in .wav format. This method is simple to implement which uses less memory space.

Keywords
TTS, Gujarati Text to Speech, Speech Corpus

1. INTRODUCTION
Text to Speech(TTS) is a system in which sequence of words are taken as input and converts them to speech. In conversion process using concatenating speech synthesis method, vowels and consonants are most important in Gujarati language. Each phonemes are made of combination of consonants and vowels. There are different concatenation methods like unit selection, diaphone or domain specific method for speech synthesis. In all these methods the voices are sampled from real recorded speech and speech synthesis is handled by computers. Here researchers main focus is to designing a database of phonemes in such a way that it speedup searching and retrieving process in Gujarati TTS.

2. LITERATURE REVIEW
Common speech synthesis systems are based on predefined units whose concatenation is obligatory. Small speech units such as diphones or demisyllables recorded from a human speaker are concatenated to build the synthetic utterance. The goal of speech synthesis is to develop a machine having an intelligible, natural sounding voice for conveying information to a user in a desired accent, language, and voice. For evaluation of TTS systems three parameters need to be evaluated: accuracy, intelligibility and naturalness [1]. There are different techniques for speech synthesis. It can be bifurcate as follow.

2.1 Different speech synthesis techniques
The following subsections describe the main principles of the three most commonly used speech synthesis methods: formant synthesis, concatenative synthesis, and articulatory synthesis[2] .

2.1.1 Formant synthesis
Formant synthesis is based on the well-known source-filter model which means that the idea is to generate periodic and non-periodic source signals and to feed them through a resonator circuit – or a filter – that models the vocal tract. The principles are thus very simple, which makes formant synthesis flexible and relatively easy to implement. Formant synthesis can be used to produce any sounds. On the other hand, the simplifications made in the modelling of the source signal and vocal tract inevitably lead to somewhat unnatural sounding result.

2.1.2 Articulatory synthesis
Compared with the other two synthesis methods, articulatory synthesis is by far the most complicated in regard to the model structure and computational burden. The idea in articulatory synthesis is to model the human speech production mechanisms as perfectly as possible. The implementation of such a system is very difficult and therefore it is not widely in use yet. Experiments with articulatory synthesis systems have not been as successful as with other synthesis systems but in theory it has the best potential for high-quality synthetic speech.

2.1.3 Concatenative synthesis
Concatenative speech synthesis is based on concatenation of pre-recorded natural sounding utterances. This is the so called cut and paste synthesis in which short segments of speech are selected from a pre-recorded database and joined one after another to produce the desired utterances[3]. This methodology has the advantage in its simplicity, i.e. there is no mathematical model involved. Speech is produced out of natural, human speech. Concatenative synthesis is based on the concatenation (or stringing together) of segments of recorded speech. Generally, concatenative synthesis produces the most natural sounding synthesized speech. There are three main subtypes of concatenative synthesis: Unit selection synthesis, Diphones synthesis, Domain-specific synthesis. PSOLA (Carpentier and Moulines, 1989) [4] method, MBROLA method are some different well-known concatenative synthesis methods.

2.2 Current Research projects in India:
Some of the institutions in India are engaged in speech synthesis. The IIT Madras has worked on a novel scheme where the "unit" is a character of written "text". The Tata Institute of Fundamental Research (TIFR), Mumbai has reported unlimited continuous speech synthesizer using formant synthesis technique. Whereas TIFR [5] and Central Electronics Engineering Research Institute (CEERI) [6] worked with formant synthesis, ISI, Kolkata[7], Indian Institute of Information Technology (IIIT), Hyderabad [8], centre for Development of Advanced Computing (CDAC), Pune and Kolkata developed concatenation-based synthesizers. Between the concatenation and formant synthesizers, the quality obtained so far is comparable. Speech

synthesizers based on Festival has been developed in languages including Hindi, Bangla, Kannada, Marathi and Tamil.

Linguistic Data Consortium for Indian Languages (LDCIL) is the Consortium responsible to create the database and shall provide forum for the researchers all over the world to develop speech application using the collected data in various domains. The LDC-IL has collected Speech databases in various Indian languages.[9]. The research that has been carried out is mostly for text to speech synthesis which uses phoneme/syllables concatenation on isolated words and is either based either on concatenative or formant synthesis techniques.

Dhvani is a Text To Speech System specially designed for Indian languages. This system has been developed by Simputer trust headed by Dr. Ramesh Hariharan at Indian Institute of Science Bangalore in year 2000. It uses diphones concatenation algorithm. Currently this system has Hindi, Malayalam, Kannada, Bengali, Oriya, Punjabi, Gujarati, Telugu and Marathi modules.

Speech group at Language Technologies Research Centre (LTRC) focuses on building Text to Speech (TTS) systems and Automated Speech Recognition (ASR) systems for Indian languages. They have built vary natural sounding speech synthesis systems for Hindi and Telugu. Reading Aid software for Visually Impaired (RAVI). Work is under progress for developing TTS engines for other Indian languages.

Vani is an Indian Language Text to speech Synthesizer for Sanskrit [10] developed using formant synthesis, in which the basic assumption is that the vocal tract transfer function can be satisfactorily modelled by simulating formant frequencies and formant amplitudes.

Text to speech synthesis for Indian languages (Acharya), is a syllable level representation of the text and each syllable directly translates into a sound that can be synthesized or simply played from a pre-recorded piece of audio [11].

3. GUJARATI CHARACTERS FEATURE

Gujarati (ગુજરાતી) 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[12]. It is spoken by over 46 million people all over the world. Gujarati is the first language of Mahatma Gandhi(fondly known as the Father of the nation in India), Mohammad Ali Jinnah, Sardar Vallabhai Patel, Morarji Desai, Dr. Vikram Sarabhai and Dhirubhai Ambani, to name a few. It is the 26th most spoken language in the world. Gujarati Script is based on abugida system rather than the alphabet system commonly used for European languages[13]. 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 [14].

Gujarati script has been descended from Brahmi which is a part of Brahmic family. There are about 34 consonants (vyañjana), 2 compound characters that are treated as consonants (not lexically though) and 12 vowels (svara) in a Gujarati language in which one is left vowel sign (e.g. ki/ કિ). A combination of vowels and consonants are used to represent a syllable in a Gujarati language. The different combinations are: C, V, CV, VC, CCV, and CVC where C is consonant and V is vowel. The major task of text processing is to convert the Gujarati text into Phonetic units to identify the above mentioned possible syllables, using syllabification rules.

4. 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 is given below.

Table 1: Consonants in Gujarati Language.

ક	**ખ**	**ગ**	**ઘ**	**ચ**	**છ**	**જ**	**ઝ**	**ટ**	**ઠ**	**ડ**	**ઢ**
ka	kha	ga	gha	ca	cha	ja	jha	ṭa	ṭha	ḍa	ḍha
ણ	**ત**	**થ**	**દ**	**ધ**	**ન**	**પ**	**ફ**	**બ**	**ભ**	**મ**	**ચ**
ṇa	ta	tha	da	dha	na	pa	pha	ba	bha	ma	ya
ર	**લ**	**વ**	**સ**	**શ**	**ષ**	**હ**	**ળ**	**ક્ષ**	**જ્ઞ**		
ra	la	va	sa	sha	śa	ha	ḷa	kṣa	jña		

Table 2: Vowels and Diacritic with ક in Gujarati Language

અ	**આ**	**ઇ**	**ઈ**	**ઉ**	**ઊ**	**એ**	**ઐ**	**ઓ**	**ઔ**	**અં**	**અઃ**
a	aa	i	ee	u	oo	e	ai	o	au	aṃ	ah
ક	**કા**	**કિ**	**કી**	**કુ**	**કૂ**	**કે**	**કૈ**	**કો**	**કૌ**	**કં**	**કઃ**
ka	kaa	ki	kee	ku	koo	ke	kai	ko	kau	kaṃ	kah

Table 3: Numerals in Gujarati Language

૦	**૧**	**૨**	**૩**	**૪**	**૫**	**૬**	**૭**	**૮**	**૯**
શુન્ચ	**એક**	**બે**	**ત્રણ**	**ચાર**	**પાંચ**	**છ**	**સાત**	**આઠ**	**નવ**

0	1	2	3	4	5	6	7	8	9

5. CREATION OF DATABASE STRUCTURE FOR GUJARATI PHONEMES AND SPEECH CORPUS.

Phones characterize any sound that can be produced by a human vocal tract, if a phone is part of a specific language; it becomes a phoneme of the language. Phonemes are the elementary sounds of a language. In this paper we are going to differentiate a character in Gujarati language scripts is close to syllable and can be typically of the following form:

Table 4. Different forms of Gujarati syllables.

Combination	Example	Explanation
V (Vowel)	ઉ	ઉ (Vowel)
C (Consonant)	ક	ક (Consonant)
C+V (Consonant + Vowel)	કી	ક (Consonant) + ઈ (Vowel) = ક + ી = કી
V+C (Vowel + Consonant)	કિ	ઈ (Vowel) + ક (Consonant) = િ + ક = કિ
C+C (Consonant + Consonant)	સ્વ	સ (Consonant) + વ (Consonant) = સ્વ
C+C+V (Consonant + Consonant + Vowel)	પ્રા	પ (Consonant) + ર (Consonant) + આ (Vowel) = પ્ + ા = પ્રા

- **Model for utterance selection and recording:**

Final intention of researcher is to generate a phoneme table which should be use as corpus database in Gujarati TTS. There are two issues concerning the generation of phoneme databases[15]. 1. Utterances selection and 2. Utterances recording.

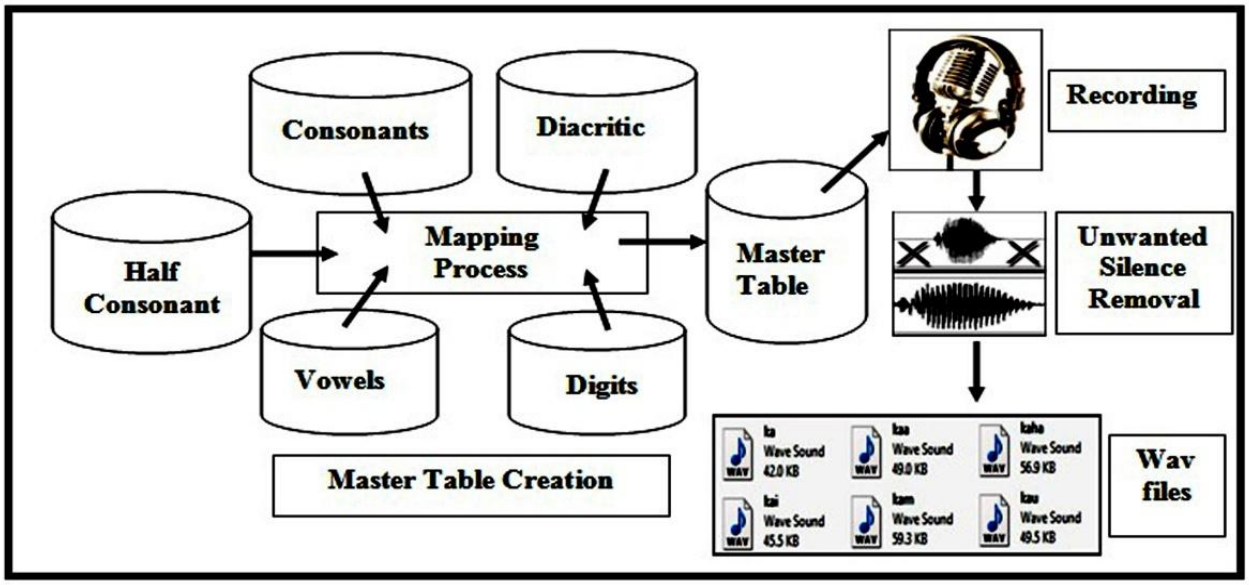

Figure 1. Utterances selection and recording

5.1 Utterances selection.

This is first phase of model 1 to identify all possible combination of character needed to record. For resolving that purpose first of all database tables are generated in which vowels and consonants are stored containing "akshar" name with its ASCII code as follow.

Database table 1 shown in table 5 contains Gujarati half consonant (dead consonant or consonant with *halant* symbol which in Gujarati called "ખોડિયો અક્ષર") . for example word /khyati/ ખ્યાતી contains half consonant.

Database table 2 shown in table 6 contains Gujarati consonants (in Gujarati called "વ્યંજન") . It is base characters for creating Gujarati "barakhadi". For creating "Barakhadi" of "અ", vowel /a/ "અ" is also included.

Database table 3 shown in table 7 contains Diacritic (Anuswara). Barakhadi of a Gujarati alphabets is combination of that alphabet with Anuswara. For example ક + ા = કા, ક + િ = કિ, ક + ી = કી, ક + ુ = કુ and so on.

Vowels are combination of " અ " with each anuswara. But there are four specially glyphic characters. Database table 5 shown in Table 9 represents such characters.

Table 5. database table 1 storing Half consonants

Sr	'Symbol'	Equivalent Gujarati 'Akshar'	ASCII
1	k	ક્	201
2	kh	ખ્	098
3	g	ગ્	117
4	gh	ઘ્	169
5	ch	ચ્	114
And so on …. total 34 half consonant			

Table 6. Database table 2 storing Consonants

Sr	'Symbol'	Equivalent Gujarati 'Akshar'	ASCII
1	av	અ	086
2	ka	ક	083
3	kha	ખ	066
4	ga	ગ	085
5	gha	ઘ	051
6	cha	ચ	082
And so on … total 34 consonant + " અ"			

Table 7. Database table 3 storing Anuswara

Sr	'Symbol'	Equivalent 'Anuswar'	ASCII
1	a	-	-
2	aa	ા	070
3	i	િ	108
4	ee	ી	076
5	u	ુ	093
6	oo	ૂ	125
7	e	ે	091
8	ai	ૈ	123
9	o	ો	077
10	au	ૌ	193
11	am	ં	092
12	aha	ઃ	111

Table 8: Database table 4 storing numerals.

Sr	'Akshara'	Equivalent Gujarati 'Ank'	ASCII
1	0	૦	095
2	1	૧	033
3	2	૨	090
4	3	૩	035
5	4	૪	036
6	5	૫	053
7	6	૬	038
8	7	૭	042
9	8	૮	040
10	9	૯	041

Numbers and digits are part of language. Sound recording of each digits is also needed in TTS system. Database table 4 shown in Table 8 represents such characters.

As per first part of the speech corpus creation model, all the database tables listed in table 5 to table 9 are combined using mapping technique to create master database table. Master table is collection of all phoneme for which sound is to be record. Number of phoneme to be record is counted in table 10.

Table 9. Database table 5 storing special glyphic vowels.

Sr	'Symbol'	Equivalent Gujarati 'Akshar'	ASCII
1	iv	ઇ	108
2	eev	ઈ	076
3	uv	ઉ	093
4	oov	ઊ	125

Table 10: Calculation of master table entries.

Database Table combination	Total combination	Final Entry
1	34 Half consonant	34
2 and 3 and 5	34 consonants + " V" x 12 Diacritic to make "Barakhadi"	35 x 12 = 420
4	10 Digits	10
2 and 3	(34 consonants + 'Rakar") x 12 Diacritic	34 x 12 = 408
	Total phonemes	872

Total number of entries in master table will be 872. All these 872 phonemes are needed to record to produce sound file.

5.2 Recording of these utterances.

Speech for each phoneme should be recorded from the native speakers of the language [15]. The recorded sound files are then named and stored by using the phoneme name itself. For example the sound file of /ka/ (ક) is named ka.wav. All the sound files recorded are named and stored in the similar way. Module for sound recording and testing is developed in JAVA. Each uttered sound is recorded at sampling rate of 44100Hz, 16 bit Stereo quality format.

The steps followed for recording the speech samples was as follows:

Step 1: Selected speakers were asked regarding any problem with reading or speaking the Gujarati phonemes

Step 2: Speakers were given the basic information about the headset used and when to speak the phoneme.

Step 3: The rate of sampling frequency set to 44100Hz, 16 bit Stereo quality format.

Step 4: The speaker was asked to speak each phoneme and the recorded sample was saved as (phoneme name).wav file.

Step 5: Step 4 was repeated for all 872 utterances that were recorded from the speaker.

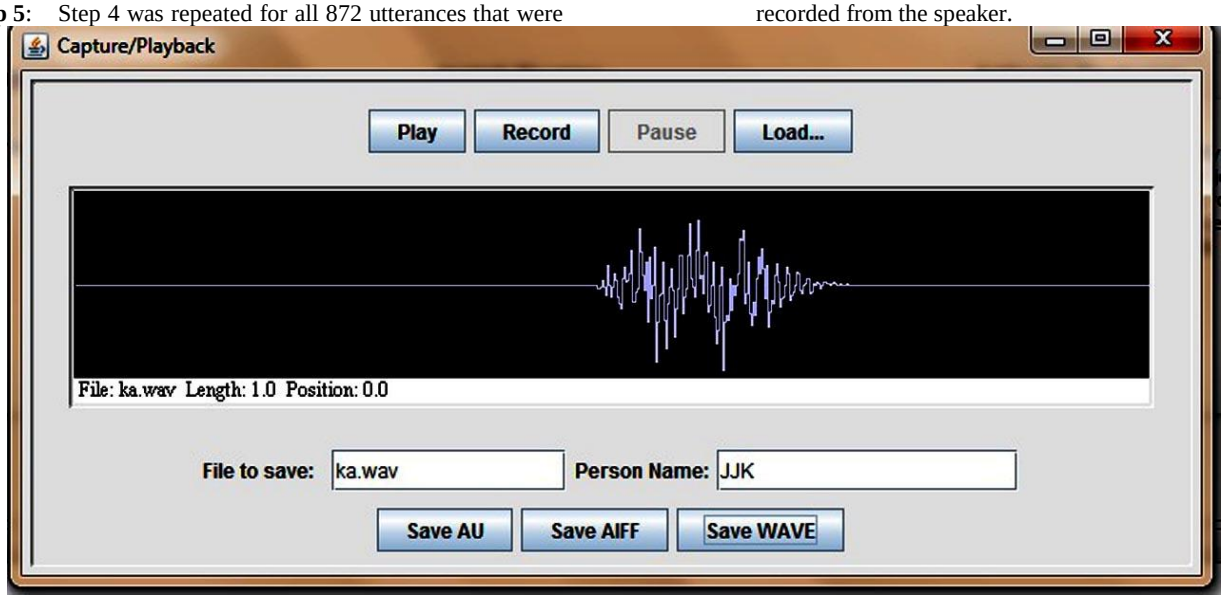

Figure 2. The acoustic graph of the recorded Gujarati phoneme /ka/ (ક)

Recording of a phoneme will contain voice segment as well as unvoiced segment. Before storing it in phoneme database, its unvoiced segments should be removed. Here Free Audio Editor 2015 is used to crop each recorded sound files to remove unwanted silence.

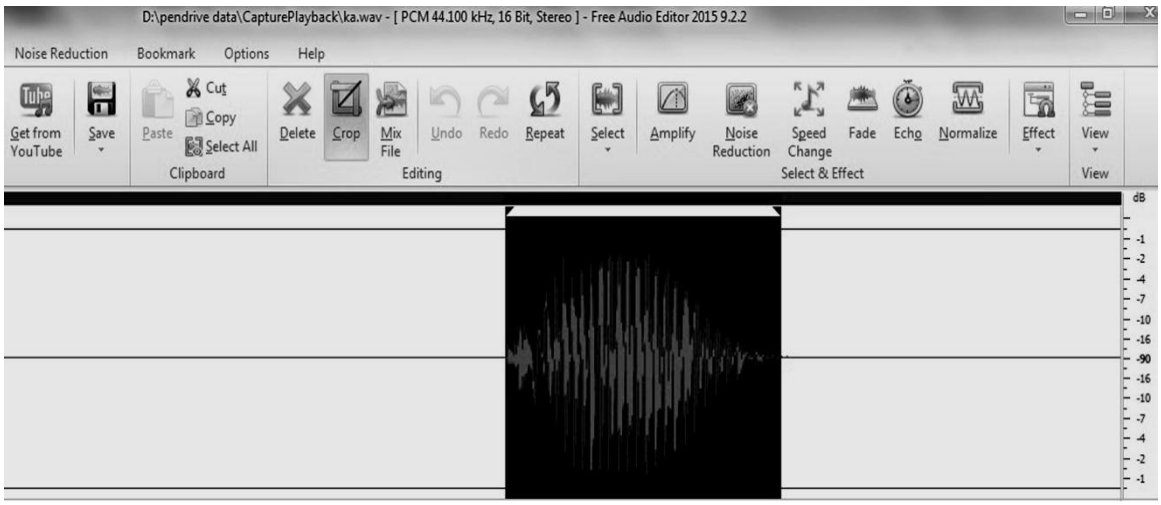

Figure 3. Wav file cropping using Free Audio Editor 2015

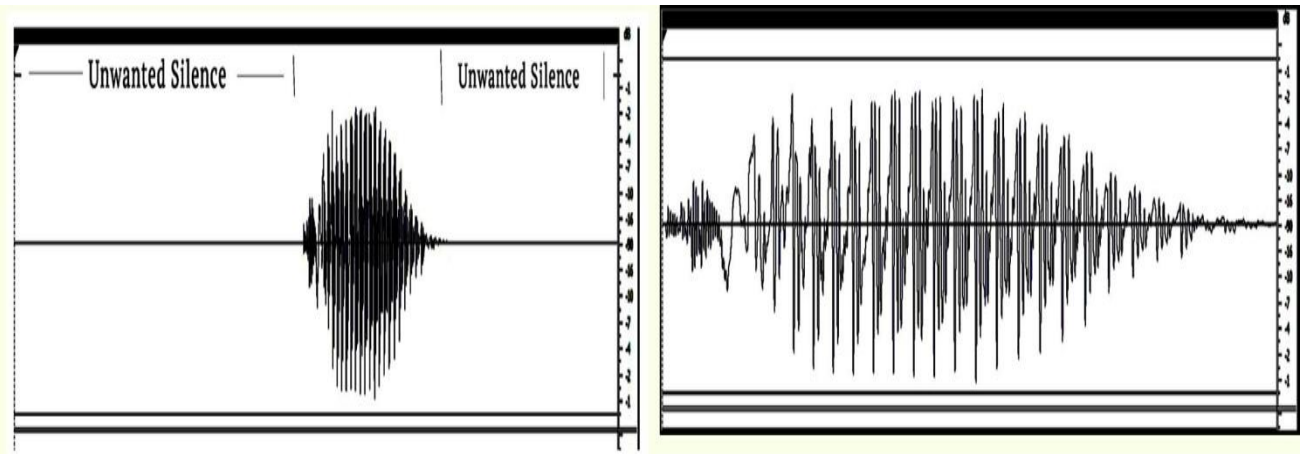

Figure 4. The acoustic graph of the Gujarati phoneme /ka/ (ક) before and after removing unwanted silence.

This is the complete process to construct database for developing Gujarati text to speech system which includes most of the phoneme variation of Gujarati language.

6. CONCLUSION

One of the factors regarding success of a good TTS system using concatenative method is depend upon how speech database is organized. Here authors have construct speech database in such a way that speech units are associated with phoneme and its ASCII code. Author found two benefits by applying this methodology:

1] phoneme matching and speech unit searching becomes very easy.

2] As Indian languages are phonetic in nature, one can easily apply same methodology for other Indian languages like Marathi, Bengali and Hindi etc. instead of Gujarati just by changing ASCII code.

Authors are on way to develop and study Gujarati TTS Synthesis System. This research will be useful as a base to recognise Guajarati text pattern in speech conversion process.

7. REFERENCES

[1] Taylor P., Text to Speech Synthesis. University of Cambridge, 2007, DRAFT available: http://mi.eng.cam.ac.uk/pat40/ttsbook_draft_2.pdf, Retrieved on Feb 01, 2009

[2] S. Lemmetty, Review of Speech Synthesis Technology, Master's Thesis, Helsinki University of Technology, 1999.

[3] http://www.cs.tut.fi/kurssit/SGN-4010/puhesynteesi_en.pdf

[4] Speech Synthesis, http://en.wikipedia.org/wiki/Speech_synthesis, Retrieved on June 24, 2008

[5] Furtado X A & Sen A, "Synthesis of unlimited speech in Indian Languages using formant-based rules"' Sadhana,1996,pp 345-362

[6] Agrawal S S & Stevens K, "Towards synthesis of Hindi consonants using KLSYN88", Proc ICSLP92, Canada, 1992, pp.177-180

[7] Dan T K, Datta A K & Mukherjee, B, "Speech synthesis using signal concatenation", J ASI, vol. XVIII (3&4), 1995, pp 141-145

[8] Kishore S. P., Kumar R & Sanghal R, "A data driven synthesis approach for Indian language using syllable as basic unit", Proc ICON 2002, Mumbai, 2002

[9] Agrawal S. S. 2010, "Recent Developments in Speech Corpora in Indian Languages: Country Report of India", O-COCOSDA, Nepal.

[10] Jain, H., Kande, V., Desikan, K.: Vani - An India Language Text to speech Synthesizer. IIT, Mumbai.

[11] Acharya Project by IIT, Madras. Multilingual computing for Literacy and Education, http://acharya.iitm.ac.in/disabilities/tts.php.

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

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

[14] 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.

[15] Singh, S. P., et al Building Large Vocabulary Speech Recognition Systems for Indian Languages , International Conference on Natural Language Processing, 1:245-254, 2004.